



ASHG 2025 Platform Abstracts

As of October 1, 2025

Instructions

1. View the Table of Contents on the next page.
 - a. The Table of Contents is sorted by ascending session number and then chronologically by subsession presentation time.
2. Click on the title of the abstract you would like to view, and you will be taken to the page the abstract is on.

Table of Contents

Session 19: Alz in the Details: Piecing Together the Alzheimer's Puzzle	20
Single-cell-resolved phenotypes for Alzheimer's disease across 599 individuals and 3.4 million cells	20
Spatial transcriptomic of hippocampal region-specific pathology in Primary Age-Related Tauopathy and Alzheimer's Disease	20
A genome-wide epitranscriptomic regulatory landscape in the aged human brains with Alzheimer's disease	21
Single-Cell Multi-Omic Profiling Reveals an Alzheimer's Disease-Associated Somatic Mutation Burden in the Human Brain.....	22
Session 20: Evolution of the Coding and Non-Coding Genome	25
Expanded models of gene constraint improve selection estimates from multi-ancestry population data	25
The cumulative effect of protein-truncating variants on human life expectancy and overall health	26
Centromere-spanning haplotypes reveal deep ancestry and evolutionary dynamics in the human genome	27
Co-evolution of lactase persistence and glucose homeostasis in ethnically diverse Africans	28
Session 21: From Identification to Consequence: Structural Variant Detection, Mapping, and Effects	29
Marker-based tumor-normal haplotypes matching for accurate genome-wide recombination identification and haplotype specific somatic structural variation discovery using GIAB tumor-normal assemblies	29
Personalized Mapping of Allele-Specific Chromatin Accessibility at Single-Cell Resolution	30
Bridging long- and short-read sequencing to accelerate population scale structural variant association studies.....	30
★ Structural and rare variants underlying psoriasis susceptibility identified by whole-genome sequencing: discovery of CERCAM as a novel risk gene	32
Session 22: From Interaction to Insight: Leveraging Gene-Environment and Multi-Omics Interactions Across Populations	34

Novel method for scalable non-additive GWAS reveals 781 new loci across 2,329 traits in FinnGen	34
Context-specific polygenic scores improve prediction and portability in complex human traits	35
Genetic Ancestry and Environmental Differences in Somatic Alterations and Clinical Outcomes for Five Common Cancers	36
★ Incorporating UV exposure into functional genomic analyses identifies additional key cancer drivers and pathways at melanoma susceptibility loci	37
Session 23: It Is What It Isn't, Non-Canonical Drivers of Disease Gene Expression.....	39
Genetic control of alternative transcriptional initiation reveals a pan-cancer atlas of susceptibility genes	39
Population-scale single-cell long-read sequencing reveals transcriptome variations and genetic effects on alternative splicing across lung cell types.....	39
Improving isoform-level eQTL and integrative genetic analyses of breast cancer risk and mammographic density using long-read RNA transcript assemblies.....	40
Topoisomerase IIb binding underlies driver alterations and frequently mutated regulatory elements in cancer genomes.....	41
Session 24: Loop Logic: 3D Chromatin Architecture and Functional Genomics	43
3D genomic approaches mapping accessible chromatin and architecture implicate causal risk variants in primary open-angle glaucoma pathogenesis.....	43
Subtype-Specific Differences in 3D Genome Architecture Reveal Regulatory Mechanisms in Medulloblastoma	44
Comprehensive Dissection of Alzheimer's Disease Genetic Variants Using a Neuron-Tuned 3D DNA Foundation Model	44
Transformer-enhanced Hi-C Upscaling Unveils 3D Genome Architecture at Low Sequencing Depth	45
Session 25: Mendelian Mysteries Through a Global Lens.....	47
A decade of Initiative on Rare and Undiagnosed Diseases (IRUD): a nation-wide diagnostic and research network in Japan	47
★ Novel coding and noncoding rare variant association and in vivo functional validation across the developmental continuum	48
★ Biallelic Variants in KMO Cause a Novel Form of Congenital NAD Deficiency	49

Advancing National Genomic Initiatives: Abu Dhabi's Model for Newborn Genetic Screening and Lifelong Precision Medicine	51
Session 26: Seq-ing Function with Single Cells and CRISPR Perturbations	53
Systematic targeting and inhibition of bone mineral density GWAS loci in single cells identifies novel regulatory networks for human bone formation	53
PerTurbo: A scalable Bayesian framework for whole-transcriptome differential expression analysis and optimal design of single-cell CRISPR screens	54
PhenoBridge: A Unified Framework for Interpreting Molecular Effects on Complex Phenotypes through Integration of Perturb-seq and Phenotypic Data	55
★ Single-cell, Trimodal Omics Atlas of Ileal and Colonic Mucosal T Cells in Crohn's Disease Highlights CD4 ⁺ Regulatory T Cell Plasticity in Inflammation Progression	56
Session 27: Advanced Technologies Uncover Dosage, Structure, and Phenotypic Complexity	58
More is more: Shared phenotypes amongst sex chromosome trisomies hints dosage-sensitive effect of pseudoautosomal regions in genetic males and females	58
Completely resolved structural variants by optical genome mapping with adaptive sampling from CNV discovery.....	59
A multi-omic diagnostic framework leveraging long-read sequencing for unsolved rare disease	60
Comparative genomic analysis of syndromic and nonsyndromic forms of orofacial clefting	61
Session 28: Advances in Long-Read Sequencing Studies	64
Novel variant discovery and interpretation from long-read sequencing on 1,027 African Americans in All of Us	64
Long story short: discovery of new complex mutations and new candidate genes in Parkinson's disease using long-read sequencing.....	65
Cue2: deep learning framework for structural variant discovery with long reads	66
Sharing valuable rare disease and multiomic data through the GREGoR Consortium	66
Session 29: Cancer Risk: Gene Discovery and Polygenicity	69
Cell type-specific risk gene discovery in colorectal cancer through single-cell transcriptome-wide association study and regulatory element integration.....	69

Single-cell Transcriptome-wide Association Study Identifies Cell-Type-Specific Genes for Lung Cancer.....	70
The most comprehensive characterisation of polygenic risk for cutaneous melanoma to date.....	71
Identification of 33 new germline risk loci for clonal hematopoiesis by integrating genetic data on blood cell counts and indices	73
Session 30: Hijacked Immunity: Navigating the Balance Between Host Defense and Immune Dysregulation	75
★ Modeling Viral Evolution and T Cell Evasion through HLA-Informed Deep Learning..	75
★ East Asian GWAS of serum immunoglobulin and subclass levels provides insights into the genetic architecture of humoral immunity implicated with immune cell type specificity	76
Using functional genomics to understand variation in response to infection in northeast Thailand.....	77
Gene regulation in T cells depends on the severity of atopic dermatitis.....	78
Session 31: Making a Connection: Pleiotropy in Neurologic and Psychiatric Conditions	80
Gene discovery and pleiotropic architecture of chronic pain in a genome-wide association study of over 1.2 million individuals.....	80
Mapping the Pleiotropic Genetic Landscape of Brain Imaging Traits Using DIMPLE-GWAS and SuSie-GeneSet	80
A cell-type-resolved deep learning atlas of human cis-regulatory evolution	81
A deep cellular atlas of the human ventral substantia nigra in Parkinson's identifies a genetic and molecular overlap with Type 2 Diabetes	82
Session 32: Precision Risk for Every Brain	84
Genome-wide multi-trait and multimodal single-cell analyses of tobacco smoking and schizophrenia reveal novel risk loci and regulatory mechanisms	84
Genome Wide Association Study of Schizophrenia identifies African-enriched variants in PTPRD and TMEM150C	85
Rare Genetic Variation Illuminates Bipolar Disorder Risk Across Diverse Populations ...	86
Leveraging genome assembly and pangenome graphs to investigate an African-origin protective haplotype for Alzheimer's Disease in APOEε4 Carriers	87

Session 33: RNA Functions Beyond Coding Sequence 89

- ★ mcDETECT: Characterizing mRNA Localization in Polarized Neuronal Compartments With Spatial Transcriptomics 89
- Sequence-dependent regulators of transposable-element-associated alternative splicing 90
- Saturation mutagenesis of 37 human splicing factor genes using pooled prime editing. 90
- Non-coding snRNA variants in RNU4-2 and RNU6 paralogues are a recurrent cause of retinitis pigmentosa 91

Session 34: Single-Cell Multiomic Dissection of Gene Regulation Across Disease

States 93

- Integrated single cell map of the human pancreas and immune-related tissues provides novel insights into diabetes risk 93
- Integrative analysis of single-cell chromatin accessibility and GWAS statistics to prioritize enhancers and partition trait heritability in lung development 94
- Enhancer-Driven Gene Regulatory Networks Reveal Distinct Transcription Factor Hierarchies in Myeloid Cell Fate Decisions in Crohn's Disease 95
- Disease Genetic Scorecard Reveals Convergent Coding and Regulatory Variants in Complex Diseases 95

Session 36: Artificial Intelligence and Machine Learning Tools Reshaping Modern

Genomics 97

- Decoding gene expression in cellular and disease states with deep learning 97
- Integrating genetics, multimodal medical imaging, and AI for high-precision risk prediction of type 2 diabetes 98
- G2DBridge: A Two-Step Machine Learning Framework for Disease Risk Prediction Using Genetically Inferred Imaging Features 99
- Machine learning-derived exosomal circRNA signature enables early lung cancer detection and reveals metabolic dysregulation 100
- Distinguishing causal variants from linked hitchhikers in the human genome using deep learning 101
- PubMind: A Large Language Model Framework for Scalable Extraction of Variant and Disease Associations from Biomedical Literature 101

Session 37: Choose Your Own Adventure: Novel Approaches in Statistical Genetics 103

Computation and resource efficient genome-wide association analysis for large-scale imaging studies	103
A phenome-wide map of pleiotropy: Connecting genetic signals to disease mechanisms and drug discovery	104
Parent-of-origin inference and its role in the genetic architecture of complex traits: evidence from ~235,000 individuals	105
mFABIO: Multi-tissue TWAS fine-mapping to prioritize causal tissues and genes for binary traits	106
Leveraging cis and trans genetic data to improve protein level prediction in proteome-wide association studies	107
causalR2M: causal Mendelian randomization mediation analysis with high-dimensional multi-omics mediators in 7000 individuals provides novel insights into BMI's casual effect on blood pressure	107
Session 38: Computational Tools for Precision Therapeutics	108
Machine learning-enabled genetic discovery and therapeutic targeting of hepatic IRS1 in MASLD.....	108
Cas9 cytosine base editor proteins for precision genome editing with high specificity and activity	108
Leveraging Open-Source Large Language Models to Identify Undiagnosed Patients with Rare Genetic Aortopathies	109
Enhanced Prediction of Antisense Oligonucleotide Therapy Eligibility for Rare Genetic Diseases Using a Semi-Automated Bioinformatics Pipeline	110
DeepDR-ADT: Multi-Modal Deep Learning for Drug Repurposing with Auxiliary Dosage-Time Task Learning	111
Building a 100 million cell single cell atlas of drug perturbations across 50 human cell models for AI driven drug discovery.....	112
Session 39: Delivering on Precision Medicine: Final Outcomes and New Insights on Returning Genetic Results Across Diverse Cohorts.....	113
The All of Us Research Program's Return of Research DNA Results Program: A Complete Review	113
RESPECT3: A randomized study evaluating return of actionable genetic research results from the Penn Medicine Biobank	114

Return of hereditary cancer research results to Black women: Results of a non-inferiority trial comparing online versus genetic counselor modalities	115
The GUARDIAN expanded newborn screening study: Short- and medium-term follow-up with the first 15,000 participants enrolled.....	116
A post-test counseling model accelerates access to cardiovascular genetic test results while maintaining psychosocial benefits of genetic counseling: Interim results of RESEQUENCE-GC (NCT05422573)	117
Implementing polygenic risk scores in clinical care: The eMERGE Network's return of genome informed risk assessment results for common conditions returned to 23,840 adults and children	117
Session 40: Expanding the RNAverse in Health and Disease	119
Deep reinforcement learning tuned self-attention for nucleotide resolution mapping of regulatory variants in long non-coding RNA	119
First Comprehensive Long-Read Transcriptome Assembly of the Human Placenta Reveals Novel Molecular Mechanisms of Maternal-Fetal Health	120
GTEx and GENCODE 47 integration expands the role of long-noncoding RNAs in complex traits and disease	121
Comprehensive long non-coding and circular RNA atlas of human brain regions across multi-ancestry cohorts	121
Identification of long non-coding RNAs potentially involved in inherited retinal disease pathogenesis	122
A novel human β cell microprotein from a lincRNA controls insulin secretion and harbors rare null variants associated with type 2 diabetes	123
Session 41: Human Diversity and Evolution	124
Is Directional Selection Rare in the Recent Human Past? Insights from Ancient DNA Time Series	124
Spatiotemporal dynamics of adaptive evolution in Neolithic East Asia.....	124
A comprehensive map of introgressed and adaptive structural variation in diverse humans	125
PHATE Derived Ancestry Coordinates Can Effectively Capture Continuous Population Structure in Human Genomic Data	127
Stable genetic risk for myopia over 15,000 years in East Asian and European populations reveals predominant environmental drivers of the modern myopia epidemic	128

Machine learning identifies the genomic and statistical features that predict cross-population replication of GWAS signals.....	128
Session 42: Massively Parallel Experimental and AI Approaches for Variant Characterization.....	
Variant pre-classification using ~280,000 variant effect measurements in 42 disease genes illustrates a realistic, scalable path to dramatically reduce variants of uncertain significance	129
T cell-specific causative variants underlying immune-mediated inflammatory disorders	129
Global identification of inflammatory bowel disease risk variant allele-dependent gene regulatory activity	130
Cell2net: A Deep Learning Framework to Decode Genotype-Specific Gene Regulatory Networks from Single-Cell Multi-Omics	131
AlphaGenome: advancing regulatory variant effect prediction with a unified DNA sequence model.....	132
Functional annotation of over 25,000 type 2 diabetes-associated variants.....	132
Session 43: Statistical Genetics: Multi-Omics to Improve Gene Discovery.....	
Integrating Genomic and Proteomic Data Improves Complex Trait Prediction in Diverse Populations.....	133
Integrative multi-omics approaches identify key molecular pathways and improve risk prediction of Late-Onset Alzheimer's Disease.....	134
Integrating multi-omics and multi-context QTL data with GWAS reveals the genetic architecture of complex traits and improves the discovery of risk genes	135
Integrating multi-omics data to identify lung-tissue-specific susceptibility proteins for lung cancer risk	136
Integrated Mendelian randomisation and multi-omic validation identify causal serum proteins in amyotrophic lateral sclerosis	137
Multiomic analysis of circadian and seasonal biomarkers	137
Session 44: Beyond the Bean: Genetic Studies of Kidney Disease	
Novel genetic insights into kidney function in countries with a high prevalence of chronic kidney disease: A meta-analysis of GWAS from 596,624 East Asians in Japan, South Korea, and Taiwan.....	139

Ancestry-specific genetic loci for kidney function identified in the largest African-ancestry meta-analysis	141
Bridging Genomic Inequities: Chronic Kidney Disease Susceptibility and Pharmacogenomic Variation in Indigenous Australians	142
Uncovering the inconsistencies and potential risks of AI: inaccuracies in nephro genetic counseling provided by five LLMs	143
Session 45: Dissecting the Biology of Neurodegenerative Disorders	145
Localizing MS GWAS SNPs to pathogenic cell types reveals convergent mechanisms of microglia dysfunction	145
Age at onset GWAS of Parkinson's Disease in the FinnGen study suggests links with genetic risk for earlier age at onset and known prodromal symptoms for Parkinson's ..	146
Large-scale transcriptomic profiling across brain regions in Progressive Supranuclear Palsy.....	147
Genomic and epigenomic insights into purkinje and granule neurons in Alzheimer's disease and related dementia using single-nucleus multiome analysis	148
Session 46: Functional Dissection of Variants at Scale	149
Multi-modal functional mapping of non-coding sequences regulating fetal hemoglobin	149
A comprehensive MPRA of both trans-ancestral genome-wide and suggestively-associated childhood obesity loci implicates a role for CBLN4 in microglia.....	150
Deciphering causal variant effects in Digestive Disease via Organoid-based allele specific open chromatin and CRISPRi	151
Multimodal CACNA1C deep mutational scanning to assess the impact of all missense variants on ion channel function	152
Session 47: Genetic Findings in African Genetic Ancestry Populations	154
Proteogenomic analysis of BMI in ~200,000 continental Africans and admixed African identifies 28 loci and enhances cross-ancestry incident obesity prediction using proteomic scores	154
Characterizing the genetic and environmental factors that impact vitamin D metabolism in sub-Saharan Africa	155
Dissecting Genetic Contributors to eGFR Bias: Effects of African-Enriched GATM Regulatory Variants on Detailed Renal Phenotypes	156

Gene expression associated with impending severe asthma exacerbations in non-Hispanic Black individuals with asthma	157
Session 48: In Situ Insight: Uncovering Cellular Regulation with Spatial-Omics	159
BRIDGE: Inferring 3D Molecular Tissue Structure from Sparse Spatial Omics and Histological Sections	159
Applying the Novel Illumina Spatial Technology in Pig-Kidney to Human Xenotransplant biopsies allows deep profiling of infiltrating human immune cells	159
Mitochondrial lineage tracing within spatially intact human tissues.....	160
Subcellular spatial transcriptomics with spatial neighborhood analysis framework reveals immune–stromal crosstalk within the synovium of patients with juvenile idiopathic arthritis	161
Session 49: Methylation Mechanisms: New Insights into DNA-Methylation in Health and Disease.....	163
★ Epigenome-Wide and Causal Mediation Analyses Identify DNA Methylation Pathways Mediating Longitudinal Glycemic Effects on Diabetic Retinopathy	163
Widespread non-Mendelian inheritance of DNA methylation patterns in mice	164
Differential Bulk and Cell-Type Specific DNA Methylation in Asthma with Severe Exacerbations in Pediatric Populations of Diverse Ethnicities.....	165
Therapeutic epigenome editing for Prader-Willi Syndrome using a novel non-viral, non-nanoparticle delivery platform for CRISPR/dCas9-based gene regulation	166
Session 50: Precision in Chaos: Exploring the Intricacies of Immune Regulation and Autoimmunity.....	168
A multi-omics resource of B cell activation reveals genetic mechanisms for immune-mediated diseases	168
Multi-Trait GWAS and Fine-Mapping of 14 Traits Identify 17 Ancestry-Specific Immune Loci in Fibrotic and Inflammatory Diseases	169
T cell receptor repertoire-wide association study identified potential biomarkers for rheumatoid arthritis.....	170
Post-GWAS analysis in autoimmune-inflammatory diseases: strength and patterns of allele-specific DNA (ASM) vary between disease-relevant cell types.....	171
Session 51: Resolving Complex Genomic Variation in Rare Disease	173
Comprehensive characterization and analysis of nuclear mitochondrial DNA segments (NUMTs) from long-read sequencing data.....	173

Identification of de novo variants from parent-proband duos via long-read sequencing	174
Improved Transcriptome Diagnostics in the Undiagnosed Diseases Network Using the Telomere-to-Telomere CHM13 Genome	175
Characterization of Rare Genomic Structural Variants Across 2,981 Genomes Reveals Significant Involvements in Recessive Conditions	176
Session 52: Autism Spectrum Disorder: From Genetic Variation to Neurobiology	178
The role of topologically associating domains in defining the pathogenicity of de novo non-coding variants on autism spectrum disorder risk	178
Rare coding variants identify novel genes and phenotypic context of autism	179
Shared and distinct genetic architectures of autism and neuropsychiatric disorders ..	180
Chd8-dependent molecular and functional vulnerabilities in the developing and adult mouse cerebellum.....	181
Session 53: Every Story Has a Beginning and an End: About UTRs and Poly(A) Tails ..	183
Genome-wide detection of human 5'UTR variants of functional significance in translation	183
MicroRNA Perturb-Seq Reveals Genome-wide Functional Targets and Deleterious 3'UTR Variants	184
Modulating gene expression in human cells via high-throughput prime editing of regulatory elements.....	184
Atlas of cell type-specific genetic impacts of alternative polyadenylation in immune-related diseases	185
Session 54: Function of Complex Variation and Repeats	187
A phenome-wide association study of CNVs genotyped from genome sequencing data in the UK Biobank.....	187
Identification of genetic modifiers of somatic instability across multiple repeat expansion disease loci using in vivo CRISPR–Cas9 genome editing	188
Computational analysis of tandem repeats for genome-wide identification of novel repeat expansions	189
Going Viral: Targeted Detection and Profiling of Endogenous Retrovirus Transcripts in the Human Genome	190
Session 55: Gene Editing and Gene Replacement for Rare Disease.....	191

Corrective editing to treat the connective tissue disorder pseudoxanthoma elasticum	191
High-Coverage Whole-Genome Sequencing Uncovers Complex CRISPR-Induced Genomic Rearrangements in Human Hematopoietic Stem Cells.....	192
A one and done therapeutic strategy to correct ELP1 splicing defect in familial dysautonomia	193
Mouse models of, and AAV gene therapy for, cobalamin C (cblC) deficiency	194
Session 56: Genetic Insights Fueling Drug Discovery and Precision Health	195
Human genetics evidence supporting ANGPTL4 inhibition to reduce remnant cholesterol and triglycerides informs successful Phase 2a trial outcomes	195
Associations between ultra-rare loss-of-function variation and 454 traits in the Mexico City Prospective Study	196
GLP1R and GIPR coding variants reveal differential impacts on metabolic and mental health – Implications for drug response in incretin receptor agonist treatment	197
Hypercholesterolemia polygenic score modulates response to statin therapy	198
Session 57: Genetic Mechanisms of Cancer Risk.....	200
Comprehensive characterization of functional pleiotropy at the 5p15.33 multi-cancer risk locus	200
Genome-wide characterization of clonal hematopoiesis reveals extensive non-coding putative driver mutations	201
Detection and quantification of cancer-associated mutagenic processes in normal human somatic cells	202
A multiplex assay of variant effect (MAVE) of MSH6 enables accurate, prospective Lynch Syndrome clinical variant interpretation	203
Session 58: Mendelian Disorders of the Epigenetic Machinery: New Cellular and Phenotypic Insights	205
Cell type-specific dysregulation of gene expression due to Chd8 haploinsufficiency during mouse cortical development.....	205
ASXL1 mutations drive metabolic alterations through mitochondrial dysfunction.....	205
Defining the source of motor deficits in KMT5B-linked syndrome.....	207
Rare genetic variants in histone modification pathways implicate epigenetic dysregulation in ADHD	208

Session 59: Shaking The Evolutionary Tree: Multi-Species Studies to Understand Human Biology	210
Single Cell Multiomics Across Nine Mammals Reveals Cell Type Specific Regulatory Conservation in the Brain	210
Genetic control of regulatory dispersion across human and chimpanzee cell types....	211
Elucidating the metabolic mechanisms of cellular homeostatic resilience	211
The human and non-human primate developmental GTEx projects reveal tissue-specific gene expression across developmental stages	212
Session 69: Advances in Cell-Free DNA Applications from Prenatal Genomics to Cancer Diagnostics	214
Noninvasive fetal exome sequencing from maternal plasma: Large-scale validation and optimization for prenatal genetic use	214
Performance of prenatal cell-free DNA screening to identify sex chromosome abnormalities: Experience of a commercial laboratory	215
Prenatal cell-free DNA screening detects fetal genetic conditions linked to uniparental disomy.....	216
Genome-Wide Variations of End Motif in Cell-Free DNA Fragments Distinguish Immunotherapy Responders from Non-Responders in Head and Neck Cancer: A Multi-Institute Prospective Study.....	217
Session 70: Contextualizing Somatic Variation: From Clonal Expansion to Molecular Mechanisms	219
★ Landscape of somatic mutations across genomes, tissues, and individuals	219
Clonal Hematopoiesis of Indeterminate Potential (CHIP) Mutations Are Enriched in Alzheimer's Disease Patients with APOE ε3/ε3 Genotype	219
Multi-ethnic Transcriptomic Analysis of Clonal Hematopoiesis Reveals Driver-Specific Pathways and Causal Associations.....	220
★ Population scale single-cell multi-omic analysis of clonal haematopoiesis	221
Session 71: From Classroom to Community: Shaping the Future of Genomics Education	224
PaRtnEring to build understanding of gEnomics Responsibly: the PREFER CHW research education program	224
Psychometric Validation of the Education and Assessment of Genetic Literacy (EAGL) Scale	225

Learning Competencies and Pedagogical Considerations for the Future Public Health Genetics Workforce	226
How Epigenetics is Communicated on YouTube: A Content Analysis of Science Communication Strategies	227
Session 72: Functional Consequences of Transcriptome Variation Across Tissues and Populations	228
Multi-modal dissection of coronary artery disease GWAS loci reveals transcript splicing-mediated risk mechanisms in vascular smooth muscle cells	228
Isoform-Level Transcriptomic Signatures of Glycemic Traits.....	229
Allele-specific transcriptome variation identification at single-cell resolution in human retina enabled by single-cell long-read RNA sequencing	229
Multi-omics in 10,000 diverse individuals expands mapping genetic mechanisms in health and disease	230
Session 73: Grey Matters: The Science of Aging	232
Linking single-cell transcriptomic and genomic changes in the aging human brain	232
Epigenetic mosaicism acquired in human embryogenesis and aging	232
Epigenome-Wide Analysis of Aging in the Largest Nanopore Methylation Cohort of the Emirati Genome Program	233
Non-linear molecular QTLs in the human brain reveal context-dependent effects on aging and neurodegenerative disorders	234
Session 74: Pan-Genome Perspectives on Cardiovascular Diseases.....	236
Uncovering Hundreds of New Loci from a Multi-Ancestry Study of Blood Pressure Traits in Over 2.5M Individuals	236
Cigarette smoking modifies genetic associations with serum lipids in cross-population meta-analyses of 920,966 individuals	237
Genetic underpinnings of stroke in Hispanic/Latino populations.....	238
Global Evaluation of Congenital Heart Disease-Associated Non-Coding Variant Effect on Cardiac Transcription Factor Regulation.....	239
Session 75: The Utility of AI in Clinical Genomics Workflows	241
Automated Phenotyping Using NLP Rivals Human Expert Curation and Enables Scalable Gene Prioritization in Clinical Diagnostic Workflows.....	241

Artificial intelligence cannot replace human intelligence: publicly available generative AI is not suitable for variant classification	242
The Community Texome Project: Advancing Rare Disease Diagnosis Through Genomic Medicine, AI, and Functional Modeling in Texas communities	243
Benchmarking genAI tools for literature retrieval and summarization for genomic variant interpretation	244
Session 76: Unraveling Disease Mechanisms Through Multi-Omics: From Genetics to Microbiome and Chromatin Regulation	246
Microbiome and host genetic effects on gene regulation in Inflammatory Bowel Disease	246
Multi-omics integration nominates disease-relevant cell types and novel effector genes for Crohn's disease and ulcerative colitis	247
Integrating GWAS and CRISPR screens reveals context-specific T cell cytokine regulation in inflammatory bowel disease	248
Resolving genome regulatory complexity with new multiomic long-read sequencing tools for chromatin state, DNA methylation, and genetic variation.....	249
Session 84: All That and a Bag of Variants: Functional Genomics of Brain Disorders. 250	
Optimally matching healthy and disease-specific adult brains onto the EN-TEp personal epigenome resource	250
★ Genetic atlas of the human brain revealed by a 2-Million-Voxel GWAS at millimeter resolution	251
★ Uncovering molecular mechanisms of neuropsychiatric risk via functional genomics and AlphaFold 3	252
SingleBrain: A Meta-Analysis of Single-Nucleus eQTLs Linking Genetic Risk to Brain Disorders	253
Session 85: Better Together: Multi-Ancestry Approaches for Gene Discovery.....	255
Multi-ancestry genome-wide and transcriptome-wide association analyses identify new risk loci and genes for Inflammatory Bowel Disease	255
★ Largest multi-ancestry genetic study of 99 infectious disease traits identifies hundreds of novel disease-associated loci across the whole genome	256
Multi-ancestry genome-wide association study of endometriosis and its clinical manifestations in ~1.4 million women: translating gene discovery into pathogenic mechanisms and therapeutic targets	257

Next generation multi-population GWAS of Coronary Artery Disease	259
Session 86: Genomic Causes and Consequences in Drug Treatment	261
Genome-wide association study for Glucocorticoid-induced Ocular Hypertension	261
High-throughput Identification and Genetic Analysis of Adverse Drug Effects in Individuals of African Ancestry	262
Genetic control of gene expression in whole blood of childhood cancer survivors	264
Genomic and transcriptomic insights into antipsychotic-induced changes in total cholesterol in a multi-ancestry cohort of 58,000 veterans	265
Session 87: Prenatal Genetics and Reproductive Health	267
Anticipating Futures: Stakeholder Perspectives on Ethical Governance of Prenatal Gene Editing	267
Utility of Genomic Autopsy for Infant Mortality	267
Prospective comparative evaluation of optical genome mapping, mate-pair genome sequencing, and chromosomal microarray analysis in prenatal genetic diagnosis of fetuses with ultrasound anomalies	269
Lifecourse Genetic Risk of Cardiovascular Disease in Women with Reproductive Health Conditions	270
Session 88: Using Disease Models to Assess Variant Affect	272
Exploring the Utility of A Knockout Mouse Resource for Insights into Human Disease .	272
Digenic variants that affect an actin-mitochondria-glutamate pathway cause epilepsy	272
Investigating cilia-autophagy crosstalk in skeletal development and FGFR3-mediated pathogenesis of achondroplasia	273
ZIC1 is a transcription factor that intersects Idiopathic Hypogonadotropic Hypogonadism and Craniosynostosis	275
Session 89: Biobank-Scale Genomic Data Resources of Diverse Populations	277
Large-scale metabolomic profiling enables genome-wide and phenome-wide association studies in the Million Veteran Program	277
Our Future Health: The World's Largest Cohort Study is Open for Genetic Research ...	278
The All of Us Research Program data release 2026 (CDR v9): Powering genomic research through All of Us	279

Uncovering genetic etiologies in 30,000 rare disease genomes from a nationwide Chinese cohort.....	280
Session 90: Cardiometabolic Traits: Insights from Multi-Omics Approaches.....	282
Large scale metabolomics study in 69,671 South Asians identifies novel rare variant associations with lipids.....	282
Connecting proteomics and genomics to identify sex-specific plasma biomarkers for calcific aortic valve stenosis.....	283
Proteomic discovery in an inception cohort of acute myocardial infarction survivors ..	284
Adipose single nucleus transcriptomic and epigenomic sequencing discovers 192 colocalized regulatory variants for eight cardiometabolic disease traits.....	285
Session 91: End Results—Telomere Resolution and Involvement in Genomic Instability	287
Telomere dysfunction promotes remote recombination events between telomeres, centromeres and ribosomal DNA that are catalyzed by long-range chromatin interactions	287
A genomic foundation model for ALT detection reveals new therapeutic vulnerabilities in pediatric cancer lacking actionable drug targets	288
FLT3 regulates telomere maintenance and suppresses senescence in acute myeloid leukemia.....	288
Determinants of chromosome-specific telomere lengths estimated from long reads among 2,573 All of Us participants.....	289
Session 92: Pathogenic Mechanisms in Neurological Disease: Gene to Network and Everything In Between	291
Dissecting MECP2 cis-regulatory architecture with massively parallel reporter assays to assess its contribution to male-biased autism	291
CYP2D6 contributes to cognitive dysfunction and promotes excessive synapse loss ..	292
Synergistic Loss of Regulatory and Protein-Coding Genes Drives Neuronal and Mitochondrial Dysregulation in Prader-Willi Syndrome	293
Transcriptomic Profiling of Extracellular Vesicles from Autism Patient-Derived Forebrain Organoids Reveals Novel Signatures	294
Session 93: The Metabolic Matrix: Untangling Complex Conditions	296
Molecular Risk Scores Enhance Type 2 Diabetes Prediction Across Clinical Contexts .	296

Male-specific genetic regulation across germline and somatic variation on the Y chromosome contributes to type 2 diabetes	297
Signatures of positive selection in Pacific-ancestry genomes reveal adaptive signals relevant to metabolic diseases	298
Phenome-wide Genetic Colocalization and Mendelian Randomization of eQTLs and pQTLs against complex traits in up to 2.5 million participants across multiple Biobanks	299

Session 19: Alz in the Details: Piecing Together the Alzheimer's Puzzle

Location: Ballroom West/Level 3, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 10:45am – 11:45am

Single-cell-resolved phenotypes for Alzheimer's disease across 599 individuals and 3.4 million cells

Subsession Time: Wednesday, October 15 at 10:45am – 11:00am

Authors: Riley J Mangan (MIT, Broad Institute, HMS), Gerard Boulad (MIT, Broad Institute, Delft University of Technology, Vrije Universiteit Amsterdam), Ahmed Mahfouz (Delft University of Technology, Leiden University Medical Centre), Marcel JT Reinders (Delft University of Technology, Leiden University Medical Centre), Manolis Kellis (MIT, Broad Institute).

Abstract: Single-cell transcriptomics can identify disease driver genes and biomarkers with high resolution, revealing heterogeneity both between and within individual donors. However, current differential expression methods typically ignore within-individual variation and focus on binary case/control phenotypes, limiting the power of single-cell data. We hypothesize that modeling individuals as mixtures of healthy and disease state cells, rather than treating disease as a binary label, can uncover more nuanced molecular variation and increase discovery power. Here, we leverage large-scale disease cohorts to develop Cell-Projected Phenotypes (CPP), a method that maps multi-donor phenotypes to transcriptional states at single-cell resolution, capturing gradients of phenotypic variation rather than traditional binary 'bulk' phenotypes. CPP links each cell to multiple transcriptional neighborhoods and corresponding phenotypic enrichments based on donor variables, capturing transcriptional and disease variation both within and between individuals. We applied CPP to scRNA profiles of the prefrontal cortex, encompassing 3.4 million cells from 599 postmortem donors with and without Alzheimer's disease (AD), and found that disease manifestation varies greatly across and within cell types among individuals. By refining coarse case/control labels into distributions of disease transcriptional states, CPP greatly increases power to detect differentially expressed genes and enriches pathway discovery. Key findings include potential compensatory mechanisms in AD such as neurotransmitter switching, elevated glucose and cortisol levels in excitatory neurons potentially linked to hyperexcitability, and a neuroinflammation-regulating process. Finally, CPP links specific phenotypes to transcriptional hallmarks and reveals patient subtypes with distinct transcriptional, clinical, pathological, and cellular signatures. Overall, CPP offers a general framework for analyzing single-cell datasets applicable to molecular manifestations of donor-level traits and diseases.

Spatial transcriptomic of hippocampal region-specific pathology in Primary Age-Related Tauopathy and Alzheimer's Disease

Subsession Time: Wednesday, October 15 at 11:00am – 11:15am

Authors: Yun Gong (Tulane University), Qilei Zhang (Central South University), Di Wu (Tulane University), Anqi Liu (Tulane University), Tianying Li (Central South University), Zhengwu Xiao (Central South University), Kuan-Jui Su (Tulane University), Xiaoxin Yan (Central South University), Hui Shen (Tulane University), Hongmei Xiao (Central South University), Hong-Wen Deng (Tulane University)

Abstract: While both Primary Age-Related Tauopathy (PART) and Alzheimer's Disease (AD) are characterized by the accumulation of hyperphosphorylated tau (pTau)-positive neurofibrillary tangles (NFTs) in the human hippocampus, the absence of β -amyloid (A β) deposition in PART is associated with potentially milder cognitive effects compared to AD. To clarify the mechanisms underlying the divergence between PART and AD, we constructed the first transcriptome-wide, high-resolution atlas of the human hippocampus from six subjects across control, PART, and AD groups, enabling the delineation of cellular and molecular dynamics linked to disease progression. Our analysis revealed that while the morphology of NFTs in PART and AD is similar, the mechanisms driving tau phosphorylation differ. In PART, it is primarily associated with DNA damage repair processes related to aging, whereas in AD, tau phosphorylation is influenced by A β deposition. Furthermore, excitatory neuron interactions across different hippocampal regions were downregulated in AD due to neuronal degradation, whereas communication related to apoptosis between glial cells and large vessels was upregulated in AD compared to control and PART groups. Additionally, we identified a P53 signalling-related gene, TP53INP2, that was uniquely upregulated in astrocytes near large vessels in AD. This suggests a potential mechanism of vessel-induced neuronal apoptosis under pathological stress in AD, a feature absent in PART. In summary, our study provides new insights into the divergent mechanisms of NFT formation between PART and AD and highlights vascular-associated neuronal degradation pathways in AD that are not observed in PART. These findings enhance our understanding of AD pathogenesis and its distinction from PART, with potential implications for targeted therapeutic strategies.

A genome-wide epitranscriptomic regulatory landscape in the aged human brains with Alzheimer's disease

Subsession Time: Wednesday, October 15 at 11:15am – 11:30am

Authors: Amit Kumar Gupta (Cleveland Clinic Genome Center, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA, Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA), Andrew A. Pieper (Department of Psychiatry, Case Western Reserve University, Cleveland, OH, USA, Brain Health Medicines Center, Harrington Discovery Institute, University Hospitals Cleveland Medical Center, Cleveland, OH, USA, Geriatric Psychiatry, GRECC, Louis Stokes Cleveland VA Medical Center; Cleveland, OH, USA, Institute for Transformative Molecular Medicine, School of Medicine, Case Western Reserve University, Cleveland, OH, USA, Department of Neurosciences, Case Western Reserve University, School of Medicine, Cleveland, OH, USA), Andrew J. Saykin (Indiana Alzheimer's Disease Research Center, Indiana University School of Medicine, Indianapolis, IN, USA, Department of Radiology and Imaging Sciences, Center for Neuroimaging, School of Medicine, Indiana University School of Medicine, Indianapolis, IN, USA, Department of Medical and Molecular Genetics, Indiana University

School of Medicine, Indianapolis, IN, USA, Stark Neurosciences Research Institute, Indiana University School of Medicine, Indianapolis, IN, USA), Feixiong Cheng (Cleveland Clinic Genome Center, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA, Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA, Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH, USA)

Abstract: RNA editing is a post-transcriptional modification comprising insertion, deletion, and base substitution within different types of RNA molecules. It plays a vital role in transcriptomic diversity, RNA stability, localization, splicing, abundance, structure, and translation implicated in neuronal dynamics, regulations and immune defense for human complex diseases, such as Alzheimer's disease (AD). Here, we deduce comprehensive genome-wide RNA editing atlas from nine brain regions comprising 4,208 samples (with 1,364 AD and 742 healthy subjects) from large scale biobanks, i.e., ROSMAP, MSBB and Mayo. We have employed transcriptome, matched whole genome sequencing (WGS, $n=3,627$), and genome-wide association study (GWAS) summary statistics data to determine key RNA editing loci, key clinical and genetic variables (disease, sex and apoe4 status), and regulatory implications in AD while adjusted for age, postmortem interval (PMI), sex, and apoe4 covariates. Later, we investigate brain-wide cis-regulatory variants (RNA editing quantitative trait loci or cis-edQTLs) and colocalization has been performed to inspect likely causal and genomic regulatory relationships associated with AD. We emphasize 127 key RNA edited genes including SYT11, KCNIP4, NRG3, ANKS1B, and RALYL highlighting role in synaptic plasticity, signaling and transmission, neuronal development and morphogenesis. Further, brain-wide tissue-specific edQTLs colocalized with AD GWAS variants exhibit likely causal significant RNA editing loci pertaining to 48 genes including CLU, BIN1, GRIN3B, NYAP1, DGKQ, CLPTM1, AGFG2, etc. Several core RNA edited gene locus include CLU (rs7982, adjusted p-value $[q] = 1.07 \times 10^{-48}$, $PP4 \geq 0.70$) and BIN1 (rs2276582, $q = 1.28 \times 10^{-36}$, $PP4 \geq 0.99$), associated with β -amyloid formation, lipid metabolism, tauopathy, and neurotransmission. Correspondingly, editing in GRIN3B (rs10417824) locus shows significant association ($q = 7.46 \times 10^{-40}$, $PP4 \geq 0.99$), a critical excitatory neurotransmitter in the brain. Similarly, NYAP1 (rs12539172) gene editing locus show substantial colocalization ($q = 6.68 \times 10^{-97}$, $PP4 \geq 0.99$), crucial for neuronal migration and morphogenesis. Similarly, an RNA editing locus (rs2734897, $q = 8.43 \times 10^{-110}$, $PP4 \geq 0.80$) in AGFG2, indicating a role in amyloid metabolism. In summary, these emerging functional and regulatory findings suggest likely implications of RNA editing signals in AD biology, proving to be potential biomarkers and therapeutic target for AD and other neurological diseases, if broadly applied.

Single-Cell Multi-Omic Profiling Reveals an Alzheimer's Disease-Associated Somatic Mutation Burden in the Human Brain

Subsession Time: Wednesday, October 15 at 11:30am – 11:45am

Authors: Swapnil Tichkule (Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, Department of Psychiatry, Icahn School of Medicine at Mount Sinai), Pengfei Dong (Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, Department of Psychiatry, Icahn School of Medicine at Mount Sinai), Clara Casey (Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, Department of Psychiatry, Icahn School of Medicine at Mount Sinai), Evelyn Hennigan

(Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, Department of Psychiatry, Icahn School of Medicine at Mount Sinai), Aram Hong (Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, Department of Psychiatry, Icahn School of Medicine at Mount Sinai), Marcela Alvia (Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, Department of Psychiatry, Icahn School of Medicine at Mount Sinai), Zhiping Shao (Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, Department of Psychiatry, Icahn School of Medicine at Mount Sinai), John F. Fullard (Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, Department of Psychiatry, Icahn School of Medicine at Mount Sinai), Gabriel E. Hoffman (Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, Department of Psychiatry, Icahn School of Medicine at Mount Sinai), Panos Roussos (Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, Department of Psychiatry, Icahn School of Medicine at Mount Sinai)

Abstract: Introduction: Alzheimer's disease (AD) is a progressive neurodegenerative disorder marked by neuronal loss and cognitive decline. While research has predominantly focused on inherited risk factors in AD, the role of somatic mutations (SMs) remains largely unexplored. Shaped by intrinsic cellular processes and lifelong environmental exposures, SMs are post-zygotic genetic alterations that accumulate over the lifespan. While they may contribute to the cellular heterogeneity and selective vulnerability observed in the AD brain, their distribution across brain regions and cell types; especially within functionally active genomic regions, remains largely overlooked. By investigating the spatial and cell-type-specific landscape of SMs, this study aimed to uncover novel mechanisms of neurodegeneration and inform more precise diagnostic and therapeutic strategies. Objective: With the goal of uncovering their potential role in disease, we sought to comprehensively characterize the burden, spatial distribution, and functional impact of SMs across distinct cell types and AD relevant brain regions. Methods: We analyzed multiomic single-nucleus data combining RNA and ATAC sequencing from 135 AD cases and 111 controls, and replicated our findings in an independent dataset from the PsychAD consortium, spanning six major neuronal and glial cell types across three brain regions (the superior temporal, entorhinal, and prefrontal cortices). To increase the likelihood of identifying functional variants, we focused on SMs located in accessible and transcriptionally active regions of the genome. SM rates were quantified using generalized linear models, mutational signatures were characterized to infer underlying processes, and SM-enriched genomic hotspots were mapped to evaluate potential functional consequences. Results: Our findings reveal an elevated burden of SMs across all examined brain regions and cell types in AD cases compared to controls, driven by AD-specific, age-related impairments in DNA mismatch repair. While regional differences were not significant, glial cells showed a markedly higher SM burden than neurons. The identified mutations preferentially impact genes involved in synaptic function, cellular homeostasis, and protein clearance. Moreover, SMs also influence differentially expressed genes in a cell-type-specific manner, underscoring the context-dependent effects of SMs in shaping the molecular landscape of AD pathogenesis. Conclusion: Our findings illuminate the functional role of SMs in AD, revealing their diverse patterns across brain regions, cell types, and genomic landscapes. This work advances our understanding of the molecular underpinnings of neurodegeneration and offers new insights into the pathogenic mechanisms driving AD.

Session 20: Evolution of the Coding and Non-Coding Genome

Location: Room 210C/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 10:45am – 11:45am

Expanded models of gene constraint improve selection estimates from multi-ancestry population data

Subsession Time: Wednesday, October 15 at 10:45am – 11:00am

Authors: Prathitha Kar (Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA), Mikhail A Moldovan (Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA), Evan Koch (Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA), Vladimir Seplyarskiy (Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA), Jeremy Guez (Broad Institute of MIT and Harvard, Cambridge, MA, USA), Konrad J. Karczewski (Broad Institute of MIT and Harvard, Cambridge, MA, USA), Shamil Sunyaev (Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA)

Abstract: Measurements of gene constraint are fundamental to understanding the roles of genes in diseases. A wide range of population-genetics methods have been developed to estimate the selection acting against heterozygous loss of function (LoF) variants of different genes (shet). Methods measuring constraint can use different aspects of the distribution of loss-of-function variants in the sample ranging from the presence/absence of segregating variants in functional sites (eg., LOEUF) to models fitting the entire site frequency spectrum (SFS). Recent theoretical advancements have made it possible to analytically estimate the SFS of rare variants for arbitrary demography while allowing for recurrent mutations. Leveraging this, we recalibrated demographic models for all ancestry labels in the data allowing us to use polymorphisms stratified by ancestry. Applying the demography-based framework to the largest exome sequencing dataset collected to date (gnomAD v4, 1.46 million haploid exomes, 6 major ancestry groups), we provide per-gene estimates of shet. This approach also incorporates LoF missannotations, benefiting from LoF variants present at high frequencies at seemingly constrained genes. We estimate ~3% of stop gain LoFs to be labeled incorrectly with this fraction varying across genes. Obtained shet estimates outperform existing constraint scores on all tested benchmarks, including a list of validated genes implicated in severe pediatric conditions (Area under Precision-Recall curve for our method = 0.1961, GeneBayes = 0.157). We then combine large language models trained on phylogenetic data with our population genetics approach to obtain the distribution of fitness effects for missense variants of each gene. These estimates are obtained completely independently of the LoF shet values allowing us to study the relationship between fitness effects of missense and LoF mutations across genes. Finally, we extend our framework to identify genes that have distorted SFS due to elevated mutation rates at LoF sites. Strikingly, all of the top 10 genes that we predict to be hypermutable at LoF sites have been recently described as drivers of clonal expansions in spermatogonia. This advanced demography-based methodology transpires into improved constraint estimates, facilitates comparison of

the effects of missense mutations and LoFs, and even reveals genes under positive selection in germline tissues.

The cumulative effect of protein-truncating variants on human life expectancy and overall health

Subsession Time: Wednesday, October 15 at 11:00am – 11:15am

Authors: Ha My Vy (Icahn School of Medicine at Mount Sinai), Daniel Balick (Icahn School of Medicine at Mount Sinai, Brigham & Women's Hospital), Daniel Jordan (Icahn School of Medicine at Mount Sinai), Ron Do (Icahn School of Medicine at Mount Sinai)

Abstract: Genetic variation is a key determinant of individual health and life expectancy. While substantial progress has been made in understanding the functional effects of specific genetic variants, the cumulative impact of deleterious variants across the genome on overall health remains poorly characterized. In our previous work, we demonstrated that the aggregated effect of protein-coding variants exhibits pleiotropic effects on a wide range of medical phenotypes. However, these analyses were constrained by the limited representation of rare variants in imputed data, thereby excluding potentially high-impact variants.

In this study, we focus our analyses on protein-truncating variants (PTVs). Previous studies have indicated that a higher number of PTVs in an individual's genome is associated with reduced lifespan and health span. We extended existing methodologies by incorporating weighting parameters to account for heterogeneity in the effect sizes of PTVs. We extended their methodology by introducing weighting parameters to account for heterogeneity in the effect sizes of PTVs. Specifically, we adapted the formula proposed by Gardner et al. (2022) to estimate the aggregated impact of PTVs in an individual's genome: , where $\text{shet}[i,g]$ represents the pre-computed selection coefficient for gene g . We further stratified our calculations by minor allele frequency (MAF) cutoffs to assess the relationship between PTVs' frequency and their functional impact. As phenotypic outcomes, we estimated disability-adjusted life years (DALYs) using electronic health records (EHRs) as a proxy for disease burden and used lifespan to test for mortality. These analyses were conducted using exome sequencing data from 366,731 individuals in the UK Biobank (UKBB) and 383,357 individuals from the All of Us Research Program. Association testing was performed within genetic ancestry groups, followed by meta-analysis. We observed strong associations when restricting to PTVs with $\text{MAF} < 10^{-5}$, whereas no significant associations were found for PTVs with $\text{MAF} > 10^{-5}$. Notably, incorporating shet weights significantly improved the magnitude of associations. The PTV load score ($\text{MAF} < 10^{-5}$) showed robust associations with both lifespan ($p = 3.93 \times 10^{-36}$) and DALYs ($p = 1.69 \times 10^{-30}$). Among the two components of DALY, the PTV load score was more strongly associated with years lived with disability (YLD; $p = 1.49 \times 10^{-114}$) than with years of life lost (YLL; $p = 3.62 \times 10^{-23}$). Individuals in the top 5 percentiles of the PTV load distribution, who carried, on average, two additional PTVs, experienced a 1.3-fold increased mortality rate and an approximate 0.18 increase in DALYs. Additionally, the PTV load score was significantly associated with multiple independent diseases, and these associations remained robust after adjusting for disease-specific polygenic risk scores. Our findings provide new insights into the genetic architecture of lifespan and disease burden and highlight

the PTV load score as an independent genetic risk factor for mortality and overall health, beyond polygenic risk.

Centromere-spanning haplotypes reveal deep ancestry and evolutionary dynamics in the human genome

Subsession Time: Wednesday, October 15 at 11:15am – 11:30am

Authors: Karen H. Miga (UC Santa Cruz Genomics Institute, University of California, Santa Cruz), Hailey Loucks (UC Santa Cruz Genomics Institute, University of California, Santa Cruz), Sasha Langley (Department of Evolution and Ecology, University of California Davis, Davis, CA) Fedor Ryabov (International Laboratory of Bioinformatics, HSE University, Moscow, Russia), Julian K. Lucas (UC Santa Cruz Genomics Institute, University of California, Santa Cruz), Julian Menendez (UC Santa Cruz Genomics Institute, University of California, Santa Cruz), Human Pangenome Reference Consortium (HPRC), Ivan A. Alexandrov (Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel), Charles Langley (Department of Evolution and Ecology, University of California Davis, Davis, CA)

Abstract: Advances in long-read sequencing and assembly now enable the generation of complete telomere-to-telomere (T2T) diploid human genomes, providing access to previously unresolved regions. A substantial fraction of newly resolved sequences lies within constitutive heterochromatin, which suppresses meiotic recombination and results in extended linkage disequilibrium (LD) across centromeres. These regions, defined as centromere-spanning haplotypes (cenhaps), are largely defined by large, satellite DNA arrays, flanking segmental duplications, and contain hundreds of genes in centromere-proximal regions, many previously associated with human traits. Due to reduced recombination, cenhaps preserve long-range haplotypes that persist across deep evolutionary timescales. Population-scale comparisons indicate that many cenhaps coalesce over a million years ago, with a subset showing evidence of Neanderthal and Denisovan introgression. These patterns reflect long-term haplotype persistence shaped by low recombination, drift, and selection. Here, we analyzed cenhap structure and diversity across 470 phased human haplotypes generated by the Human Pangenome Reference Consortium (HPRC), identifying thousands of cenhaps using fully resolved diploid assemblies. These long-range haplotypes are strongly associated with specific centromeric satellite array structures. Leveraging both alignment-based and k-mer-based approaches, we establish a cenhap-informed framework for centromere annotation and functional genomic analysis. This includes applications to genome-wide association studies (GWAS), enabling improved variant detection and interpretation in pericentromeric regions. In doing so, we identify rare recombination events within peri/centromeric regions and estimate local mutation and repeat turnover rates. Selection signals are detected in specific haplotype backgrounds, and gene content within archaic cenhaps includes loci previously implicated in disease. These results demonstrate that cenhaps offer a tractable system for evolutionary inference, capturing ancestral structure, mutational dynamics, and long-range haplotype persistence, many representing deep lineages that are common in extant human populations.

Co-evolution of lactase persistence and glucose homeostasis in ethnically diverse Africans

Subsession Time: Wednesday, October 15 at 11:30am – 11:45am

Authors: Michael McQuilan (University of Pennsylvania), Anya Jayanthi (University of Pennsylvania), Alessia Ranciaro (Sapienza University of Rome), Yuanqing Feng (University of Pennsylvania), Jasmin Rees (University of Pennsylvania), Daniel Harris (University of Pennsylvania), William Beggs (University of Pennsylvania), Daniel Hui (University of Pennsylvania), Matthew E. B. Hansen (University of Pennsylvania), Dawit Wolde Meskel (Addis Ababa University), Gurja Belay (Addis Ababa University), Thomas Nyambo (Kairuki University), Sununguko Wata Mpoloka (University of Botswana), Gaonyadiwe George Mokone (University of Botswana), Charles Fokunang (University of Yaoundé), Alfred K. Njamnshi (University of Yaoundé), Sarah A. Tishkoff (University of Pennsylvania), NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium.

Abstract: Lactase persistence (LP)—the continued ability to metabolize lactose into glucose and galactose as adults—varies widely across human populations and is a classic example of recent human adaptation. In Africa, LP is most common in populations practicing pastoralism, suggesting dairy consumption may have driven selection at the lactase (LCT) locus. However, the genetic basis and evolutionary history of LP in African populations remains incompletely understood, with previously known LP-associated variants explaining no more than ~45% of the trait variation. Here, we perform lactose tolerance tests (LTTs) for over 1,200 ethnically and geographically diverse Africans representing a range of subsistence strategies and integrate these results with genome-wide genotype data. Study participants include >100 self-identified ethnic groups representing all four major African language families (Niger-Congo, Afroasiatic, Nilo-Saharan, Khoisan) derived from Ethiopia, Tanzania, Kenya, Cameroon, Botswana, and Sudan. Using GWAS on whole-genome sequence data (WGS), we confirm that known LP-associated variants in the MCM6 enhancer region upstream of LCT are strongly associated with the rise in blood glucose following lactose ingestion. We also identify several additional independent genome-wide significant ($p < 5 \times 10^{-8}$) loci both near the LCT/MCM6 locus and elsewhere in the genome. Notably, these novel loci are located in or near genes with known roles in insulin signaling (e.g. FBN3, CCL25, PAQR3, CSMD1, CERS4), pancreatic beta cell function (e.g. PDIA4), and diabetes (e.g. RASGRP1, ZRANB3). Overlap with publicly available open chromatin regions and expression quantitative trait loci (eQTL) datasets indicate potential gene-regulatory functions for many of these loci. These findings suggest that lactose digestion and glucose homeostasis may have co-evolved, highlighting a more complex evolutionary history at the LCT locus and its relationship to the emergence of pastoralism in Africa. This work expands our understanding of diet-related adaptation and its implications for cardiometabolic health, particularly in populations historically underrepresented in genetic research. This work is funded by NIH grant R35 GM134957-01 and the American Diabetes Association Pathway to Stop Diabetes grant #1-19-VSN-02.

Session 21: From Identification to Consequence: Structural Variant Detection, Mapping, and Effects

Location: Room 253ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 10:45am – 11:45am

Marker-based tumor-normal haplotypes matching for accurate genome-wide recombination identification and haplotype specific somatic structural variation discovery using GIAB tumor-normal assemblies

Subsession Time: Wednesday, October 15 at 10:45am – 11:00am

Authors: Chunlin Xiao (NCBI/NLM/NIH), Justin Wagner (Material Measurement Laboratory, National Institute of Standards and Technology), Jennifer McDaniel (Material Measurement Laboratory, National Institute of Standards and Technology), Françoise Thibaud-Nissen (NCBI/NLM/NIH), and Justin M. Zook (Material Measurement Laboratory, National Institute of Standards and Technology)

Abstract: Accurate identification of somatic mutations is critical to early diagnosis of cancers and optimal treatments to cancer patients. Genome-in-a-Bottle (GIAB) consortium has publicly released a wide variety of sequencing data, including multiple short- and long-read technologies, Hi-C, and optical mapping for a new broadly-consented tumor-normal paired samples derived from the same individual (HG008), aiming for development of high-quality somatic mutation benchmarks to advance the development of sequencing technologies and analytical methods. Leveraging the latest advancements in long read technologies and assembly algorithms, we have used the above data to generate high-quality chromosomal-scale haplotype-resolved assemblies for both normal tissue and tumor cell lines, thus enabling the direct comparison of the tumor haplotype genome to its corresponding normal haplotype genome for accurate somatic mutation detection. Here we report that we implemented an integrated analysis workflow for accurate detection of haplotype-specific somatic structural variations in paired tumor-normal samples by adopting genetic marker identification, haplotype assembly-to-haplotype assembly mapping (SyRI, svim-asm, PAV) and read-alignment approaches (Severus, Sniffles2, GRIDSS2 etc). We first identified high-quality haplotype-specific genetic markers so that each of the chromosomes from two sets of haplotype assemblies in the tumor sample could be matched with their corresponding haplotype chromosomes in the normal sample. These genetic markers also allowed us to explore and identify genome-wide recombination events in this cancer sample, such as a complex series of inverted duplications on chr19 that is attached to chr22 on one end and to the opposite haplotype of chr19 on the other end. We then generated two comprehensive somatic SV sets by incorporating multiple lines of evidence from matched tumor-normal haplotypes and long/short read mappings (including PacBio HiFi, ONT, and Illumina), when two haplotype-resolved assemblies from normal sample were used as reference, respectively. We curated each of the selected SVs using IGV, and compared them with GIAB's curated draft somatic SV benchmark (v0.3, GRCh38-based). While most somatic SVs were identical between the two callsets, our assembly-based approach was complementary to the mapping-based

benchmark in certain repetitive and complex regions. Finally, we assigned each of the somatic SVs to its correct haplotype and generated final haplotype-specific somatic structural variation callset. While we continue to improve the haplotype-based somatic SVs, these analyses are already providing more insights into the ongoing refinement of high-quality somatic SV benchmarks for this tumor-normal pair.

Personalized Mapping of Allele-Specific Chromatin Accessibility at Single-Cell Resolution

Subsession Time: Wednesday, October 15 at 11:00am – 11:15am

Authors: Yuhang Chen (Computational Biology and Bioinformatics, Yale University), Matthew Jensen (Dept of Molecular Biochemistry and Biophysics, Yale University), Mark Gerstein (Dept of Molecular Biochemistry and Biophysics, Yale University)

Abstract: Understanding how genetic variation shapes chromatin accessibility is central to uncovering the cis-regulatory mechanisms underlying gene expression and disease. However, conventional analyses using a single reference genome fail to capture individual-specific and allele-specific regulatory features, limiting the resolution of variant interpretation. Here, we present a novel framework for mapping single-cell ATAC-seq (scATAC-seq) data to haplotype-resolved personal genomes, enabling systematic identification of allele-specific peaks (ASPs) across tissues, individuals, and cell types.

Leveraging the EN-TEEx dataset—including phased genomes and scATAC-seq from four donors across 30 human tissues—we aligned reads to each individual's diploid genome and applied a customized pipeline combining ArchR and AlleleSeq2 for peak calling and allelic imbalance analysis. Across 98 annotated cell types, we identified on average ~13,000 ASPs per cell type, with hundreds of peaks missed by traditional reference-based approaches. These "haplo-novel" peaks often represent regulatory elements revealed only through personal genome alignment.

We found that ASPs are significantly enriched in promoters, enhancers, and transcription factor binding motifs, and show strong correspondence with known eQTLs and allele-specific gene expression patterns. To further explore the sources of allelic imbalance, we also intersected ASPs with structural variant annotations from matched personal genomes, identifying regulatory elements that may be shaped by larger-scale genetic variation.

To our knowledge, this is the first large-scale integration of scATAC-seq data with fully phased personal genomes, offering a cell-type- and allele-resolved view of human chromatin accessibility. By moving beyond the limitations of reference genomes, our approach expands the resolution of functional genomics to the level of individual genetic backgrounds, advancing efforts in personalized regulatory annotation and variant interpretation.

Bridging long- and short-read sequencing to accelerate population scale structural variant association studies

Subsession Time: Wednesday, October 15 at 11:15am – 11:30am

Authors: Xuefang Zhao (Massachusetts General Hospital, Broad Institute of MIT and Harvard, Harvard Medical School), Samuel K. Lee (Broad Institute of MIT and Harvard), Daniel Ben-Isvy (Massachusetts General Hospital, Broad Institute of MIT and Harvard, Harvard Medical School), Wayne E. Clarke (New York Genome Center, Outlier Informatics Inc.), Jana Ebler (Heinrich Heine University), Qiuhui Li (Johns Hopkins University), Jan Korbel (European Molecular Biology Laboratory (EMBL)), Charles Lee (The Jackson Laboratory for Genomic Medicine), Evan E. Eichler (University of Washington School of Medicine, University of Washington), Michael C. Schatz (Johns Hopkins University), Kiran Garimella (Broad Institute of MIT and Harvard), Harrison Brand (Massachusetts General Hospital, Broad Institute of MIT and Harvard, Harvard Medical School), Tobias Marschall (Heinrich Heine University), Michael C. Zody (New York Genome Center), Michael E. Talkowski (Massachusetts General Hospital, Broad Institute of MIT and Harvard, Harvard Medical School)

Abstract: Structural variants (SVs) shape the genome more profoundly than single nucleotide changes and influence phenotypes ranging from rare congenital anomalies to common cardiometabolic disease. Yet, genome-wide association studies (GWAS) of SVs remain constrained by two major challenges: (1) Virtually all existing population-scale biobanks and trait association studies involve microarray, exome, or short-read genome sequencing (srGS), technologies that can detect only ~12,000 SVs per genome, whereas long-read genome sequencing (lrGS) can resolve > 25,000 SVs per genome. (2) Application of srGS presents significant blind spots for variant discovery, particularly in highly mutable segmental duplications and simple repeat sequences, as well as bias against insertion events, a limitation exacerbated by the merging of imprecise SV breakpoints across large cohorts. While lrGS offers superior SV discovery power with the ability to assemble long haplotypes and precisely map SV breakpoints, its sequencing cost, DNA requirement, and compute burden remain prohibitive for population cohorts.

We have optimized a scalable pipeline on the Terra cloud platform that transfers lrGS discovery power into any srGS cohort via joint genotyping and imputation. The open-source workflow integrates three complementary algorithms—KAGE, PanGenie, and GLIMPSE—to generate high quality, phased SV calls at biobank scale.

A reference panel of ~31 million (M) variant alleles was assembled from 1,074 PacBio HiFi genomes in the All of Us (AoU) and Human Pangenome Reference Consortium (HPRC) projects. Among those, 308,690 SVs have a minor allele frequency (MAF) of $\geq 0.5\%$ and are estimated to capture 99.4% of common SVs (MAF $\geq 0.5\%$). Notably, 51.4% ($n = 158,579$) of these common SVs were absent from a srGS SV catalog of over 97,000 srGS genomes in AoU. Validation with 201 African trios from the 1000 Genomes Project demonstrated a Mendelian violation rate of 8.5% and a cost of just \$0.70 for genotyping and imputing 31M variants in each genome—orders of magnitude more cost-effective than lrGS.

Using this approach, we are now comprehensively genotyping and imputing SVs across the NHGRI Centers for Common Disease Genomics (CCDG) and the AoU research program. This effort will empower SV-GWAS at unprecedented scale and enable routine interrogation of complex variation in studies of both rare and common diseases.

★ Structural and rare variants underlying psoriasis susceptibility identified by whole-genome sequencing: discovery of CERCAM as a novel risk gene

Subsession Time: Wednesday, October 15 at 11:30am – 11:45am

Authors: Kyuto Sonehara (University of Tokyo, Osaka University, RIKEN Center for Integrative Medical Sciences), Rei Watanabe (Osaka University, Juntendo University), Yutaka Matsumura (Osaka University), Yuichi Mitsui (Science Tokyo), Yosuke Ogawa (University of Tokyo), Kaori Odomari (Osaka University), Saori Sakaue (Osaka University, Harvard Medical School), Shinichi Namba (University of Tokyo, Osaka University, RIKEN Center for Integrative Medical Sciences), Tomomitsu Hirota (Jikei University), Noriko Arase (Osaka University), Yuumi Nakamura (Osaka University), Kimiko Nakajima (Kochi University), Kenichi Yamamoto (Osaka University), Ken Suzuki (University of Tokyo, Osaka University), Toshihiro Kishikawa (Osaka University), Ryuya Edahiro (Osaka University, RIKEN Center for Integrative Medical Sciences), Yuya Shirai (Osaka University), Tatsuhiko Naito (Osaka University, RIKEN Center for Integrative Medical Sciences), Noah Sasa (Osaka University, RIKEN Center for Integrative Medical Sciences), Koichi Matsuda (University of Tokyo), Yayoi Tada (Teikyo University), Tatsuyoshi Kawamura (University of Yamanashi), Shinji Shimada (University of Yamanashi), Emiko Noguchi (University of Tsukuba), Shigetoshi Sano (Kochi University), Mayumi Tamari (Jikei University), Takashi Satoh (Science Tokyo), Manabu Fujimoto (Osaka University), Akimichi Morita (Nagoya City University), Yukinori Okada (University of Tokyo, Osaka University, RIKEN Center for Integrative Medical Sciences)

Abstract: Psoriasis vulgaris (PsV) is a polygenic, immune-mediated inflammatory skin disorder with approximately 75% heritability. Previous genome-wide association studies (GWAS) of PsV have predominantly targeted common small variants in Europeans, leading to gaps in understanding the contributions of other variant classes, particularly rare variants (RV; minor allele frequency <0.5%) and structural variants (SV; >50bp in size). These variants remain largely unexplored but can be effectively captured through whole-genome sequencing (WGS).

To address these gaps, we performed a comprehensive WGS study involving 1,415 PsV cases and 3,968 controls of Japanese ancestry in a nationwide collaborative effort. This study uniquely leveraged the Japanese population, where nearly 95% of PsV cases lack HLA-C*06:02, the strongest genetic driver of PsV that is prevalent in half of European PsV cases. This predominance of HLA-C*06:02-negative cases provided a valuable opportunity to identify novel genetic determinants. Our analysis identified 22,783 SVs, comprising 11,127 deletions, 7,324 insertions, 3,433 duplications, and 899 inversions. Notably, we discovered a 3.3kb non-coding deletion at the IFNLR1 locus exhibiting a genome-wide significant protective effect ($P=1.1 \times 10^{-9}$; $\beta=-0.40$). Statistical fine-mapping highlighted this deletion as a potential causal variant (PIP=0.32). Functional annotation indicated the deletion disrupts an epithelial tissue-specific enhancer linked to IFNLR1 and exerts a skin-specific downregulating eQTL effect on IFNLR1, supported by EpiMap and GTEx. Gene-based collapsing analyses for damaging RV identified CERCAM as a novel PsV risk gene ($P=4.1 \times 10^{-7}$; $\beta_{\text{burden}}=1.10$). We replicated this finding in UK Biobank ($n=343,764$). Moreover, an enrichment of damaging RVs in IL36RN ($P=1.2 \times 10^{-4}$; $\beta_{\text{burden}}=0.61$) was observed, primarily driven by known East Asian mutations (rs397514630 [p.Arg10X] and rs148755083 [p.Arg10ArgX1]) associated with generalized pustular psoriasis, a rare and severe multi-system

inflammatory disease. Finally, in-vivo validation using Cercam knockout ($Cercam^{-/-}$) mice demonstrated exacerbated inflammatory dermatitis in the imiquimod-induced psoriasis model, characterized by increased subepidermal T-cell infiltration and elevated PsV-associated cytokines.

Our comprehensive WGS study advances our understanding of the roles played by rare and structural variants in PsV, uncovering novel genetic contributors previously overlooked by conventional GWAS.

Session 22: From Interaction to Insight: Leveraging Gene-Environment and Multi-Omics Interactions Across Populations

Location: Room 210AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 10:45am – 11:45am

Novel method for scalable non-additive GWAS reveals 781 new loci across 2,329 traits in FinnGen

Subsession Time: Wednesday, October 15 at 10:45am – 11:00am

Authors: Ivan Molotkov (Nationwide Children's Hospital, The Ohio State University College of Medicine, Institute for Molecular Medicine Finland), Mitja Kurki (Institute for Molecular Medicine Finland, Broad Institute), FinnGen (Institute for Molecular Medicine Finland), Aarno Palotie (Institute for Molecular Medicine Finland), Mark J Daly (Institute for Molecular Medicine Finland, Broad Institute), Mykyta Artomov (Nationwide Children's Hospital, The Ohio State University College of Medicine, Institute for Molecular Medicine Finland)

Abstract: Drug candidates supported by genetic evidence are more likely to succeed in clinical trials, with genome-wide association studies (GWAS) providing a key source of such evidence. Standard GWAS approaches assume additive effects of alleles on the phenotype, but non-additive models have also successfully identified novel associations across various traits. Despite their potential, the large-scale application of non-additive GWAS across thousands of phenotypes in biobanks has been limited by high computational costs.

To address this challenge, we developed a cost-efficient strategy for non-additive GWAS by leveraging the observation that variants lacking any additive signal are unlikely to reach genome-wide significance in non-additive analyses. By identifying and excluding such variants from analysis, our method retains nearly all novel non-additive associations while reducing the computational cost by several orders of magnitude. Applying this approach to FinnGen data across 2,329 phenotypes and 500,329 individuals, we identified 781 novel non-additive associations that were not genome-wide significant in the additive model. The entire analysis cost was under \$40, compared with an approximately \$27,000 required for exhaustive non-additive GWAS of the same 2,329 phenotypes using the identical computing environment. Despite the >700-fold increase in cost efficiency, our approach still identified more than 90% of the true associations, as demonstrated by simulations and empirical analyses of 19 selected phenotypes. Overall, the proposed pipeline transforms non-additive analyses into a routine supplement to the standard additive GWAS, recovering hundreds of previously missed associations for a tiny fraction of computational cost.

We validated the new 781 non-additive loci using the Million Veteran Program (MVP) data, fine-mapped them to identify causal variants, and then linked the loci to genes by colocalizing our signals with >500 datasets. This integrative approach identified, for example, rs1788100 – a high-confidence causal variant

(PIP = 1.0) for autoimmune disorders whose signal colocalized with CD226 expression in immune cells, corroborating rare-variant and TWAS evidence that implicated CD226 in autoimmune conditions such as gout and lupus. We have made all results publicly available for follow-up research.

Context-specific polygenic scores improve prediction and portability in complex human traits

Subsession Time: Wednesday, October 15 at 11:00am – 11:15am

Authors: Renée Fonseca (Department of Human Genetics, University of Chicago, Chicago, IL, USA, Section of Genetic Medicine, University of Chicago, Chicago, IL, USA), Christa Caggiano (Institute for Genomic Health, Icahn School of Medicine, Mount Sinai, New York City, NY, USA) Manuela Costantino (Section of Genetic Medicine, University of Chicago, Chicago, IL, USA, Committee on Genetics, Genomics and Systems Biology, University of Chicago, Chicago, IL, USA), Ophelia Dominguez (Section of Genetic Medicine, University of Chicago, Chicago, IL, USA), Jerome Freudenberg (Department of Bioinformatics, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA), Eimear E. Kenny (Institute for Genomic Health, Icahn School of Medicine, Mount Sinai, New York City, NY, USA), and Andy Dahl (Section of Genetic Medicine, University of Chicago, Chicago, IL, USA)

Abstract: Polygenic scores (PGS) are a main deliverable of modern genetic studies and are poised to improve clinical care. However, genetic effects can vary based on an individual's context, such as their sex or treatments, which reduces PGS accuracy. To address this, we developed a novel PGS method called PGSC that exploits gene-by-context (GxC) interaction effects captured via GxC GWAS to improve prediction. Simulations establish that PGSC improves over PGS in a range of realistic GxC models. Conversely, a simpler method that we call ampPGS—which allows the ordinary PGS to have context-specific effect sizes—only outperforms PGS when genetic effects are highly correlated across contexts. We evaluated PGSC in unrelated white British individuals in the UK Biobank across 40 phenotypes. We first use sex as the context measure, as it is known to modify PGS effects, and find that PGSC outperforms PGS on average across phenotypes (average %R² gain: 7%) and for phenotypes with known sex-specificity, such as testosterone and BMI-adjusted waist-hip ratio. We then expand to study age and statins usage as context measures in PGSC and find similar results (average %R² gain = 4% & 6%, resp). As expected, different contexts are relevant to different phenotypes: for example, PGSC using age improved prediction for bilirubin and cystatin-C (23% & 16%, resp), and PGSC using statins improved prediction for cholesterol-related measures (11-29%). By contrast, the simpler ampPGS only outperforms PGS for sex as a context, and PGSC outperforms ampPGS across all contexts (average %R² gain = 4-6%). We investigate log-transformed phenotypes, because GxC effects may depend on scale, and find that PGSC continues to outperform PGS for most tested contexts and phenotypes whereas ampPGS no longer outperforms PGS for any context or phenotype. Finally, we replicate in external populations, which is crucial for establishing confidence in statistical GxC signals. We find that PGSC improves portability to African and Asian individuals in UK Biobank (average %R² gain = from 31-84%, resp across contexts). We tested replication for PGSC with sex as a context in BioMe and found PGSC significantly improved performance for 14/18 tested traits (average %R² improvement = 34%). Overall, PGSC is a straightforward

and computationally efficient method enabling practitioners to improve PGS performance and portability, improving accuracy and equity in clinical use of PGS.

Genetic Ancestry and Environmental Differences in Somatic Alterations and Clinical Outcomes for Five Common Cancers

Subsession Time: Wednesday, October 15 at 11:15am – 11:30am

Authors: Jiawei Tu (University of Texas Health Science Center at Houston), Kodi Taraszka (Dana-Farber Cancer Institute), Stefan Groha (GSK.ai), Chirag M. Lakhani (Knowles Lab), Alicia R. Martin (Massachusetts General Hospital), David C. Qian (MD Anderson Cancer Center), Chirag J. Patel (Harvard University), Alexander Gusev* (Dana-Farber Cancer Institute), Yixuan He* (University of Texas Health Science Center at Houston)

Abstract: While cancer disparities are well established, genomic differences and their impact on outcomes remain poorly understood. We leverage the largest multi-ancestry cancer cohort to quantify the influences of genetic ancestry on recurrent somatic alterations and survival.

We estimated the ancestry of 17721 individuals in Dana Farber Cancer Institute's Profile cohort using somatic sequencing data. Our analysis focused on five prevalent cancers in the cohort: colorectal (N=3806), breast (N=3372), glioma (N=3072), pancreatic (N=1819), and non-small cell lung cancer (NSCLC, N=5652). To characterize the ancestry-specific genomic landscape, we performed univariate associations between 1900 mutation phenotypes (somatic single-point mutations and copy number alterations [CNA] with >10 carriers) and each ancestry by cancer. We then conducted multivariable analyses to identify independent effects, additionally adjusting for socioeconomic status (SES) and air pollution to evaluate environmental influences. Lastly, we assessed the effect of ancestry-associated alterations on survival through individual and joint modeling.

All patients were assigned one of five ancestries: African (AFR, 3.5%), Admixed American (AMR, 5.8%), East Asian (EAS, 2.8%), European (EUR, 86.5%), and South Asian (SAS, 1.4%). Univariate analysis identified 80 significant (FDR adjusted $P < 0.05$) ancestry-associated alterations across cancers. Subsequent joint analysis revealed 8, 1, 10, 4, and 2 independent associations with AFR, AMR, EAS, EUR, and SAS, respectively; 6/25 involved clinically actionable genes. All associations remained robust after adjusting for SES and pollution. We replicated known findings, e.g. EGFR's association with EAS (65.3%, OR=4.72, $P < 0.001$) in NSCLC, and identified novel ones including AKT2 CNA with AFR (6.2%, OR=6.71, $P < 0.001$) and EXT1 CNA with EAS (18.2%, OR=4.87, $P < 0.001$) in breast cancer, and SMAD2 CNA with AMR (13.1%, OR=5.54, $P < 0.001$) in colorectal cancer. Ancestry was significantly ($P < 0.05$) associated with survival and modified prognostic associations of mutations. We plan to replicate our results in MD Anderson's cohort.

In conclusion, we conducted the largest multi-population analysis of genomic alterations and ancestry across cancers to date, expanding the study size beyond prior studies by several-fold. Our results highlight the impact of genetic ancestry on cancer incidence and progression, emphasizing the need for ancestry-aware approaches to precision oncology.

★ Incorporating UV exposure into functional genomic analyses identifies additional key cancer drivers and pathways at melanoma susceptibility loci

Subsession Time: Wednesday, October 15 at 11:30am – 11:45am

Authors: Rohit Thakur (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Mai Xu (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), G J M Shanika R Jayasinghe (Population Health, QIMR Berghofer Medical Research Institute), Jianxin Shi (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Diptavo Dutta (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Phuc H Hoang (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Mathias Seviiri (Population Health, QIMR Berghofer Medical Research Institute), Christopher I Amos (Epidemiology & Population Science, Baylor College of Medicine), Andrew Bakshi (Department of Epidemiology and Preventive Medicine, Monash University), Anne E Cust (The Daffodil Centre, Sydney School of Public Health, The University of Sydney), Florence Demenais (INSERM, Université Paris Diderot), David L Duffy (Genetic Epidemiology Laboratory, QIMR Berghofer Medical Research Institute), Lars G Fritsche (Department of Biostatistics, School of Public Health, University of Michigan), Jiali Han (IU Simon Cancer Center, Indiana University Indianapolis), Nicholas K Hayward (Oncogenomics, QIMR Berghofer Medical Research Institute), Kiarash Khosrotehrani (Frazer institute, Dermatology Research Centre, The University of Queensland), Rajiv Kumar (German Cancer Research Center), Serigne Lo (Melanoma Institute Australia), Stuart MacGregor (Statistical Genetics, Population Health, QIMR Berghofer Medical Research Institute), Miguel Renteria (QIMR Berghofer Medical Research Institute), Diane T Smelser (Department of Genomic Health, Geisinger Clinic, Geisinger Health System), Sarah V Ward (School of Population and Global Health, The University of Western Australia), Maria Concetta Fargnoli (Istituto Dermatologico San Gallicano - IRCCS), Paola Ghiorzo (IRCCS Ospedale Policlinico San Martino, University of Genoa), Alisa M Goldstein (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Georgina V Long (Melanoma Institute Australia, The University of Sydney, and Royal North Shore Hospital), Chiara Menin (Immunology and Molecular Oncology Unit, Veneto Institute of Oncology - IOV IRCCS), David Millan-Esteban (Department of Dermatology, Universidad Catolica de Valencia San Vicente Mártir; Fundación Instituto Valenciano de Oncología), Eduardo Nagore (Department of Dermatology, Fundación Instituto Valenciano de Oncología; Universidad Católica de Valencia San Vicente Mártir), Cristina Pellegrini (Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila), Susana Puig (Dermatology Department, Universitat de Barcelona), Alex Stratigos (Department of Dermatology-Venereology, National and Kapodistrian University of Athens), David C Whiteman (Cancer Control Group, Population Health, QIMR Berghofer Medical Research Institute), Melanoma Meta-Analysis Consortium, Mark M Iles (Leeds Institute for Data Analytics, University of Leeds), Rebecca I Hartman (Dermatology Section, VA Boston Healthcare System), Lee E Wheeles (Tennessee Valley Healthcare System VA Medical Center), Maria Teresa Landi (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Matthew H Law (Population Health, QIMR Berghofer Medical Research Institute), Kevin M Brown (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health)

Abstract: Cutaneous melanoma risk is highly heritable with the additional contribution of a well-established environmental exposure, ultraviolet radiation (UV). Melanoma heritability is enriched near genes specifically expressed by human melanocytes (Landi 2020). Using new H3K27ac ChIP time-course data incorporating in vitro UV exposure in melanocytes, we observed further enrichment of heritability in regulatory regions following UV exposure, suggesting a likely role of risk variants in UV-response. Based on these data, we generated melanocyte H3K27ac-HiChIP time-course data (no-UV and 8/24/72/336h post-UV) and applied them to identify potential causal genes and pathways from a new melanoma GWAS of 70,966 cases. We fine-mapped 116 loci and performed integrative variant-to-gene mapping using HiChIP data. Incorporating HiChIP data without UV exposure resulted in 506 genes nominated as candidates at 106 loci. Incorporation of post-UV data nominated an additional 80 genes as targets at 107 loci, yielding a small but substantial contribution to candidate gene nomination. While incorporation of post-UV timepoints showed similar enriched pathways as the no-UV timepoint, enrichment of key pathways involved in nevus and melanoma formation increased, notably MAPK signaling ($P=0.0002$ vs $P=0.02$) and PTEN regulation ($P=0.001$ vs $P=0.007$), suggesting a potential role for variation in post-UV oncogenic signaling in melanoma risk.

Notably, post-UV HiChIP data specifically identified long-range (>1Mb) enhancer to promoter interactions between fine-mapped variants and established UV-responsive cancer driver genes. For example, we observed a 1.8 Mb loop between a locus over TBL1XR1 and the established driver gene PIK3CA at 24h post UV exposure. A role for PIK3CA is further supported by a second locus with short range looping to PIK3CA at all timepoints. In addition, we observed a long-range (1.2 Mb) UV-specific interaction between a locus over HELZ to the established cancer driver gene PRKAR1A. Taken together with the simultaneous nomination of multiple PIK3CA/PTEN and PKA/MITF effector genes, respectively, these data suggest germline variation in PIK3 and cAMP/PKA signaling can mediate risk. Overall, these findings demonstrate that incorporating disease-relevant exposures (e.g. UV in melanoma) enhances the interpretation of disease risk loci. For melanoma, this nominated novel susceptibility genes and highlighted the convergence of genetic and environmental risk in melanomagenesis.

Session 23: It Is What It Isn't, Non-Canonical Drivers of Disease Gene Expression

Location: Room 205ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 10:45am – 11:45am

Genetic control of alternative transcriptional initiation reveals a pan-cancer atlas of susceptibility genes

Subsession Time: Wednesday, October 15 at 10:45am – 11:00am

Authors: Hui Chen(Shenzhen Bay Laboratory,Institute of Systems and Physical Biology), Xudong Zou(Shenzhen Bay Laboratory,Institute of Systems and Physical Biology), Wei Wang(Shenzhen Bay Laboratory,Institute of Cancer Research), Shuxin Chen(Shenzhen Bay Laboratory,Institute of Systems and Physical Biology), Chen Yu(Shenzhen Bay Laboratory,Institute of Cancer Research), Lei Li(Shenzhen Bay Laboratory,Institute of Systems and Physical Biology)

Abstract: Alternative transcriptional initiation (ATI) is a widespread gene regulatory mechanism, characterized by multiple transcription start sites (TSSs) for a single gene, generating functionally distinct isoforms. ATI plays crucial regulatory roles in mediating tissue-specific gene expression and contributes significantly to the dysregulated transcriptomes observed in cancer. However, the genetic architecture underlying ATI variation and its mechanistic links to cancer susceptibility remain largely unclear. Here, we present the first comprehensive characterization of the genetic regulation of ATI spanning 49 human normal tissues and 33 tumor tissues. We identified 9,075 5'UTR alternative transcription initiation trait loci (5'aQTLs), encompassing approximately 0.41 million common genetic variants associated with the usage of distal TSS of 5,436 genes, 32.1% of which were overlooked by eQTLs. We found that 7.6% of disease variants are colocalized with 5'aQTL signals, and 74.0% of them were overlooked by eQTLs. By integrating our 5'aQTL with well-powered GWAS datasets through transcriptome-wide association studies (TWAS), we identified 156 cancer susceptibility genes, including MAFF and MLLT10, as well as novel candidates. Collectively, our study reveals ATI as a critical mechanism linking non-coding variants to cancer risk, providing new insight to interpret the regulatory mechanisms of cancer risk loci.

Population-scale single-cell long-read sequencing reveals transcriptome variations and genetic effects on alternative splicing across lung cell types

Subsession Time: Wednesday, October 15 at 11:00am – 11:15am

Authors: Bolun Li (National Cancer Institute), Thong Luong (National Cancer Institute), Elelta Sisay (National Cancer Institute), Jinhu Yin (National Cancer Institute), Zixuan Eleanor Zhang (Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California), Ju Hye Shin (Department of Internal Medicine, Yonsei University College of Medicine), Jinyoung Byun (University of New Mexico Comprehensive Cancer Center), Yoon Soo Chang (Department of Internal

Medicine, Yonsei University College of Medicine), Maria Teresa Landi (National Cancer Institute), Nathaniel Rothman (National Cancer Institute), Erping Long (Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College), Qing Lan (National Cancer Institute), Christopher Amos (University of New Mexico Comprehensive Cancer Center), Tongwu Zhang (National Cancer Institute), Jianxin Shi (National Cancer Institute), Nicholas Mancuso (Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California), Haoyu Zhang (National Cancer Institute), Jin Gu Lee (Department of Thoracic and Cardiovascular Surgery, Yonsei University College of Medicine), Eun Young Kim (Department of Internal Medicine, Yonsei University College of Medicine), Jiyeon Choi (National Cancer Institute)

Abstract: Background: Full-length transcriptome atlas is essential to understand alternative splicing, especially in the context of cell-type-specific gene regulation and cancer development. Although splicing quantitative trait loci (sQTL) contributes to genome-wide association studies (GWAS) signals, sQTLs using bulk short-read sequencing fail to capture full-length and novel isoforms or cell type specific splicing events.

Methods: To uncover cellular transcriptome variations and genetic effects on isoform regulation, we performed PacBio single cell MAS-seq on tumor-distant normal lung tissues from 129 never-smoking Korean women. Isoform signatures of each lung cell type were identified by differential analysis. We mapped isoform-level QTLs (isoQTLs) using jaxQTL negative binomial model of pseudo-bulk counts. Colocalization and transcriptome-wide association study (TWAS) with lung cancer GWAS were performed to prioritize susceptibility isoforms.

Results: We identified 325,865 post-QC full-length isoforms (21,616 genes), where 83% is novel. We annotated 360,133 cells into 37 cell types, including 9 epithelial populations accounting for 45% of cells. Across lung cell types, 19,595 isoforms were differentially expressed with 67.2% being more specific than gene levels. Mapping isoQTLs within 33 abundant cell types, we identified 2,016 isoforms (1,562 genes) with 37.5% of lead isoQTL-isoform pairs specific to one cell type. Among these pairs, 31.64% were driven by cell type specific isoform expression. We identified 921 genes not detected in GTEx v8 bulk lung sQTL, of which 54.7% was cell type specific and 29.4% was for novel isoforms. Colocalization found that 46% of isoQTLs were independent from single-cell eQTLs in matching cell types. Moreover, isoQTLs were enriched in splicing relevant protein binding motifs and included those overlapping 4bp splice sites. Notably, integrating isoQTL and lung cancer GWAS signals identified 9 colocalized isoforms, including those independent from eQTL signals (e.g., PPIL6). TWAS of isoQTLs identified 74 lung cancer susceptibility isoforms, including 11 new loci not significant in GWAS, mainly in the epithelial group.

Conclusion: We established an isoform-level lung cell atlas and detected isoQTLs that are cell-type-specific and distinct from eQTLs. Integration with GWAS prioritized susceptibility isoforms relevant to lung cancer etiology. The resource will support future studies of lung cancer and lung disease mechanisms.

Improving isoform-level eQTL and integrative genetic analyses of breast cancer risk and mammographic density using long-read RNA transcript assemblies

Subsession Time: Wednesday, October 15 at 11:15am – 11:30am

Authors: Taylor Head (MD Anderson Cancer Center), Sean Bresnahan (MD Anderson Cancer Center), Tabitha Harrison (University of Washington), Sara Lindstroem (University of Washington), Arjun Bhattacharya (MD Anderson Cancer Center)

Abstract: Understanding the genetic architecture of gene regulation and alternative splicing is essential for uncovering the biological mechanisms underlying complex traits such as breast cancer and its intermediate phenotype, mammographic density (MD). To assess how transcript annotation affects expression quantification and downstream integrative analyses, we quantified gene- and isoform-level expression in breast tumor tissue (TCGA, N = 1078) using both the standard GENCODEv45 transcriptome and extended annotations incorporating long-read RNA-seq-derived assemblies. Using these transcript references, we performed expression quantitative trait locus (eQTL) mapping, followed by fine-mapping and colocalization analyses with GWAS loci. Across quantification strategies, we observed both shared and distinct eQTLs at gene and isoform levels. Of 47,873 isoforms detected, 17,273 (36%) were identified exclusively through the inclusion of long-read-derived transcripts. Fine-mapping results also demonstrated differences based on annotation strategy: 11,405 (38%) of 29,880 tumor isoforms with at least one high-confidence causal eQTL were unique to analyses using expanded isoform catalogs from long-read annotations. When integrating these findings with GWAS summary statistics for overall and subtype-specific breast cancer risk, as well as MD traits (percent MD, dense area, and non-dense area), we further found that transcript annotation choices substantially influenced colocalization outcomes. Notably, we observed colocalization of isoforms with four additional GWAS loci for MD phenotypes using long-read-informed annotations. This included isoforms in SGSM3, a locus previously implicated in breast cancer GWAS and transcriptomic studies, and a novel isoform in MRTFA, a gene at an established breast cancer risk locus on 22q13. These tumor-based analyses highlight the critical role of transcript-level resolution and choice of reference annotation in expression quantification, QTL discovery, and integrative analyses. Ongoing analyses in non-cancerous breast tissue will provide further insight into tissue-specific regulatory mechanisms. These findings provide methodological guidance for transcriptomic studies and reinforce the value of comprehensive isoform representation for understanding the splicing mechanisms driving breast cancer risk.

Topoisomerase IIb binding underlies driver alterations and frequently mutated regulatory elements in cancer genomes

Subsession Time: Wednesday, October 15 at 11:30am – 11:45am

Authors: Jüri Reimand (Ontario Institute for Cancer Research, University of Toronto), Christian A. Lee (Ontario Institute for Cancer Research, University of Toronto), Robin H. Oh (Lunenfeld-Tanenbaum Research Institute, University of Toronto), Zoe P. Klein (Ontario Institute for Cancer Research, University of Toronto), Nina Adler (Ontario Institute for Cancer Research, University of Toronto), Sana Akhtar Alvi (The Hospital for Sick Children), Ellen Langille (Lunenfeld-Tanenbaum Research Institute, University of Toronto), Elisa Pasini (University Health Network), Kevin C. L. Cheng (Ontario Institute for Cancer Research, University of Toronto), Diala Abd-Rabbo (Ontario Institute for Cancer Research), Huayun Hou

(The Hospital for Sick Children), Ricky Tsai, (Lunenfeld-Tanenbaum Research Institute), Mamatha Bhat (University Health Network, University of Toronto), Daniel Schramek (Lunenfeld-Tanenbaum Research Institute, University of Toronto), Michael D. Wilson (The Hospital for Sick Children, University of Toronto), Liis Uusküla-Reimand (The Hospital for Sick Children)

Abstract: Type-II topoisomerases resolve topological stress in DNA by inducing and repairing double-strand breaks in a coordinated fashion. TOP2A is highly expressed in proliferating cells and a target of common cancer chemotherapeutics such as doxorubicin. In contrast, TOP2B is ubiquitously expressed in many tissues and is also known a potential off-target of chemotherapies. Here, to understand the role of TOP2B in mutagenesis and cancer driver mechanisms, we generated unique DNA-binding maps of TOP2B, CTCF, and RAD21 in human liver cancer samples using chromatin immunoprecipitation sequencing (ChIP-seq). Next, we conducted a systematic analysis of these genome-wide maps for cancer driver mutations and somatic mutational processes across a dataset of 6500 whole cancer genomes representing 18 major cancer types. We show that TOP2B binds ubiquitously active promoter and enhancer elements that associate with higher gene expression and chromatin looping activity, many of which are evolutionarily conserved sites of TOP2B binding or show in vitro double strand break activity, connecting TOP2B binding with hallmarks of mutagenesis. Systematic mutational analyses indicate that TOP2B-CTCF-RAD21 and TOP2B-RAD21 sites are highly enriched in somatic small mutations (SNVs, indels) as well as structural variants (SVs), for which specific mutational processes such as APOBEC are implicated. When focusing on evolutionarily selected genes using a systematic driver analysis, TOP2B binding appears to underlie SVs and hotspot mutations in cancer-driving genes such as TP53, MYC, FOXA1, and VHL, and hundreds of cis-regulatory elements as candidate non-coding drivers. We show that the TOP2B-bound mutational hotspot at RMRP is a novel non-coding driver mutation that causes tumor initiation and growth in vivo and leads to transcriptional deregulation of migration and cell adhesion pathways. Together, these data highlight TOP2B as a protector of the genome from topological challenges whose aberrant activity promotes driver and passenger mutations in cancer genomes.

Session 24: Loop Logic: 3D Chromatin Architecture and Functional Genomics

Location: Room 206AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 10:45am – 11:45am

3D genomic approaches mapping accessible chromatin and architecture implicate causal risk variants in primary open-angle glaucoma pathogenesis

Subsession Time: Wednesday, October 15 at 10:45am – 11:00am

Authors: Vrathasha Vrathasha (Scheie Eye Institute, University of Pennsylvania), Mathew C. Pahl (Center for Spatial and Functional Genomics, The Children's Hospital of Philadelphia), James A. Pippin (Center for Spatial and Functional Genomics, The Children's Hospital of Philadelphia), Jie He (Scheie Eye Institute, University of Pennsylvania), Rebecca Salowe (Scheie Eye Institute, University of Pennsylvania), Mina Halimitabrizi (Scheie Eye Institute, University of Pennsylvania), Struan F. A. Grant (Center for Spatial and Functional Genomics, The Children's Hospital of Philadelphia), Joan M. O'Brien (Scheie Eye Institute, University of Pennsylvania)

Abstract: Primary open-angle glaucoma (POAG) is the leading cause of irreversible blindness worldwide and displays higher prevalence and severity in individuals of African ancestry. This heritable disease has a strong established genetic component. Genome-wide association studies (GWAS) have implicated over 329 single-nucleotide polymorphisms (SNPs) to be associated with POAG and its related traits. However, the vast majority of GWAS-implicated signals are located in non-coding regions of the genome. Furthermore, the assumption that the nearest gene to the GWAS-implicated signal is the true effector gene is often incorrect. Here, we integrated the variants implicated by POAG GWASs with gene expression data, accessible chromatin peaks, and three-dimensional chromatin architecture identified in two glaucoma-specific cell types: human trabecular meshwork cells (hTMCs) and retinal ganglion cells derived from induced pluripotent stem cells (iPSC-RGCs). Our goal was to identify SNP-to-target gene relationships and better elucidate how non-coding variants contribute to POAG pathogenesis.

We used two public datasets: 1000 Genomes (1kg) and TOPMED LD. A subset of the linkage disequilibrium-expanded variant lists was matched to those in open chromatin regions (peaks from ATAC-seq) that contact a gene promoter (significant chromatin contacts called from Hi-C). Our analysis implicated 24 genes in hTMCs and 36 genes in iPSC-RGCs. Of these genes, ARHGEF12 was previously implicated in our GWAS study (Cell, 2024). Using qRT-PCR and western blot techniques, we confirmed significant differences in the expression of a few select risk variants and genes in hTMCs and iPSC-RGCs derived from POAG case and control subjects. Furthermore, calcium staining and whole-cell patch clamping of iPSC-RGCs carrying the ARHGEF12 variant showed differences in functional activity when compared to control iPSC-RGCs.

Integrating GWAS findings with ATAC-seq, Hi-C, and RNA-seq approaches enabled us to implicate causal variants in two ocular tissues that confer POAG susceptibility. The resulting list of putative variants and genes requires additional functional verification; hence, we are studying their gene regulation and mechanism in hTMCs and iPSC-RGCs derived from glaucoma cases. Applying structural genomic approaches to POAG databases will provide researchers with additional novel risk variants and targets to investigate their role in glaucoma pathogenesis.

Subtype-Specific Differences in 3D Genome Architecture Reveal Regulatory Mechanisms in Medulloblastoma

Subsession Time: Wednesday, October 15 at 11:00am – 11:15am

Authors: Samantha Gonzales (Florida International University), Dhruv Battula (Allen D. Nease High School), Jane Zizhen Zhao (University of North Carolina at Chapel Hill), Sicong Xie (University of Michigan), Yun Li (University of North Carolina at Chapel Hill), Xuexia Wang (Florida International University)

Abstract: Medulloblastoma, the most common malignant pediatric brain tumor, presents clinically distinct subtypes, with Group 3 (G3) associated with a poorer prognosis compared to Group 4 (G4). While genomic profiling has revealed important subtype characteristics, the contribution of 3D genome organization to tumor biology remains underexplored. We analyzed chromatin conformation data from 21 pediatric medulloblastoma tumor samples (8 G3, 13 G4) using LAWS-HiC, a spatially adaptive multiple testing framework that accounts for local correlation structures to characterize subtype-specific chromatin interactions. We identified over 2.3 million and 2.2 million significant chromatin interactions in G3 and G4 samples, respectively (Benjamini–Hochberg false discovery rate (FDR) corrected p-value < 0.01). Among these, 42,226 interactions were unique to G3, and 18,074 were unique to G4. G3-specific interactions demonstrated a slightly higher frequency of enhancer–promoter contacts (45.5% in G3 vs. 43.8% in G4). Functional enrichment analysis of G3-specific interactions revealed significant overrepresentation of genes involved in nucleosome assembly, chromatin remodeling, and protein–DNA complex formation. Notably, key chromatin regulators including ARID2, SMARCD1, and SSRP1 showed highly significant promoter-associated contacts in G3 tumors. These genes are implicated in chromatin remodeling complexes, and some (e.g., SSRP1) are emerging as potential therapeutic targets in aggressive medulloblastoma. Conversely, G4-specific interactions did not yield significantly enriched gene ontology terms, suggesting a less distinct regulatory network at the chromatin level. Findings were validated using independent datasets GSE267471 for G3 and GSE240410 for G4, confirming the robustness of subtype-specific regulatory landscapes. These results uncover distinct 3D genomic landscapes in G3 versus G4 medulloblastoma and highlight the potential for chromatin architecture to inform biologically relevant pathways and therapeutic strategies, particularly for the more aggressive G3 subtype.

Comprehensive Dissection of Alzheimer’s Disease Genetic Variants Using a Neuron-Tuned 3D DNA Foundation Model

Subsession Time: Wednesday, October 15 at 11:15am – 11:30am

Authors: Yunxiao Ren (Cleveland Clinic, Lerner Research Institute), Xiaoyu Yang (Cleveland Clinic, Lerner Research Institute), Ming Hu (Cleveland Clinic, Lerner Research Institute), Yang Eric Li (Washington University School of Medicine, Department of Neurosurgery and Genetics), Feixiong Cheng (Cleveland Clinic, Lerner Research Institute)

Abstract: Background: The 3D organization of the genome regulates gene expression in brain cells and is increasingly implicated in Alzheimer's disease (AD) and related dementias (ADRD). While genome-wide association studies (GWAS) have identified a set of single-nucleotide polymorphisms (SNPs) linked to AD risk, emerging evidence suggests that structural variations (SVs) also contribute significantly to the heritability of AD. SVs disrupt chromatin architecture, such as through the deletion of topologically associating domain (TAD) boundaries, potentially leading to aberrant gene regulation. Despite their potential impact, the effects of AD- and ADRD-associated SVs on 3D genome structure remain poorly understood. Experimental methods like Hi-C reveal SV-induced changes in chromatin conformation, but these approaches are prohibitively time-consuming and costly for large-scale analysis.

Methods: To address this gap, we developed and applied a neuron-specific 3D DNA foundation model trained on neural progenitor cell-derived Hi-C data. We curated a set of 116 AD/ADRD-associated SVs and performed in silico mutagenesis to predict their effects on chromatin conformation and gene regulation. SVs were ranked by their predicted impact on 3D genome organization, with a focus on those disrupting known AD risk genes. Top candidates were experimentally validated in human iPSC-derived neurons using CRISPR assays.

Results: We identified 16 SVs that significantly alter chromatin folding near known AD-associated loci (including PLCG2, BIN1, PSEN1, SLC24A4, ITPR2) using this neuron-specific 3D DNA foundation model. Notably, an 80-kb deletion at chr16:81,860,089–81,994,050 was predicted to enhance local chromatin contacts around the PLCG2 gene, a known AD risk locus involved in protective roles in microglia. Additional top-ranked SVs were identified to disrupt enhancer–promoter interactions involving genes such as MAPT and BIN1, highlighting potential mechanisms by which SVs contribute to transcriptional dysregulation underlying neuropathology in AD.

Conclusion: This study presents a 3D DNA foundation model tailored to neuronal contexts in AD. Our results highlight the functional consequences of SVs on genome folding and demonstrate a scalable strategy for prioritizing noncoding variants likely to perturb regulatory chromatin interactions in AD and other neurological diseases if broadly applied.

Transformer-enhanced Hi-C Upscaling Unveils 3D Genome Architecture at Low Sequencing Depth

Subsession Time: Wednesday, October 15 at 11:30am – 11:45am

Authors: Chong Li (Department of Computer and Information Sciences at Temple University, Institute for Genomics and Evolutionary Medicine at Temple University), Mohammad Erfan Mowlaei (Department of Computer and Information Sciences at Temple University, Institute for Genomics and Evolutionary

Medicine at Temple University), Lingbin Ni (Department of Genome Sciences at University of Washington School of Medicine), Human Genome Structural Variation Consortium (HGSVC), HGSVC Functional Analysis Working Group, Mark J.P. Chaisson (Department of Quantitative and Computational Biology at University of Southern California), Evan E. Eichler (Department of Genome Sciences at University of Washington School of Medicine), Vincenzo Carnevale (Institute for Genomics and Evolutionary Medicine at Temple University, Institute for Computational Molecular Science at Temple University), Sudhir Kumar (Department of Computer and Information Sciences at Temple University, Institute for Genomics and Evolutionary Medicine at Temple University, Department of Biology at Temple University), Xinghua Shi (Department of Computer and Information Sciences at Temple University, Institute for Genomics and Evolutionary Medicine at Temple University)

Abstract: Accurate 3D maps of chromatin are crucial for understanding gene regulation and disease etiology, yet conventional Hi-C demands prohibitive sequencing depth to resolve fine-scale topological features, such as topologically associating domains (TADs) and chromatin loops. Here, we present TRUHiC, a deep learning framework that converts sparse Hi-C contact maps into high-resolution representations. TRUHiC fuses a transformer-augmented U²-Net with a multi-head self-gated attention transformer block to capture both local and long-range genomic interactions captured in Hi-C data. Unlike existing methods, which are primarily based on convolutional neural networks (CNNs) or generative adversarial networks (GANs), TRUHiC excels at maintaining global contact context while enhancing key structural features, even from sparsely sampled Hi-C datasets.

In this work, we also introduce TRUHiC-LCL, a specialized TRUHiC model for augmenting lymphoblastoid cell line (LCL) Hi-C data that is pre-trained using population-scale LCL Hi-C datasets (n=43) from the Human Genome Structural Variation Consortium (HGSVC). This real-world training confers robustness across diverse genetic backgrounds, underscoring the model's translational relevance.

Comprehensive benchmarking across multiple human and mouse cell lines (GM12878, K562, IMR90, NHEK, CH12-LX) demonstrates that our model consistently outperforms baseline methods across both technical metrics in evaluating imaging quality and biological reproducibility scores. Notably, TRUHiC recovers more true-positive TAD boundaries and loops while minimizing false-positives on both down-sampled and real low-resolution Hi-C data, further attesting to its robustness for downstream interpretation. Additionally, TRUHiC maintains superior performance under varying levels of data sparsity (1/50 and 1/100 downsampling), highlighting its resilience in highly degraded Hi-C datasets. We also examined the correlations between key visual quality metrics and biologically meaningful features, reinforcing the relevance of these metrics in evaluating structural fidelity in enhanced Hi-C data.

In summary, TRUHiC and TRUHiC-LCL deliver high-fidelity 3D chromatin reconstruction from modest sequencing inputs, setting a new benchmark for computational 3D genomics and empowering both basic discovery and translational applications.

Session 25: Mendelian Mysteries Through a Global Lens

Location: Room 156ABC/Level 1, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 10:45am – 11:45am

A decade of Initiative on Rare and Undiagnosed Diseases (IRUD): a nation-wide diagnostic and research network in Japan

Subsession Time: Wednesday, October 15 at 10:45am – 11:00am

Authors: Yuji Takahashi (Department of Neurology, National Center Hospital, National Center of Neurology and Psychiatry), Hidetoshi Date (Medical Genome Center, National Center of Neurology and Psychiatry), Yuka Hama (Department of Genomic Medicine, National Center Hospital, National Center of Neurology and Psychiatry), Hideki Oi (Department of Clinical Data Science, Clinical Research and Education Promotion Division, National Center of Neurology and Psychiatry), Shinji Kosugi (Department of Medical Ethics and Medical Genetics, Kyoto University School of Public Health), Tadashi Kaname (National Center for Child Health and Development), Naomichi Matsumoto (Department of Human Genetics, Yokohama City University Graduate School of Medicine), Kenjiro Kosaki (Center for Medical Genetics, Keio University School of Medicine), Hidehiro Mizusawa (National Center of Neurology and Psychiatry), IRUD consortium

Abstract: Background

IRUD was launched in 2015 as a nation-wide project in Japan to construct a comprehensive medical and research system for establishing diagnosis, discovering causes and ultimately providing cures for rare and undiagnosed diseases, supported by Agency for Medical Research and Development (AMED). IRUD covers the entire geographic areas and the whole fields of specialties, leveraging the healthcare system in Japan.

Purpose

To present the accomplishment of IRUD over the past decade and to illustrate the current landscape of rare and undiagnosed diseases in Japan.

Methods

IRUD consists of clinical centers (IRUD-CLs) to operate diagnostic committees managing the process from the enrollment to the final diagnosis, analysis centers (IRUD-ACs) to conduct whole exome sequencing analysis (WES) , disease model centers (IRUD-dMCs) to promote functional analysis using model organisms. a data center (IRUD-DC) for data sharing via IRUD Exchange based on Human

Phenotype Ontology (HPO), a resource center (IRUD-RC) to reposit genomic DNA samples and cell lines and the coordinating center (IRUD-CC) for the governance of the entire system.

Results

Forty IRUD-CLs, 13 semi-clinical centers, 5 IRUD-ACs, 4 IRUD-dMCs, 1 IRUD-DC, IRUD-RC and IRUD-CC have been established. As of March 2025, 9908 pedigrees including 28207 individuals have been enrolled in IRUD. WES was completed in 8090 pedigrees leading to final genetic diagnosis in 3894 (48.3%). Novel genes or new disease entities were identified in 50 pedigrees. Genomic DNA samples of 9930 individuals and cell lines of 5521 probands were stored in IRUD-RC. Phenotypic and genomic data of 7738 pedigrees were registered in IRUD-Exchange. Functional analysis was conducted in 65 genes in IRUD-dMC, resulting in the demonstration of loss or gain of function in 13 genes.

Conclusion

IRUD has established a nation-wide comprehensive diagnosis and research network for rare and undiagnosed diseases, enabling accurate diagnosis, novel gene/disease entity discovery, data-sharing, resource repository and functional validation. Continued efforts are warranted to elucidate the remaining unsolved patients and to advance therapeutic development through international collaboration.

★ Novel coding and noncoding rare variant association and in vivo functional validation across the developmental continuum

Subsession Time: Wednesday, October 15 at 11:00am – 11:15am

Authors: Arthur S. Lee (Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard), Alba Sanchis-Juan (Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard), Eren Shin (Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard), Nehir E. Kurtas (Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard), Laura E. Cook (Lawrence Berkeley National Laboratory), Michael Kosicki (Lawrence Berkeley National Laboratory), Neil Slaven (Lawrence Berkeley National Laboratory), Jack M. Fu (Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard), Xuefang Zhao (Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard), International Fetal Genomics Consortium, Genomics Research to Elucidate the Genetics of Rare Diseases GREGoR Consortium, Angie Jelin (The Johns Hopkins Medical Institutions), Terri H. Beaty (Johns Hopkins Bloomberg School of Public Health), Frances A. High (Massachusetts General Hospital, Harvard Medical School), Patricia K. Donahoe (Massachusetts General Hospital, Harvard Medical School), Stephanie M. Ware (Indiana University School of Medicine), Ian D. Krantz (The Children's Hospital of Philadelphia, Perelman School of Medicine at University of Pennsylvania, Hofstra University), Joseph Gleeson (University of California San Diego, Howard Hughes Medical Institute), Azeez Butali (University

of Iowa), Elizabeth C. Engle (Boston Children's Hospital, Howard Hughes Medical Institute), Ali Gharavi (Columbia University), Anne O'Donnell-Luria (Boston Children's Hospital, Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard), Heidi L. Rehm (Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard), Wendy Chung (Boston Children's Hospital, Harvard Medical School), Elizabeth J. Leslie-Clarkson (Emory University School of Medicine), Mary L. Marazita (University of Pittsburgh), Axel Visel (Lawrence Berkeley National Laboratory), Len A. Pennacchio (Lawrence Berkeley National Laboratory), Harrison Brand (Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard), Michael E. Talkowski (Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard)

Abstract: Despite a wealth of rare disease genomic data >50% of cases remain unsolved. Pathogenic variant discovery remains limited by siloed cohorts restricted to individual diseases that limit power to detect recurrent variants; incomplete genotyping of disease-relevant variation; limited vantage of pleiotropic effects; and difficulty interpreting and validating candidates at scale. To overcome these limitations, we created CrossDEV, a genome-phenotype aggregation resource for rare disease, incorporating 261,307 total sequenced samples (including 69,443 WES and 33,102 WGS cases with parental sequencing), and case-control microarray data from 950,278 rare disease samples.

CrossDEV encompasses diverse congenital phenotypes across many organ systems, from stillbirth and structural birth defects to neurodevelopmental disorders. We annotated all variants with >84 million combinations of annotations across 69 functional categories. To disentangle variant effects across organ systems and celltypes, we translated patient phenotypes to 1,440 unique HPO terms and applied a battery of cross-disorder variant discovery strategies. Pilot analysis of 137,435 samples (53% of sequenced cohort) identified pathogenic variation in over 700 coding genes. To further dissect the genetic architecture of rare diseases at sequence resolution, we evaluated the noncoding contribution to disease risk in CrossDEV genomes. We performed a hypothesis-free 1kb sliding window analysis and found recurrent de novo mutations in known noncoding disease loci RNU4-2 and RNU2-2—as well as 5 additional loci harboring ≥ 20 de novo case variants absent in controls. In parallel, we generated per-gene regulatory mutational burden estimates by connecting noncoding variants to their cognate genes via Activity-by-Contact enhancer predictions, revealing 16 genes with ≥ 20 de novo case-only enhancer variants, of which 5 (31%) qualified as constrained when relying exclusively on coding-based constraint metrics.

We tested 13 noncoding case variants across varied phenotypes—including autism, epilepsy, atrial septal defect, and GI dysfunction—with in vivo functional enhancer assays. Remarkably, 5 (38%) demonstrated reproducible and tissue-specific loss of reference enhancer activity caused by the prioritized variant. These results highlight the promise of aggregated disease association approaches and principled functional prioritization in uncovering pathogenic variation throughout human development.

★ Biallelic Variants in KMO Cause a Novel Form of Congenital NAD Deficiency

Subsession Time: Wednesday, October 15 at 11:15am – 11:30am

Authors: Nathalie M. Aceves-Ewing (Department of Molecular and Human Genetics, Baylor College of Medicine (BCM), Houston, TX, USA), Nanbing Li-Villarreal (Department of Integrative Physiology, BCM, Houston, TX, USA), Xiaohui Li (Department of Molecular and Human Genetics, BCM, Houston, TX, USA), Seema Lalani (Department of Molecular and Human Genetics, BCM, Houston, TX, USA; Texas Children's Hospital, Houston, TX, USA), Jill A. Rosenfeld (Department of Molecular and Human Genetics, BCM, Houston, TX, USA), Angelina Gaspero (Department of Molecular and Human Genetics, BCM, Houston, TX, USA), Denise Lanza (Department of Molecular and Human Genetics, BCM, Houston, TX, USA), Audrey E. Christiansen (Department of Integrative Physiology, BCM, Houston, TX, USA), Bao Tran (Metabolomics Core Facility, Department of Bioinformatics & Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA), Philip L. Lorenzi (Metabolomics Core Facility, Department of Bioinformatics & Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA), Lin Tan (Metabolomics Core Facility, Department of Bioinformatics & Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA), Tara L. Rasmussen (Department of Integrative Physiology, BCM, Houston, TX, USA), Haonan Zhouyao (The University of Texas MD Anderson Cancer Center, Houston, TX, USA), John Seavitt, Mary E. Dickinson (The Jackson Laboratory, Bar Harbor, ME, USA), Brendan Lee (Department of Molecular and Human Genetics, BCM, Houston, TX, USA; Texas Children's Hospital, Houston, TX, USA), Ronit Marom (Department of Molecular and Human Genetics, BCM, Houston, TX, USA; Texas Children's Hospital, Houston, TX, USA), Undiagnosed Diseases Network, Hartmut Cuny (Victor Chang Cardiac Research Institute, Sydney, New South Wales, Australia; School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales, Sydney, New South Wales, Australia), Sarah Elsea (Department of Molecular and Human Genetics, BCM, Houston, TX, USA), Sally Dunwoodie (Victor Chang Cardiac Research Institute, Sydney, New South Wales, Australia; School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales, Sydney, New South Wales, Australia; Faculty of Science, University of New South Wales, Sydney, New South Wales, Australia), Baylor College of Medicine Center for Precision Medicine Models, Jason Heaney (Department of Molecular and Human Genetics, BCM; Dan L. Duncan Comprehensive Cancer Center, BCM, Houston, TX, USA), Chih-Wei Hsu (Optical Imaging and Vital Microscopy Core, BCM; Department of Integrative Physiology, BCM, Houston, TX, USA), Lindsay C. Burrage (Department of Molecular and Human Genetics, BCM, Houston, TX, USA; Texas Children's Hospital, Houston, TX, USA)

Abstract: Deficiencies in key enzymes of the kynurenine pathway (KP) have been associated with a spectrum of clinical phenotypes, including congenital anomalies, recurrent pregnancy loss, and systemic nicotinamide adenine dinucleotide (NAD⁺) deficiency. The KP is responsible for synthesizing NAD⁺ from tryptophan, although the Preiss-Handler and salvage pathways can synthesize NAD⁺ from sources of niacin/nicotinamide. Biallelic variants in KYNU, HAAO, and NADSYN1, genes encoding KP enzymes, have been linked to congenital NAD deficiency disorder (CNDD). Through the Undiagnosed Diseases Network at Baylor College of Medicine (BCM), we identified two siblings with biallelic variants in kynurenine 3-monooxygenase (KMO), which encodes another KP enzyme. One sibling is deceased. The living individual presented with vertebral, cardiac, and inner ear abnormalities, short stature, elevated levels of kynurenine and other precursors, and reduced NAD⁺ levels. The case was nominated to the BCM Center for Precision Medicine Models for further investigation. We hypothesize that KMO deficiency represents a novel CNDD and increases the risk for congenital anomalies. To test this, we generated Kmo knockout

(Kmo^{-/-}) mice. Western blot and gene expression analyses confirmed the deletion, and only Kmo^{-/-} mice showed elevated serum kynurenine. While these mice are viable and fertile on a standard diet, unlike wild-type (Kmo^{+/+}) mice, they fail to thrive on a low-niacin diet. This result suggests a functional KMO deficiency and reliance on the Preiss-Handler and salvage pathways for NAD⁺ synthesis. Our embryonic studies revealed that Kmo^{-/-} embryos derived from Kmo^{-/-} dams fed a low-niacin diet had lower whole-body NAD⁺ levels at E12.5 and a higher incidence of congenital anomalies compared to Kmo^{+/+} littermates at E18.5. We observed vertebral, cardiac, renal, and other anomalies, which recapitulate those seen in patients with CNDD. Importantly, dietary niacin supplementation ameliorated these developmental defects. Collectively, these findings support KMO deficiency as a new cause of CNDD and highlight the importance of NAD⁺ metabolism and gene-environment interactions during embryonic development.

Advancing National Genomic Initiatives: Abu Dhabi's Model for Newborn Genetic Screening and Lifelong Precision Medicine

Subsession Time: Wednesday, October 15 at 11:30am – 11:45am

Authors: Sahar Almarzooqi (Department of Health, Abu Dhabi), Wadha Mohamed (Department of Health, Abu Dhabi), Mohamed Salem Al Ameri (Department of Health, Abu Dhabi), Azza Attia (M42), Nicholas Hoque (Kanad Hospital), Huraa Hasan Mohamed (Abu Dhabi Public Health Center)

Abstract: Background: The Department of Health (DOH), Abu Dhabi, is pioneering a genomics-driven strategy to address the high prevalence of genetic disorders in the United Arab Emirates (UAE). Building on the mandatory national newborn metabolic screening program (covering ~34 conditions) and the premarital genetic screening program for Emirati couples, DOH has launched an advanced newborn genetic screening program. This initiative leverages the Emirati Genome Program, the national whole genome sequencing (WGS) program, to detect early-onset, actionable genetic disorders, thereby enabling precision-based early intervention.

Methodology: The program offers prospective parents the option for WGS-based screening of 733 genes associated with over 815 childhood-onset genetic conditions. Cord blood was selected as the specimen of choice, enabling optional public donation for stem cell research and transplant through the Public Cord Blood Bank.

Gene Panel Curation: The gene panel was curated through global horizon scanning of existing newborn screening models and refined via multi-disciplinary workshops involving clinical specialists across Abu Dhabi. This collaborative model ensured relevance to local clinical practice and operational feasibility across health systems.

Workforce Capacity Building: To address the shortage of clinical geneticists and genetic counselors, DOH launched a certified training program for pediatricians and neonatologists. Sixteen physicians from three hospitals have completed training, with plans to train an additional 50 by 2025 to meet program expansion needs.

Results:The program is the first in the region to implement whole genome sequencing at birth at scale, utilizing the Emirati Genome Program. This enables ongoing data reanalysis as clinical knowledge evolves. The pilot phase has been successfully launched in two hospitals, with rollout to additional sites in 2025.

Conclusion and Significance:This DOH-led initiative represents a transformational model in public health genomics. By integrating lifelong-accessible genomic data with actionable newborn screening, Abu Dhabi is significantly shortening the diagnostic journey and enabling timely therapeutic interventions. This initiative is a landmark in population-scale precision medicine, offering a replicable framework for other nations with similar health system ambitions.

Session 26: Seq-ing Function with Single Cells and CRISPR Perturbations

Location: Room 258ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 10:45am – 11:45am

Systematic targeting and inhibition of bone mineral density GWAS loci in single cells identifies novel regulatory networks for human bone formation

Subsession Time: Wednesday, October 15 at 10:45am – 11:00am

Authors: John A. Morris (University of Toronto), Larry D. Mesner (University of Virginia), Mohammadjavad Kahlani (University of Toronto), Veda Madhu (University of Virginia), Lucia Wagner (University of Virginia), Ana C. Mello (New York Genome Center), Samuel Ghatan (New York Genome Center), Tuuli Lappalainen (New York Genome Center), Neville E. Sanjana (New York Genome Center), Charles R. Farber (University of Virginia)

Abstract: Bone mineral density (BMD) is a heritable, causal risk factor for osteoporosis, a debilitating age-related disease. Genome-wide association studies (GWAS) of BMD have identified over 500 loci (20% heritability explained) and known osteoporosis drug targets. 90% of BMD GWAS variants are noncoding, with unknown effects, complicating target gene identification. We developed Systematic Targeting and Inhibition of Noncoding GWAS loci with single-cell sequencing (STING-seq) for high-throughput target gene discovery at noncoding loci, leveraging biobank-scale GWAS data and pooled CRISPR screens. We applied STING-seq to study BMD GWAS loci in the context of human bone formation using the fetal human osteoblast (hFOB) model. We prioritized BMD GWAS loci likely to regulate transcription factors (TFs), given their potential to mediate gene regulatory networks in trans.

We performed ATAC-seq in hFOBs across early osteogenic differentiation (days 0, 2, and 4) to identify candidate cis-regulatory elements (cCREs) harboring fine-mapped noncoding BMD GWAS variants. We targeted 361 cCREs from 132 loci, where at least one of the three closest genes was a TF, using STING-seq. We performed multimodal single-cell sequencing and differential expression testing to link cCRE inhibition to target genes. We identified 83 CRE-gene pairs in 37% of targeted loci, including 12 timepoint-specific hits, and validated select hits with CRISPR inhibition/activation, targeted siRNA knockdowns, in vitro mineralization assays, and mitochondrial phenotyping, as mineralization is an energy-intensive process. We identified a promoter CRE that bidirectionally regulated two genes, DAP3 and YY1AP1, neither previously linked to human bone formation; targeting this CRE decreased mineralization. DAP3 siRNA knockdown increased mitochondrial fragmentation and led to decreased mineralization, while YY1AP1 knockdown had no effect. 21 CRE-gene pairs also impacted downstream genes in trans. TRIP11 regulated a network of 104 genes; TRIP11 CRISPR inhibition increased mineralization, whereas CRISPR activation decreased mineralization, revealing a novel osteoblast gene network that can be tuned to promote bone formation.

We discovered target genes at 37% of targeted noncoding GWAS loci across osteogenic timepoints and characterized novel mechanisms in bone formation. This work offers a new paradigm for common disease drug target discovery and functional study of noncoding GWAS loci.

PerTurbo: A scalable Bayesian framework for whole-transcriptome differential expression analysis and optimal design of single-cell CRISPR screens

Subsession Time: Wednesday, October 15 at 11:00am – 11:15am

Authors: Logan Blaine (Harvard Medical School, MGH, Broad Institute), Jiaqi Lu (Helmholtz Munich), Martin Jankowiak (Broad Institute), Jayoung Ryu (Harvard Medical School, MGH, Broad Institute), Lucas Ferreira da Silva (MGH, Broad Institute), Matthias Heinig (Helmholtz Munich, Technical University Munich), Luca Pinello (MGH, Harvard Medical School, Broad Institute)

Abstract: Pooled single-cell CRISPR screens, such as Perturb-seq, link genetic perturbations to transcriptomic responses at scale. These platforms have revolutionized functional genomics by enabling large-scale perturbation studies of genes, enhancers, and genetic variants. Yet current computational analysis approaches suffer from critical limitations. Existing fast approaches produce many false positives, but more robust alternatives are markedly slower, making transcriptome-wide testing for differentially expressed genes difficult.

To address these challenges, we developed PerTurbo: a scalable, fully Bayesian analysis framework for large-scale Perturb-seq experiments. PerTurbo incorporates stochastic variational inference and GPU acceleration to efficiently analyze perturbation effects in datasets comprising millions of cells and thousands of perturbations. Crucially, unlike previous tools, it also models CRISPR guide efficiency as a source of technical variability, allowing more accurate effect size estimation.

We benchmarked PerTurbo using simulated datasets, where it performed competitively with existing methods in detecting true effects and estimating ground-truth effect sizes. We next reanalyzed a published screen targeting candidate enhancers and found that PerTurbo identifies highly similar sets of significant cis enhancer–gene pairs compared with existing robust methods, while running >10x faster. PerTurbo’s scalability enabled us to exhaustively test for trans effects on distal genes, revealing enhancers of GATA1 and LMO2 whose repression affected hundreds of genes.

To underscore PerTurbo’s scalability, we reanalyzed a genome-wide Perturb-seq screen with ~2 million cells and 9,866 perturbations targeting all expressed genes in K562 cells. PerTurbo completed the analysis in under two hours on a single machine, replicated key findings from the original analysis, and discovered effects of “weak” perturbations overlooked in the previous analysis.

Finally, we performed simulation-based power analyses to examine the impact of sample size, sequencing depth, and multiplicity of infection on statistical power. We used these results to develop a user-friendly calculator to estimate optimal sequencing depth and numbers of cells to profile given a fixed budget. PerTurbo democratizes robust analysis and optimized design of pooled single-cell CRISPR

screens, meeting a pressing need as these screens become more widespread in variant characterization efforts.

PhenoBridge: A Unified Framework for Interpreting Molecular Effects on Complex Phenotypes through Integration of Perturb-seq and Phenotypic Data

Subsession Time: Wednesday, October 15 at 11:15am – 11:30am

Authors: Romain Lopez* (Stanford, Genentech), Emma Dann* (Stanford), Jeffrey Spence (Stanford), Mineto Ota (Stanford), Aviv Regev (Genetech), Alex Marson (UCSF), Jonathan Pritchard (Stanford)

Abstract: Understanding how genetic variants affect complex traits through intermediate molecular mechanisms remains a fundamental challenge in human genetics. While rare-variant burden analyses and CRISPR screens identify genes that influence phenotypes, the underlying molecular pathways mediating these effects are often unclear. Recently, a proof-of-concept using simple linear models uncovered how genes regulate blood traits like mean corpuscular hemoglobin through specific programs including hemoglobin synthesis and cell cycle control in K562, using Perturb-seq data.

Building on this approach, we present PhenoBridge, a unified computational framework that integrates Perturb-seq data with large-scale phenotyping experiments to identify gene programs that explain perturbation effects on phenotypes.

PhenoBridge extends supervised variational autoencoders to jointly model gene expression profiles from Perturb-seq experiments and phenotypic effects of matched genetic perturbations. The model learns latent gene programs that both reconstruct expression patterns and predict phenotypic outcomes, enabling mechanistic interpretation of how regulators affect traits through specific transcriptional programs. By incorporating phenotypic information during training, PhenoBridge identifies more phenotype-relevant gene groupings compared to unsupervised approaches.

We evaluate PhenoBridge on multiple datasets, including K562 and Jurkat genome-wide Perturb-seq integrated with CRISPR fitness screens and loss-of-function burden estimates from UK Biobank. The model successfully reconstructs phenotypic effects from learned gene programs, with performance improving when cellular context matches trait relevance. Application to primary T cell CRISPRa screens reveals distinct molecular mechanisms underlying cytokine production: we distinguish IL-2 and IFN- γ regulators controlling TCR stimulation programs from regulators controlling Th1 polarization, as well as TNF- α regulators involved in NF- κ B signaling.

Our framework provides a principled approach for mechanistic interpretation of genetic effects, bridging molecular perturbation data with organism-level phenotypes. By identifying shared and distinct pathways across different perturbations and traits, PhenoBridge offers insights into convergent mechanisms and context-specific effects that could inform therapeutic target identification and precision medicine approaches.

★ **Single-cell, Trimodal Omics Atlas of Ileal and Colonic Mucosal T Cells in Crohn's Disease Highlights CD4⁺ Regulatory T Cell Plasticity in Inflammation Progression**

Subsession Time: Wednesday, October 15 at 11:30am – 11:45am

Authors: Xiangyu Ye (Department of Pediatrics, University of Pittsburgh, PA, USA), Site Feng (Department of Medicine, University of Pittsburgh, PA, USA, School of Medicine, Tsinghua University, Beijing, China), Elisa Heidrich-O'Hare (Department of Medicine, University of Pittsburgh, PA, USA), Haoran Hu (Department of Pediatrics, University of Pittsburgh, PA, USA, Department of Biostatistics and Health Data Science, University of Pittsburgh, Pittsburgh, PA, USA), Shiyue Tao (Department of Pediatrics, University of Pittsburgh, PA, USA, Department of Biostatistics and Health Data Science, University of Pittsburgh, Pittsburgh, PA, USA), Carl Samaha (Department of Medicine, University of Pittsburgh, PA, USA), Logan Slade (Department of Medicine, University of Pittsburgh, PA, USA), Wei Chen (Department of Pediatrics, University of Pittsburgh, PA, USA, Department of Biostatistics and Health Data Science, University of Pittsburgh, Pittsburgh, PA, USA), Richard H. Duerr (Department of Medicine, University of Pittsburgh, PA, USA)

Abstract: Background: Crohn's disease (CD), one of two major forms of inflammatory bowel disease, can affect any part of the gastrointestinal tract but most commonly involves the ileum and/or the colon. Genome-wide association studies (GWAS) have identified hundreds of CD-associated loci, but how most of these loci confer risk for CD remains unclear.

Methods: We profiled ~270,000 mucosal immune cells from 62 paired sets of colonoscopic biopsies of macroscopically inflamed or uninfamed ileal or colonic mucosal regions in 57 CD patients using DOGMA-seq, a single-cell multiomics assay that simultaneously characterizes antibody-derived cell surface protein tags, gene expression, and chromatin accessibility. Comparisons between ileal and colonic locations and inflammation states identified cell types, modality features, gene regulation networks, and cell communication events associated with active inflammation. S-LDSC, together with cell type-specific chromatin accessibility, identified enrichment of CD GWAS signals in specific cell types.

Results: By using three modalities, we annotated these cells into 84 phenotypic and functional distinct subtypes. GWAS signals were enriched in multiple CD4⁺ T resident memory (TRM) and regulatory (Treg) subtypes. Treg subtypes were consistently expanded in samples from macroscopically inflamed compared to uninfamed mucosa in both ileum and colon, while CD4⁺ TRM subcluster alterations were more heterogenous. Cell-cell communication analysis showed that Treg/Treg-like subtypes send increased adhesion signals and receive increased chemotaxis signals in the setting of active inflammation. We also observed transcriptional plasticity characterized by increased activity of PPAR γ , RORC, and RORA activator regulons in a CD4⁺ IKZF2^{low} Treg subcluster from inflamed regions of both ileal and colonic mucosa, which also showed positive correlations with sample histological score.

Conclusion: We built a multiomics-based mucosal immune cell atlas for CD across ileal and colonic locations and inflammation states. Single-cell and GWAS evidence highlighted CD4⁺ TRM and Treg

subtypes. Immunosuppressive states, represented by expansion of Treg subtypes and increased interactions between Treg and other CD4+ T subtypes, were observed in macroscopically inflamed compared to uninflamed mucosal regions in both ileum and colon. One CD4+ IKZF2^{low} Treg subcluster exhibited transcriptional plasticity, potentially contributing to the progression of inflammation.

Session 27: Advanced Technologies Uncover Dosage, Structure, and Phenotypic Complexity

Location: Room 258ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 1:30pm – 2:30pm

More is more: Shared phenotypes amongst sex chromosome trisomies hints dosage-sensitive effect of pseudoautosomal regions in genetic males and females

Subsession Time: Wednesday, October 15 at 1:30pm – 1:45pm

Authors: Aoxing Liu (Broad Institute of MIT and Harvard, Massachusetts General Hospital, FIMM), Giulio Genovese (Broad Institute of MIT and Harvard, Harvard Medical School), Wenhan Lu (Broad Institute of MIT and Harvard, Massachusetts General Hospital), Mary P. Reeve (FIMM), Julia Goodrich (Broad Institute of MIT and Harvard, Massachusetts General Hospital), Craig C. Teerlink (VA Salt Lake City Health Care System, University of Utah School of Medicine), Masahiro Kanai (Broad Institute of MIT and Harvard), Zhili Zheng (Broad Institute of MIT and Harvard), Juha Mehtonen (FIMM), Sanni Ruotsalainen (FIMM), FinnGen, Julie A. Lynch (VA Salt Lake City Health Care System, University of Utah School of Medicine), Richard Hauger (VA San Diego Healthcare System, University of California San Diego), Mitja I. Kurki (Broad Institute of MIT and Harvard, Massachusetts General Hospital, FIMM), Konrad Karczewski (Broad Institute of MIT and Harvard, Massachusetts General Hospital), Mark J. Daly (Broad Institute of MIT and Harvard, Massachusetts General Hospital, FIMM)

Abstract: Genetic males and females differ biologically in many ways. Nearly all the time, comparisons searching for sex differences are made directly between XX females and XY males, with observed differences, either in behaviors, biological processes, or diseases, attributed to the apparent differences in sex gonadal hormones and genetically, the dosage effect of X chromosome or the testis-determining factor gene residing on the Y chromosome. Unlike autosomal aneuploidy, which usually causes spontaneous abortion, sex chromosome trisomies (SCTs) are relatively tolerated and have a considerable prevalence among general populations such as biobank participants. Indeed, among the 1.5 million individuals of FinnGen, UKB, and MVP, 1319, 1108, and 342 are genetically inferred to have 47,XXY, 47,XYY, and 47,XXX, respectively, of which 73.8%, 98.6%, and 93.6% are unaware of their karyotypes. Introducing these trisomies into the analysis can disrupt the perfect correlation between genetic sex and sex chromosome karyotypes, thereby providing a unique perspective on the phenotypic consequences of sex chromosomes. A cross-cohort SCT pheWAS meta-analysis reveals numerous associations across all examined trait categories, except for cancers. Surprisingly, many phenotypic associations are shared amongst all three SCT subtypes, including height, asthma, and vascular diseases. We further ask whether the origin of the extra sex chromosome would have any impact on the phenotypic consequence; we test it in XXY and see no differences regarding both the parent of origin and the identical or homologous status of the two X chromosomes. Collectively, these findings suggest that the shared phenotypes

amongst SCTs stem from having an extra copy of the chromosome, and more specifically, the 3 Mb of pseudoautosomal regions (PAR) shared between chromosomes X and Y. We, therefore, performed pheWAS for each PAR SNP; among the >2000 phenotypes examined, height and asthma are the traits with the most significant associations in FinnGen, UKB, MVP, and All of Us ($P < 5.0 \times 10^{-8}$). Interestingly, rare-variant analysis in UKB shows that the loss-of-function variants of the SHOX gene reduces human height ($P < 1.5 \times 10^{-52}$), which converges to the increased height from physically gaining an extra copy of the sex chromosome in SCTs. Although the gene-dosage effect is complex to model and primarily influenced by the genomic regions and traits examined, leveraging the thousands of SCTs in large biobanks of diverse ancestries, our work suggests a broad dosage-sensitive effect of pseudoautosomal regions for traits such as height and asthma.

Completely resolved structural variants by optical genome mapping with adaptive sampling from CNV discovery

Subsession Time: Wednesday, October 15 at 1:45pm – 2:00pm

Authors: Li Fu (Department of Human Genetics, Yokohama City University Graduate School of Medicine, Department of Obstetrics and Gynecology, Juntendo University Faculty of Medicine), Naomi Tsuchida (Department of Human Genetics, Yokohama City University Graduate School of Medicine, Department of Clinical Genetics, Yokohama City University Hospital), Yuri Uchiyama (Department of Human Genetics, Yokohama City University Graduate School of Medicine, Department of Clinical Genetics, Yokohama City University Hospital), Eriko Koshimizu (Department of Human Genetics, Yokohama City University Graduate School of Medicine), Kohei Hamanaka (Department of Human Genetics, Yokohama City University Graduate School of Medicine), Satoko Miyatake (Department of Human Genetics, Yokohama City University Graduate School of Medicine, Department of Rare Disease Genomics, Yokohama City University Hospital, Yokohama Japan), Takeshi Mizuguchi (Department of Human Genetics, Yokohama City University Graduate School of Medicine), Atsushi Fujita (Department of Human Genetics, Yokohama City University Graduate School of Medicine), Naomichi Matsumoto (Department of Human Genetics, Yokohama City University Graduate School of Medicine, Department of Clinical Genetics, Yokohama City University Hospital, Department of Rare Disease Genomics, Yokohama City University Hospital)

Abstract: Structural variants (SVs) range in size from 50 bp to over a million bp, and include duplications, deletions, inversions, translocations and insertions. They may play an important role in human phenotypic diversity. However, SVs have various combinations, making the analysis of SVs complex. Optical genome mapping (OGM) is a technology that uses ultra-high molecular weight DNA (>150 kbp) fluorescently labeled at specific 6-nucleotide sequence (CTTAAG) and can detect large and complex SVs based on the labeling pattern of the long-molecule DNA. This study aimed to fully characterize constitutional SVs using OGM, which are associated with previously identified copy number variants (CNVs) by exome sequencing (ES).

We performed OGM in 31 cases with ES-based CNVs (including 17 deletions, 7 duplications and 7 deletions and duplications), disclosing additional SVs in 14 cases. The newly discovered SVs by OGM included 3 translocations, 3 tandem duplications, 2 inverted duplications, 1 duplication with inversion, 3

intrachromosomal translocations with inverted duplications, 1 reciprocal translocation and 1 chromothripsis. Various types of SVs were newly found in cases containing duplications, whereas only one new SV was found in cases with a deletion. We found aberrant genes associated with phenotypes in the additionally identified SVs in 4 cases. To confirm the breakpoints of newly found SVs at the nucleotide level, we performed Oxford nanopore targeted long-read sequencing using adaptive sampling. In our case with chromothripsis, the long-read sequencing data analyzed using the dnarrange software were inconsistent between chromosomal structure and copy number. Therefore, we manually connected the breakpoints of long-read sequencing data, considering the orientation of SVs/CNVs called by OGM.

All SVs of 14 cases could not be detected by ES alone because of SVs located outside the coding region of ES, or the difficulty in distinguishing between tandem duplications, inverted duplications, or insertions only by ES. We should especially suspect the presence of complex SVs in cases containing duplications or a combination of duplication and deletion and consider further analysis including OGM. Furthermore, in cases of complex chromosomal rearrangements involving multiple chromosomal segments and breakpoint connections, it was advantageous to use OGM to overview the entire SVs first, followed by targeted long-read sequencing to confirm the breakpoint in detail.

A multi-omic diagnostic framework leveraging long-read sequencing for unsolved rare disease

Subsession Time: Wednesday, October 15 at 2:00pm – 2:15pm

Authors: Chloe Munderloh (Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine), Nikhita Gogate (Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine), Medhat Mahmoud (Human Genome Sequencing Center, Baylor College of Medicine), Shalini N. Jhangiani (Human Genome Sequencing Center, Baylor College of Medicine), Moez Dawood (Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine), Christopher M. Grochowski (Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine), Shaghayegh Beheshti (Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine), Sairam Behera (Human Genome Sequencing Center, Baylor College of Medicine), Siyu Wang (Human Genome Sequencing Center, Baylor College of Medicine), Ryan Cheng (Human Genome Sequencing Center, Baylor College of Medicine), Rupesh Kesharwani (Human Genome Sequencing Center, Baylor College of Medicine), Donna M. Muzny (Human Genome Sequencing Center, Baylor College of Medicine), Vernon R. Sutton (Department of Molecular and Human Genetics, Baylor College of Medicine; Texas Children's Hospital), Pengfei Liu (Department of Molecular and Human Genetics, Baylor College of Medicine; Baylor Genetics Laboratory, Houston, TX), James R. Lupski (Department of Molecular and Human Genetics, Human Genome Sequencing Center, Department of Pediatrics, Baylor College of Medicine; Texas Children's Hospital), Richard Gibbs (Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine), Fritz Sedlazeck (Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine), Daniel Calame (Department of Molecular and Human Genetics,

Department of Pediatrics, Neurology Section, Baylor College of Medicine; Texas Children's Hospital), Jennifer E. Posey (Department of Molecular and Human Genetics, Baylor College of Medicine; Department of Pediatrics, Department of Medicine, Columbia University Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center)

Abstract: Despite advances in genome sequencing, a significant proportion of rare disease cases remain unsolved, often due to technical limitations in detecting complex genomic variants and incomplete integration of complementary omics data. Long-read genome sequencing (LR-GS) offers the potential to more accurately resolve variant classes such as repeat expansions, structural variants, and methylation profiles. We applied LR-GS to 1,100 individuals from 369 families, including 311 trios, using both PacBio and Oxford Nanopore Technologies platforms. Our comprehensive analysis pipeline integrated multiple variant classes (single-nucleotide variants, indels, copy number variants, structural variants, repeat expansions, methylation) including in-house variant callers. Short read sequencing, transcriptomic, and metabolomic data were included when available. Orthogonal validation was performed to confirm candidate variants and elucidate disease mechanisms.

Our integrated approach enabled systematic prioritization of candidate variants in challenging cases. Notable examples include a 3-year-old with ataxia and encephalopathy found to have biallelic GGGCC repeat expansions in the NAXE promoter, supported by RNA-seq and metabolomic data. Two additional unrelated individuals were identified to have monoallelic repeat expansions in DIP2B and ATN1, respectively, with transcriptomic and methylation data providing additional mechanistic clarity. In a fourth family, monozygotic twins diagnosed with propionic acidemia through metabolomic testing were found by short-read sequencing to have a homozygous, deep intronic splice-altering variant in PCCB (c.966+1334A>G) shared with their unaffected mother. LR-GS demonstrated the mother to be hemizygous for this variant, which was in trans with a benign, intronic 5.2 kb deletion. This additional analysis thus supported the pathogenicity of the splice variant in the affected twins. Four further cases were likely solved by coding variants in candidate (RPTOR, CCDC82) or established (PTPN1, PTEN) disease genes for which reduced penetrance and apparent genotype/phenotype discrepancy complicated diagnostic ascertainment. Our study demonstrates that integration with transcriptomics, metabolomics, and expert multidisciplinary interpretation can maximize LR-GS diagnostic yield. We propose that such a comprehensive multi-omics framework be considered for diagnostic laboratory implementation to improve resolution of unsolved rare disease cases.

Comparative genomic analysis of syndromic and nonsyndromic forms of orofacial clefting

Subsession Time: Wednesday, October 15 at 2:15pm – 2:30pm

Authors: Nehir E. Kurtas (Massachusetts General Hospital, the Broad Institute of MIT, Harvard Medical School), Alba Sanchis-Juan (Massachusetts General Hospital, the Broad Institute of MIT, Harvard Medical School), Eren Shin (Massachusetts General Hospital, the Broad Institute of MIT, Harvard Medical School), Sarah Curtis (Emory University), Arthur Lee (Massachusetts General Hospital, the Broad Institute of MIT, Harvard Medical School), Emily M. Liao (Massachusetts General Hospital, the Broad Institute of MIT,

Harvard Medical School), Kelsey Robinson (Emory University), Azeez A. Alade (University of Iowa), Xuefang Zhao (Massachusetts General Hospital, the Broad Institute of MIT, Harvard Medical School), Jack Fu (Massachusetts General Hospital, the Broad Institute of MIT, Harvard Medical School), Kimberly-Diaz-Perez (Emory University), Lord J. J. Gowans (Kwame Nkrumah University of Science and Technology), Mekonen A. Eshete (Addis Ababa University), Wasiu L. Adeyemo (University of Lagos), Carmen J. Buxó (Dental and Craniofacial Genomics Core, University of Puerto Rico School of Dental Medicine), Carmencita D. Padilla (University of the Philippines Manila), Fernando A. Poletta (Laboratorio de Epidemiología Genética), George L. Wehby (University of Iowa), Jacqueline T. Hecht (The University of Texas Health Science Center at Houston), John R. Shaffer (University of Pittsburgh), Lina M. Moreno Uribe (University of Iowa), Nandita Mukhopadhyay (University of Pittsburgh), Seth M. Weinberg (University of Pittsburgh), Jeffrey C. Murray (University of Iowa), Terri H. Beaty (John Hopkins University), Azeez Butali (University of Iowa), Michael Talkowski (Massachusetts General Hospital, the Broad Institute of MIT, Harvard Medical School), Mary L. Marazita (University of Pittsburgh), Elizabeth J. Leslie-Clarkson (Emory University), Harrison Brand (Massachusetts General Hospital, the Broad Institute of MIT, Harvard Medical School)

Abstract: Orofacial clefts (OFC) are among the most common birth defects worldwide and are characterized by the abnormal formation of the lip and/or palate. Genetic studies have traditionally classified OFC cases as either syndromic, involving OFCs alongside other congenital anomalies, or nonsyndromic, which represent the majority of cases and occur in isolation. However, emerging evidence from genomic studies indicates a growing number of genes (e.g. TFAP2A) traditionally associated with syndromic forms of OFC can also harbor variants contributing to isolated cases, challenging the notion of a strict dichotomy between the two categories. In this study, we systematically investigate the extent of this pleiotropy by leveraging comprehensive exome and genome sequencing data from syndromic (n = 913), nonsyndromic (n = 1,598), and control trios without OFC (n = 5,534). We observed a striking enrichment of de novo protein-truncating and damaging missense variants (MPC score ≥ 2) in OFC cases compared to controls (OR: 2.1, $p = 2.4 \times 10^{-31}$). This enrichment persisted across multiple biologically relevant gene sets including OFC genes (OR: 13.2, $p = 8.9 \times 10^{-41}$), dominant genes for OMIM (OR: 5.8, $p = 4.7 \times 10^{-57}$), constrained genes (LOEUF ≤ 0.35) (OR: 3.9, $p = 1.6 \times 10^{-50}$), and OFC genes discovered by GWAS (OR: 3.8, $P = 0.002$). Dissecting these findings by syndromic status showed greater enrichment across all categories in syndromic cases, except for GWAS genes, which were more enriched in nonsyndromic cases (OR: 3.4, $p = 0.05$), highlighting the subtler impact of disrupting these weaker effect genes. Building upon these results, we extended our analysis to include a cell-type enrichment assessment using expression data from mouse embryos. Across 38 cell types, we observed significant enrichment of genes expressed in granule neurons (OR = 5.9, $p = 3.4 \times 10^{-5}$) and postmitotic premature neurons (OR: 12.3, $p = 4.5 \times 10^{-3}$) among syndromic cases, while genes expressed in epithelial cells showed nominal enrichment in nonsyndromic cases (OR = 4.0, $p = 0.06$). This aligns with expectations, as many syndromic cases show an OFC-neurodevelopmental comorbidity, whereas nonsyndromic are more often limited to craniofacial tissues, consistent with the enrichment of epithelial cells. Importantly, these insights could enhance prenatal diagnostics by enabling earlier risk stratification. Overall, these findings highlight the distinct genetic architectures underlying syndromic and nonsyndromic OFC.

Session 28: Advances in Long-Read Sequencing Studies

Location: Room 210C/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 1:30pm – 2:30pm

Novel variant discovery and interpretation from long-read sequencing on 1,027 African Americans in All of Us

Subsession Time: Wednesday, October 15 at 1:30pm – 1:45pm

Authors: Qiuhui Li (Department of Computer Science, Johns Hopkins University), Kiran V. Garimella (Data Sciences Platform, Broad Institute of MIT and Harvard), Fritz J. Sedlazeck (Human Genome Sequencing Center, Baylor College of Medicine, Department of Computer Science, Rice University), Michael C. Schatz (Department of Computer Science, Department of Biology, Johns Hopkins University), Michael E. Talkowski (Center for Genomic Medicine, Massachusetts General Hospital, Medical and Population Genetics, Broad Institute of MIT and Harvard, Harvard Medical School), Evan E. Eichler (Department of Genome Sciences, University of Washington School of Medicine, Howard Hughes Medical Institute, University of Washington), All of Us Long Read Working Group

Abstract: The All of Us Research Program (AoU) is building a comprehensive national biobank comprised of one million participants with matched genomic, trait, and health data, aiming to advance genomic medicine in the U.S. While short-read sequencing (SRS) data are available for >400,000 individuals, short-reads are limited in resolving complex variants, particularly in highly repetitive regions.

Addressing this need, we report on phase I long-read sequencing (LRS) of 1,027 individuals self-identifying as African American or Black. Using these data, we assembled a comprehensive variant resource, including SNVs, indels, structural variants (SVs), repeat expansions, and more. We identified nearly one million SVs, including >190,000 previously undetected variants. Using CADD-SV, we prioritized 473 likely pathogenic SVs overlapping 85 high-priority genes. LRS-based tandem repeat analysis revealed multiple mutations in FMR1 and other clinically relevant regions, including motif changes not observed in prior SRS data. Additionally, we detected repeat recombination events affecting BRCA2, TP53, GSTM1, and ALS2.

We then genotyped AoU SVs in 3,202 samples from the 1000 Genomes Project, identifying 8,249 SVs enriched in African populations. By integrating matched RNA sequencing data, we identified 986 new SV-eQTLs, including for PIK3R1, a known tumor suppressor, and SH2B3, a gene implicated in autoimmunity and cardiovascular disease. Moreover, we found thousands of AoU SVs in strong linkage disequilibrium with significant variants from NHGRI-EBI GWAS catalog, indicating associations with diverse traits and diseases, including several African-prevalent diseases (e.g., hypertension, type 2 diabetes mellitus, and glaucoma).

Finally, to further assess clinical impact of genomic variants, we genotyped our variants in 10,000 short-read African American AoU participants with linked electronic health records, enabling a phenome-wide

association study. We revealed 507 diseases associated with SVs in medically relevant genes, including protein-altering SVs (e.g., in CLCN7 and COL14A1) supported by AlphaFold and prioritized as causal over nearby SNPs. Notably, some SVs absent from the paired AoU short-read dataset showed strong associations with diseases such as atherosclerosis of aorta and heart failure. This phase I dataset, along with the phase II of >10,000 long-read samples available later this year, serves as a valuable resource for advancing the development of precision medicine.

Long story short: discovery of new complex mutations and new candidate genes in Parkinson's disease using long-read sequencing

Subsession Time: Wednesday, October 15 at 1:45pm – 2:00pm

Authors: Guillaume Cogan (Paris Brain Institute), Amel Zidoum (Paris Brain Institute), Kensuke Daida (National Institutes of Health), Kimberley J. Bilingsley (National Institutes of Health), Kim Paquette (National Institutes of Health), Ludmila Jornea (Paris Brain Institute), Sylvie Forlani (Paris Brain Institute), Christelle Tesson, Andrew B. Singleton (National Institutes of Health), Suzanne Lesage (Paris Brain Institute), Cornelis Blauwendraat (National Institutes of Health), Alexis Brice (Paris Brain Institute)

Abstract: About 20 genes are known to be causal for Parkinson's disease (PD). Nevertheless, about 90% of early-onset and familial PD cases remain undiagnosed after the sequencing of these genes. This may be due to the limitations of short-read sequencing (SRS). Long-read sequencing (LRS) outperforms SRS in the identification of structural variants and repeat expansions.

In this study, we performed whole genome LRS in a cohort of 140 early-onset and familial PD cases (78 probands) left undiagnosed after exome sequencing using Oxford Nanopore Technologies. The cohort included 78 probands: 32 isolated cases with a mean age at onset of 34 years old and 46 pedigrees with multiple affected individuals.

Few causal mutations were identified. Briefly, LRS revealed large compound heterozygous deletion encompassing exon 3 and 4 (162kb) and duplication encompassing exon 3 (104kb) in two siblings with a PD phenotype suggestive of PRKN. Second, LRS revealed compound heterozygous variants in the new PD gene EPG5 in two siblings with EOPD: a 1.9kb deletion encompassing exon 17 transmitted by the father and a missense pathogenic mutation p.Lys2129Glu transmitted by the mother in two affected siblings. The phenotype was compatible with that of EPG5-related disorders. Third, LRS found a missense mutation classified as severe according to the GBA1 PD browser in GBA1, a locus difficult to sequence for SRS because of its pseudogene GBA1LP (96% sequence homology). Additionally, to the identification of complex mutations invisible for SRS in known-PD genes, LRS identified a mutation in a new candidate PD gene in large pedigree including 13 affected members with PD and PD related disorders (Lewy body dementia, schizophrenia and isolated tremor). The mutation is a 65 kb heterozygous frameshift deletion of exon 3 (out-of-frame) initially found in the PDE10A gene in four cases. Subsequent segregation analysis revealed that a total of 11 affected cases were carriers of the deletion. While PDE10A has not yet been reported to be involved in PD, multiple pieces of evidence from the literature emphasize that PDE10A is involved in PD pathogenesis.

Overall, LRS solved 6.4% (n=5/78) of exome negative early-onset and familial PD (PRKN, EPG5, VPS13C, ATXN2, GBA1), and 10.3% (n=8/78) if we include candidate variants (FMR1, MADD, PDE10A). Of note, none of these mutations were visible for conventional methods, showing the relevance of LRS in undiagnosed PD cases.

Cue2: deep learning framework for structural variant discovery with long reads

Subsession Time: Wednesday, October 15 at 2:00pm – 2:15pm

Authors: Victoria Popic (Broad Institute of MIT and Harvard)

Abstract: Structural variants (SVs) are the greatest source of genetic diversity in the human genome and a key driver of disease. While countless advances in sequencing technologies, and especially the recent advent of long-read platforms, have unlocked our ability to faithfully capture and read out genomic sequences at scale and with high accuracy, computational methods for SV discovery that can accurately convert this information into insight are still pressingly needed. In large part, this is due to the intractability of accurately modeling the full range of SV signatures arising in the genome using manually-engineered heuristic approaches, which have been predominantly developed and employed to date. Here we motivate the development of data-driven learning-based approaches for genome analysis, and introduce Cue2, an image-based deep learning method for SV discovery for PacBio and ONT long reads. Similar to our prior method Cue (Popic et al, Nature Methods 2023), which focuses on short reads, Cue2 casts SV discovery as a multi-class keypoint localization task in custom multi-channel images derived from read alignments to a reference genome. Cue2 includes 6 additional image channels specifically designed for long reads, significantly expanding the set of alignment signals used to comprehensively capture the diversity of SV signatures in the genome. Importantly, it also introduces new strategies for SV signal extraction from pervasive alignment artifacts. Cue2 PacBio and ONT models were trained on >1 million examples, capturing a variety of SV types, sizes, genome contexts, and coverage regimes. When tested on extensive synthetic data benchmarks, both models significantly outperformed state-of-the-art SV callers in the discovery of deletions, duplications, and inversions of size greater than 50bp. For example, we found that Cue2 recalled >50% more duplications and >30% more inversions than Sniffles2 across 5-30x coverage regimes using PacBio and ONT inputs. It also had the lowest rate of false positive calls, especially in benchmarks where complex SVs were present in the genome. We found that existing methods consistently misidentified signals from complex SVs as simple SVs (e.g., predicting deletion events for breakpoints involved in dispersed duplications). We also benchmarked Cue2 using the Genome in a Bottle SV truthset. We found that Cue2 either matched or slightly outperformed other methods in deletion calling, while reporting a significantly higher number of duplications. Cue2 is freely available at: <https://github.com/PopicLab/cue2>.

Sharing valuable rare disease and multiomic data through the GREGoR Consortium

Subsession Time: Wednesday, October 15 at 2:15pm – 2:30pm

Authors: Marsha M. Wheeler (University of Washington), Seth Berger (Children's National Hospital), Jessica X. Chong (University of Washington), Matthew P. Conomos (University of Washington), Stephanie M. Gogarten (University of Washington), Moez Dawood (Baylor College of Medicine), Benjamin D. Heavner (University of Washington), Shalini N. Jhangiani (Baylor College of Medicine), Shruti Marwaha (Stanford University), Stephen Montgomery (Stanford University), Anne O'Donnell-Luria (Broad Institute of MIT and Harvard University), Paul Petrowski (Stanford University), Sarah Stenton (Broad Institute of MIT and Harvard University), Adrienne M. Stilp (University of Washington), Catherine Tong (University of Washington), Matthew T. Wheeler (Stanford University), Jay Yu (Stanford University), Jimmy Zhen (Stanford University), Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) Consortium

Abstract: The Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) Consortium aims to advance genetic disease research and diagnosis through the application of omic technologies and diagnostic approaches that are not routine in clinical practice. This ongoing effort includes rapid consortium-wide data sharing generated by participating Research Centers, Partner Members, and Research Grant Awardees and controlled-access release of a rich dataset to the broader scientific community.

The GREGoR Consortium Dataset is broadly consented for research and is available via controlled access on NHGRI's AnVIL platform, a secure and scalable cloud environment for data storage and analysis. Researchers may request access to GREGoR Consortium Data by submitting a Data Access Request through dbGaP or DUOS. Additionally, the GREGoR Consortium has developed a publicly available open-access variant browser to enable searches for variants or genes of interest across the Consortium joint callset. The GREGoR Consortium Variant Browser is available at <https://variants.gregorconsortium.org>.

As of June 2025, the GREGoR Consortium Dataset consists of data generated for over 9000 participants from more than 3,600 families. Currently, the GREGoR Dataset contains short-read DNA sequencing (exome and whole genome) for the majority of participants and substantial short-read RNA-sequencing and long-read whole genome datasets. This Dataset also contains short-read ATAC-seq and will soon include long-read RNA-sequencing and metabolomics data. The GREGoR Dataset includes detailed pedigree and phenotype information, as well as candidate genetic findings curated by participating Research Centers. Short-read whole genome sequencing data is uniformly reprocessed by the GREGoR Data Coordinating Center and includes a Consortium joint callset for single nucleotide variants and indels. Mitochondrial and de novo callsets are also available. The GREGoR Dataset is organized by an expandable data model, which serves as a robust framework to facilitate data interoperability across participating Research Centers and supports secondary data use by the broader scientific community. This data model is publicly available at https://github.com/UW-GAC/gregor_data_models.

We will present additional details about the GREGoR Dataset, including additional information about how to obtain access, a range of tools available to support rare disease research, as well as our experience with cloud-based genomic data reprocessing and analysis

Session 29: Cancer Risk: Gene Discovery and Polygenicity

Location: Room 253ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 1:30pm – 2:30pm

Cell type-specific risk gene discovery in colorectal cancer through single-cell transcriptome-wide association study and regulatory element integration

Subsession Time: Wednesday, October 15 at 1:30pm – 1:45pm

Authors: Qing Li (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center), Chao Li (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center), Jungyoon Choi (Division of Oncology and Hematology, Department of Internal Medicine, Korea University Ansan Hospital, Korea University College of Medicine), Jifeng Wang (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center), Jongwon Lee (Department of Biomedical Sciences, Korea University College of Medicine), Dong-Won Lee (Core Research and Development Center, Korea University Ansan Hospital), Hyun Woo Kwon (Department of Nuclear Medicine, Korea University Ansan Hospital, Korea University College of Medicine), Shuai Xu (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center), Quanhui Sheng (Department of Biostatistics, Vanderbilt University Medical Center), Qi Liu (Department of Biostatistics, Vanderbilt University Medical Center, Department of Biomedical Informatics, Vanderbilt University Medical Center), Victor Moreno (Oncology Data Analytics Program, Catalan Institute of Oncology (ICO), Colorectal Cancer Group, ONCOBELL Program, Institut de Recerca Biomedica de Bellvitge (IDIBELL), Department of Clinical Sciences, Faculty of Medicine and Health Sciences, Universitat de Barcelona Institute of Complex Systems (UBICS), University of Barcelona (UB), Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP)), Wanqing Wen (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center), Xiang Shu (Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center), Yaohua Yang (Department of Public Health Sciences, Center for Public Health Genomics, University of Virginia), Jirong Long (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center), Qi Dai (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center), Xiao-Ou Shu (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center), Bhuminder Singh (Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, Vanderbilt University School of Medicine), Martha J. Shrubsole (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center), Qiuyin Cai (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center), Ulrike Peters (Public

Health Sciences Division, Fred Hutchinson Cancer Center, Department of Epidemiology, University of Washington School of Public Health), Wei Zheng (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center), Ken S Lau (Department of Cell and Developmental Biology, Vanderbilt University School of Medicine), Xingyi Guo (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Department of Biomedical Informatics, Vanderbilt University Medical Center)

Abstract: Most colorectal cancers (CRC) originate from colorectal polyps, evolving through either the conventional adenoma (AD) or serrated polyp (SP)–carcinoma sequence. While transcriptome-wide association studies (TWAS) using bulk normal colon tissues have identified over 400 putative CRC risk genes, tissue heterogeneity may obscure cell type–specific signals. Moreover, TWAS based on gene expression data of polyps—an intermediate but critical stage—remains underexplored. Advances in scRNA-seq and scATAC-seq allow single-cell resolution of gene expression and regulation, offering new insights into CRC molecular basis. We analyzed scRNA-seq data from 53 normal colon (NC) tissues, 82 colorectal polyps (combining 46 AD and 36 SP), and 67 carcinomas from European-ancestry individuals in the Colorectal Molecular Atlas Project (COLON MAP), along with their genotyping data generated from Illumina MEGA. Cell type-specific cis-regulatory variants were inferred from scATAC-seq open chromatin peaks in NC and AD/SP tissues from a single donor. We constructed genetically predicted gene expression models for seven epithelial cell types from NC and nine from AD/SP, including adenoma-specific cells (ASC) and serrated-specific cells (SSC). We applied these models into CRC GWAS summary statistics (78,473 cases and 107,143 controls of European ancestry; 100,204 cases and 154,587 controls of a multi-ancestry) to identify risk genes. Differential expression analyses comparing ASC to carcinomas and SSC to carcinomas were conducted to assess gene expression changes along the AD–carcinoma and SP–carcinoma sequence. At a Bonferroni-corrected threshold of $P < 0.05$, we identified 276 genes associated with CRC risk, including 133 from NC and 166 from polyps (28 from ASC and 30 from SSC). Among these, 213 genes had not been previously implicated in CRC susceptibility, including 60 genes located at loci not reported in prior GWAS. Of 276 genes, 156 (57%) were differentially expressed, including 114 from the AD–carcinoma and 130 from the SP–carcinoma, supporting functional roles in CRC development. Notably, we identified key genes involved in established CRC signaling pathways, including MYC and YAP1 (WNT/ β -catenin), TEAD1, FOSL1, and PTPN1 (EGFR/MAPK), TGFBI (TGF- β), and KLF4, KLF5, and TEAD3 (Hippo and PI3K–AKT). Our study large numbers of novel CRC risk genes with cell type–specific regulation, highlighting distinct pathways of tumorigenesis and advancing our understanding of CRC biology.

Single-cell Transcriptome-wide Association Study Identifies Cell-Type-Specific Genes for Lung Cancer

Subsession Time: Wednesday, October 15 at 1:45pm – 2:00pm

Authors: Yaohua Yang (University of Virginia), Qingying Chen (University of Virginia), Stephen S. Rich (University of Virginia), Ani W. Manichaikul (University of Virginia), David F. Kashatus (University of Virginia), Todd W. Bauer (University of Virginia), Xiaobo Zhou (University of Virginia), Yongjun Liu

(University of Washington), Rashmi M. Agni (University of Wisconsin - Madison), Jingjing Yang (Emory University), Marcus C.B. Tan (Vanderbilt University Medical Center), Shuai Xu (Vanderbilt University Medical Center), Xingyi Guo (Vanderbilt University Medical Center), Wei Zheng (Vanderbilt University Medical Center), Xiao-Ou Shu (Vanderbilt University Medical Center), Qing Lan (National Cancer Institute), Qiuyin Cai (Vanderbilt University Medical Center), Jirong Long (Vanderbilt University Medical Center)

Abstract: Background: Transcriptome-wide association studies (TWAS) have uncovered over 100 putative risk genes for lung cancer. However, most of these studies relied on bulk RNA-seq data with limited resolution of cell-type-specific genetic effects on gene expression and lung cancer risk.

Methods: We conducted a single-cell TWAS by integrating genotype and single-cell RNA-seq data from lung tissues of 112 non-cancer individuals with genome-wide association study (GWAS) data from ~1.7 million subjects (72,586 lung cancer cases). For each gene in each cell type across four major cell subgroups (immune, epithelial, endothelial, and mesenchymal), we built expression prediction models using elastic net and Group LASSO, retaining the model with better predictive accuracy. While both methods used the target cell type data, Group LASSO additionally leveraged information from other cell types with correlated genetically regulated expression. We applied these models to GWAS data using S-PrediXcan to test associations between genetically predicted gene expression and lung cancer risk, with Bonferroni correction applied within each cell type.

Results: Of 6,995 genes analyzed across 38 cell types, we built models for 6,758 (97%) genes, with a median of 2,592 models per cell type (interquartile range [IQR]: 2,465-2,746). At Bonferroni-corrected $P < 0.05$, a median of 16 genes per cell type (IQR: 15-19) were associated with lung cancer risk, totaling 213 unique gene-cancer associations, 180 (84.5%) of which showed heterogeneous effects across cell types (Bonferroni-corrected Cochran's $Q < 0.05$). Of these 213 associations, 109 (51%) and 31 (15%) showed moderate ($PP.H4 > 0.50$) and strong ($PP.H4 > 0.80$) evidence of colocalization, respectively. Notably, 147 (69%; 62 loci) of the 213 genes were not identified in prior bulk-tissue TWAS, including 30 genes at 21 putative novel loci (> 2 Mb from any GWAS-identified risk variant), such as TSG101 (11p15.1), HAUS4 (14q11.2), HSD3B7 (16p11.2), and GOSR1 (17q11.2), each showing $PP.H4 > 0.80$ in specific epithelial or immune cell types. Single-cell disease relevance score (scDRS) analyses further revealed five cell types with significantly enriched expression of lung cancer-associated genes, including CD4, CD8/natural killer T, plasma, and proliferating immune cells, and proliferating epithelial cells.

Conclusion: This study identified novel risk genes for lung cancer, many with cell-type-specific effects, advancing our understanding of lung cancer etiology.

The most comprehensive characterisation of polygenic risk for cutaneous melanoma to date

Subsession Time: Wednesday, October 15 at 2:00pm – 2:15pm

Authors: G J M Shanika R Jayasinghe (QIMR Berghofer Medical Research Institute), Rohit Thakur (National Cancer Institute, National Institutes of Health), Lee E Wheeles (Vanderbilt University Medical

Center), Huanwei Wang (QIMR Berghofer Medical Research Institute), Anne E Cust (University of Sydney), Sarah V Ward (University of Western Australia), Rajiv Kumar (German Cancer Research Center), Florence Demenais (Université Paris Cité, Inserm), Christopher I Amos (Baylor College of Medicine), Georgina V Long (University of Sydney), Serigne Lo (University of Sydney), Alisa M Goldstein (National Cancer Institute, National Institutes of Health), Nicholas K Hayward (QIMR Berghofer Medical Research Institute), Diane T Smelser (Geisinger Health System), Andrew Bakshi (Monash University), Jiali Han (Indiana University), David L Duffy (QIMR Berghofer Medical Research Institute), Miguel Renteria (QIMR Berghofer Medical Research Institute), Lars G Fritsche (University of Michigan), Paola Ghiorzo (University of Genoa), Eduardo Nagore (Instituto Valenciano de Oncología), Alex Stratigos (University of Athens), Susana Puig (University of Barcelona), Chiara Menin (Veneto Institute of Oncology - IOV IRCCS), Maria Concetta Fargnoli (University of L'Aquila), Cristina Pellegrini (University of L'Aquila), David Millan-Esteban (Universidad Católica de València San Vicente Mártir, València), Mai Xu (National Cancer Institute, National Institutes of Health), Phuc H Hoang (National Cancer Institute, National Institutes of Health), Kiarash Khosrotehrani (University of Queensland), Diptavo Dutta (National Cancer Institute, National Institutes of Health), David C Whiteman (QIMR Berghofer Medical Research Institute), Stuart MacGregor (QIMR Berghofer Medical Research Institute), Mark M Iles (University of Leeds), Mathias Seviiri (QIMR Berghofer Medical Research Institute), Jianxin Shi (National Cancer Institute, National Institutes of Health), Rebecca I Hartman (VA Boston Healthcare System), Kevin M Brown (National Cancer Institute, National Institutes of Health), Maria Teresa Landi (National Cancer Institute, National Institutes of Health), Matthew H Law (QIMR Berghofer Medical Research Institute)

Abstract: Cutaneous melanoma (CM) is the deadliest form of skin cancer. Although the heritability of CM is high (twin $h^2 = 58\%$), only ~54 risk loci have been identified to date, and much of the heritability for CM remains to be characterised. In addition to the benefits of a greater understanding of the biology of CM that comes with identifying new risk loci and genes, larger sample sizes enable better estimation of polygenic risk scores (PRS). Improved PRS for CM are of interest given a growing desire for targeted screening approaches to manage this deadly skin cancer.

We have expanded on previously included datasets and added new genome-wide association studies from cohorts from Australia (including from QIMR Berghofer, Melanoma Institute Australia and Asprey), the US (including the National Cancer Institute, Million Veteran Program, Geisinger, Michigan Genomics Initiative, All of Us, and the Partners biobank), Canada (CLSA) and Europe (MelaNostrum, FinnGen). Combining 35 European ancestry GWAS from 29 cohorts, we have more than doubled previous sample sizes (final total 70,966 cases and 1,597,619 controls).

This new analysis identifies 116 independent risk loci, and 52 of the 54 previously reported loci are replicated in this analysis. A follow-up multi-trait analysis of GWAS (MTAG) combining key CM risk factors, including nevus count and skin colour, added a further 30 potential risk loci. In an independent replication sample, the 30 multi-trait identified loci showed concordant effect sizes ($r=0.76$, 95% CI 0.55-0.88, $P=1.3e-6$). Amongst the new loci, we observe hits within or near multiple cancer driver genes, including NOTCH2 ($P=1.51e-08$), MDM4 ($P=1.31e-14$), TP63 ($P=6.89e-17$), PIK3CA ($P=4.79e-11$), and CDKN1B ($P=9.08e-18$). In addition, we observe a new signal over the HLA-A locus ($1.88e-09$), adding to evidence for an immune component to melanoma risk.

We validate a new PRS derived from this larger meta-analysis using sbayesRC, in independent sets from the US and Australia (AllofUS [US], PLCO [US], QSkin [Australia]). The new PRS significantly improved AUC above baseline models that include age and sex (AUC increase 0.06 ~ 0.08). This suggests that the PRS performs equally well across different at-risk European ancestry populations with different UV exposure. Excitingly, our predictive strength for CM (AUC 0.70) rivals or exceeds multifactorial models that use traditional risk factors (e.g. skin colour, tanning ability) recommended for use in melanoma screening.

Identification of 33 new germline risk loci for clonal hematopoiesis by integrating genetic data on blood cell counts and indices

Subsession Time: Wednesday, October 15 at 2:15pm – 2:30pm

Authors: George Richenberg (MRC Integrative Epidemiology Unit, University of Bristol, Population Health Sciences, Bristol Medical School, University of Bristol), Andrew J. Bass (Department of Medicine, University of Cambridge), Chris Wallace (Department of Medicine, University of Cambridge, MRC Biostatistics Unit, University of Cambridge), George Vassiliou (Cambridge Stem Cell Institute, University of Cambridge), Tom R. Gaunt (MRC Integrative Epidemiology Unit, University of Bristol, Population Health Sciences, Bristol Medical School, University of Bristol), Siddhartha P. Kar (MRC Integrative Epidemiology Unit, University of Bristol, Population Health Sciences, Bristol Medical School, University of Bristol, Early Cancer Institute, University of Cambridge)

Abstract: Clonal hematopoiesis (CH), the selective expansion of hematopoietic stem and progenitor cells harboring specific somatic mutations in leukemogenic driver genes is a major risk factor for blood cancers and linked to other diseases of aging. Over 90% of germline common variant risk loci for CH reported so far are associated with at least one blood cell count or index at $P < 5 \times 10^{-8}$. Here we leveraged this observed pleiotropy and applied the powerful new surrogate functional false discovery rate (sfFDR) method to conduct a multi-trait genome-wide association study (GWAS) of CH with 15 blood cell counts or indices. Specifically, we combined GWAS of CH in the UK Biobank (25,657 cases, including 16,219 DNMT3A-, 3,918 TET2- and 2,007 ASXL1-mutant CH cases, and 342,869 controls) with GWAS of 15 blood cell counts and indices ($n=562,132$) to identify 33 new genome-wide significant ($P < 5 \times 10^{-8}$) CH risk loci > 500 kb away from any previously reported CH risk locus. These included 13 with DNMT3A-, TET2-, or ASXL1-specific associations and 20 loci associated with overall CH risk. We replicated 20 of the 33 new risk loci in GWAS of CH in TOPMed (USA; 3,831 cases/61,574 controls) and/or MCPS (Mexico; 4,234 cases/132,167 controls). At nominal significance ($P < 0.05$), 22 of the 33 new CH risk loci were associated with leukocyte telomere length, a key etiological mechanism underpinning CH; 10 of the 33 were associated with risk of developing at least one myeloid cancer downstream of CH; 26 of 33 were associated with circulating levels of proteins involved in progression to myeloid cancers (CEBPA, FLT3, FLT3LG, KIT, KITLG, TP53); and 23 of 33 were associated with risk of acquiring hematopoietic autosomal and/or sex chromosomal mosaicism. Variant-to-gene annotation implicated genes involved in the regulation of hematopoiesis (CTCF, JARID2, NOTCH4), leukemia pathogenesis (AFF1, KITLG), DNA damage repair (HELLS, DNMT2) and cell growth/proliferation (PIK3R3, NDFIP1) at the new CH risk loci and proposed BCL2 inhibition as a potential therapeutic strategy for CH given that one of the new CH risk loci is at BCL2. Our results demonstrate that integrating blood cell count/index GWAS with CH GWAS using

the new sfFDR method dramatically boosts CH risk locus discovery, revealing novel germline genetic determinants of CH. These insights may inform the precision prevention and early interception of hematological malignancies in those at high risk.

Session 30: Hijacked Immunity: Navigating the Balance Between Host Defense and Immune Dysregulation

Location: Room 206AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 1:30pm – 2:30pm

★ Modeling Viral Evolution and T Cell Evasion through HLA-Informed Deep Learning

Subsession Time: Wednesday, October 15 at 1:30pm – 1:45pm

Authors: David J. Hamelin (Montreal Heart Institute, Université de Montréal, Montréal, Quebec, Canada, Mila - Quebec AI Institute, Montréal, Quebec, Canada, Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montréal, Quebec, Canada), Jean-Christophe Grenier (Montreal Heart Institute, Université de Montréal, Montréal, Quebec, Canada), Raphaël Poujol (Montreal Heart Institute, Université de Montréal, Montréal, Quebec, Canada), Benoîte Bourdin (CHU Sainte-Justine Research Center, Université de Montréal, Montréal, Quebec, Canada), Bastien Paré (CHU Sainte-Justine Research Center, Université de Montréal, Montréal, Quebec, Canada), Shawn Simpson (CHU Sainte-Justine Research Center, Université de Montréal, Montréal, Quebec, Canada), Martin Smith (Ramaciotti Centre for Genomics, School of Biotechnology and Biomolecular Sciences, Faculty of Science, UNSW Sydney, Kensington, Australia, RNA Institute, UNSW Sydney, Kensington, Australia, Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montreal, Canada, Cancer and Immune Diseases Axis, CHU Sainte-Justine Research Centre, Montreal, Canada), Hélène Decaluwe (CHU Sainte-Justine Research Center, Université de Montréal, Montréal, Quebec, Canada, Microbiology, Infectiology and Immunology Department, Faculty of Medicine, Université de Montréal, Montreal, Canada, Department of Pediatrics, Faculty of Medicine, Université de Montréal, Montréal, QC, Canada), Etienne Caron (Yale Center for Immuno-Oncology, Yale Center for Systems and Engineering Immunology, Yale Center for Infection and Immunity, Yale School of Medicine, New Haven, CT, USA, CHU Sainte-Justine Research Center, Université de Montréal, Montréal, Quebec, Canada), Julie Hussin (Montreal Heart Institute, Université de Montréal, Montréal, Quebec, Canada, Mila - Quebec AI Institute, Montréal, Quebec, Canada, Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montréal, Quebec, Canada, Department of Medicine, Faculty of Medicine, Université de Montréal, Montréal, QC, Canada)

Abstract: Understanding how viral evolution interacts with human immunity is key to future pandemic preparedness. Human genetic variation, particularly in HLA class I genes, plays a critical role in shaping T cell responses to viral infections. T cells play a critical role in clearing infections by recognizing viral epitopes presented by HLA molecules, a process shown to be essential during the COVID-19 pandemic. Due to HLA variability, T cell responses to infection vary widely across individuals. Accumulating evidence shows that viral mutations can disrupt epitope presentation and enable immune escape in an HLA-dependent manner. Our group previously demonstrated this using predictive models and binding assays.

Several Omicron variants, for instance, evade T cell recognition in specific HLA contexts, highlighting how host genetics can shape viral evolution.

We present an AI-driven framework to study the interplay between viral mutations and HLA variability across populations. First, we created a machine learning framework to assess and visualize HLA class I haplotype diversity across populations, visualizing immunogenetic clustering across global cohorts. Next, we developed a framework to estimate HLA-specific susceptibility to T cell evasion. Using published and custom algorithms, we quantified epitope loss (ie. how viral variants disrupt T cell epitope) across six HLA class I alleles in the UK Biobank and 1000 Genomes cohorts. We identified up to six immunogenic epitopes affected by escape mutations in individuals with HLA-B07:02 and A03:01. These predictions were validated in the RECOVER-2 cohort (n = 576), fully HLA-typed via nanopore sequencing, and confirmed functionally by ELISpot assays. Finally, to forecast T cell epitope diversification, we developed an epistasis-aware deep learning framework based on protein language models (ESM-2), intra-host mutation data, and known evolutionary drivers. This model predicts emerging mutations that may impact T cell epitope recognition across diverse HLA profiles .

By integrating immunogenetics, population-scale HLA diversity, and viral evolution, our framework offers a new lens for understanding host-pathogen dynamics. This work also supports more inclusive public health strategies and vaccine design, grounded in human global genetic diversity.

★ **East Asian GWAS of serum immunoglobulin and subclass levels provides insights into the genetic architecture of humoral immunity implicated with immune cell type specificity**

Subsession Time: Wednesday, October 15 at 1:45pm – 2:00pm

Authors: Kei Sugahara (the University of Tokyo, Development of Genome Informatics), Kyuto Sonehara (the University of Tokyo, Development of Genome Informatics), Yoshifumi Uwamino (Keio University, Department of Laboratory Medicine), Shiori Yoshikawa (Keio University, Department of Infectious Diseases), Ryuya Edahiro (Osaka University, Department of Statistical Genetics), Tatsuhiko Naito (Osaka University, Department of Statistical Genetics), Yukinori Okada (the University of Tokyo, Development of Genome Informatics), Ho Namkoong (Keio University, Department of Infectious Diseases)

Abstract: Introduction: Immunoglobulins (Igs) are key components of humoral immunity, playing essential roles in protection against pathogens, regulation of autoimmunity, allergy, and cancer immunity. Serum Ig levels vary substantially between individuals, partly due to genetic factors. However, genome-wide association studies (GWAS) focusing on serum Ig levels, particularly IgG subclasses, in non-European populations remain limited. This study aimed to identify genetic loci associated with serum Ig levels in Japanese cohorts and to elucidate cell-type-specific regulatory mechanisms using single-cell transcriptomic data. Methods: We performed GWAS of serum levels of IgM, IgA, IgG, and IgG subclasses (IgG1–4) in 2,129 Japanese individuals. Additionally, we constructed ratio-based composite traits (IgA/IgM, IgG/IgM, IgA/IgG) and conducted GWAS for these derived phenotypes. We assessed

colocalization between GWAS signals and expression quantitative trait loci (eQTLs) from single-cell RNA sequencing (scRNA-seq) data of peripheral blood mononuclear cells (PBMC) to identify cell type-specific regulatory associations. We further integrated GWAS results with PBMC scRNA-seq data using the single-cell disease relevance score (scDRS) to evaluate enrichment of genetic signals in specific immune cell populations. A multi-ancestry meta-analysis integrating previous European results was performed using IgG subclass GWAS data (n=11,033). Results: We identified eight genome-wide significant variants associated with serum Ig levels, including one novel variant of IgM at 19p13. One variant is in the human leukocyte antigen (HLA) region and three is in the Ig heavy chain (IGH) locus. GWAS for composite traits revealed three additional significant variants that were not detected in any of the individual Ig traits. Colocalization analyses with PBMC scRNA-seq eQTLs showed that five loci colocalized with eQTLs in B cells, monocytes, and dendritic cells (PP.H4>0.8), suggesting cell-type-specific effects. scDRS analysis demonstrated significant enrichment of GWAS signals in B cells and dendritic cells. A multi-ancestry meta-analysis of IgG subclass GWAS identified three additional novel loci. Conclusion: Our findings provide a comprehensive view of the genetic architecture underlying serum Ig levels in East Asians and highlight the contribution of specific immune cell types to their regulation.

Using functional genomics to understand variation in response to infection in northeast Thailand

Subsession Time: Wednesday, October 15 at 2:00pm – 2:15pm

Authors: Emma E Davenport (Wellcome Sanger Institute), Phumrapee Boonklang (Wellcome Sanger Institute, Mahidol Oxford Tropical Medicine Research Unit), Arin Wongprommoon (Wellcome Sanger Institute, Mahidol Oxford Tropical Medicine Research Unit), Wanseon Lee (Wellcome Sanger Institute), Katie L Burnham (Wellcome Sanger Institute), Claire Cheewapreecha (Wellcome Sanger Institute, Mahidol Oxford Tropical Medicine Research Unit)

Abstract: Background: Melioidosis is a neglected infectious disease caused by the intracellular bacterium *Burkholderia pseudomallei*. It is endemic across the tropics, mainly affecting low-to-middle-income countries. The disease results in heterogeneous clinical outcomes, including sepsis. We have previously identified transcriptomic sepsis response signature (SRS) endotypes in UK adults with sepsis that are associated with immune response and mortality but have not assessed their presence in other populations and hospital settings. To determine the influence of host genetics on the variable response to melioidosis infection, we recruited a cohort from northeast Thailand through the BurkHostGen study.

Methods: We performed whole genome sequencing (WGS, 15x) and bulk RNA-sequencing on a cohort of adult patients with melioidosis (n=371), other infections (n=158) and non-infected comparators (n=492). We integrated the WGS data with global data from the 1,000 Genome Project and performed principal component analysis, estimated ancestry with ADMIXTURE and calculated population-stratified minor allele frequencies (MAF). HLA typing was conducted with HLA*LA. For the RNA-sequencing data, we assigned SRS endotypes (Sepstratifier), performed differential gene expression analysis (limma) and inferred immune cell proportions (CIBERSORTx).

Results: As expected, this northeast Thai population is most genetically similar to the 1,000 Genomes East Asian subpopulation. However, some variants show a highly divergent MAF. We also observed a high frequency of the HLA-B*13:01 allele, which has been reported to be associated with severe cutaneous adverse reactions to co-trimoxazole.

From the transcriptomic data, patients assigned to SRS1 had more severe infection (Kruskal-Wallis test: SOFA $p=1.1e-11$; NEWS2 $p=2.5e-14$) and a higher mortality rate than non-SRS1 patients (log-rank test $p=6.58e-06$) replicating our findings from our UK cohort. Similarly, the 4,502 differentially expressed genes (SRS1 vs non-SRS1) were highly correlated ($r=0.73$) with the previously identified differences and we replicated the associations with inferred cell proportions, including expansion of immature neutrophils in the SRS1 endotype.

Conclusions: Here we increase genetic data representation from northeast Thailand and successfully stratify patients in a non-European bacterial infection cohort. Future work will integrate these datasets to understand the genetic variants driving the varied response to infection.

Gene regulation in T cells depends on the severity of atopic dermatitis

Subsession Time: Wednesday, October 15 at 2:15pm – 2:30pm

Authors: Cristian Valencia (J&J Innovative Medicine), Connor Hill (J&J Innovative Medicine), Hurley Li (J&J Innovative Medicine), Christopher Sisk (J&J Innovative Medicine), Monica Leung (J&J Innovative Medicine), Lam C Tsoi (Department of Dermatology, University of Michigan), Allison C Billi (Department of Dermatology, University of Michigan), Johann E Gudjonsson (Department of Dermatology, University of Michigan), Emanuel Maverakis (Department of Dermatology, University of California, Davis), Marta E Polak (J&J Innovative Medicine), Dawn Waterworth (J&J Innovative Medicine).

Abstract: Atopic dermatitis (AD) is a complex skin disease with features of eczema, dry skin, and itching and affects 10-20% of the general population. AD may be exacerbated by environmental factors and infection, and cell-types, such keratinocytes and T cells, are involved in its pathogenesis. Genome-wide association studies (GWAS) have identified over 60 AD-linked loci, and integrating expression quantitative trait loci (eQTLs) has helped prioritize disease susceptibility genes. However, most eQTL studies are done in healthy tissue and previous work has demonstrated that eQTLs can be context-dependent on factors such as tissue or disease. Furthermore, clinically AD is heterogeneous and severity is assessed using continuous clinical scores such as Scoring AD (SCORAD) index and Eczema Area and Severity Index (EASI), but molecular markers of disease heterogeneity remain limited. Single cell RNA-seq (scRNA-seq) can assess disease-state heterogeneity at single cell resolution. We show that studying eQTLs with scRNA-seq from skin, accounting for disease-state heterogeneity, could help prioritize genes to facilitate better understanding of AD. We integrated genetic and scRNA-seq data from skin biopsies of 63 AD and 26 healthy donors, to study the effect of continuous disease-states on gene regulation in 91,177 T cells. Each cell was decomposed based on a reference of continuous gene expression programs (GEPs) using consensus non-negative matrix factorization (Tcat). We then used stepwise regression to

identify GEPs associated with AD (AD-GEPs). Notably, the GEP CD40LG, indicative of CD4+T cell activation for antibody production, had higher values in AD vs healthy ($P=8.51e-02$) samples. Furthermore, the CD8+T tissue resident memory (CD8.Trm) GEP ($P=1.88e-06$) and CD172K.Mertk GEP (associated with an immunosuppressive environment) were lower in AD compared to healthy ($P=5.52e-09$) samples. These GEPs correlated with disease severity (e.g., CD40LG with EASI, $r = 0.68$). These findings support use of AD-GEPs as molecular surrogates of the continuous disease-state of a cell. To find eQTLs with different magnitudes of effect depending on disease-state, we fitted a Poisson mixed effect model of gene expression in single cells as a function of genotype and genotype interactions with the AD-GEPs. We found that 43% of eQTLs (328/763) had disease-state-dependent effects. Among the most significant results, the eQTL effect of rs583887 for CTSW was stronger in more severe disease (eQTLPercentile25_CD40LG=0.35 $P=1.92e-07$; eQTLPercentile75_CD40LG=0.79, $P=1.24e-14$) (LRT $P=1.44e-141$). Similarly, the eQTL rs2468402 for IL22 was only detected at high values of the CD40LG GEP (eQTLPercentile85_CD40LG=1.15, $P=2.10e-02$) (LRT $P=3.69e-158$), suggesting that IL22 may be a more relevant target in severe AD. Our results highlight the importance of studying disease-dependent gene regulation. This approach may aid in identification of genes involved in disease severity.

Session 31: Making a Connection: Pleiotropy in Neurologic and Psychiatric Conditions

Location: Room 210AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 1:30pm – 2:30pm

Gene discovery and pleiotropic architecture of chronic pain in a genome-wide association study of over 1.2 million individuals

Subsession Time: Wednesday, October 15 at 1:30pm – 1:45pm

Authors: Sylvanus Toikumo (Crescenz VAMC, University of Pennsylvania), Christal Davis (Crescenz VAMC, University of Pennsylvania), Zeal Jinwala (University of Pennsylvania), Yousef Khan (University of Pennsylvania), Mariela Jennings (University of California, San Diego), Lea Davis (Vanderbilt University), Sandra Sanchez-Roige (University of California, San Diego), Rachel L. Kember (Crescenz VAMC, University of Pennsylvania), Henry R. Kranzler (Crescenz VAMC, University of Pennsylvania)

Abstract: Chronic pain is highly prevalent worldwide, and genome-wide association studies (GWAS) have identified a growing number of chronic pain loci. To further elucidate its genetic architecture, we leveraged data from 1,235,695 European ancestry individuals across three biobanks. In a meta-analytic GWAS, we identified 343 independent loci for chronic pain, 92 of which are new. To investigate potential sex-specific genetic effects, we performed separate meta-analyses for males and females. Among 583,066 male participants, we identified 115 independent loci linked to chronic pain, 12 of which were novel. In contrast, the analysis of 241,266 female participants revealed 12 associated loci, including two novel ones. Multi-omics gene prioritization analyses, which incorporate data on gene expression and regulation in both brain and blood tissues, identified a total of 490 genes that likely contribute to chronic pain pathophysiology. Loci associated with increased risk for chronic pain were also associated with increased risk for multiple other traits, and Mendelian randomization analyses showed that chronic pain was causally associated with psychiatric disorders, substance use disorders, and levels of C-reactive protein—an inflammatory biomarker. Chronic pain variants overlapped genetically with brain structural traits, particularly cortical surface area, supporting a neurobiological role in chronic pain risk. This study expands our knowledge of the genetics of chronic pain and its pathogenesis, highlighting the importance of its pleiotropy with psychiatric disorders and inflammation and elucidating its multi-omic pathophysiology.

Mapping the Pleiotropic Genetic Landscape of Brain Imaging Traits Using DIMPLE-GWAS and SuSie-GeneSet

Subsession Time: Wednesday, October 15 at 1:45pm – 2:00pm

Authors: Max Lam (LKC Medicine, Feinstein Institutes for Medical Research), Upasana Bhattacharyya (Feinstein Institutes for Medical Research), Jibin John (Feinstein Institutes for Medical Research), Zhong Yuanxin (MGH Harvard Medical School), Chia-Yen, Chen (Merck Inc.), Tian, Ge (Harvard Medical School), Todd Lencz (Feinstein Institutes for Medical Research, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell)

Abstract: Identifying shared genetic architecture across high-dimensional neuroimaging traits remains a key challenge in genomics. We present a novel pipeline—DIMPLE-GWAS (Dimensionality reduction and Integrated Multi-Phenotype Landscape Explorer for GWAS)—which integrates UMAP and HDBSCAN clustering with genetic principal component analysis to identify pleiotropic loci across imaging-derived phenotypes (IDPs). We analyzed GWAS summary statistics from 2,315 IDPs across multiple MRI modalities (T1w, DTI, fMRI) from the UK Biobank, identifying 24 genetically homogeneous clusters. These clusters recapitulated known lobar brain anatomy and revealed previously unappreciated genetic dissociations across dorsal/ventral cortical streams, white matter tracts, and functional networks.

Genomic PCA-based GWAS of each cluster improved heritability estimates (h^2 range = 0.0717–0.4099; median = 0.267) relative to univariate IDP GWAS (median = 0.185), and identified 422 genome-wide loci (86 shared, 189 unique). To interpret the underlying biology, we developed SuSie-GeneSet, a fine-mapping-informed gene set analysis tool leveraging SuSiE posterior inclusion probabilities. Together with FLAMES, MAGMA, and SMR, we prioritized putative causal genes and identified enriched pathways including immune regulation, synaptic signaling, and mitochondrial function.

Mendelian randomization analyses revealed putative causal effects of specific imaging clusters on metabolic (BMI), growth (height), and cognitive traits (reaction time, cognitive task performance), while LDSC-based genetic correlations implicated shared genetic architecture with stroke, Parkinson's disease, educational attainment, and sleep phenotypes. Single-cell transcriptomic enrichment across 461 brain cell types (Duncan et al., 2025) revealed 13 independently associated populations. Oligodendrocyte lineage cells were strongly associated with white matter clusters, while astrocytes and vascular cells were linked to occipital and subcortical volumes.

This integrative framework reveals a multidimensional pleiotropic structure underlying brain imaging traits, offering scalable tools for mechanistic discovery at the interface of brain structure, function, and complex disease.

A cell-type-resolved deep learning atlas of human cis-regulatory evolution

Subsession Time: Wednesday, October 15 at 2:00pm – 2:15pm

Authors: Nikitha Thoduguli (MIT), Riley J. Mangan (MIT, Broad Institute, HMS), Jayashabari Shankar (MIT), Yuru Lin (MIT), Zunpeng Liu (MIT, Broad Institute), Manolis Kellis (MIT, Broad Institute)

Abstract: As humans and great apes are distinguished by extensive phenotypic changes with relatively few protein-coding differences, deciphering the regulatory consequences of sequence divergence is essential to understanding the molecular basis of human-specific traits and disease susceptibilities. Although millions of derived alleles distinguish modern humans from great apes, only a small fraction

are likely to influence human-specific traits. Previous studies have focused on regions with high sequence divergence, based on the idea that rapid molecular evolution reflects functional adaptation. However, individual regulatory mutations that influence gene expression may evade detection in divergence-based scans. To address these limitations, we developed a deep learning-based framework to reconstruct cis-regulatory changes across 6 million years of hominin evolution. Using personalized reference-guided genome assemblies, we applied the pre-trained Enformer model to predict 3,400 genome-wide chromatin accessibility profiles across 100 distinct cell and tissue epigenomic contexts for 34 modern human, archaic hominin, and great ape haplotypes. We predict 37,548 human-ape differentially accessibility (DA) events (in 10,918 regions), and 695 DA modern-archaic events (in 231 regions). We validate our predicted DA elements using experimentally-observed interspecies regulatory differences and proximity to human-ape differentially expressed genes. Our predicted DA elements tend to be shorter, less conserved, and less pleiotropic, with many occurring in regions with limited sequence divergence that would be overlooked in divergence-based approaches. Consistent with known phenotypic transitions, our predicted DA events are enriched near genes in neurogenic and developmental pathways. We highlight specific examples where predicted regulatory divergence is attributable to alterations in transcription factor binding sites, including a hominin-specific neurodevelopmental intronic enhancer in the *KCNS1* locus, which encodes a voltage-gated potassium channel involved in nociceptive regulation. Overall, our results reveal extensive cis-regulatory divergence in the human lineage and establish a computational framework for prioritizing variants underlying human-specific traits and disease risk.

A deep cellular atlas of the human ventral substantia nigra in Parkinson's identifies a genetic and molecular overlap with Type 2 Diabetes

Subsession Time: Wednesday, October 15 at 2:15pm – 2:30pm

Authors: Viola Volpato (UKDRI, Cardiff University), David A. Menassa (Nuffield Department of Clinical Neurosciences, University of Oxford), Preethi Sheshadri (UKDRI, Cardiff University), Stefania Giussani (Kavli Institute for Nanoscience Discovery, University of Oxford), Michal Rokicki (UKDRI, Cardiff University), Lucia F. Cardo (UKDRI, Cardiff University), Marirena Bafaloukou (UKDRI, Imperial College London), Ann-Kathryn Schalkamp (UKDRI, Imperial College London), Agata Zaremba (UKDRI, Cardiff University), Jimena Monzón-Sandoval (UKDRI, Cardiff University), Ngoc-Nga Vinh (Centre for Neuropsychiatric Genetics and Genomics, Cardiff University), Joanne Morgan (Centre for Neuropsychiatric Genetics and Genomics, Cardiff University), Michele T.M. Hu (Oxford Parkinson's Disease Centre, University of Oxford), Scott Miners (Cerebrovascular Dementia Group, University of Bristol), Richard Wade-Martins (Kavli Institute for Nanoscience Discovery, University of Oxford), Cynthia Sandor (UKDRI, Imperial College London), Laura Parkkinen (Nuffield Department of Clinical Neurosciences, University of Oxford), Caleb Webber (UKDRI, Cardiff University)

Abstract: Parkinson's disease (PD) is a neurodegenerative disorder marked by the loss of specific dopaminergic neurons (DaNs) in the ventral substantia nigra pars compacta (SNpc). In particular, DaNs marked by AGTR1+ expression are highly vulnerable to loss. Epidemiological studies link Type 2 Diabetes (T2D) to an increased PD risk, suggesting shared pathological mechanisms, with >50% of PD patients

demonstrating insulin resistance. However, the underlying molecular mechanisms remain poorly understood, hindering effective treatment development.

Here, we generate the first deep (SMARTseq) single cell atlas characterising a neurological disease providing unparalleled resolution of gene expression, and combine this atlas with novel GWAS to reveal shared genetic and molecular mechanisms with T2D that underlie the fundamental vulnerability of nigral dopaminergic neurons in PD.

By sequencing nuclei from human ventral SNpc across PD Braak stages, we find that PD genetic risk converges onto three cell types: AGTR1+ dopaminergic neurons, perineuronal (non-myelinating) oligodendrocytes and oligodendrocyte precursor cells. Both AGTR1+ neurons and perineuronal oligodendrocytes are lost in PD, validated through independent neuropathology. Disease-associated isoform switching proposes alterations in the OPC to ODC subtype differentiation away from perineuronal oligodendrocytes. Notably, PD risk is concentrated within cellular interactions between these three cell types, highlighting a vulnerable and targetable network disrupted early in PD.

AGTR1+ dopaminergic neurons represent a metabolically stressed state, not a distinct cell type, characterized by activation of the renin-angiotensin (RAS) and MAPK pathways, oxidative stress, mitochondrial dysfunction, and express targets of metabotropic drugs including angiotensin receptor blockers and GLP1R/GIPR agonists. We demonstrate this state can be invoked within iPSC-derived dopaminergic neurons. This state is also associated with T2D genetic risk. A GWAS of comorbid PD and T2D reveals genome-wide significant associations within the AGTR1 and TCF7L2 loci, providing a mechanistic link between diseases. Comorbid PD/T2D risk is associated with multiple PD rare variant genes that cause non-Lewy body PD, mechanistically stratifying PD.

These findings identify previously underappreciated cell types and intercellular networks central to PD pathogenesis, revealing a neuronal vulnerability associated to T2D that appears therapeutically targetable.

Session 32: Precision Risk for Every Brain

Location: Room 156ABC/Level 1, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 1:30pm – 2:30pm

Genome-wide multi-trait and multimodal single-cell analyses of tobacco smoking and schizophrenia reveal novel risk loci and regulatory mechanisms

Subsession Time: Wednesday, October 15 at 1:30pm – 1:45pm

Authors: Meghan J. Chenoweth (University of Toronto, Centre for Addiction and Mental Health), Suzan Farhang-Sardroodi (University of Toronto), Saori Sakaue (Brigham and Women's Hospital), Kathryn Weinand (Brigham and Women's Hospital), Soumya Raychaudhuri (Brigham and Women's Hospital, Broad Institute, Harvard Medical School), Jennie G. Pouget (University of Toronto, Centre for Addiction and Mental Health), Rachel F. Tyndale (University of Toronto, Centre for Addiction and Mental Health)

Abstract: Background: Tobacco smoking and schizophrenia (SCZ) are highly heritable and comorbid disorders that cause significant morbidity and mortality. We hypothesized that genome-wide multi-trait analysis would identify novel risk loci not captured in existing GWAS, and that candidate regulatory mechanisms could be elucidated using multimodal single-cell data. Methods: We applied multi-trait analysis of GWAS (MTAG) to GWAS data from Europeans: SCZ (53,386 cases; 77,258 controls), cigarettes per day (CigDay; n=618,489), smoking cessation (SmkCes; n=1,147,272), and two smoking biomarkers (cotinine, COT; and nicotine metabolite ratio, NMR; n=5,185 each). We queried the “druggability” of novel mapped genes using the Drug-Gene Interaction Database (DGIdb). To functionally characterize novel risk loci, we applied SCENT to paired transcriptional (sn-RNAseq) and epigenetic (sn-ATACseq) profiles from dorsolateral prefrontal cortex (DLPFC) to construct cell type-specific enhancer-gene maps across six brain cell types. Fine-mapped MTAG SNPs with $PIP > 0.03$ were then intersected with significant peak-gene pairs identified by SCENT. Results: MTAG identified 46 novel loci for CigDay and 28 for SmkCes [genome-wide significant in MTAG ($P < 5 \times 10^{-8}$) but not in GWAS], with no new loci identified for SCZ, COT, or NMR. Three CigDay genes (TSPAN5, EPHA3, MTHFD1) and two SmkCes genes (TAC1, HIST1H2BD) had known drug interactions. One of the genes, TAC1, interacted with naltrexone. Overall, five of the fine-mapped MTAG SNPs were prioritized using cell type-specific enhancer gene mapping, including SNPs located within putative inhibitory neuron-specific enhancer regions for ST3GAL3 (rs839764; significant for SmkCes) and SYNE2 (rs12879827; significant for CigDay), located 457kb 5' and 278kb 3' of the respective genes. Conclusions: This multi-trait GWAS and multimodal single-cell analysis revealed novel genes implicated in tobacco smoking, the leading cause of preventable illness and death in the United States. The interaction between TAC1 and naltrexone, a treatment for alcohol and opioid use disorder, suggests this medication may be repurposed for tobacco use disorder. Rare damaging variants in ST3GAL3 and SYNE2, genes prioritized using cell type-specific enhancer-gene maps, have been previously associated with neurodevelopmental disorders, lending orthogonal support for their relevance to brain-related traits and the regulatory mechanisms we uncovered.

Genome Wide Association Study of Schizophrenia identifies African-enriched variants in PTPRD and TMEM150C

Subsession Time: Wednesday, October 15 at 1:45pm – 2:00pm

Authors: Lerato Majara (Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA, Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA), Toni Boltz (Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA, Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA), Dickens Akena (Department of Psychiatry, School of Medicine, College of Health Sciences, Makerere University, Kampala, Uganda), Lukoye AtwoliBizu (Department of Mental Health and Behavioural Sciences, School of Medicine, Moi University College of Health Sciences, Eldoret, Kenya), Gelaye (Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA), Symon Kariuki (Neurosciences Unit, Clinical Department, KEMRI-Wellcome Trust Research Programme-Coast, Kilifi, Kenya), Charles R.J. Newton (Department of Psychiatry, University of Oxford, Oxford, UK), Dan Stein (Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa), Anne Stevenson (Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA, Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA), Rocky E Stroud (Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA), Lori Chibnik (Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA), Solomon Tefera (Department of Psychiatry, School of Medicine, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia), Zukiswa Zingela (Executive Dean's Office, Faculty of Health Sciences, Nelson Mandela University, Gqebera, South Africa), Karestan C. Koenen (Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA), Alicia R. Martin (Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA, Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA) & NeuroGAP-Psychosis Study Team

Abstract: Schizophrenia is a highly heritable severe mental illness consistently affecting ~1% of populations worldwide with devastating effects. Genetic studies of schizophrenia have largely involved European and East Asian ancestry individuals to date. These studies have shown remarkably similar genetic architecture between these populations, indicating that their genetic etiology is consistent. However, African populations, despite being the most genetically diverse, are vastly underrepresented in large-scale studies. Here, we present results from the Neuropsychiatric Genetics in African Populations (NeuroGAP) - a study from five sites: Ethiopia, South Africa, Uganda, Kenya (two sites - KEMRI and Moi) including ~42,000 Africans with schizophrenia (SCZ), bipolar disorder (BD) and controls. We performed genome-wide association study (GWAS) in each NeuroGAP site separately, and then combined them using a fixed effects meta-analysis using inverse-variance weighting. We also conducted cross-ancestry meta-analyses with summary statistics from Psychiatric Genomics Consortium (PGC) for schizophrenia and bipolar disorder. Additionally, we evaluated the prediction accuracy of PRS across ancestries using the PGC summary statistics and within-AFR ancestry per site, measuring accuracy by genetic divergence.

Our analyses identified associations for variants rs377703402 ($\beta=0.46$, PNGP-META= $1.22e-07$, gnomAD MAF=0.0003 in AFR and absent elsewhere) in TMEM150C and rs190966045 ($\beta = 0.72$, PNGP-META = $1.79e-07$, gnomAD MAF=0.02 in AFR and ≤ 0.002 elsewhere) in PTPRD. We also found consistent direction of effect between the summary statistics from the PGC for schizophrenia ($\beta=0.88$, $P=0.00045$) and bipolar disorder ($\beta=1.23$, $P=0.0164$) for the independent genome-wide significant variants in PGC, providing further evidence for shared biology across ancestries. However, the cross-ancestry meta-analyses between PGC with NeuroGAP yielded fewer genome-wide associated independent loci, suggesting potential phenotypic heterogeneity or underlying population differences. PRS prediction accuracy was generally low and varied across sites, with Ethiopia showing the highest accuracy (adjusted $R^2=0.03$, $P=1.84e-113$ for SCZ, and adjusted $R^2=0.006$, $P=0.25$ for BD) and Uganda the lowest (adjusted $R^2=0.00$, $P=7.13e-24$ for SCZ and adjusted $R^2=0.001$, $P=0.11$ for BD). These results mirrored genetic divergence estimates – Ethiopia had the smallest F_{ST} with Europeans ($F_{ST} = 0.045$) and Uganda ($F_{ST} = 0.12$) the largest. Within-ancestry PRS improved prediction accuracy for BD but not for SCZ, suggesting that phenotypic heterogeneity and sample size impact prediction accuracy. Large genetic studies from diverse populations provide the best starting point to disentangle molecular insights into severe mental illness etiology. The inclusion of African data from NeuroGAP-Psychosis has revealed novel loci and uncovered polygenic risk prediction complexity, ultimately contributing to better global health outcomes in mental health.

Rare Genetic Variation Illuminates Bipolar Disorder Risk Across Diverse Populations

Subsession Time: Wednesday, October 15 at 2:00pm – 2:15pm

Authors: Cal Liao (Broad Institute of MIT and Harvard), Bipolar Exome Consortium (BipEx, Broad Institute of MIT and Harvard), Hilary Finucane (Broad Institute of MIT and Harvard), Michael Talkowski (Massachusetts General Hospital), Benjamin Neale (Broad Institute of MIT and Harvard)

Abstract: Bipolar disorder (BD) is a highly heritable yet genetically complex psychiatric condition. However, the landscape of rare coding variation influencing BD risk remains incompletely understood. Here, the Bipolar Exome (BipEx) Consortium presents a large-scale effort—expanding sample size fivefold and enhancing ancestral diversity—to map the constellation of rare genetic risk factors contributing to BD. We analyzed exome sequencing data from 195,257 individuals, including 45,479 BD cases and 149,778 controls. Forty percent of participants were of non-European ancestry, with new recruitment spanning Latin America, Africa, and Asia. Ultra-rare protein-truncating variants (PTVs) and damaging missense mutations (minor allele count <10) were prioritized through stringent quality control. Rare exonic CNVs (>3 exons) were called using GATK-gCNV. To identify convergent biological signals, we mapped missense variants onto protein 3D structures using AlphaFold, assessing spatial clustering across the proteome. We identified 29 genes surpassing FDR significance, including AKAP11, SHANK1, ATP2B2, KDM5B, and DOP1A. These genes cluster into early and late neurodevelopmental expression patterns and implicate pathways linked to treatment response, including clozapine pharmacodynamics. Damaging missense variants in BD cases selectively clustered within constrained regions of ATP2B2, while 3D modeling revealed significant missense hotspots in ATP2A2, KDM5B, and G3BP2. CNV analyses identified

BPNT2—a lithium-responsive gene—as the strongest deletion signal (OR: 9.04, $P = 2.67 \times 10^{-9}$) and BIRC6 as the top duplication hit (OR: 5.04, $P = 2.49 \times 10^{-14}$). A reciprocal CNV association at 15q11.2 was observed, with duplications elevating BD risk and deletions depleted among cases.

Leveraging the largest and most ancestrally diverse exome dataset for BD to date, BipEx 2.0 reveals new rare variant associations spanning PTVs, CNVs, and spatially convergent damaging missense variants. These findings refine the genetic architecture of BD and highlight novel biological pathways for future mechanistic and therapeutic investigation.

Leveraging genome assembly and pangenome graphs to investigate an African-origin protective haplotype for Alzheimer's Disease in APOEε4 Carriers

Subsession Time: Wednesday, October 15 at 2:15pm – 2:30pm

Authors: Luciana Bertholim Nasciben (John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA), Derek Van Booven (John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA), Sofia Moura (John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA), Aura M. Ramirez (John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA), Griffin S. Parker (John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA), Karen Nuytemans (John P. Hussman Institute for Human Genomics and Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA), Derek M. Dykxhoorn (John P. Hussman Institute for Human Genomics and Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA), Farid Rajabli (John P. Hussman Institute for Human Genomics and Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA), Liyong Wang (John P. Hussman Institute for Human Genomics and Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA), William K. Scott (John P. Hussman Institute for Human Genomics, Dr. John T. Macdonald Foundation Department of Human Genetics, and Brain Endowment Bank, Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA), David A. Davis (Brain Endowment Bank, Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA), Regina T Vontell (Brain Endowment Bank, Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA), Katalina McInerney (Department of Neurology, Miller School of Medicine, Miami, FL, USA), Michael Cuccaro (John P. Hussman Institute for Human Genomics and Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA), Goldie Byrd (Maya Angelou Center for Health Equity, Wake Forest University, Winston-Salem, NC, USA), Jonathan Haines (Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, Ohio, USA), Larry Adams (John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA), Margaret A Pericak-Vance (John P. Hussman Institute for Human Genomics and Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA), Juan I Young (John P. Hussman Institute for Human Genomics and Dr. John T. Macdonald

Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA), Anthony J. Griswold (John P. Hussman Institute for Human Genomics and Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA) and Jeffery M. Vance (John P. Hussman Institute for Human Genomics and Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA)

Abstract: Background: We have previously reported (Rajabli et al., 2022) a statistical interaction between the African ancestry specific A allele at rs10423769 and APOE ϵ 4, where the presence of the A allele is associated with a reduction in AD risk up to 75% in APOE ϵ 4 homozygotes. The mechanism by which this variant confers protection could provide insights into new therapeutics for APOE ϵ 4 carriers. However, the genomic context of this protective variant is complex. The protective locus is located 2 MB upstream from APOE in an area of segmental duplications (SD), within a cluster of PSG genes (pregnancy-specific glycoproteins) and a long non-coding RNA. Here, we have used long-read sequencing to investigate the region between rs10423769_A and APOE to gain further insights into potential AD-protective mechanisms.

Methods: We used high-coverage PacBio Revio whole genome sequencing (three runs) on 9 samples of rs10423769_A/A, rs10423769_A/G, and rs10423769_G/G carriers. We performed genome assembly with Hifiasm to identify contigs spanning the whole region between the protective allele and APOE. We compared the contigs using Minigraph and Bandage. The area of segmental duplication from carriers and non-carriers of the protective locus was analyzed with Biser.

Results: The average coverage for the 5 MB region (chr19:39,980,000–45,146,000), including the protective locus and APOE, was $60.5 \pm 10.8X$. Genome assembly and sequence alignment revealed the co-occurrence of an expanded VNTR containing multiple MEF2D transcription factor binding sites within the protective haplotype. A higher number of structural variants (SVs) in the SD region surrounding the rs10423769_A protective haplotype were observed compared to the non-protective haplotype and are currently under investigation.

Conclusions: Our results show that the protective haplotype carries an expanded VNTR that is specific to this haplotype and contains multiple transcription factor binding sites. Differences in SVs could affect chromatin structure or regulatory elements in the region, potentially contributing to the protective mechanism of this African-origin locus. These insights may ultimately pave the way for therapeutic strategies targeting APOE ϵ 4 carriers.

Session 33: RNA Functions Beyond Coding Sequence

Location: Room 205ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 1:30pm – 2:30pm

★ **mcDETECT: Characterizing mRNA Localization in Polarized Neuronal Compartments With Spatial Transcriptomics**

Subsession Time: Wednesday, October 15 at 1:30pm – 1:45pm

Authors: Chenyang Yuan (Emory University), Krupa Patel (Emory University), Hongshun Shi (Emory University), Hsiao-Lin V. Wang (Emory University), Feng Wang (Emory University), Ronghua Li (Emory University), Yangping Li (Emory University), Victor G. Corces (Emory University), Hailing Shi (Emory University), Sulagna Das (Emory University), Jindan Yu (Emory University), Peng Jin (Emory University), Bing Yao (Emory University), Jian Hu (Emory University)

Abstract: Background: Neurons are highly polarized cells with distinct distal compartments—such as synapses, dendrites, and axons—that enable connectivity and communication. Thousands of mRNAs have been shown to localize to these distal sites, supporting local protein synthesis critical for their plasticity and function. The diversity and regulation of these mRNAs are closely linked to neuronal function and are implicated in various neurodegenerative diseases. However, conventional RNA sequencing methods, such as single-cell/nucleus RNA sequencing, are limited to capturing nuclear and somatic transcripts, leaving the mRNA content of distal compartments unexplored. Recent advances in in situ spatial transcriptomics (iST) allow subcellular measurement of gene expression by capturing individual mRNAs in their spatial context. This resolution creates new opportunities to study compartmentalized transcripts. As most analytical pipelines still focus on mRNAs located in the cell body, the primary challenge lies not in the technology itself, but in the development of appropriate analytical methods.

Methods: We present mcDETECT, a machine learning framework for studying mRNA localization in polarized neuronal compartments using iST data. mcDETECT first applies density-based clustering to identify RNA granules within distal subcellular regions. It then integrates information from nearby RNA granules to reconstruct compartment-specific expression profiles for individual neurons. These profiles reveal neuronal cell states that complement conventional cell type classifications based on nuclear mRNA.

Results: We applied mcDETECT to mouse brain datasets generated using various state-of-the-art iST platforms, including Xenium 5K, MERSCOPE, and CosMx. mcDETECT successfully identified RNA granules within polarized neuronal compartments and further classified them into distinct subtypes, each associated with specific functional roles across brain regions. Leveraging compartment-specific expression profiles, mcDETECT consistently uncovered previously unrecognized neuronal states across all different brain regions. In an Alzheimer's disease (AD) mouse model, mcDETECT revealed alterations in

both compartmentalized RNA patterns and neuronal states prior to observable neuronal loss, potentially serving as molecular markers for early AD diagnosis and therapeutic intervention.

Conclusion: mcDETECT is the first method to systematically characterize mRNAs for polarized neuronal compartments, offering novel molecular insights and revealing potential therapeutic targets within these distal regions for neurodegenerative diseases.

Sequence-dependent regulators of transposable-element-associated alternative splicing

Subsession Time: Wednesday, October 15 at 1:45pm – 2:00pm

Authors: Jayoung Ku (Boston Children's Hospital, Broad Institute of MIT and Harvard, Harvard Medical School), Eunjung Alice Lee (Boston Children's Hospital, Broad Institute of MIT and Harvard, Harvard Medical School)

Abstract: Transposable elements (TEs) are a major component of the human genome and are abundant in introns. Although most TE-derived sequences are removed during pre-mRNA splicing, antisense Alu elements harbor latent 3' and 5' splice sites (SSs) and long uridine tracts that can be aberrantly exonized, generating cryptic exons implicated in cancers and other diseases. Several RNA-binding proteins (RBPs), such as hnRNPC, hnRNPM, and SAFB, suppress TE-associated alternative splicing (TEAS), yet a genome-wide landscape of TEAS regulation is still lacking. Here, we engineered dual-fluorescent bidirectional reporters that incorporate endogenous intronic sequences in which mEGFP fluorescence decreases when an Alu exon is spliced in. Using these reporters, we performed genome-wide CRISPR knockout screens to identify endogenous regulators of Alu exonization. The screens validated hnRNPC as a global Alu exonization suppressor and uncovered SRSF2 as a novel selective suppressor of a distinct subset. hnRNPC-dependent events are enriched in right-arm Alu exons across all subtypes, whereas SRSF2 preferentially inhibits evolutionarily older AluJ exons, whose high basal inclusion renders them constitutive exons. Motif analysis showed that hnRNPC repression correlates with proximal U-tract length, consistent with prior studies, while SRSF2 sensitivity was linked to shorter U-tract length and a low local density of SSNG, SRSF2 binding motifs. Our findings delineate two mechanistically distinct RBP circuits that safeguard the transcriptome from cryptic Alu exonization. These findings provide a predictive framework for interpreting splice-altering variants within TEs and lay the groundwork for therapeutic targeting of TE-associated exons in cancer and genetic disease.

Saturation mutagenesis of 37 human splicing factor genes using pooled prime editing

Subsession Time: Wednesday, October 15 at 2:00pm – 2:15pm

Authors: Michael Herger (Francis Crick Institute), Joachim De Jonghe (Francis Crick Institute), Chloé Terwagne (Francis Crick Institute), Christina M. Kajba (Francis Crick Institute), Gregory M. Findlay (Francis Crick Institute)

Abstract: Multiplexed assays of variant effect (MAVEs) efficiently map genotypes to phenotypes by assessing thousands of variants per experiment. However, well-established MAVEs are limited either by testing variants exogenously or by constraints of genome editing. Recently developed pooled prime editing (PE) platforms enable functional interrogation of virtually any short variant anywhere in the genome. Here, we leveraged a highly optimized pooled PE platform to test variant effects across human splicing factor genes. Mutations in components of the spliceosome can drive diverse cancers and neurodevelopmental disorders, yet the functional consequences of most observed variants are unknown, precluding accurate genetic diagnoses and development of targeted therapies.

Using a library of 119,693 PE guide RNAs (pegRNAs), we attempted saturation mutagenesis of the coding regions of 37 splicing factors associated with cancer. All targeted genes are essential in the human haploid line, HAP1, such that variants impacting function lead to reduced cell growth, enabling pegRNA dropout screening. We extensively optimized PE efficiency in our platform, yielding an average pegRNA editing efficiency of 75% by day 5. Further, we introduced replicate barcodes to pegRNA cassettes, which improved variant coverage and discrimination between neutral and deleterious variants. Overall, these improvements allowed us to score 96,991 of 114,813 variants included in library design. We benchmarked our scores by independently assaying over 500 variants using a well-established saturation genome editing (SGE) protocol, revealing loss-of-function (LoF) variants to be identified with high precision. Across all splicing factors, we found 3,908 (4.3%) LoF missense variants. By mapping these to gene and protein structures, we observed that 58% of LoF missense variants are located within well-known functional protein domains. However, we also reveal clusters of deleterious variants outside of known motifs and in disordered regions, suggesting putative roles in human disease.

We expect our high-resolution variant effect maps for 37 splicing factors to serve as a valuable resource by highlighting regions of mechanistic importance for RNA splicing and by enabling improved clinical variant interpretation. More broadly, our results highlight the power of pooled PE to identify functionally impactful variants across many human genes.

Non-coding snRNA variants in RNU4-2 and RNU6 paralogues are a recurrent cause of retinitis pigmentosa

Subsession Time: Wednesday, October 15 at 2:15pm – 2:30pm

Authors: Carlo Rivolta (Institute of Molecular and Clinical Ophthalmology Basel, University of Basel, University of Leicester), Mathieu Quinodoz (Institute of Molecular and Clinical Ophthalmology Basel, University of Basel, University of Leicester), Kim Rodenburg (Radboud University Medical Center), Zuzana Cvakova (Institute of Molecular Genetics of the Czech Academy of Sciences), Karolina Kaminska (Institute of Molecular and Clinical Ophthalmology Basel, University of Basel), Suzanne de Bruijn (Radboud University Medical Center), Ana Belén Iglesias-Romero (Institute of Molecular and Clinical Ophthalmology Basel, University of Basel), Erica G M Boonen (Radboud University Medical Center, Rotterdam Ophthalmic Institute), Mukhtar Ullah (Institute of Molecular and Clinical Ophthalmology Basel, University of Basel), Nick Zomer (Radboud University Medical Center), Marc Folcher (Institute of Molecular and Clinical Ophthalmology Basel, University of Basel), Jacques Bijon (Vitreous Retina Macula

Consultants of New York, Rothschild Foundation Hospital), Lara K Holtes (Radboud University Medical Center), Stephen H Tsang (Columbia University Irving Medical Center, New York-Presbyterian Hospital), Zelia Corradi (Radboud University Medical Center), K Bailey Freund (Vitreous Retina Macula Consultants of New York, NYU Grossman School of Medicine), Stefanida Shliaga (Radboud University Medical Center), Daan Panneman (Radboud University Medical Center), Rebekkah Hitti-Malin (Radboud University Medical Center), Chaim Landau (Bonei Olam - Center for Rare Jewish Genetic Diseases), Allan I Jacob (Physicians Dialysis), Siying Lin (Moorfields Eye Hospital, Saint Mary's Hospital), Frans Cremers (Radboud University Medical Center), Winston Lee (Columbia University Irving Medical Center), Jamie M Ellingford (The University of Manchester, Genomics England Ltd), David Stanek (Institute of Molecular Genetics of the Czech Academy of Sciences), Susanne Roosing (Radboud University Medical Center)

Abstract: The major spliceosome is a macromolecular complex composed of five small nuclear RNAs (snRNAs)—U1, U2, U4, U5, and U6—and associated proteins, all essential for pre-mRNA splicing. While de novo variants in the RNU4-2 gene, encoding the U4 snRNA, have recently been implicated in neurodevelopmental disorders (ReNU syndrome), no causal relationship with retinal disease was previously known. Here, we report both inherited and de novo heterozygous variants in RNU4-2 and four U6 paralogous genes (RNU6-1, RNU6-2, RNU6-8, and RNU6-9) as the cause of autosomal dominant retinitis pigmentosa (adRP), a progressive inherited blindness. These variants were identified through a large-scale analysis of 4,722 RP cases lacking a genetic diagnosis and were present in 67 families (153 individuals). Strikingly, all pathogenic changes clustered within the three-way junction of the U4/U6 duplex, a region critical for the interaction with tri-snRNP proteins such as PRPF3 and PRPF31, themselves established RP genes. Structural modeling and biochemical assays suggested that these variants may destabilize the U4/U6 RNA duplex and selectively enhance the association with di-snRNP factors (e.g., SART3, PRPF31), consistent with a delay in snRNP biogenesis rather than defective splicing. Transcriptome profiling of affected individuals revealed no global changes in mRNA splicing or expression in blood leukocytes. Notably, these RP-linked U4 and U6 variants are distinct from those causing ReNU syndrome, which reside in regions involved in splice-site recognition and yield different molecular consequences. Based on prevalence estimates, we infer that such snRNA variants account for approximately 1.4% of all unsolved RP cases and up to 3.0% of undiagnosed adRP. These findings highlight the pathogenic relevance of non-coding RNA genes in monogenic diseases, reveal pleiotropy within RNU4-2, and emphasize the need to include snRNA loci in future diagnostic pipelines. They also suggest that non-coding variants can lead to tissue-specific diseases by subtly altering the dynamics of universally essential molecular complexes.

Session 34: Single-Cell Multiomic Dissection of Gene Regulation Across Disease States

Location: Ballroom West/Level 3, Thomas M. Menino Convention and Exhibition Center

Session Time: Wednesday, October 15 at 1:30pm – 2:30pm

Integrated single cell map of the human pancreas and immune-related tissues provides novel insights into diabetes risk

Subsession Time: Wednesday, October 15 at 1:30pm – 1:45pm

Authors: Weston Elison (UCSD Biomedical Science Program), Rebecca Melton (UCSD Biomedical Science Program, UCSD Department of Pediatrics), Abigail Howell (UCSD Department of Pediatrics), Carolyn McGrail (UCSD Biomedical Science Program), Emily Griffin (UCSD Department of Pediatrics), Sierra Corban (UCSD Department of Pediatrics), Mickey Lorenzini (UCSD Biomedical Science Program, Salk Integrative Biology Laboratory), Micheal Miller (UCSD Center for Epigenomics), Qian Yang (UCSD Center for Epigenomics), Melanie Shapiro (University of Florida Department of Pathology, Immunology and Laboratory Medicine, University of Florida Diabetes Institute), Todd Brusko (University of Florida Department of Pathology, Immunology and Laboratory Medicine, University of Florida Diabetes Institute, University of Florida Department of Pediatrics), Allen Wang (UCSD Center for Epigenomics), Graham McVicker (Salk Integrative Biology Laboratory), Kyle Gaulton (UCSD Department of Pediatrics)

Abstract: The pancreas plays a critical role in both endocrine and exocrine functions in the human body, and dysregulation of the pancreas is the major cause of type 1 and type 2 diabetes (T1D, T2D) and other diseases which together affect over 40 million Americans. Improved understanding of cell type regulation and function in the pancreas is paramount to uncovering disease mechanisms and identifying novel therapeutic targets. In this study we created an integrated map of gene expression and accessible chromatin in the human pancreas, pancreatic lymph nodes (pLNs) and spleen comprised of 2,314,215 cells from single cell assays of 278 human donors. We defined endocrine, exocrine, immune and other cell types and states, where the expanded profiling of cell, donors and tissues substantially improved annotation particularly in immune sub-types in the pancreas. Within each cell type we defined gene regulatory networks (GRNs) including cis-regulatory elements (cREs), expressed genes, and key transcriptional regulators. Cell type GRNs revealed both known regulators of cell type activity such as GATA4 and SMAD4 in acinar cells as well as novel regulators of cell type identity. We examined transcriptional and epigenomic changes in cell type regulatory programs across multiple axes including diabetes and pre-diabetes status, T1D/T2D genetic risk score (GRS), the presence of islet autoantibodies, and many other phenotypes, with revealed genes and pathways altered in pre-disease states and at-risk individuals. For example, individuals who were autoantibody positive without T1D had altered expression of genes related to insulin secretion relative to non-diabetic donors. In addition, we mapped thousands of cell type-specific QTLs for gene expression (eQTLs) and chromatin accessibility (caQTLs). Fine-mapping and colocalization of QTLs with diabetes association signals revealed loci that impact

pancreatic cell type regulation and highlighted novel genes and mechanisms underlying disease risk. Together this work establishes a foundational resource for understanding cell type-specific gene regulation in the pancreas and its role in diabetes risk and pathogenesis.

Integrative analysis of single-cell chromatin accessibility and GWAS statistics to prioritize enhancers and partition trait heritability in lung development

Subsession Time: Wednesday, October 15 at 1:45pm – 2:00pm

Authors: Heena Rijhwani (Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School), Yunzhe Qian (Harvard T.H. Chan School of Public Health), Courtney Tern (Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School), Michael H. Cho (Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School)

Abstract: Genome-wide association studies of chronic obstructive pulmonary disease (COPD) and reduced lung function have implicated hundreds of loci. However, effector genes and cell types for most of these regions are not well defined. While these loci are from studies of adults, substantial evidence indicates that many of these loci may be involved in lung development. Thus, to bridge this knowledge gap, we leveraged existing single-cell ATAC-seq data from fetal lung to quantify contributions to trait heritability, identify high-confidence regulatory elements, perform enhancer-gene linking, and integrate this data with genome-wide association studies (GWAS) of COPD and its defining trait of reduced forced expiratory volume in 1 second to forced vital capacity (FEV1/FVC) ratio.

First, we performed stratified LD Score Regression (LDSC) to quantify the contribution of overall and cell type-specific regulatory regions to lung-related traits. We identified significant heritability over all fetal lung regions, as well as in AT1 and AT2 cell types. To provide locus-specific links, we also applied the recently developed scE2G method to link to putative target genes using the fetal lung single-cell ATAC-seq data. In preliminary analysis, we intersected fine-mapped GWAS summary statistics with scE2G results, identifying putative functional variants (posterior inclusion probability > 0.30) linked to WNT2B, RREB1 and BRDT in AT1 and AT2 cell types. In addition, to optimize peak calling, particularly in rarer cell types, we applied gkmQC, a sequence-based quality control framework leveraging gapped k-mer support vector machines (gkm-SVMs) to evaluate and optimize scATAC-seq peak calls by assessing their overall regulatory potential. This approach optimizes peak selection by recovering peaks missed due to low coverage and produces a high-confidence set of regulatory elements for downstream analyses. In preliminary analysis across 16 cell types, gkmQC increased the number of peaks identified by ~15-20%. We identified new overlaps of fine-mapped GWAS summary statistics with these gkmQC-optimized peaks near LTBP1, TGFB2, and MST1P2 in AT1 and AT2 cell types, providing a potential functional context to prioritize enhancers with potential trait relevance. This integrative approach provides a comprehensive framework for linking regulatory elements to potential target genes and heritability, facilitating the functional interpretation of trait-associated loci in lung development. Extension to other cell types is ongoing.

Enhancer-Driven Gene Regulatory Networks Reveal Distinct Transcription Factor Hierarchies in Myeloid Cell Fate Decisions in Crohn's Disease

Subsession Time: Wednesday, October 15 at 2:00pm – 2:15pm

Authors: Jiayu Zhang (Icahn School of Medicine at Mount Sinai), Michelle Bao (Icahn School of Medicine at Mount Sinai), Ling-shiang Chuang (Icahn School of Medicine at Mount Sinai), Chris Tastad (Icahn School of Medicine at Mount Sinai), Judy H. Cho (Icahn School of Medicine at Mount Sinai), and the NIDDK IBD Genetics Consortium

Abstract: Background: Crohn's disease (CD) is a type of inflammatory bowel disease (IBD) characterized by chronic inflammation of the gastrointestinal tract. Monocytes play a crucial role in IBD by regulating inflammatory responses, and can differentiate into myeloid cells (macrophages, dendritic cells--DCs) and stromal cells, contributing to the disease's pathogenesis. Multiple transcription factors are IBD-associated, and we sought to define the cell fate decision using direct ex-vivo, multi-omic single-nucleus data.

Methods: We applied SCENIC+ to construct enhancer-driven gene regulatory networks (eRegulons) from integrated snRNA-seq and snATAC-seq data (n=96,976 cells) obtained from 6 CD gut tissues. We quantified eRegulon activity using Regulon Specificity Scores (RSS) and compared standard and custom-trained Celltypist models, and our eRegulon-based clustering approach to overcome sparsity of single-nucleus RNA data in myeloid sub-clusters annotation.

Results: Analysis of the myeloid-stromal subset (n=23,212 cells) identified 50 high-quality eRegulons, confirming cell type- and pioneering transcription factors (BNC2, ETS1, NFKB1, SPI1) and revealing dual functionality of key transcription factors (NFKB1, ETS1, REL) as both activators and repressors across different cell clusters. We identified three key eRegulons with distinct roles in IBD pathogenesis: ETV3 (activator, $r=0.68$, $p=9.54e-144$), ETV6 (repressor, $r=-0.57$, $p=2.24e-70$), and CREM (activator, $r=0.65$, $p=3.46e-99$). These networks showed cell type-specific activity patterns and regulated multiple IBD-associated genes and AP1 components, with ETV3 targeting TET2, STAT5A, and NFKB1, while ETV6 regulating FOS and CREM regulating ELF1 and FOSL2. Notably, our eRegulon-based clustering distinguished myeloid subsets undetectable by conventional snRNA-seq analysis, identifying specific eRegulon markers for macrophages (BNC2, RUNX2), monocytes (STAT4, NFKB1), and DCs (POU2AF1, REL). This approach outperformed auto-annotation using Celltypist, successfully identifying two distinct states each of DCs and macrophages, plus one monocyte cluster.

Conclusion: eRegulon-based clustering outperformed conventional annotation methods for myeloid cell identification in sparse nuclei RNA data. Our approach revealed distinct transcriptional regulatory networks in myeloid subpopulations, providing deeper insights into cellular heterogeneity and regulatory mechanisms governing myeloid cell fate decisions in CD pathogenesis.

Disease Genetic Scorecard Reveals Convergent Coding and Regulatory Variants in Complex Diseases

Subsession Time: Wednesday, October 15 at 2:15pm – 2:30pm

Authors: Hongbo Liu (University of Pennsylvania, University of Rochester), Katalin Susztak (University of Pennsylvania)

Abstract: Background: Genome-wide association studies (GWAS) have identified numerous DNA sequence variants associated with complex human diseases. However, over 90% of disease-associated variants reside in noncoding genome regions, and their functions in complex diseases remain largely unknown—a problem often referred to as the ‘variant-to-function’ problem. Methods: To link non-coding variants to functions in human diseases, we developed a Disease Genetic Scorecard that integrates various genomic and epigenomic datasets, including methylation quantitative trait loci (meQTL), allele-specific expression (ASE), and allele-specific expression accessibility (ASA), at the single-cell level. In particular, we developed a statistical model, Open4Gene, to link non-coding variants to their target genes using single-cell multiome data, which simultaneously profiles DNA accessibility and gene expression within the same cell. Results: Firstly, we applied the Disease Genetic Scorecard to annotate the multi-ancestry GWAS of kidney function mapping in over 2 million individuals. By integrating kidney function GWAS and other omics (comprising a total of 32 types), we created the first Disease Genetic Scorecard. We found over 600 genes targeted by both regulatory variants and coding variants. The integration of single-cell multiomes elucidates the disease-causing cell types and the transcription factors involved. In particular, we identified the convergence of coding and regulatory variants on 161 key disease genes, critical cell types, transcriptional regulators, and potential drug targets for kidney disease. Next, we constructed the Disease Genetic Scorecard for various traits collected from the GWAS Catalog. We built a comprehensive atlas of convergent coding and regulatory variants associated with complex human diseases. At last, we applied massively parallel reporter assays and epigenetic editing systems to characterize the functional variants and elements that impact transcriptional regulation in dynamic cellular systems. Compared to traditional GWAS annotation methods, the Disease Genetic Scorecard enables the identification of disease-causal genes with higher confidence. Overall, the Disease Genetic Scorecard presents a novel framework for identifying convergent coding and regulatory variants with causal effects, as well as their associated molecular mechanisms in complex human diseases.

Session 36: Artificial Intelligence and Machine Learning Tools Reshaping Modern Genomics

Location: Ballroom West/Level 3, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 8:30am – 10:00am

Decoding gene expression in cellular and disease states with deep learning

Subsession Time: Thursday, October 16 at 8:30am – 8:45am

Authors: Avantika Lal (Genentech), Alexander Karollus (Technical University of Munich, Munich Center for Machine Learning, Genentech), Laura Gunsalus (Genentech), David Garfield (Genentech), Surag Nair (Genentech), Alex M Tseng (Genentech), M Grace Gordon (Genentech), John Blischak (Genentech), Bryce van de Geijn (Genentech), Tushar Bhangale (Genentech), Jenna L Collier (Genentech), Nathaniel Diamant (Genentech), Tommaso Biancalani (Genentech), Hector Corrada Bravo (Genentech), Gabriele Scalia (Genentech), Gokcen Eraslan (Genentech)

Abstract: Deep learning models that predict gene expression from DNA sequence have transformed our understanding of gene regulation. However, these models are primarily trained on bulk data from healthy tissues, and so cannot model gene expression in specific cell types or disease contexts. In contrast, single-cell RNA sequencing (scRNA-seq) datasets profile expression at high resolution across diverse cell types, states, and diseases, yet revealing regulatory mechanisms from such data remains challenging.

We present Decima, a sequence-to-function model trained on atlas-scale single-cell expression data. Decima is based on the Borzoi model (Linder et al. 2023), and is trained on pseudobulked scRNA-seq data from over 22 million cells,. Given the 500 kb sequence surrounding an unseen gene, Decima can predict the expression level of that gene in hundreds of cell types, across diverse cellular states, tissues, and disease conditions.

We demonstrate that Decima identifies key regulatory elements, including distal enhancers, and learns motifs for transcription factors (TFs) that define cell type identities. The model also predicts subtle changes in gene expression between the same cell type across different tissues, states, or diseases, and reveals the regulatory mechanisms driving such changes. For example, it identified TF motifs associated with fibroblasts in specific tissues and highlighted cooperative TF interactions driving fibroblast-specific disease responses.

Over 90% of variants identified by GWAS studies are non-coding, and current tools struggle to link these to phenotype via cell type-specific mechanisms. Decima addresses this by directly predicting non-coding variant effects at cell type resolution. Using single-cell eQTL and GWAS datasets, we showed that Decima achieves state-of-the-art performance in identifying expression-altering variants, predicting the magnitude and direction of their effect, and identifying the affected cell types. Moreover, its attributions yield testable mechanistic hypotheses to explain the effects of such variants.

Finally, we show that Decima can design regulatory elements that drive cell type specific and even disease-specific expression, which are crucial in gene therapy. This may enable novel gene therapies that are activated specifically in diseased cells. Altogether, Decima combines the utility of sequence models with the richness of single-cell datasets, unlocking new applications in regulatory biology.

Integrating genetics, multimodal medical imaging, and AI for high-precision risk prediction of type 2 diabetes

Subsession Time: Thursday, October 16 at 8:45am – 9:00am

Authors: Hsin-Chou Yang (Institute of Statistical Science, Academia Sinica; Institute of Public Health, National Yang-Ming Chiao-Tung University), Yi-Jia Huang (Institute of Statistical Science, Academia Sinica), Yun-Jer Shieh (Institute of Public Health, National Yang-Ming Chiao-Tung University), Chun-houh Chen (Institute of Statistical Science, Academia Sinica)

Abstract: Background: Genetic and medical imaging data have each been studied for their contributions to the risk evaluation and early detection of Type 2 Diabetes (T2D). However, key gaps remain—particularly in understanding the mutual information between T2D, genetic factors, and imaging phenotypes, as well as the integrated contribution of these modalities to T2D risk prediction.

Materials: We leveraged multimodal data from the Taiwan Biobank (TWB), encompassing up to 131,048 Taiwanese Han Chinese participants. The dataset included genome-wide imputation data (9,814,944 loci), raw pixel (over hundreds of thousands DICOM image files), and tabular image-derived phenotypes (152 IDPs) from four medical imaging modalities (abdominal ultrasound, vertebral artery ultrasound, bone density scan, and electrocardiogram). Extensive demographic, behavior, and environmental exposure data were also incorporated. Furthermore, we integrated our dataset with genetic information from 1.4 million participants in large-scale cross-ancestry T2D studies to construct polygenic risk scores (PRS).

Methods: We established a comprehensive analytical framework integrating quality-controlled genetic and imaging data. Genome-Wide Association Studies (GWAS) were performed to identify T2D susceptibility loci, complemented by Quantitative Trait Locus Mapping (QTL) to prioritize genetic loci associated with glucose markers and IDPs. Pleiotropic variants linking IDPs and T2D were identified, and significant signals were functionally annotated using ANNOVAR. CYTOSCAPE was employed to visualize the interaction network of the genetic associations between IDPs and T2D.

To infer causal relationships, we applied a two-sample bidirectional Mendelian Randomization (MR) framework using internal GWAS-derived genetic instruments for IDPs and external large-scale GWAS data for T2D. For risk modeling, Polygenic Risk Scores (PRS) were constructed via Bayesian statistical learning. Multi-Image Risk Scores (MRS) were derived using deep learning (Densely Connected Convolutional Networks; DenseNet) on raw image pixels, with graph neural networks capturing intra-individual dependencies across imaging modalities. Annotated tabular IDPs were modeled using the machine learning (eXtreme Gradient Boosting; XGBoost) algorithm. Finally, fusion models combining

PRS, MRS, and other demographic and risk factors were developed using XGBoost to produce robust and scalable risk prediction models.

Findings:Phenome-wide association analyses identified 14,222 significant QTLs, 303 T2D-moderating QTLs, and 220 T2D-suppressor QTLs associated with IDPs. Pleiotropic SNPs were discovered linking IDPs and T2D, such as four SNPs on LCORL and one SNP on NCAPG, with a pleiotropic effect on T2D and IDPs of bone density scans. Bidirectional MR analysis confirmed significant causal relationships between imaging phenotypes and T2D risk. However, various imaging modalities exhibit different correlation mechanisms with T2D. MR analysis suggests complex causal relationships, such as the bidirectional causality observed between IDPs of abdominal ultrasound and T2D, and trait-specific causal patterns in IDPs of bone density scans.

Our integrated AI-driven approach demonstrated superior predictive performance. Models utilizing raw imaging pixels outperformed those relying on human-annotated features in terms of accuracy, automation, and cost-effectiveness. The fusion model of genetic, medical imaging, and demographic variables achieved an Area Under the Curve (AUC) of 0.944, with an accuracy of 0.875, a sensitivity of 0.882, and a specificity of 0.875, significantly enhancing the model's performance, which had previously only included part of the predictors. Strong associations between both PRS and MRS with T2D enabled the stratification of high-risk subgroups.

Interpretation:This study introduces a novel AI-powered framework that integrates multimodal data to advance precision medicine for T2D. By integrating genome-wide genetic data, multimodal medical imaging, and demographic information, we provide a high-performance, scalable strategy for predicting T2D risk. Beyond prediction, the study sheds light on the biological mechanisms underlying T2D through causal inference and pleiotropy mapping. Importantly, we demonstrate that raw image analysis via deep learning circumvents the need for labor-intensive image annotation, addressing a key bottleneck in clinical implementation. This work sets a new benchmark for AI-based precision health, with potential applications across a range of complex diseases.

G2DBridge: A Two-Step Machine Learning Framework for Disease Risk Prediction Using Genetically Inferred Imaging Features

Subsession Time: Thursday, October 16 at 9:00am – 9:15am

Authors: Okan Bilge Ozdemir(Cedars Sinai Medical Center), Ruowang Li(Cedars Sinai Medical Center),

Abstract: Background:Polygenic Risk Scores (PRS) have been widely used to estimate disease risk from genome-wide genetic associations. However, genetic variation represents only a distal layer of disease etiology and is several steps removed from clinical manifestation. In contrast, imaging-derived phenotypes (IDPs), such as those from MRI, are more proximal to disease processes and have demonstrated superior predictive power for various complex diseases, including Alzheimer's disease (AD). Despite their effectiveness, IDPs are often unavailable in large cohorts due to limited access and high imaging costs. To address this limitation, we introduce G2DBridge (Bridging Gene and Disease via

Imaging), a machine learning framework that leverages genetic data to infer IDPs and improve disease risk prediction in populations lacking neuroimaging data.

Methods: G2DBridge is a two-step framework: it first trains a model to predict IDPs from genotypes using datasets with paired genetic and imaging data. Then it uses the inferred IDPs to build downstream disease prediction models. These genetically inferred IDPs serve as additional features to enhance risk stratification in external cohorts with genotype-only data. The framework is further optimized using cohorts with available IDPs.

Results: We applied G2DBridge to Alzheimer's disease using data from ADNI (training and validation) and ADSP (testing). On the ADNI dataset, models based solely on genetic data achieved AUCs ranging from 0.5945 to 0.6239. In contrast, G2DBridge, by inferring and augmenting IDPs, achieved a substantially higher average AUC of ~ 0.6778 . External validation on ADSP confirmed the framework's robustness. Benchmarking against 10 conventional PRS methods (AUCs: 0.5180–0.6400, avg. 0.5825), G2DBridge consistently outperformed all methods, achieving AUCs ranging from 0.6415 to 0.6680, avg 0.6563, across 5-fold cross-validation. Notably, like PRS, G2DBridge only utilized genetic data in ADSP.

Conclusions: Augmenting genetically inferred IDPs significantly improves disease risk prediction beyond traditional genetic risk prediction models, such as PRS. G2DBridge maintains the scalability of PRS while incorporating richer, imaging-informed intermediate phenotypes and holds promise for broader application across other complex diseases.

Machine learning-derived exosomal circRNA signature enables early lung cancer detection and reveals metabolic dysregulation

Subsession Time: Thursday, October 16 at 9:15am – 9:30am

Authors: Yuanyuan Fu (University of Hawaii John A. Burns School of Medicine), Qiongyu Hao (Charles Drew University of Medicine and Science), Masaki Nasu (University of Hawaii John A. Burns School of Medicine), Hua Yang (University of Hawaii John A. Burns School of Medicine), Zhanwei Wang (University of Hawaii John A. Burns School of Medicine), Youping Deng (University of Hawaii John A. Burns School of Medicine).

Abstract: Background: Lung cancer remains the leading cause of cancer-related mortality worldwide, with a 5-year survival rate of only 15%, largely due to late-stage diagnosis. Emerging evidence underscores the pivotal role of exosomal circular RNAs (circRNAs) in cancer development and metabolic regulation, positioning them as promising biomarkers.

Methods: A supersensitive human plasma exosome isolation and purification as well as exosome RNA sequencing and metabolomic profiling were conducted on 436 subjects (204 early-stage lung cancer [90% Stage I, 10% Stage II] and 232 non-cancer controls [131 benign, 101 normal]). Using a four-phase design (discovery, in-house testing, independent validation, experimental validation), we developed a circRNA signature (circSig) via 12 machine learning algorithms and 111 algorithm combinations in the discovery cohort (n=261). Validation included an in-house set (n=175) and five external datasets (TCGA

[n=1,078] and four GEO cohorts [n=485 total]). Parallel metabolomics and in-vitro assays explored the functional mechanistic links.

Results: A total of 7519 circRNAs were initially annotated, and differentially expression analysis identified 667 circRNAs up-regulated and 1598 down-regulated in cancer compared with non-cancer groups. By diagnostic modeling in the discovery dataset (n=261), a panel of circRNAs were identified and developed as circSig score, with exceptional diagnostic accuracy: AUCs of 97.0% (cancer vs. non-cancer), 96.7% (cancer vs. benign), and 97.3% (cancer vs. normal). Subtype analysis revealed AUCs of 97.07% for adenocarcinoma (AD) and 98.13% for squamous cell carcinoma (SCC). Notably, circSig achieved 99.40% AUC for Stage I SCC in Caucasian smokers. Multi-cohort validation confirmed robustness. Mechanistic analysis linked circSig to glutamine metabolism. Metabolomic profiling revealed 40 significantly altered features, with enrichment in the glutamate/glutamine pathway. Key enzymes (GLUD1, GLS) were identified as upstream regulators and further validated the biological network through experiments.

Conclusion: Our study established a machine learning-derived exosomal circSig as effective non-invasive biomarker and risk stratification score system for early lung cancer detection and monitoring, and reveals its therapeutic potential through glutamine pathway regulation. These findings open avenues for both liquid biopsy development and metabolism-targeted therapies.

Distinguishing causal variants from linked hitchhikers in the human genome using deep learning

Subsession Time: Thursday, October 16 at 9:30am – 9:45am

Authors: Arya S. Rao (Harvard Medical School, Broad Institute of MIT and Harvard), Tian Xia (Yale University), Sager J. Gosai (Broad Institute of MIT and Harvard), Joseph J. Vitti (Broad Institute of MIT and Harvard), Ilya Shlyakhter (Broad Institute of MIT and Ha

Abstract: Human migration within and out of Africa exposed populations to strong evolutionary pressures in the form of unique climates, diets, and pathogens, driving adaptive differences among groups. Selection acts both on de novo and standing genetic variation vi

PubMind: A Large Language Model Framework for Scalable Extraction of Variant and Disease Associations from Biomedical Literature

Subsession Time: Thursday, October 16 at 9:45am – 10:00am

Authors: Peng Wang (Children's Hospital of Philadelphia, Center for Cellular and Molecular Therapeutics), Kai Wang (Children's Hospital of Philadelphia, Center for Cellular and Molecular Therapeutics)

Abstract: Background: Despite a surge in biomedical publications, much of the knowledge on human genetic variants remains buried in unstructured text. Several databases have attempted to address this gap. ClinVar suffers from variable data quality, lacks experimenta

Session 37: Choose Your Own Adventure: Novel Approaches in Statistical Genetics

Location: Room 258ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 8:30am – 10:00am

Computation and resource efficient genome-wide association analysis for large-scale imaging studies

Subsession Time: Thursday, October 16 at 8:30am – 8:45am

Authors: Zhiwen Jiang (UNC), Jason Stein (UNC), Tengfei Li (UNC), Ethan Fang (Duke University), Yun Li (UNC), Patrick Sullivan (UNC), Hongtu Zhu (UNC)

Abstract: Most existing imaging genome-wide association studies (GWASs) are limited to a few hundred image-derived phenotypes (IDPs), which are usually generated by aggregating local signals, potentially discarding substantial local variations. Alternatively, some IDPs are derived from complex models; the interpretation of genetic associations linked to these IDPs is ambiguous, affecting clinical usage. Voxel-level GWAS on the original images is a solution to these issues, because it can clearly demonstrate genetic effects on arbitrary regions, but it is challenged by intractable computation burden due to high-dimensional imaging and genetic data.

We propose a Representation learning-based Voxel-level Genetic Analysis (RVGA) framework for conducting GWAS on biomedical images at voxel/vertex level, as well as sharing GWAS summary statistics for the entire image across the whole genome by reducing the computational and storage burden by 2-3 orders of magnitude. RVGA proposes a novel strategy that generates low-dimensional representations (LDRs) for images using functional principal component analysis (FPCA), conducts GWAS on them, and uses functional bases to reconstruct voxel-level summary statistics. This strategy significantly reduces computational burden and removes random noise in images. Additionally, RVGA proposes a unified estimator for voxel heritability, genetic correlations between pairs of voxels, and cross-trait genetic correlations between voxels and non-imaging phenotypes, only using the shared data and a population-matched LD matrix.

We analyzed 6.3 million common variants across 62,217 voxels/vertices in hippocampus shape and white matter (WM) microstructure using 53,454 subjects in the UK Biobank. RVGA took less than eight hours to scan the whole genome for 15,000 vertices in the left hippocampus. We identified 72 and 526 loci for hippocampus shape and WM microstructure ($n=33,324$), respectively, and 39 and 275 of them are novel. Among the total 598 loci, 432 of them can be replicated using an independent dataset ($n=20,130$). The shared data was only 55 GB, compared to 12,611 GB if saving all variant-voxel pairs. We additionally discovered regional heterogeneity of heritability within hippocampus and WM tracts. Moreover, we targeted local regions that were genetically correlated with brain-related phenotypes/disorders, such as

between the presubiculum and educational attainment, and between subregions of the genu of the corpus callosum and bipolar disorder.

A phenome-wide map of pleiotropy: Connecting genetic signals to disease mechanisms and drug discovery

Subsession Time: Thursday, October 16 at 8:45am – 9:00am

Authors: Anna Hutchinson (GSK), Zhana Kuncheva (bioXcelerateAI), Soroosh Afyouni (bioXcelerateAI), Manju Dissanayake (bioXcelerateAI), Daniel McCartney (bioXcelerateAI), Peter Scanlon (bioXcelerateAI), Daniel Crouch (GSK), Adrian Cortes (GSK), Chris Foley (bioXcelerateAI)

Abstract: As genetic datasets continue to grow in scale and phenotypic diversity, the key challenge is to move beyond the identification of isolated genome-wide association (GWAS) signals toward a mechanistic understanding of how genetic variation shapes human biology. Two commonly used approaches - phenome-wide association studies (PheWAS) and colocalisation - offer partial insights: PheWAS can identify variant-trait associations but is confounded by linkage disequilibrium (LD), while colocalisation provides stronger evidence of shared causal variants but is limited to a few traits at a time due to computational demand.

Here, we present a genome- and phenome-wide map of pleiotropy - likely one of the most comprehensive views to date of shared causal architecture across the human phenome. We analysed over 3,200 traits from whole-genome sequencing (WGS) data in the UK Biobank and genotyping data from the VA Million Veteran Program using a novel, scalable pipeline purpose-built for this task. To overcome bottlenecks in conditional analysis, fine-mapping and summary statistic imputation, we developed Switch-Step, a novel fine-mapping algorithm as accurate as SuSiE but >80x faster, and Imp-Map, a summary statistic imputation method comparable in accuracy to Imp-G but >50x faster. These methods, combined with cloud-native infrastructure (Nextflow, GCP), enabled genome-wide multi-trait colocalisation with HyPrColoc (Foley et al. 2021) across >500,000 genetic signals with suggestive associations ($P < 1 \times 10^{-6}$).

Our resulting map identifies over 35,000 pleiotropic loci with a high posterior probability (>0.7) of a shared causal variant, confirmed by post-hoc sensitivity analysis. The median number of colocalising traits was 5 (IQR 3 to 13). By integrating GWAS with expression, proteomic and metabolomic QTLs, we prioritise likely causal genes and biomarkers, shedding light on the mechanisms that link molecular traits to complex diseases.

This scalable framework opens new possibilities for leveraging human genetics in drug development: expanding the value of existing assets through indication extension and repurposing; predicting adverse associations to improve probability of success; revealing mechanisms for first-in-class targets with differential biology; and identifying biomarkers for patient stratification. Our approach shows how method innovation, paired with large-scale data and scalable computational infrastructure, can transform pleiotropy from a confounder into a discovery tool.

Parent-of-origin inference and its role in the genetic architecture of complex traits: evidence from ~235,000 individuals

Subsession Time: Thursday, October 16 at 9:00am – 9:15am

Authors: Robin J. Hofmeister (Department of Computational Biology, University of Lausanne, Lausanne, Switzerland and Estonian Genome Centre, Institute of Genomics, University of Tartu, Estonia), Théo Cavinato (Department of Computational Biology, University of Lausanne, Lausanne, Switzerland), Roya Karimi (Mohn Center for Diabetes Precision Medicine, Department of Clinical Science, University of Bergen, Bergen, Norway), Adriaan van der Graaf (Department of Computational Biology, University of Lausanne, Lausanne, Switzerland),

Fanny-Dhelia Pajuste (Estonian Genome Centre, Institute of Genomics, University of Tartu, Estonia), Jaanika Kronberg (Estonian Genome Centre, Institute of Genomics, University of Tartu, Estonia), Nele Taba (Estonian Genome Centre, Institute of Genomics, University of Tartu, Estonia), Reedik Mägi (Estonian Genome Centre, Institute of Genomics, University of Tartu, Estonia), Marc Vaudel (Mohn Center for Diabetes Precision Medicine, Department of Clinical Science, University of Bergen, Bergen, Norway), Simone Rubinacci (Institute for Molecular Medicine Finland, Helsinki, Finland), Stefan Johansson (Mohn Center for Diabetes Precision Medicine, Department of Clinical Science, University of Bergen, Bergen, Norway), Lili Milani (Estonian Genome Centre, Institute of Genomics, University of Tartu, Estonia),

Olivier Delaneau (Regeneron Genetics Center, Tarrytown, NY, USA), Zoltán Kutalik (Department of Computational Biology, University of Lausanne, Lausanne, Switzerland)

Abstract: Traditional genome-wide association studies (GWAS) assume that alleles inherited from the mother and father contribute equally to the examined trait. However, some sequence variants can have distinct phenotypic effects depending on whether they are maternally or paternally inherited, a phenomenon known as parent-of-origin effects (POEs). Traditionally linked to genomic imprinting, these effects are believed to have evolved from parental conflict over resource allocation to offspring, which results in opposing parental genetic influences. Despite their potential importance, POEs remain heavily understudied in complex traits, largely due to the lack of parental genomes.

Here, we present a multi-step approach to infer the parent-of-origin of alleles without parental genomes, leveraging inter-chromosomal phasing, mitochondrial whole-genome sequences and chromosome X data, and sibling-based crossover inference. This method was initially applied to the UK Biobank, where we inferred parent-of-origin for 123,716 individuals (discovery cohort). For replication, we extended it to the Estonian Biobank (97,346 individuals) and included an additional 45,402 offspring from the Norwegian Mother, Father, and Child Cohort Study cohort. This combined dataset of 266,464 individuals - representing the largest dataset of its kind - enabled a systematic investigation of POEs across a broad range of human traits, including anthropometric characteristics, blood biomarkers, diseases, and protein levels.

GWAS scans in the UK Biobank for more than 60 complex traits and over 14,000 additive protein Quantitative Trait Loci contrasting maternal and paternal effects identified over 30 novel POEs and confirmed more than 50% of testable known associations. Notably, approximately half of our POEs exhibited a bi-polar pattern, where maternal and paternal alleles exert conflicting effects. These effects were particularly prevalent for traits related to growth (e.g., IGF-1, height, fat-free mass) and metabolism (e.g., type 2 diabetes, triglycerides, glucose). Replication in the Estonian Biobank and MoBa validated over 87% of testable associations, notably highlighting dynamic POEs on infant BMI at different time points (6 weeks to 8 years of age).

Overall, our findings shed new light on the influence of POEs on diverse complex traits and align with the parental conflict hypothesis, providing compelling evidence for this understudied evolutionary phenomenon.

mFABIO: Multi-tissue TWAS fine-mapping to prioritize causal tissues and genes for binary traits

Subsession Time: Thursday, October 16 at 9:15am – 9:30am

Authors: Haihan Zhang (University of Michigan - Ann Arbor), Kevin He (University of Michigan - Ann Arbor), Lam C. Tsoi (University of Michigan - Ann Arbor), Xiang Zhou (University of Michigan - Ann Arbor)

Abstract: Recent developments in transcriptome-wide association study (TWAS) fine-mapping enable joint modeling of multiple genes to improve causal gene prioritization. However, these fine-mapping methods face two major limitations: first, they have been developed primarily for quantitative traits, leading to potentially reduced performance for binary disease traits; second, they use gene expression profiles from a single tissue and may not capture the complex mechanisms spanning multiple tissues. Here, we present mFABIO, a multi-tissue TWAS fine-mapping method specifically designed for binary traits. mFABIO employs a probit model to directly link the genetically regulated expression (GREx) of genes across multiple tissues to a binary outcome, while fully accounting for the GREx correlation structure among tissue-gene pairs within a specific region of interest. Consequently, mFABIO achieves substantial power gains for binary traits and maintains robust control of false discovery rates (FDR), outperforming existing fine-mapping approaches. We validated the performance of mFABIO through extensive simulations and applied it to the in-depth analysis of five binary disease traits (asthma, breast cancer, gout, prostate cancer, and rheumatoid arthritis) in the UK Biobank, using summary statistics of expression quantitative trait loci (eQTL) data from 38 Genotype-Tissue Expression (GTEx) tissues. mFABIO identified an average of 35 causal genes and 49 tissue-gene pairs per disease (FDR < 0.05), with 59% of the genes and 79% of the tissue-gene pairs supported by TWAS or GWAS evidence. This represented an 8%-42% increase in evidence-supported genes and a 10%-54% increase in evidence-supported tissue-gene pairs, compared to existing approaches. mFABIO was also able to reduce the number of causal candidates by at least 44% for genes, and 40% for tissue-gene pairs, compared to existing single-tissue approaches. Leveraging its improved power, mFABIO successfully prioritized multiple potentially causal tissue-gene pairs consistent with prior biological knowledge for these diseases. Examples include IL-18R1 in lung tissue for asthma, CYBRD1 in breast mammary tissue for breast cancer, and CCDC122 in prostate

tissue for prostate cancer. Overall, mFABIO serves as an effective tool for multi-tissue TWAS fine-mapping of binary disease traits.

Leveraging cis and trans genetic data to improve protein level prediction in proteome-wide association studies

Subsession Time: Thursday, October 16 at 9:30am – 9:45am

Authors: Suzanne M. Leal (Columbia University), Derek Lamb (Columbia University), Rui Dong (Columbia University), Gao Wang (Columbia University), Andrew DeWan (Yale University)

Abstract: Genetic effects on complex traits and diseases are mediated through protein pathways, making proteome-wide association studies (PWAS) increasingly important for understanding disease etiology. However, many studies lack proteomic data. We developed a comp

causalR2M: causal Mendelian randomization mediation analysis with high-dimensional multi-omics mediators in 7000 individuals provides novel insights into BMI's casual effect on blood pressure

Subsession Time: Thursday, October 16 at 9:45am – 10:00am

Authors: Peng Wei (The University of Texas MD Anderson Cancer Center), Zhichao Xu (The University of Texas MD Anderson Cancer Center)

Abstract: Background:Causal mediation analysis is a powerful tool for uncovering biological mechanisms that link risk factors such as obesity to health outcomes like cardiovascular disease and traits, but its validity depends on strong, often untestable assumptions

Session 38: Computational Tools for Precision Therapeutics

Location: Room 206AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 8:30am – 10:00am

Machine learning-enabled genetic discovery and therapeutic targeting of hepatic IRS1 in MASLD

Subsession Time: Thursday, October 16 at 8:30am – 8:45am

Authors: Hari Somineni (insitro), Santhosh Satapati (insitro), Vaishaali Natarajan (insitro), Tanaya Walimbe (insitro), Sumit Mukherjee (insitro), David Amar (insitro), Xueya Zhou (insitro), insitro research team (insitro), Nav Ranu (insitro), Saradha Venkatachalapathy (insitro), Anna Shcherbina (insitro), Arijit Bhowmick (insitro), Emily Fox (insitro), Daphne Koller (insitro), Mike Rothenberg (insitro), David Lloyd (insitro), Colm O'Dushlaine (insitro)

Abstract: Therapeutic targets with human genetic support are more likely to succeed in clinical development, yet genetic discovery in MASLD has been hindered by limited access to scalable, precise phenotypes. To address this, we used machine learning (ML) to impute liver fat and adipose tissue depots – visceral (VAT), subcutaneous (ASAT), and gluteofemoral (GFAT) – from four widely available data modalities in UK Biobank: MRI (GWAS effective sample size, $n = 25,474$), DXA (48,494), plasma metabolites (82,138), and blood-based biomarkers (262,927). GWAS of liver fat adjusted for adipose volumes uncovered 480 loci, including known MASLD genes and novel associations such as the first at FGF21. Liver fat predicted across modalities showed strong genetic correlation ($r_g = 0.79 - 0.97$) and with clinically-diagnosed MASLD ($r_g = 0.73 - 0.87$). Loci supported across multiple modalities demonstrated stronger associations, and higher relevance to clinical MASLD, highlighting the robustness of convergent multimodal signals. Among our novel associations were several de novo lipogenesis (DNL) genes, including IRS1. A non-coding variant near IRS1 demonstrated the strongest association with increased liver fat ($P = 1.8 \times 10^{-82}$) and was replicated across modalities ($P < 5 \times 10^{-8}$) and in meta-analysis of 16,888 clinically-diagnosed MASLD cases and 1.2 million controls ($P = 5.1 \times 10^{-6}$). The risk allele increased IRS1 expression in HepG2 cells, and pooled optical screening with ML-based lipid phenotyping confirmed IRS1 as a functional effector. We developed liver-targeted GalNAc-siRNAs against IRS1 ($IC_{50} \leq 0.06$ nM), which significantly reduced (~40% vs. vehicle) lipid accumulation in primary human hepatocytes. In diet-induced obese mice, a single dose (0.3 – 10 mg/kg) reduced hepatic IRS1 (85%), triglycerides (45%), histologic NAS score (2.2 vs. 4.3, $P < 0.0001$), and DNL (59%) after 1 month. In cynomolgus monkeys, IRS1 siRNA (30 – 300 mg/kg) reduced hepatic expression up to 70% and DNL without affecting plasma glucose. These findings identify IRS1 as a promising therapeutic target and highlight the power of integrating ML-based phenotype imputation and multimodal integration, with genetic discovery and functional validation to accelerate drug development for MASLD.

Cas9 cytosine base editor proteins for precision genome editing with high specificity and activity

Subsession Time: Thursday, October 16 at 8:45am – 9:00am

Authors: Katherine Mechling¹, Erin Brettmann², Stephen Beishir², Xiao Chen², Xiaokang Zhang³, Rajagopalan Lakshmi Narasimhan⁴, Ajaya Kumar Kotekar Manjunath⁴, Daniel Taglicht⁵, Jiajian Liu², Fuqiang Chen², 1. MilliporeSigma, Burlington, MA, USA, 2. MilliporeSigma, Saint Louis, MO, USA, 3. MilliporeSigma, Shanghai, China, 4. MITC, MilliporeSigma, Bangalore, India, 5. MilliporeSigma, Jerusalem, Israel

Abstract: Base editing technologies allow for the precise and efficient installation of defined nucleotide substitutions into a target genome without double strand breaks. These editing methods can be used to generate cell lines and animal models for the study of a diverse range of topics, from molecular mechanisms of pathogenesis to evolutionary genetics. To support this work, we have developed a suite of recombinant protein-based cytosine base editor (CBE) tools. This suite includes two separate CBE proteins with complementary editing windows, a protein format uracil glycosylase inhibitor (UGI), and a guide design tool to facilitate experimental planning and computationally predict editing outcomes. The first CBE, “Flexible,” has a wide editing window, enabling editing over a large fraction of the target sequence and facilitating flexible placement of the guide RNA relative to the target cytosine residue. The second CBE, “Precision,” has a narrowed editing window, enabling discrimination between the target cytosine and nearby, nontarget cytosines. We show that co-transfection with the UGI protein increases the proportion of substitutions that are C-to-T and the ratio of C-to-T editing to indel formation, thus reducing undesired editing outcomes. We further show that delivery with a transfection enhancer results in an increase in overall editing rates, and that the transfection enhancer and UGI can be combined to maximize the rate of C-to-T editing while minimizing undesired outcomes. We use in vitro editing experiments to characterize the editing profiles of our CBEs and identify a preference for cytosine residues preceded by thymine (TC dinucleotides) and unmethylated cytosine residues. Finally, we show that RNP delivery of Flexible and Precision CBEs results in lower rates of mismatch editing and genome-wide nonspecific editing than plasmid delivery of the same editors, demonstrating their improved specificity profiles. To facilitate use of these editors, we have developed a guide design tool that will identify candidate guide RNAs in a user-provided sequence, identify potential off-target sites, and design primers to amplify the on- and off-target sites. The tool further uses a machine learning model trained on experimental data to predict which cytosine residues are most likely to be edited, enabling the user to predict their editing outcomes before performing the experiment. Together, these tools will facilitate precision genome editing experiments.

Leveraging Open-Source Large Language Models to Identify Undiagnosed Patients with Rare Genetic Aortopathies

Subsession Time: Thursday, October 16 at 9:00am – 9:15am

Authors: Pankhuri Singhal (University of Pennsylvania), Zilinghan Li (Argonne National Laboratory), Ze Yang (University of Illinois Urbana-Champaign, Tarak Nandi (Argonne National Laboratory, The University of Chicago), Colleen Morse (University of Pennsylvania), Zachary

Rodriguez (University of Pennsylvania), Alex Rodriguez (Argonne National Laboratory), Theodore Drivas (University of Pennsylvania), Ravi Madduri (Argonne National Laboratory, The University of Chicago), Anurag Verma (University of Pennsylvania)

Abstract: Background: Rare genetic aortopathies are frequently undiagnosed due to high levels of phenotypic heterogeneity in the patient population. Delayed diagnosis can lead to fatal cardiac outcomes. While genetic testing can enable early proactive interventions, the current system relies on primary care physicians to recognize a genetic basis for symptoms and then refer patients to clinical genetics for further evaluation. Broad-scale screening methods are needed to capture cases that do not fit an obvious diagnostic pattern. Free-text clinical notes contain observations and narratives that could support the automated flagging of patients for recommended genetic testing. Given the strength of large language models in processing unstructured text, this study explores their potential for the identification of undiagnosed patients by the automated evaluation of patient notes. Methods: We developed an open-source LLM-enabled genetic testing recommendation pipeline for rare genetic aortopathies. The pipeline leverages retrieval augmented generation on curated genetic aortopathy-related corpora to utilize relevant clinical knowledge for predicting which patients would benefit from genetic testing. The pipeline was validated using 22,510 patient progress notes from 500 individuals in the **** BioBank. Our approach couples the powerful capabilities of base LLMs with the addition of customized corpora to make predictions for ambiguous cases. Results: In validation experiments, our recommendation pipeline successfully identified 425 out of 499 patients (one case requiring further clinician evaluation due to incomplete information). The model achieved a patient-level recommendation accuracy of 0.852, precision of 0.889, recall of 0.803, F1 score of 0.844, and F3 score of 0.811. These results demonstrate the pipeline's potential to automate genetic testing recommendations and accelerate disease identification and early detection of rare genetic aortopathies. Conclusions: Our open-source LLM-enabled workflow integrates RAG with curated clinical domain-specific corpora to effectively analyze free-text clinical notes to recommend genetic testing for rare genetic aortopathies. The high accuracy and strong performance metrics demonstrate its potential to automate early disease identification and improve patient outcomes. Moreover, our generalizable workflow can be adapted to any other disease for broader application and requires no pre-processing of patient notes.

Enhanced Prediction of Antisense Oligonucleotide Therapy Eligibility for Rare Genetic Diseases Using a Semi-Automated Bioinformatics Pipeline

Subsession Time: Thursday, October 16 at 9:15am – 9:30am

Authors: Arash Salmaninejad (Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA), Emily Blake (Department of Quantitative Health Science, Division of Computational Biology, Mayo Clinic, Rochester, MN, USA), Filippo Pinto e Vairo (Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA), Eric W. Klee (Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA), Cousin Margot (Department of Molecular Medicine, Mayo Clinic, Rochester, MN, USA)

Abstract: Over 7,000 rare diseases have been identified, many of which are severe, begin in childhood, and lack effective treatments, resulting in devastating impacts on patients and families. Less than 5% have approved therapies, and fewer than 1% have curative options, highlighting a substantial unmet therapeutic need. Despite their individual rarity, rare diseases collectively affect more than 263 million people worldwide, underscoring the urgent demand for scalable and targeted therapeutic solutions. Antisense oligonucleotides (ASOs) have emerged as a promising modality for individualized treatment of genetic disorders. However, assessing the suitability of specific pathogenic variants for ASO intervention remains a complex and predominantly manual process. To address this challenge, we developed a modular, semi-automated computational pipeline that systematically evaluates the ASO targetability of genetic variants. This framework integrates variant- and gene-level annotations from curated genomic databases and public resources; applies bioinformatics tools to predict splicing impacts and pathogenic mechanisms; and incorporates published guidelines for assessing ASO treatment eligibility. It supports multiple ASO strategies, including splice correction, canonical exon skipping, transcript knockdown, and targeted augmentation of nuclear gene output (TANGO). TANGO-based strategies encompass poison exon skipping, modulation of naturally occurring antisense transcripts (NATs), and targeting upstream open reading frames (uORFs) to enhance translation and restore gene function. The pipeline incorporates key gene-level features—such as disease association, inheritance patterns, and haploinsufficiency—and transcript/exon-level attributes, including MANE transcript selection, reading frame preservation, and protein domain localization. Variant-level assessments including predicted splicing effects and encoded protein effect (e.g., loss-of-function, gain-of-function, dominant negative, or nonsense-mediated decay potential). Designed for both single-variant interpretation and high-throughput batch analysis, the platform is adaptable for clinical, research, and translational settings. To validate the pipeline, we analyzed 100 cases which were manually curated according to N1C VARIANT guidelines (PMID: 40139194). The platform demonstrated high concordance [EK1] with expert manual review, confirming its utility in streamlining variant assessment and therapeutic prioritization. By operationalizing ASO eligibility guidelines, this platform enables efficient and equitable identification of candidates for ASO therapy accelerating the development of individualized treatments for rare genetic disorders.

DeepDR-ADT: Multi-Modal Deep Learning for Drug Repurposing with Auxiliary Dosage-Time Task Learning

Subsession Time: Thursday, October 16 at 9:30am – 9:45am

Authors: Shaoyi Zhang (PSU), Sen Yang (PSU), Lida Wang (Harvard), Chachrit Khunsriraksakul, Laura Carrel (PSU), Bibo Jiang (PSU), Dajiang Liu (PSU).

Abstract: Background: Drug repurposing accelerates drug discovery by identifying new therapeutic uses for existing medications. Traditional approaches often overlook complex drug-disease relationships and dosage-dependent effects. Here, we propose a deep learning

Building a 100 million cell single cell atlas of drug perturbations across 50 human cell models for AI driven drug discovery

Subsession Time: Thursday, October 16 at 9:45am – 10:00am

Authors: Joey Pangallo (Parse Biosciences), Vuong K Tran (Parse Biosciences), Efthymia Papalexi (Parse Biosciences), Simone Marujo (Parse Biosciences), Bryan Hariadi (Parse Biosciences), Crina Curca (Parse Biosciences), Olivia Kaplan (Parse Biosciences), Sarah Sc

Abstract: High performance machine learning methods for drug discovery demand training data that capture transcriptomic responses across thousands of compound-cell combinations. However, when those transcriptomic measurements are generated across multiple experimen

Session 39: Delivering on Precision Medicine: Final Outcomes and New Insights on Returning Genetic Results Across Diverse Cohorts

Location: Room 210C/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 8:30am – 10:00am

The All of Us Research Program's Return of Research DNA Results Program: A Complete Review

Subsession Time: Thursday, October 16 at 8:30am – 8:45am

Authors: Scott Topper (Color Health), Hannah Hoban (Color Health), Richard Gibbs (Baylor College of Medicine), Eric Venner (Baylor College of Medicine), Stacey Gabriel (Broad Institute of MIT and Harvard), Heidi L. Rehm (Broad Institute of MIT and Harvard), Kim Doheny (Johns Hopkins University), Philip E. Empey (University of Pittsburgh), Gail P. Jarvik (University of Washington), Chia-Lin Wei (University of Washington), Josh Smith (University of Washington), Tara Dutka (NIH All of Us Research Program), Rhonda Moore (NIH All of Us Research Program), Anjene Musick (NIH All of Us Research Program), Anastasia Wise (NIH All of Us Research Program)

Abstract: Background/Objectives: The All of Us Research Program (AoU) aims to build a longitudinal resource of genomics and health data from over 1M people. Since its inception, returning research DNA results directly to participants has been a core element of the engagement strategy. Now complete, this large-scale implementation provides a technical roadmap and a unique opportunity to evaluate engagement, health-related population genomics data, and insights into participant follow-up.

Methods: A return of results protocol was developed to provide engaging non-clinical genetic content (genetic ancestry and traits) and meaningful health information (ACMG Secondary Findings v2.0 & PGx) to adult research participants under an Investigational Device Exemption protocol. Results were derived from research data generated for AoU. With input and oversight from IRB and FDA, risk-mitigations were incorporated including: confirmation for positive ACMG SF, no VUS reporting, genetic counseling (GC) support, and subsidized clinical confirmation. From 2020-24, the program offered >315,000 participants access to DNA results. Participants opted-in to and viewed results through an online, digital platform.

Results: In total, >70% of eligible participants engaged with the return of results program, including participants from every US state and territory, and varied socio-economic backgrounds. About 85% of participants come from groups that have had less opportunity to contribute to, or benefit from, medical research, including individuals living in geographically rural areas and from lower socioeconomic and educational backgrounds. 2.8% of result recipients had a disease-causing variant in one of the analyzed genes and were offered clinical confirmation testing. Over 4,700 participants spoke with a GC, with 93% stating that the consultation helped them understand their results. 86% of PGx result recipients had a

medication-interaction phenotype in one of seven PGx genes (CPIC Level A, FDA). Participant feedback about the results return process was overall positive via Likert-scaled questions and free text responses.

Conclusion: AoU has established a new standard for human subjects research, having directly returned research results to one of the largest cohorts to date. Participants were engaged with their research results, suggesting they valued these results, which may help drive recruitment and retention, ultimately increasing the value of the program's research database. We have also demonstrated the feasibility of using research participation as an opportunity to provide personally meaningful and impactful health-related genetic information.

RESPECT3: A randomized study evaluating return of actionable genetic research results from the Penn Medicine Biobank

Subsession Time: Thursday, October 16 at 8:45am – 9:00am

Authors: Angela R. Bradbury (University of Pennsylvania), Elisabeth M. Wood (University of Pennsylvania), Anuja R. Godbole (University of Pennsylvania), Sanjana Shastri (University of Pennsylvania), Brian Egleston (Fox Chase Cancer Center, Temple University), Lily Hoffman-Andrews (University of Pennsylvania), Sarah Brown (University of Pennsylvania), Demetrios Ofidis (University of Pennsylvania), Justin Feng (University of Pennsylvania), Anjali Owens (University of Pennsylvania), Susan Domchek (University of Pennsylvania), Bryson Katona (University of Pennsylvania), Staci Kallish (University of Pennsylvania), Giorgio Sirugo (University of Pennsylvania), Joellen Weaver (University of Pennsylvania), Linda Fleisher (Fox Chase Cancer Center, Temple University), Kuang-Yi Wen (Jefferson University), Elena Elkin (Columbia University), Katherine Nathanson (University of Pennsylvania), Daniel Rader (University of Pennsylvania).

Abstract: Background: Research participants have reported interest in receiving genetic research results. How best to return actionable genetic research results remains unclear. Unlike clinical genetic testing, research participants may not be prepared for actionable results, particularly if the results are unrelated to the reason they enrolled in research, or if they enrolled as healthy unselected participants. Methods: RESPECT3 is a randomized trial comparing return of actionable genetic research results through a digital intervention versus a genetic counselor (GC) in the Penn Medicine biobank. In Step 1, biobank participants with (cases) and without an actionable result (procedural controls) are informed of plans for return of results (ROR), invited to digital pre-disclosure education and provided options for opting out of ROR. This step allows participants to learn about the ROR and opt out before notification of an actionable result. In Step 2, those with an actionable genetic research result (ACMG v3.2) are randomized to return by a GC or the digital intervention (goal 225 results returned/arm). Results: At the interim analysis, 969 participants were sent the Step 1 notification. Mean age was 59.7 (24-99 YO), 30.5% were non-white race. Eight participants (1%) opted out of ROR. 475 participants with actionable research results (40.2% cancer, 59.8% cardiovascular results) were sent the Step 2 notification, and 47 (10%) opted out. After Step 1, 96 (10%) accessed education by web, 23 (2%) accessed by chatbot and 3 spoke with a GC. Factors associated with accessing web education were female sex ($p=0.031$) and White race ($p=0.002$). Older age was associated with requesting GC education ($p=0.03$). 20% accessed digital

education after Step 2 notification, and they were more likely to be White ($p=0.037$). Among 473 randomized participants, 107 (46%) have received their actionable results in the GC and 88 (37%) in the digital arm. There were no significant differences between arms in uptake, knowledge or distress (interim analysis). 60.5% of participants have completed confirmation testing. Conclusion: Digital tools may be acceptable for return of actionable results, although addressing low uptake of pre-disclosure education, receipt of results and confirmation testing may be important to realize the benefit of returning actionable genetic research results in biobanks not designed specifically for return of genetic research results.

Return of hereditary cancer research results to Black women: Results of a non-inferiority trial comparing online versus genetic counselor modalities

Subsession Time: Thursday, October 16 at 9:00am – 9:15am

Authors: Catharine Wang (Boston University), Kimberly Bertrand (Boston University), Michelle Trevino-Talbot (Boston University), Maureen Flynn (MGH Institute of Health Professions), Maggie Ruderman (Boston University), Mihika Shinde (Boston University), Howard J. Cabral (Boston University), Julie R. Palmer (Boston University)

Abstract: The return of genetic research results is a key challenge in precision medicine, particularly in underrepresented populations, and has led to efforts to evaluate the effectiveness of digital formats for results return. This trial, embedded within the Black Women's Health Study (BWHS), compares two modalities for returning hereditary cancer genetic results: conventional genetic counseling (GC) over the telephone and self-guided online disclosure. In particular, we aimed to evaluate the non-inferiority of the online modality with regard to test result uptake, cancer genetics knowledge, and psychosocial outcomes, including test-specific distress and anxiety/depression, at 6-week follow-up. A conservative margin of 0.3 standard deviation worse than the GC modality was used for all measures. Data were analyzed using intention-to-treat principles, with one-sided hypothesis tests at the 0.01 significance level.

A total of 807 participants were randomized to receive cancer genetic research results (positive/negative only, from a CLIA-certified laboratory) via either a GC or online through a secure website. Participants were Black women; mean (SD) age was 67 (8.5) and 69% had a college degree or higher. Over 90% of participants decided to learn their genetic results. At 6 weeks post-disclosure, cancer genetic knowledge was high and levels of anxiety, depression, test-specific distress, and decision regret were low. Online results return was noninferior to a GC results return for uptake (learning) of genetic results (95% vs 91%, $p<0.0001$). Our analyses further confirmed noninferiority for knowledge (78% vs 80% correct, $p<0.0001$), anxiety [$p<0.0001$], depression [$p<0.0007$], genetic testing distress [$p<0.008$], and decision regret [$p<0.0001$], but did not meet the noninferiority threshold for uncertainty [$p<0.08$].

Results from this study demonstrate that return of positive or negative genetic research results via an online self-guided modality is an effective option compared to telephone disclosure by a genetic counselor. Online disclosure of these types of genetic results could be a scalable alternative to traditional genetic counseling, potentially improving the accessibility of genetic information from research studies.

The GUARDIAN expanded newborn screening study: Short- and medium-term follow-up with the first 15,000 participants enrolled

Subsession Time: Thursday, October 16 at 9:15am – 9:30am

Authors: Brenna M Boyd (Columbia University Irving Medical Center), Carrie Koval-Burt (Columbia University Irving Medical Center), Katherine Langley (GeneDx), Laura Amendola (Ambry Genetics), Amber Begtrup (GeneDx), Jennifer Bradley (New York State Department of Health Newborn Screening Program), Candace Cameron (Boston Children's Hospital), Onassis Castillo Ceballo (Columbia University Irving Medical Center), Winnie Chen (Columbia University Irving Medical Center), Alison J. Coffey (Illumina), Lilian L. Cohen (NewYork-Presbyterian Weill Cornell Medical Center), Joseph M. Devaney (GeneDx), Beata Dygulska (NewYork-Presbyterian Methodist Hospital), Bethany Friedman (GeneDx), Mehmet Bugrahan Duz (Boston Children's Hospital, Harvard Medical School), Jamie Galarza-Cornejo (Columbia University Irving Medical Center), Adam Guenzel (GeneDx), Awura Gyimah (Columbia University Irving Medical Center), Rebecca Hernan (Boston Children's Hospital), Kathleen S. Hruska (GeneDx), Zhanzhi Hu (Columbia University Irving Medical Center), Mederic Jeanne (Boston Children's Hospital, Harvard Medical School), Guanjin Jin (Columbia University Irving Medical Center), D. Aaron Johnson (GeneDx), Rachel Klein (Boston Children's Hospital), Rudolph L. Leibel (Columbia University Irving Medical Center), Bobbi McGivern (GeneDx), Kirsty McWalter (GeneDx), Kristin G Monaghan (GeneDx), Leanne Owens (GeneDx), Yeyson Quevedo (Columbia University Irving Medical Center), Samantha Ratner (Columbia University Irving Medical Center), Natasha Shapiro (NewYork-Presbyterian Methodist Hospital), Robert J. Sicko (New York State Department of Health Newborn Screening Program), Samuel Strom (Fabric Genomics), Alban Ziegler (Columbia University Irving Medical Center), Rebekah Zimmerman (GeneDx), Ryan J Taft (Genetic Alliance), Denise M. Kay (New York State Department of Health Newborn Screening Program), Michele Caggana (New York State Department of Health Newborn Screening Program), Paul Kruszka (University of Virginia), Wendy K Chung (Boston Children's Hospital, Harvard Medical School)

Abstract: GUARDIAN (Genomic Uniform-screening Against Rare Disease in All Newborns) seeks to evaluate the implementation and impact of expanded genomic newborn screening (gNBS) at six New York City hospitals.

Over 28 months, 15,000/20,478 (73.2%) parents approached postnatally enrolled. NBS dried blood spots were used for gNBS. All enrollees were tested for up to 321 treatable Group 1 conditions and could opt in to 144 Group 2 neurodevelopmental conditions (gene list expanded from 237 to 455 in March 2024). The majority (92.6%) selected both groups.

Among parents who enrolled their first eligible baby, 87% enrolled their second, suggesting study comprehension and satisfaction. Of parents who declined for their first baby, 35% enrolled the next.

Genetic counselors returned 475 (3.2%) positive screens: 420 (2.8%) Group 1 and 55 (0.4%) Group 2. Confirmatory testing revealed 64 (13.5%) false positives. Notable results included G6PD (n=302), long QT syndrome (LQTS; n=18), SCN1A (n=17; ten inherited with no known family seizure history), Wilson disease (n=10), Rett syndrome (n=4), Fabry disease (n=3), severe combined immunodeficiency (n=2; one

received a bone marrow transplant), and achondroplasia/hypochondroplasia (n=2). Babies with LQTS are avoiding contraindicated medications; three began recommended medications, and one mother was diagnosed and initiated treatment. Ten Group 2 babies are symptomatic and receiving neurodevelopmental care and interventions. Diagnoses have motivated some families to utilize IVF and embryo screening for subsequent pregnancies.

Outcomes are assessed longitudinally via a post-results survey, annual health assessment, medical chart review, and interview. Approximately 30% of all families completed surveys with the highest response by Group 2 positives (~50%). The Decision Regret Scale (DRS) revealed low regret for enrollment; the Feelings About genomic Testing Results (FACToR) Scale showed low to moderate impact of results. Those with positive results reported more regret (DRS) and greater impact (FACToR) than negatives, and Group 2 positives more than Group 1. A notable decrease in Group 2 scores at 12 months on both scales suggests adjustment to the diagnosis and benefit of onboarding to care.

Results indicate low regret and high appreciation for participation both short- and longer-term, including in families receiving difficult diagnoses. Follow-up shows significant, sometimes life-saving impact on infant and family medical management.

A post-test counseling model accelerates access to cardiovascular genetic test results while maintaining psychosocial benefits of genetic counseling: Interim results of RESEQUENCE-GC (NCT05422573)

Subsession Time: Thursday, October 16 at 9:30am – 9:45am

Authors: Cynthia A. James (Division of Cardiology, Department of Medicine, Johns Hopkins University), Brittney Murray (Division of Cardiology, Department of Medicine, Johns Hopkins University), Crystal Tichnell (Division of Cardiology, Department of Medicine, John

Abstract: Post-test genetic counseling (GC) can incorporate genetic test (GT) results and may be optimal for adults seeking efficient GT to guide medical management.

RESEQUENCE-GC is a 3-arm trial testing two approaches to post-test GC for adults with a cardiovas

Implementing polygenic risk scores in clinical care: The eMERGE Network's return of genome informed risk assessment results for common conditions returned to 23,840 adults and children

Subsession Time: Thursday, October 16 at 9:45am – 10:00am

Authors: Lucinda P. Lawson (Cincinnati Children's Hospital Medical Center), Cynthia A. Prows (Cincinnati Children's Hospital Medical Center), Gail Jarvik (University of Washington), Iftikhar Kullo (Mayo Clinic), Hana Bangash (Mayo Clinic), Josh Cortopassi (Univers

Abstract: Incorporating genetic risk factors into disease screening is essential to advance precision medicine. The Electronic Medical Records and Genomics (eMERGE) Network conducted a large-scale study returning genome-informed risk assessments (GIRAs) to 23,840

Session 40: Expanding the RNAverse in Health and Disease

Location: Room 253ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 8:30am – 10:00am

Deep reinforcement learning tuned self-attention for nucleotide resolution mapping of regulatory variants in long non-coding RNA

Subsession Time: Thursday, October 16 at 8:30am – 8:45am

Authors: Kashish Kumar (Columbia University, Department of Microbiology & Immunology), Sankar Ghosh (Columbia University, Department of Microbiology & Immunology)

Abstract: Functional interpretation of variants linked to disease in long noncoding RNAs remains challenging because causal nucleotides are difficult to resolve. Reinforcement learning can optimize neural self-attention to improve regulatory predictions and localize disease-associated hotspots in lncRNAs.

lncRNA transcript/gene sets (30,436/17,139) were annotated with per-base features: GWAS SNPs with $p \leq 5.0 \times 10^{-8}$ for 485,783 variants, phastCons conservation, JASPAR motifs (~17,000 matrices, $p < 1 \times 10^{-4}$), ChIP-seq histone-mark overlaps, ATAC-seq accessibility, RNAfold structure, and RNA or protein binding target sequence. The model encodes each transcript with a 1D convolutional network, then applies a softmax-normalized self-attention layer to produce per-base attention weights and aggregate them into a scalar regulatory score. In a deep Q-learning framework, an agent observes a seven-element state: (1) normalized SNP density, (2) mean conservation, (3) TF motif density, (4) histone-mark overlap fraction, (5) ATAC-seq overlap fraction, (6) RNA or protein binding target density, and (7) attention-entropy metric, applying discrete adjustments to attention scaling and additive bias. The reward combined prediction accuracy, attention concentration, and overlap with disease-associated variants. Models were trained for 500 episodes; evaluation used ~3,000/1600 held-out transcript/gene sets under a fixed policy.

Compared to uniform attention, reinforcement tuning reduced normalized error from 0.32 to 0.18 and increased attention concentration from 0.47 to 0.88. The ROC AUC for SNP classification by attention rose from 0.59 to 0.81. Spearman's ρ averaged 0.42 ($p < 1 \times 10^{-10}$) between attention frameworks. The fraction of disease-associated variants in high-attention regions increased by 35 percent. By episode 300, 90 percent of runs achieved error < 0.2 and concentration > 0.8 in fewer than 150 steps. Sequence segments underlying attention hotspots were significantly enriched (FDR $p < 0.05$) across protein binding domains, including PRC2 subunits and spliceosomal snRNP assembly factors.

These findings suggest that a deep Q-learning model can identify functionally important nucleotides and align predictions with genetic and epigenetic evidence. The identified hotspots offer prioritized targets for CRISPR base editing, RNA-protein interaction assays, and fine-mapping causal noncoding variants, advancing mechanistic annotation of lncRNA regulatory roles in human disease.

First Comprehensive Long-Read Transcriptome Assembly of the Human Placenta Reveals Novel Molecular Mechanisms of Maternal-Fetal Health

Subsession Time: Thursday, October 16 at 8:45am – 9:00am

Authors: Sean T. Bresnahan (The University of Texas MD Anderson Cancer Center, Department of Epidemiology), William Wu (Rice University, Department of BioSciences), Hannah Yong (Agency for Science, Technology and Research, Institute for Human Development and Potential), Sierra Lopez (The University of Hawai'i at Mānoa, Department of Public Health Sciences), Jerry Kok Yen Chan (KK Women's and Children's Hospital, Duke-NUS Medical School), Shiao-yng Chan (Agency for Science, Technology and Research, Institute for Human Development and Potential, National University of Singapore, Yong Loo Lin School of Medicine, Department of Obstetrics & Gynecology), Jonathan Y. Huang (The University of Hawai'i at Mānoa, Department of Public Health Sciences, Duke-NUS Medical School, National University of Singapore, Yong Loo Lin School of Medicine, Department of Obstetrics & Gynecology), Arjun Bhattacharya (The University of Texas MD Anderson Cancer Center, Department of Epidemiology)

Abstract: Background: The placenta mediates all maternal-fetal interactions yet has been notably absent from major genomic initiatives, creating a significant knowledge gap in understanding human development. Current analyses rely on adult tissue annotations that miss nearly 40% of placental transcript diversity, obscuring the molecular mechanisms underlying pregnancy complications and developmental origins of disease.

Methods: We generated the largest placental long-read RNA sequencing dataset (N = 72) to create a comprehensive transcript reference, then demonstrated its utility by analyzing gestational diabetes mellitus (GDM) effects in two multi-ancestry birth cohorts (N = 344) and PFAS exposure effects using fetal-facing placenta samples. We performed isoform-resolved differential expression and mediation analyses to identify molecular pathways linking maternal exposures to fetal outcomes.

Results: Our assembly yielded 37,661 high-confidence isoforms across 12,302 genes, including 14,985 novel transcript variants and 2,759 previously unannotated genes, with remarkable complexity concentrated in critical endocrine regulators including CSH1 and GH2. This improved reference reduced inferential uncertainty in quantification of short-read datasets by 1.15-fold relative to GENCODE annotations and identified associations undetected at the gene level. We discovered novel population-specific CSH1 isoforms mediating 5-10% of GDM effects on birth weight, providing transcript-level evidence of developmental programming mechanisms. For PFAS toxicants, we identified maternal-fetal transfer mechanisms and placental endocrine mediators essential for understanding how early-life chemical exposures program disease risk.

Conclusions: This work demonstrates that isoform-level placental transcriptomics analyses provide enhanced sensitivity for understanding developmental programming compared to gene-level approaches. Our findings reveal previously undetected molecular associations through which maternal exposures affect fetal development, with broad implications for maternal-fetal medicine. We provide this valuable resource publicly, establishing a foundation that will advance investigation of regulatory

variants, evolutionary innovations, and population-specific mechanisms across developmental genetics and precision pregnancy medicine.

GTEX and GENCODE 47 integration expands the role of long-noncoding RNAs in complex traits and disease

Subsession Time: Thursday, October 16 at 9:00am – 9:15am

Authors: Kathryn A. Lawrence (Stanford University Department of Genetics), Daniel Nachun (Stanford University School of Medicine Department of Pathology), Tamara Perteghella (Barcelona Institute of Science and Technology, Centre for Genomic Regulation (CRG), Universitat Pompeu Fabra), Francois Aguet (Broad Institute), Kristin Ardlie (Broad Institute), Roderic Guigo (Barcelona Institute of Science and Technology Centre for Genomic Regulation (CRG), Universitat Pompeu Fabra), Stephen B. Montgomery (Stanford University School of Medicine Department of Pathology, Stanford University Department of Genetics)

Abstract: Long non-coding RNAs (lncRNAs) have been challenging to definitively link to specific functions, but the recent GENCODE 47 release—which adds roughly 18,000 newly annotated lncRNA genes—presents a unique opportunity to investigate how genetic effects on these transcripts may shape complex traits. To that end, we re-aligned and re-quantified the complete V10 Genotype-Tissue Expression (GTEx) multi-tissue transcriptomics dataset with GENCODE 47. We quantified tissue-specific expression patterns for 34,914 (18,307 newly annotated) lncRNAs across 49 GTEx tissues, finding 20,180 (8,612 newly annotated) lncRNAs to be expressed in at least one tissue at levels sufficient for QTL mapping. We mapped expression and splicing quantitative trait loci (e/sQTL), identifying 33% (6,106) of newly annotated lncRNAs as eGenes and 4.9% (887) as sGenes. To explore complex trait and disease relevance, we colocalized lncRNA e/sQTL with genome-wide association study (GWAS) loci, identifying new colocalizations with GWAS loci, independent of nearby protein-coding genes. Because lncRNAs often share regulation with nearby protein-coding genes, incorporating lncRNAs into integrative multi-gene analyses improved detection of QTLs and trait colocalizations for nearby protein-coding genes. Generalizing this to protein-coding genes which also share regulation; for protein-coding genes with correlated expression, multi-gene analyses increased GWAS colocalizations by 34% over single-gene methods. Combined, this work provides an updated GTEx QTL resource for the community, highlights the importance of moving beyond conventional single-gene QTL methods, and demonstrates novel associations and trait-relevance among 18,000 novel lncRNAs annotated in GENCODE 47.

Comprehensive long non-coding and circular RNA atlas of human brain regions across multi-ancestry cohorts

Subsession Time: Thursday, October 16 at 9:15am – 9:30am

Authors: Alexander Knyshov (Department of Medicine, Boston University), Junming Hu (Department of Medicine, Boston University), Mintao Lin (Department of Medicine, Boston University), Liyuan Hu (Department of Medicine, Boston University), Jenny A. Empawi (Department of Medicine, Boston University), Sambhavi Puri (Department of Anatomy & Neurobiology, Boston University), Jenna L. Libera

(Department of Anatomy & Neurobiology, Boston University), Thor D. Stein (Department of Pathology and Laboratory Medicine, Boston University), Lindsay A. Farrer (Department of Medicine, Boston University), Benjamin Wolozin (Department of Anatomy & Neurobiology, Boston University), and Xiaoling Zhang (Department of Medicine, Boston University)

Abstract: Long non-coding RNA (lncRNA) is the largest class of human non-protein-coding transcripts, most commonly subdivided from other non-coding RNA (ncRNA) based on their length of at least 200 nucleotides. Circular RNA (circRNA) is another important class of ncRNAs, characterized by covalently closed-loop structures formed from exonic back-splicing events. To date, a comprehensive atlas of these ncRNAs in human brain tissues is lacking, primarily due to their low expression levels and high tissue- and region-specificity, which hinder their detection and characterization. Nevertheless, both lncRNAs and circRNAs play critical roles in various regulatory processes, and growing evidence implicates them in the pathogenesis of brain disorders. CircRNAs, in particular, are highly abundant in brain tissues, making the brain an ideal target for circRNA profiling. In this study, we analyzed the largest total/bulk RNA-seq dataset (ribo-depleted, short-read) from aged human brain samples, spanning multiple brain regions, to systematically profile lncRNAs and circRNAs expression. Our discovery dataset contains ~3,200 samples from seven brain regions, derived from three multi-ancestry participants (European ancestry, African American, and Hispanic). An independent validation dataset includes 1,140 samples from eight brain regions across three additional cohorts. Preliminary analyses detected ~15,100 lncRNAs and ~9,200 circRNAs passing quality control, including ~7,300 lncRNAs and ~1,500 circRNAs consistently expressed across all brain regions. To assess evolutionary conservation, we compared these findings with brain RNA-seq data from six non-human primate species. We then analyzed four publicly available RNase R-treated long-read nanopore RNA-seq samples from human brain tissue to investigate circRNA isoforms. Finally, we examined signatures of circRNA biogenesis, including enrichment of RNA-binding protein (RBP) motifs and inverted repeat elements/distribution, and integrated N6-methyladenosine (m6A) modification sites derived from our in-house DART-seq profiling of 38 human cortex samples to ascertain their post-transcriptional regulation. Collectively, our study provides a synoptic resource for understanding the expression patterns and regulation of lncRNAs and circRNAs in the aged human brain. This atlas lays a foundation for future functional studies to elucidate their roles in cognitive decline and brain neurodegenerative diseases.

Identification of long non-coding RNAs potentially involved in inherited retinal disease pathogenesis

Subsession Time: Thursday, October 16 at 9:30am – 9:45am

Authors: Sandro Banfi (University of Campania "Luigi Vanvitelli", Telethon Institute of Genetics and Medicine), Dalila Capasso (University of Campania "Luigi Vanvitelli", Telethon Institute of Genetics and Medicine), Maria Zawadzka (Telethon Institute of Genetics and Medicine)

Abstract: Introduction

Inherited retinal diseases (IRDs) are a group of disorders that cause vision impairment via photoreceptor (PR) cell death. They display high genetic heterogeneity, which represents an obstacle to an effective molecular diagnosis and renders d

A novel human β cell microprotein from a lincRNA controls insulin secretion and harbors rare null variants associated with type 2 diabetes

Subsession Time: Thursday, October 16 at 9:45am – 10:00am

Authors: Frédéric Oger (Inserm 1283), Bénédicte Toussaint (Inserm 1283), Judith Merrheim (Inserm 1283), Mehdi Derhourhi (Inserm 1283), Arno R. Van der Silk (Leiden University Medical Center), Jérôme Delplanque (Inserm 1283), Raphaël Scharfmann (Inserm 1016), Phili

Abstract: Emerging evidence indicates that some long non-coding RNAs (lncRNAs) encode functional microproteins through translation of small open reading frames (sORFs). Although these non-canonical peptides are increasingly implicated in cellular regulation and in

Session 41: Human Diversity and Evolution

Location: Room 156ABC/Level 1, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 8:30am – 10:00am

Is Directional Selection Rare in the Recent Human Past? Insights from Ancient DNA Time Series

Subsession Time: Thursday, October 16 at 8:30am – 8:45am

Authors: Ali Akbari (Harvard University), David Reich (Harvard University)

Abstract: Natural selection has been extensively studied using present-day human genomes. These analyses suggest that classic selective sweeps, a specific form of directional selection in which a de novo beneficial mutation arises and reaches fixation, have been rare over the past 250,000 years (Hernandez et al. 2011). Patterns of neutral diversity and fixed differences among continental populations (e.g., Africans, Europeans, and Asians) show limited evidence for such sweeps (Coop et al. 2009), with most variation shaped by admixture and genetic drift (Simon and Coop 2024). Genome-wide diversity is largely influenced by background selection acting on linked deleterious variants (Murphy et al. 2022), leaving little detectable contribution from directional selection.

To provide new insights into this question, we analyzed time series data from 15,836 ancient genomes spanning the Holocene and 6,438 modern samples from Western Eurasia. We identified 488 loci with genome-wide significant signals of directional selection. Although directional selection accounts for only about 2% of genome-wide allele frequency changes, consistent with most variation being driven by admixture and drift (Simon and Coop 2024), its presence across hundreds of loci highlights how directional selection has, in fact, been widespread.

Among the 488 loci, 29% involve selection on ancestral alleles, and 36% on derived alleles that originated more than 50,000 years ago, suggesting that selective pressure on these alleles likely fluctuated over time. Many of the remaining 35% appear to represent ongoing sweeps that may never reach fixation, unlike classic selective sweeps that go fully to fixation. Classic selective sweeps can take tens of thousands of years to complete, requiring a beneficial mutation to arise, survive the early stochastic phase, and gradually reach fixation. Changing environmental conditions can lead to shifts in selective pressure during this process, altering the trajectory or halting the sweep entirely.

We also find that selective pressure intensified over the past 10,000 years. Selection on immune and cardiometabolic alleles increased significantly during the Bronze Age compared to the Mesolithic and Neolithic periods and has remained elevated since. This acceleration coincides with cultural and environmental changes, including the rise of agriculture, urbanization, and increased population density, all of which likely introduced new selective pressures on human biology.

Spatiotemporal dynamics of adaptive evolution in Neolithic East Asia

Subsession Time: Thursday, October 16 at 8:45am – 9:00am

Authors: Dan Ju (Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences), Xiaotian Feng (Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences), Qiaomei Fu (Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences)

Abstract: The advent of farming introduced a suite of new selective pressures related to changes in diet, mobility, and pathogen exposure. Previous studies of selection related to agriculture have mostly examined ancient genomes from West Eurasia, whereas few such studies have been done in East Asia. Taking advantage of the recent growth in the number of ancient East Asian genomes, we assembled the largest dataset of published ancient genomes from East Eurasia, the majority of from the last 12k years, to investigate the timing and strength of selection (s) of adaptive alleles at known and novel loci implicated from our selection scan. Given the low coverage of ancient DNA, we performed imputation using GLIMPSE2 with the high coverage 1000 Genomes reference panel to increase statistical power. After filtering the ancient data for coverage >0.2x, removing close relatives, and merging with modern HGDP East Asians, we had available 1495 individuals for our time series analysis.

One of the strongest signals of selection in East Asia was ADH1B ($P=2.12 \times 10^{-16}$, $s=0.017$), a well-established adaptive locus involved in the oxidation of ethanol. Stratifying our dataset by latitude based on the Qinling Mountains, the traditional demarcation between northern and southern China, we observed an earlier onset of selection in the south that coincides with the domestication of rice around 11k years before present (yBP). In the north, however, the derived allele began increasing in frequency around 6k yBP, which previous archaeobotanical research has indicated to be around the time of expansion of the farming of a variety of rice more suitable for temperate conditions. In addition, we show that another key variant downstream in the alcohol metabolism pathway, rs671 (ALDH2), rose in frequency much later, likely in the past two thousand years. A derived allele of FADS2, an important player in fatty acid metabolism, showed a striking increase in frequency ($P=9.08 \times 10^{-11}$, $s=0.0065$), which began prior to the Neolithic. Interestingly, in the south, changes in ancestry could largely explain the change in frequency, whereas in the north, selection was significantly stronger based on modelling the interaction between location and date ($P=3.48 \times 10^{-4}$). Finally, we found a novel candidate of selection related to immunity at TMEM229B ($P=2.65 \times 10^{-9}$, $s=0.0037$) that was increasing in frequency since the Upper Paleolithic.

A comprehensive map of introgressed and adaptive structural variation in diverse humans

Subsession Time: Thursday, October 16 at 9:00am – 9:15am

Authors: PingHsun Hsieh (Department of Genetics, Cell Biology, and Development, University of Minnesota, Twin Cities, MN, USA), Natthapon Soisangwan (Bioinformatics and Computational Biology Graduate Program, University of Minnesota, Twin Cities, MN, USA), David Gordon (Department of Genetics, Cell Biology, and Development, University of Minnesota, Twin Cities, MN, USA), Athef Javidh (Bioinformatics and Computational Biology Graduate Program, University of Minnesota, Twin Cities, MN,

USA), William Harvey (Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA), David Porubsky (Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA), Kendra Hoekzema (Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA), Carl A. Baker (Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA), Katherine M. Munson (Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA), Christopher Kinipi (Strand of Anthropology, Sociology and Archaeology, School of Humanities and Social Sciences, University of Papua New Guinea, University 134, National Capital District, Papua New Guinea), Matthew Leavesley (Strand of Anthropology, Sociology and Archaeology, School of Humanities and Social Sciences, University of Papua New Guinea, University 134, National Capital District, Papua New Guinea), Nicolas Brucato (Centre de Recherche sur la Biodiversité et l'Environnement, Université de Toulouse, CNRS, IRD, Toulouse INP, Université Toulouse 3 – Paul Sabatier, Toulouse, France), Murray P. Cox (College of Sciences, Massey University, Palmerston North, New Zealand), François-X Ricaut (Centre de Recherche sur la Biodiversité et l'Environnement, Université de Toulouse, CNRS, IRD, Toulouse INP, Université Toulouse 3 – Paul Sabatier, Toulouse, France), Irene Gallego Romero (Human Genomics and Evolution, St Vincent's Institute of Medical Research, Fitzroy, Australia), Evan E. Eichler (Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA)

Abstract: Genetic introgression is ubiquitous across the tree of life and contributes to evolutionary novelty. In humans, 2–4% of non-African genomes derive from Neanderthal and/or Denisovan DNA; some of which are associated with various adaptations. However, the role of introgressed structural variants (SVs, ≥ 50 bp), which can strongly impact function, remains unclear. We integrate four high-quality, nearly telomere-to-telomere Papuan New Guinean (PNG) assemblies with 94 published assemblies of diverse ancestry to map archaic SVs. Including PNG genomes is crucial because Oceanian populations are absent from recent pangenome efforts and helps capture missing variation. Notably, we identify large SVs at biomedically important loci, such as 22q11.2 (linked to DiGeorge syndrome), that are rare or absent elsewhere, helping to fill major gaps in human genomics. Our map shows that the PNG diploid genomes carry two-fold more DNA derived from archaic hominin compared to other non-African genomes (mean fraction: 4.10% in PNG, 1.34% in admixed American ($n=16$), 2.13% in East Asian ($n=5$), 1.19% in European ($n=1$), and 1.72% in South Asian ($n=1$). The PNG also carries the highest amount of introgressed SVs (mean: 1,778 SVs), which together affect 1.9 million base pairs in the genome. In contrast, Admixed Americans have the least amount of introgressed SVs (mean: 521 SVs) affecting ~ 0.5 million base pairs. Introgressed SVs are overall enriched in genes (44%, $n=1,592$), including critical genomic disorder regions, and most abundant in PNG. Our map also enables us to infer the roles of SVs in human adaptations. Using Pangenome genotyping of 1,363 samples, we identified 16 candidate adaptive SVs, many linked to immune-related genes and expression. Our population genetics inferences suggest that these may reflect increased selective pressure from pathogen transmission following intensified population movements driven by agriculture in Papua New Guinea $\sim 6,000$ years ago. Notably, we report what is, to our knowledge, the first evidence of introgressed centromeres from archaic hominins in our Papuan assemblies. These fully resolved, multimillion-base centromeric sequences represent some of the largest archaic introgressed loci identified in modern humans. Along with many

new SVs reported in this study, our findings offer unique insights into the structure and organization of ancient genomes and their roles in human evolution and biology.

PHATE Derived Ancestry Coordinates Can Effectively Capture Continuous Population Structure in Human Genomic Data

Subsession Time: Thursday, October 16 at 9:15am – 9:30am

Authors: Matthew Scicluna (Montreal Heart Institute, Research Center, QC, Canada, Département de biochimie et médecine moléculaire, Université de Montréal, QC, Canada, Mila – Quebec AI Institute, Montreal, QC, Canada), Shuang Ni (Mila – Quebec AI Institute, Montreal, QC, Canada, Département d’informatique et de recherche opérationnelle, Université de Montréal, QC, Canada), César Miguel Valdez Córdova (Mila – Quebec AI Institute, Montreal, QC, Canada, Quantitative Life Sciences, McGill University, QC, Canada), Jean-Christophe Grenier (Montreal Heart Institute, Research Center, QC, Canada), Raphael Poujol (Montreal Heart Institute, Research Center, QC, Canada), Kevin R. Moon, PhD (Department of Mathematics and Statistics, Utah State University, UT, United States), Guy Wolf, PhD (Mila – Quebec AI Institute, Montreal, QC, Canada, Département de mathématiques et de statistique, Université de Montréal, QC, Canada), Sebastien Lemieux, PhD (Département de biochimie et médecine moléculaire, Université de Montréal, QC, Canada, Mila – Quebec AI Institute, Montreal, QC, Canada), Smita Krishnaswamy, PhD (Department of Genetics, Yale University, CT, United States, Department of Computer Science, Yale University, CT, United States), Julie Hussin, PhD (Montreal Heart Institute, Research Center, QC, Canada, Département de biochimie et médecine moléculaire, Université de Montréal, QC, Canada, Mila – Quebec AI Institute, Montreal, QC, Canada, Département de Médecine, Université de Montréal, QC, Canada)

Abstract: Current visualization methods can misrepresent the continuous nature of human genetic diversity. As biobanks expand in scale and diversity, these visualizations play a growing role in interpreting population structure. Principal Component Analysis has traditionally been used, but struggles to capture localized patterns due to its linear nature. Nonlinear methods like t-SNE and UMAP better preserve local structure but often distort global relationships, complicating insights into cross-population patterns such as admixture and migration. Often, these methods isolate populations into eye-catching “blobs” whose shapes can be sensitive to the choice of parameters. These representations risk oversimplifying the complexities of human diversity, falsely implying sharp boundaries between groups, an issue with real-world implications in biomedicine and ancestry interpretation. Here, we developed a framework based on PHATE, which captures local and global relationships by simulating heat diffusion and projecting data into a lower-dimensional space. Additionally, we developed novel metrics to quantify how well our embeddings capture both admixture and geography, providing a rigorous means to assess robustness to parameter choices. Across simulated data and major cohorts, including the Thousand Genomes Project (1KGP), Human Genome Diversity Project (HGDP), UK Biobank, All of Us and Canadian Biobanks, our framework consistently reveals geographic and admixture signals at multiple scales. In particular, PHATE captures admixture gradients in populations with mixed ancestry, organizing them according to admixture proportions—an aspect that t-SNE and UMAP fail to represent. PHATEs visualization of All of Us preserves the dense European–African–Amerindigenous continuum seen in the

Americas. Our work introduces PHATE as a robust tool for exploring population structure at scale, offering new perspectives for studying human evolutionary history. These improved representations of worldwide genetic variation support more equitable approaches to genomic research.

Stable genetic risk for myopia over 15,000 years in East Asian and European populations reveals predominant environmental drivers of the modern myopia epidemic

Subsession Time: Thursday, October 16 at 9:30am – 9:45am

Authors: Xikun Han (Peking University), Zhongying Liu (Chengdu University of Traditional Chinese Medicine), Riley J. Mangan (Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, The Broad Institute of MIT and Harvard, Genetics T

Abstract: The prevalence of myopia has risen to > 80% among East-Asian adolescents, but remains far less common in Europeans (~ 25–30%). Twin and family studies show that refractive error is highly heritable (50–90%), and recent GWAS have mapped >400 loci, establish

Machine learning identifies the genomic and statistical features that predict cross-population replication of GWAS signals

Subsession Time: Thursday, October 16 at 9:45am – 10:00am

Authors: Joseph Lachance (Georgia Institute of Technology), Ujani Hazra (Georgia Institute of Technology),

Abstract: Genome-wide association studies have identified hundreds of thousands of trait-associated variants, yet many of these associations fail to replicate across continental populations. Although the challenges facing the portability of polygenic scores are well

Session 42: Massively Parallel Experimental and AI Approaches for Variant Characterization

Location: Room 210AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 8:30am – 10:00am

Variant pre-classification using ~280,000 variant effect measurements in 42 disease genes illustrates a realistic, scalable path to dramatically reduce variants of uncertain significance

Subsession Time: Thursday, October 16 at 8:30am – 8:45am

Authors: Douglas M. Fowler (University of Washington), Malvika Tejura (University of Washington), Daniel Zeiberg (Northeastern University), Abbye E. McEwen (University of Washington), Pankhuri Gupta (University of Washington), Yile Chen (University of Washington), Shawn Fayer (University of Washington), Jeremy Stone (University of Washington), Ben Capodanno (University of Washington), Anne O'Donnell-Luria (Broad Institute), Predrag Radivojac (Northeastern University), Vikas Pejaver (Icahn School of Medicine at Mount Sinai), Alan F. Rubin (Walter and Eliza Hall Institute), Lea M. Starita (University of Washington), and The IGVF Coding Variant Focus Group

Abstract: Approximately 90% of missense variants in the ClinVar database are variants of uncertain significance (VUS), which arise due to a lack of evidence for variant classification and pose a central challenge to genomic medicine. Functional data from multiplexed assays of variant effect, and computational predictions from variant effect predictors are a scalable, laboratory-generated source of evidence. Major barriers to using functional data to resolve VUS include the lack of data for most genes, inadequate curation of existing data, and missing infrastructure to enable clinicians to use the data. As part of the NHGRI's Impact of Genetic Variation on Function consortium (IGVF), we addressed these challenges for ~1% of the human clinical genome. We used saturation genome editing and VAMP-seq to generate functional data for ~60,000 variants in 9 genes and curated a further ~280,000 variant effect measurements in 33 genes generated by the community. We used these data to reassess the 77,000 VUS in these 43 genes, where we were able to reclassify ~65% of VUS. When classified using only functional and predictive evidence, ~1% of known pathogenic or benign variants were misclassified, highlighting the robustness of our approach. We also used functional and predictive evidence to pre-classify previously unseen variants in these genes, providing enough evidence for 27% (150,000 variants) to be immediately classified as likely pathogenic or likely benign when observed in the clinic. Because functional evidence is being generated at scale and predictive evidence is already available for nearly all variants, this approach provides a realistic, scalable path to dramatically reduce VUS.

T cell-specific causative variants underlying immune-mediated inflammatory disorders

Subsession Time: Thursday, October 16 at 8:45am – 9:00am

Authors: Paige Elizabeth Bond (Penn State College of Medicine), Jeniece Regan (Penn State College of Medicine), Dajiang Liu (Penn State College of Medicine), Laura Carrel (Penn State College of Medicine)

Abstract: Immune-mediated inflammatory disorders (IMIDs) are characterized by an aberrant immune response against healthy cells, in part due to loss of T cell tolerance. IMIDs are common, affecting up to 5% of the Western population. A better understanding of underlying genetic loci is essential to accurately assess risk and develop more effective treatments. Although GWAS have identified hundreds of associated loci, many of which are shared across IMIDs, they only explain a portion of heritability. To improve cell-type specific eQTL analyses and subsequent TWAS, we developed JOint model viewing of Bk-eQTL as a weighted sum of Sc-eQTLs (JOBS) that combines sc-RNaseq and bulk RNaseq to increase power (Cell Genomics 2025,5:100820). Applying JOBS to immune cell expression data identified cell-type specific loci in IMIDs, including a large number of T cell loci. To identify causative variants at these loci that could contribute to the loss of T cell tolerance and aberrant activation seen in autoimmunity, we selected 147 novel loci across 12 immune-mediated disorders. We screened 9,461 variants that were in LD with lead variants at these loci and overlapped functionally annotated regulatory elements. To evaluate risk and reference alleles of these variants, we performed a massively parallel reporter assay (MPRA) for 23,419 candidate regulatory sequences to identify allelic enhancer effects on expression in Jurkat CD4 T cells. Of particular interest is a chr11 locus that is associated with T cells effects in three autoimmune disorders, including systemic lupus erythematosus and rheumatoid arthritis. Both JOBS and autoimmune GWAS identified the CCDC88B gene at this locus. To understand the role of CCDC88B in T-cell function, we performed a CRISPR knockout of CCDC88B in a Jurkat CD4 T cell line. Subsequent RNaseq highlighted 189 differentially expressed genes. Consistent with in vivo mouse studies, we observed pleiotropic effects on CD4 T activation and function, supporting the involvement of this locus in autoimmunity. Allelic variants at this locus likely impact CCDC88B regulation and explain association of this locus across many autoimmune disorders.

Global identification of inflammatory bowel disease risk variant allele-dependent gene regulatory activity

Subsession Time: Thursday, October 16 at 9:00am – 9:15am

Authors: Lois G. Parks (Cincinnati Children's Hospital Med. Ctr., University of Cincinnati), Carmy Forney (Cincinnati Children's Hospital Med. Ctr.), Xiaoting Chen (Cincinnati Children's Hospital Med. Ctr.), Sreeja Parameswaran (Cincinnati Children's Hospital Med. Ctr.), Marissa Granitto (Cincinnati Children's Hospital Med. Ctr., University of Cincinnati), Omer Donmez (Cincinnati Children's Hospital Med. Ctr.), Cailing Yin (Cincinnati Children's Hospital Med. Ctr.), Matthew T. Weirauch (Cincinnati Children's Hospital Med. Ctr., University of Cincinnati), Leah C. Kottyan (Cincinnati Children's Hospital Med. Ctr., University of Cincinnati)

Abstract: Inflammatory bowel diseases (IBD) are incurable conditions characterized by chronic gastrointestinal inflammation and progressive tissue damage. IBD etiology is complex, with both genetic and environmental factors contributing to disease susceptibility. Genome-wide association studies have identified hundreds of IBD susceptibility loci, but the biological processes underlying the genetic risk of

IBD remain unclear. IBD genetic risk loci predominantly contain non-coding variants, with regulation of gene expression implicated as the likely causal disease mechanisms. To systematically identify variants with allele-dependent gene regulatory activity, we constructed massively parallel reporter assays (MPRAs) encompassing 20,017 IBD risk variants and their flanking genomic sequences. In addition to an MPRA that uses a minimal promoter, we built an MPRA using a promoter with strong basal activity (RSV), allowing the identification of silencing regulatory elements. Application of the MPRAs in Jurkat T cells with and without T cell receptor stimulation revealed extensive context-dependent allelic transcriptional activity at IBD risk variants. In total, we nominate 209 variants with allelic enhancer activity and 510 variants with allelic silencer activity at 176 distinct IBD risk loci. Ongoing efforts are focused on establishing a lentivirus-based MPRA to examine variants within the chromatin context of Jurkat T cells and THP-1 macrophages. Additionally, we are performing single-cell CRISPR interference assays to link the identified regulatory variants to putative target genes. Collectively, this work will provide insights into the genotype-dependent biological processes underlying the complex genetic risk of IBD.

Cell2net: A Deep Learning Framework to Decode Genotype-Specific Gene Regulatory Networks from Single-Cell Multi-Omics

Subsession Time: Thursday, October 16 at 9:15am – 9:30am

Authors: Zhijian Li (Broad Institute, Massachusetts General Hospital), Jiecong Lin (Broad Institute, Massachusetts General Hospital), Eugenio Mattei (Broad Institute), Noam shoshitaishvili (Broad Institute), Bradley Bernstein (Broad Institute, Dana Farber Cancer Institute, Harvard Medical School), Jason Buenrostro (Broad Institute, Harvard University) and Luca Pinello (Broad Institute, Massachusetts General Hospital, Harvard Medical School)

Abstract: Gene transcription is a fundamental process that governs cellular identity and function. It is tightly regulated by complex gene regulatory networks (GRNs), which represent interactions between transcription factors (TFs), regulatory DNA elements, and target genes. Advances in single-cell multi-omic technologies have enabled the inference of GRNs at unprecedented resolution. However, current computational models often fail to incorporate genetic variants as an input parameter, even though they have been shown to modulate TF binding, chromatin accessibility, and ultimately gene expression.

To address this gap, we developed Cell2net, a deep learning framework designed to predict gene expression at single-cell resolution by integrating individual-specific cis-regulatory DNA sequences, chromatin accessibility, and TF expression. By leveraging state-of-the-art machine learning interpretability techniques, Cell2net quantifies the contribution of genetic variants, TF expression and chromatin state to transcriptional regulation, inferring genotype-specific GRNs.

We benchmarked Cell2net on four public single-cell multi-omic datasets and demonstrated its outperformance over existing methods in predicting gene expression. Moreover, we generated a unique dataset of “population villages” of human induced pluripotent stem cells (iPSCs), collected from 270 genotyped donors. These cells were differentiated into hepatocytes and neural progenitor cells and profiled using SHARE-seq, a single-cell multi-modal technology that jointly captures chromatin accessibility and gene expression. We applied Cell2net to this dataset and revealed how genetic variants

shape transcriptional programs during the cell differentiation, highlighting key regulatory mechanisms and TFs that mediate inter-individual variability in gene expression. In summary, Cell2net provides a scalable, interpretable, and genotype-aware framework for dissecting transcriptional regulation at single-cell resolution.

By bridging genomic sequence variation with regulatory output, Cell2net enables systematic reconstruction of GRNs across individuals and cell types. This work offers powerful new capabilities for exploring the genetic basis of gene regulation and its implications in complex traits, disease susceptibility, and personalized functional genomics.

AlphaGenome: advancing regulatory variant effect prediction with a unified DNA sequence model

Subsession Time: Thursday, October 16 at 9:30am – 9:45am

Authors: Natasha Latysheva (Google DeepMind), Žiga Avsec (Google DeepMind), Jun Cheng (Google DeepMind), Guido Novati (Google DeepMind), Kyle R. Taylor (Google DeepMind), Tom Ward (Google DeepMind), Clare Bycroft (Google DeepMind), Lauren Nicolaisen (Google DeepMi

Abstract: Deep learning models that predict functional genomic measurements from DNA sequence are powerful tools for deciphering the genetic regulatory code. Existing methods trade off between input sequence length, prediction resolution and modality coverage, ther

Functional annotation of over 25,000 type 2 diabetes-associated variants

Subsession Time: Thursday, October 16 at 9:45am – 10:00am

Authors: Adelaide Tovar (University of Michigan), Romy Kursawe (The Jackson Laboratory for Genomic Medicine), Amy S. Etheridge (The University of North Carolina at Chapel Hill), Jonathan D. Rosen (The University of North Carolina at Chapel Hill), Kirsten Nishino (

Abstract: Genome-wide association studies have identified >660 signals associated with type 2 diabetes (T2D), with risk variants enriched in regulatory regions active in insulin-producing pancreatic islets and other relevant metabolic tissues including adipose, liv

Session 43: Statistical Genetics: Multi-Omics to Improve Gene Discovery

Location: Room 205ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 8:30am – 10:00am

Integrating Genomic and Proteomic Data Improves Complex Trait Prediction in Diverse Populations

Subsession Time: Thursday, October 16 at 8:30am – 8:45am

Authors: Wenbo Wang (University of North Carolina at Chapel Hill) , Jacob Williams (Division of Cancer Epidemiology and Genetics, National Cancer Institute), Joseph G. Ibrahim (University of North Carolina at Chapel Hill), Haoyu Zhang (Division of Cancer Epidemiology and Genetics, National Cancer Institute), Xihao Li (University of North Carolina at Chapel Hill)

Abstract: Introduction

Polygenic risk scores (PRS) are widely used to quantify the genetic risk for complex traits. Meanwhile, the growing availability of large-scale proteomic data has enabled new opportunities to include proteomic biomarkers for disease and trait prediction. However, the development and evaluation of proteomic risk scores (ProRS), particularly in combination with PRS, remains limited. In particular, the predictive utility and transferability of these integrated scores across diverse populations and complex traits have not been comprehensively studied.

Methods

We developed an ensemble modeling framework to assess the total and mediated prediction effects from genomic and proteomic data. After imputing missing proteomic data by the EM algorithm and adjusting for batch effects, ProRS models were constructed using superlearning, an ensemble of statistical and machine learning methods. ProRS models were developed using a training-validation-testing random split of 36,903 unrelated UK Biobank participants with plasma proteomic data across four ancestry groups: European (EUR), African (AFR), Admixed American (AMR) and South Asian (SAS). For PRS, we used cutting-edge PRSs from the PGS Catalog excluding UK Biobank participants. The combined PRS+ProRS model was evaluated across five binary and six continuous traits using the same testing dataset. We assessed prediction performance and the extent to which ProRS mediated PRS effects. For binary traits, we also compared prediction accuracy based on whether biomarkers were collected before or after disease diagnosis.

Results

The combined PRS+ProRS significantly improved across traits and ancestries. On average, PRS alone yielded R² of 3.6% for continuous traits and AUC of 60.7% for binary traits. With PRS+ProRS, average R² increased to 45.3%, and AUC to 73.5%. Relative to PRS-only models, R² improved by an average of 42-

fold, and AUC by 22.7%. Notable gain included prostate cancer in SAS (AUC: 47.4% -> 79.6%), asthma in EUR (AUC: 67.7% -> 80.5%), type 2 diabetes in EUR (AUC: 63.3% -> 84.4%) and breast cancer in non-EUR (AUC: 67.0% -> 75.0%). Among cases diagnosed within 10 years of biomarker collection, prediction was stronger when proteins were measured after rather than before disease diagnosis (AUC 83.2% vs. 71.8%). Mediation analysis showed that 1.5% - 23.2% of the PRS effect was mediated through ProRS for binary traits, and 25.0% - 71.2% for continuous traits in EUR. The binary and continuous traits corresponding to these largest mediation effects were type 2 diabetes and total cholesterol, respectively.

Summary

Our modeling approach provides a comprehensive analytical framework integrating genomic and proteomic data in complex trait prediction, offering accurate and interpretable results across diverse populations.

Integrative multi-omics approaches identify key molecular pathways and improve risk prediction of Late-Onset Alzheimer's Disease

Subsession Time: Thursday, October 16 at 8:45am – 9:00am

Authors: Rasika Venkatesh (University of Pennsylvania), Katie M. Cardone (University of Pennsylvania), Yuki Bradford (University of Pennsylvania), Anni K. Moore (University of Pennsylvania), Rachit Kumar (University of Pennsylvania), Molly A. Hall (University of Pennsylvania), Jason H. Moore (Cedars-Sinai Medical Center, Center for Artificial Intelligence Research and Education), Li Shen (University of Pennsylvania, Institute for Biomedical Informatics), Dokyoon Kim (University of Pennsylvania, Institute for Biomedical Informatics), Marylyn D. Ritchie (University of Pennsylvania, Institute for Biomedical Informatics)

Abstract: Alzheimer's Disease (AD) is the most prevalent condition impacting the aging population, with no effective treatment currently available. Characterizing the heritable risk of late-onset AD is challenging, as genetic models like polygenic risk scores (PRS) often fail to capture disease heterogeneity, resulting in limited predictive accuracy. This study used univariate and multivariate approaches to develop novel machine learning-based integrative risk models (IRMs) that capture context-specific biological information from imputed multi-omics data, improving AD risk prediction. We performed genome-, transcriptome-, and proteome-wide association studies, followed by pathway enrichment, on 15,480 individuals from the Alzheimer's Disease Sequencing Project (ADSP), a large, comprehensively phenotyped, multi-ancestry AD dataset of 40+ global cohorts. We then developed interpretable linear and nonlinear IRMs to predict AD risk using elastic-net logistic regression and random forest (RF) classifiers to integrate imputed gene and protein expression with clinical covariates and 5 principal components (PCs) accounting for genetic ancestry. IRM performance was evaluated against baseline covariate models and 2 PRS derived from the International Genomics of Alzheimer's Project (IGAP; n=54,162) and iterative subsets of ADSP (n=12,384) under the area under the receiver operating characteristic (AUROC), precision-recall curve (AUPRC), R², balanced accuracy, and F1-score. We identified 104 genomic, 319 transcriptomic, and 17 proteomic significant associations, including 114 not previously reported for AD. Gene-set enrichment revealed novel associations were significantly enriched

in interleukin signaling (IL1RL2, IL7R), cholesterol metabolism (CBLC, STARD5), and other AD-relevant pathways, reinforcing experimental evidence for lipid metabolism and immune system dysregulation in AD. All multi-omics IRMs significantly outperformed PRS (AUROC=0.608, AUPRC=0.491). Transcriptome + clinical covariates RF models achieved the best predictive performance (AUROC=0.703, AUPRC=0.622). Feature importance analysis of the best models revealed tissue-specific variation in gene expression contributions, emphasizing the value of context-aware modeling in risk prediction. We present a novel, scalable multi-modal framework for AD risk prediction that integrates genetic architecture with functional biology, offering potential therapeutic targets and a model adaptable to other complex diseases.

Integrating multi-omics and multi-context QTL data with GWAS reveals the genetic architecture of complex traits and improves the discovery of risk genes

Subsession Time: Thursday, October 16 at 9:00am – 9:15am

Authors: Sheng Qian (University of Chicago, Human Genetics), Xiaotong Sun (University of Chicago, Human Genetics), Kaixuan Luo (University of Chicago, Human Genetics), Wesley Crouse (University of Chicago, Human Genetics), Jing Gu (University of Chicago, Human Genetics), Lifan Liang (University of Chicago, Human Genetics), Siming Zhao (Dartmouth College, Biomedical Data Science), Matthew Stephens (University of Chicago, Human Genetics), Xin He (University of Chicago, Human Genetics)

Abstract: A common strategy to interpret GWAS findings is the use of expression QTLs. However, recent studies show limited colocalization between eQTLs and GWAS loci, and eQTLs explain only a small fraction of complex trait heritability. A natural strategy is to incorporate molecular QTLs beyond expression, across diverse tissues and cell types. Yet, integrating these heterogeneous QTLs presents analytical challenges, as molecular traits often share causal variants or are in high LD, complicating the attribution of GWAS signals to specific molecular traits or contexts.

To address this, we leveraged our recently developed method, causal-TWAS (cTWAS, Zhao et al. Nature Genetics, 2024). cTWAS jointly models variants and gene expression in a GWAS region, leading to better control of false discoveries than popular methods. Building on cTWAS, we developed multi-group cTWAS (M-cTWAS), a scalable framework for integrating multi-omics QTLs across contexts to fine-map causal molecular traits. M-cTWAS is able to determine the molecular mechanisms and contexts through which genetic variations act on the phenotype, and at the same time combines information of the same genes across modalities and contexts to increase power.

We applied M-cTWAS to GWAS of common traits, using expression, splicing, RNA stability QTLs from multiple tissues. While single-tissue eQTLs explain ~3–18% of trait heritability, cross-tissue analysis increases this to 10-50%. Heritability contributions from splicing and RNA stability are comparable to expression. M-cTWAS found many genes that were missed by single-tissue eQTL analysis. Compared to colocalization analysis, which implicated ≥ 2 genes in half of loci, M-cTWAS prioritized a single gene in most loci. Many of these findings were driven by multiple molecular traits in different genes with QTLs being in high LD, highlighting the challenge of the colocalization analysis for complex QTL datasets. We further applied M-cTWAS to postmortem brain datasets for psychiatric traits, incorporating eQTLs and

epigenetic QTLs (chromatin accessibility and DNA methylation). We found that epigenetic QTLs explain substantial heritability not captured by eQTLs, suggesting regulatory effects during early development that would be missed by eQTLs from adult samples.

In summary, our results highlighted the promise of using molecular QTLs of multiple modalities across diverse contexts to advance our understanding of genetics of complex traits. Our method, M-cTWAS, provides a powerful framework for this type of analysis.

Integrating multi-omics data to identify lung-tissue-specific susceptibility proteins for lung cancer risk

Subsession Time: Thursday, October 16 at 9:15am – 9:30am

Authors: Shuai Xu* (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA), Jiajun Shi* (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA), Yaohua Yang (Department of Genome Sciences, UVA Comprehensive Cancer Center, School of Medicine, University of Virginia, Charlottesville, VA, USA), Wanqing Wen (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA), Ran Tao (Department of Biostatistics, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA), Xiao-Ou Shu (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA), Xingyi Guo (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA), Bingshan Li (Department of Molecular Physiology and Biophysics, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA), Wei Zheng (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA), Jirong Long# (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA), Qiuyin Cai# (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA)

Abstract: Background: Lung cancer is one of the most common and lethal cancer types worldwide. Genome-wide association studies (GWAS) have identified ~100 loci for lung cancer risk. However, causal genes for most of these loci are still unknown.

Methods: We conducted the first lung tissue-based proteome-wide association study (PWAS). We profiled proteomics in adjacent normal tissues (N=200) and then built genetic prediction models. The models were then applied to GWAS data of 55,174 cases and 1,294,174 controls to identify proteins associated with lung cancer. For promising proteins, we conducted colocalization analyses and compared

the protein levels in paired tumor and adjacent normal tissues (N=220). We also examined the mRNA expression levels with lung cancer risk by integrating RNA-seq data from adjacent normal tissues (N=314) and lung cancer GWAS data. We also performed cell-type specificity enrichment analyses by using scRNAseq data in 47,453 lung epithelial cells from 16 patients diagnosed with primary lung adenocarcinoma.

Results: We identified 28 proteins significantly associated with overall risk of lung cancer and five additional proteins with risk of lung cancer specific subtype at an FDR<5%, including CLDN18 associated with decreased risk of lung cancer ($P=2.2 \times 10^{-4}$). Among these 33 proteins, 18 proteins were differentially expressed between tumor and adjacent normal tissues with the same direction of PWAS at $P<0.05$. Notably, CLDN18 was downregulated in both lung squamous cell carcinoma ($P=2.2 \times 10^{-35}$) and adenocarcinoma ($P=5.0 \times 10^{-32}$) tumor tissues. Genetically predicted mRNA levels of 15 corresponding genes were associated with lung cancer risk at $P<0.05$ with the same direction of PWAS. For example, mRNA level of CLDN18 was significantly associated with decreased lung cancer risk ($P=2.38 \times 10^{-6}$). Interestingly, eight of the 33 proteins showed some evidence of colocalization with lung cancer at $PP.H4>0.50$. Especially, CLDN18 showed high evidence of colocalization ($PP.H4=0.96$). Further scRNAseq analyses showed expressions of 7 of the 33 genes were enriched in specific epithelial cell types. The CLDN18 gene was highly expressed in alveolar type 1 cell ($\log_2$ fold change=2.46; FDR= 7.58×10^{-21}) compared with other epithelial cell types.

Conclusion: By integrating proteomics and other omics data, we found that CLDN18 may be causally related to lung cancer risk. Our study offers novel insights into the biology of lung carcinogenesis.

Integrated Mendelian randomisation and multi-omic validation identify causal serum proteins in amyotrophic lateral sclerosis

Subsession Time: Thursday, October 16 at 9:30am – 9:45am

Authors: Elham Alhathli (Sheffield Institute for Translational Neuroscience, University of Sheffield, Faculty of Nursing, Taif University), Andrew Strange (Sheffield Institute for Translational Neuroscience, University of Sheffield), Alexander G Thompson (Nuffield

Abstract: Background: Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disorder characterised by motor-neuron loss, which currently lacks a diagnostic biomarker. Serum proteome changes have been observed in ALS, but a well-powered systematic evaluati

Multitomic analysis of circadian and seasonal biomarkers

Subsession Time: Thursday, October 16 at 9:45am – 10:00am

Authors: Lea Urpa (Institute for Molecular Medicine Finland), Nasa Sinnott-Armstrong (Fred Hutchinson Cancer Center, Institute for Molecular Medicine Finland), Hanna M Ollila (Institute for Molecular Medicine Finland, Broad Institute of Harvard and MIT, Massachuse

Abstract: Biological processes exhibit dynamic regulation over time, particularly across circadian (daily) and seasonal cycles. However, the genetic and temporal architecture of such oscillatory processes in human clinical biomarkers remains poorly characterized. L

Session 44: Beyond the Bean: Genetic Studies of Kidney Disease

Location: Room 206AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 10:45am – 11:45am

Novel genetic insights into kidney function in countries with a high prevalence of chronic kidney disease: A meta-analysis of GWAS from 596,624 East Asians in Japan, South Korea, and Taiwan

Subsession Time: Thursday, October 16 at 10:45am – 11:00am

Authors: Chin-Chi Kuo(Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan; Department of Biomedical Informatics, College of Medicine, China Medical University, Taichung, Taiwan; Division of Nephrology, Department of Internal Medicine, China Medical University Hospital, China Medical University, Taichung, Taiwan; College of Medicine, China Medical University, Taichung, Taiwan), Che-Yi Chou(Division of Nephrology, Asia University Hospital, Taichung, Taiwan; Department of Post-baccalaureate Veterinary Medicine, Asia University, Taichung, Taiwan), Hung-Lin Chen (Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan; Department of Biomedical Informatics, College of Medicine, China Medical University, Taichung, Taiwan), Myunghee Hong(Department of Biomedical Informatics, School of Medicine, CHA University, Seongnam, Republic of Korea; Institute for Biomedical Informatics, School of Medicine, CHA University, Seongnam, Republic of Korea), Hsiu-Yin Chiang(Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan; Department of Biomedical Informatics, College of Medicine, China Medical University, Taichung, Taiwan), Yu-Ting Lin(Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan; Department of Biomedical Informatics, College of Medicine, China Medical University, Taichung, Taiwan), Che-Chen Lin(Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan; Department of Biomedical Informatics, College of Medicine, China Medical University, Taichung, Taiwan), Min-Yu Chiang (Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan), Ya-Chi Lin(Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan; Department of Biomedical Informatics, College of Medicine, China Medical University, Taichung, Taiwan), Michael Anekson Widjaya (Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan), David Ray Cheng(Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan; Department of Biomedical Informatics, College of Medicine, China Medical University, Taichung, Taiwan; Division of Nephrology, Department of Internal Medicine, China Medical University Hospital, China Medical University, Taichung, Taiwan; College of Medicine, China Medical University, Taichung, Taiwan; Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, USA) , Hung-Chieh Yeh(Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan; Department of Biomedical Informatics, College of Medicine, China Medical University, Taichung, Taiwan; Division of Nephrology, Department of Internal Medicine, China Medical University Hospital,

China Medical University, Taichung, Taiwan; College of Medicine, China Medical University, Taichung, Taiwan), I-Wen Ting(Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan; Department of Biomedical Informatics, College of Medicine, China Medical University, Taichung, Taiwan; Division of Nephrology, Department of Internal Medicine, China Medical University Hospital, China Medical University, Taichung, Taiwan; College of Medicine, China Medical University, Taichung, Taiwan), Amrita Chattopadhyay(Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan; Institute of Health Data Analytics and Statistics, Department of Public Health, College of Public Health, National Taiwan University, Taipei, Taiwan), Tzu-Pin Lu(Institute of Epidemiology and Preventive Medicine, Department of Public Health, College of Public Health, National Taiwan University, Taipei, Taiwan), Sung Kweon Cho(Department of Pharmacology, Ajou University School of Medicine, Suwon, Republic of Korea), Fuu-Jen Tsai(School of Chinese Medicine, College of Chinese Medicine, China Medical University, Taichung, Taiwan; Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan ; Division of Medical Genetics, China Medical University Children's Hospital, Taichung, Taiwan; Department of Medical Laboratory Science & Biotechnology, Asia University, Taichung, Taiwan)

Abstract: Background: Chronic kidney disease (CKD) poses a major global health challenge worldwide, with East Asia experiencing particularly high prevalence. To address this need, we conducted the largest genome-wide association study (GWAS) meta-analysis of kidney function in East Asian populations, focusing on countries with a high prevalence of CKD, including Japan, Taiwan, and Korea.

Methods: We conducted a meta-analysis of eGFR GWAS summary statistics from Biobank Japan (n=154,633), Korean Genome and Epidemiology Study (n=72,291), Taiwan Biobank (n=124,858), and the CMUH-iHi discovery cohort (n=244,842), reaching a combined East Asian cohort with 596,624 samples. To prioritize causal variants, causal genes, and cell-types associated with eGFR, we performed functional annotation, fine-mapping, colocalization, Mendelian randomization (MR), gene set and cell-type enrichment. We also established a polygenic risk score (PRS) for eGFR to predict CKD incidence.

Results: The GWAS meta-analysis revealed 461 lead SNPs across 365 genomic loci, including 22 novel genomic loci, associated with eGFR. The most significant genetic variant (rs10917155, $p < 2.32 \times 10^{-17}$) was located near WNT4, which is a sex-determinant gene for women. Cell-type enrichment analysis showed that the top three tissues associated with eGFR were the kidney medulla, kidney cortex, and thyroid. Gene set enrichment identified the FTO obesity variant mechanism and RNA regulation as key pathways. Colocalization and MR linked eGFR to 341 tubule- and 351 glomerulus-expressed genes. The five most consistent genes across the different analyses were WDR72, DGKH, FAM131C, FGF5, and ZNF468. The PRS in predicting CKD incidence achieved an area under the receiver operating characteristic curve (AUROC) of 0.825 (95% CI, 0.816–0.834). In a time-to-event analysis, individuals in the highest PRS group (PRS > 2SD) reached 10% cumulative CKD incidence rate 7.8 years earlier (57.1 years old) compared to those in the lowest PRS group (PRS < -2SD) (64.9 years old).

Conclusion: Our study identified 22 novel genomic risk loci for the kidney function index, eGFR, in countries with a high prevalence of CKD. Our identification of WNT4 provides a possible sex-specific genetic mechanism underlying the observed sex differences in CKD risk, where more women than men

have CKD, and more men than women reach end-stage kidney failure. Although further biological validation is required, our results shed light on the etiology of kidney disease in Asian populations.

Ancestry-specific genetic loci for kidney function identified in the largest African-ancestry meta-analysis

Subsession Time: Thursday, October 16 at 11:00am – 11:15am

Authors: Oyesola O. Ojewunmi (Queen Mary University of London, UK), Chisom Soremekun (School of Biomedical Sciences, Makerere University College of Health Sciences, Kampala, Uganda), Mine Koprulu (Queen Mary University of London, UK), Abram Kamiza (Medical Research Council, Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine), Tinashe Chikowore (Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA), Christopher Kintu (Queen Mary University of London, UK), Jean-tristan Brandenburg (Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa), Dorothea Nitsch (London School of Hygiene and Tropical Medicine, UK), Cristian Pattaro (Institute for Biomedicine, Eurac Research, Bolzano, Italy), Moffat Nyirenda (Medical Research Council, Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine), Mia Crampin (Malawi Epidemiology and Intervention Research Unit, Lilongwe, Malawi), Eleftheria Zeggini (Institute of Translational Genomics, Helmholtz Munich, 85764 Neuherberg, Germany), Charles Rotimi (Center for Research on Genomics and Global Health, National Institute of Health, Bethesda, MD), Nora Franceschini (The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States), Jennifer Asimit (MRC Biostatistics Unit, University of Cambridge, Cambridge, UK), Adeyemo Adebawale (Center for Research on Genomics and Global Health, National Institute of Health, Bethesda, MD), Andrew P Morris (Centre for Genetics and Genomics Versus Arthritis, University of Manchester, Manchester, UK), Claudia Langenberg (Queen Mary University of London, UK), Michele Ramsay (Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa), June Fabian (School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa), Segun Fatumo (Queen Mary University of London, UK and Medical Research Council, Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine)

Abstract: The estimated glomerular filtration rate (eGFR) derived from creatinine measurements is an essential biomarker for assessing kidney function decline. Hundreds of genetic loci - mostly identified in individuals of European ancestry - have been associated with eGFR in genome-wide association study (GWAS) meta-analyses. However, populations of African ancestry, who are disproportionately affected by chronic kidney disease (CKD), remain largely underrepresented in genetic research. This disparity limits our understanding of genetic risk factors relevant to kidney function in African ancestry populations.

In this study, we conducted the largest GWAS meta-analysis of eGFR based on creatinine measurements in over 200,000 individuals of African ancestry to discover previously unidentified loci and prioritise variants that may be relevant to kidney function in health and disease.

We performed a GWAS meta-analysis of eGFR in 213,763 individuals of African ancestry, comprising 28,209 continental Africans and 185,554 admixed African individuals. Metal analysis was implemented in METAL, while functional mapping and tissue expression analysis were performed in Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA).

We identified 81 loci attaining genome-wide significance ($p < 5 \times 10^{-8}$), with the strongest signals mapping to GATM (rs1706801, $p = 6.9 \times 10^{-92}$), APOL1 (rs73885319, $p = 4.0 \times 10^{-48}$), and HBB (rs334, $p = 5.0 \times 10^{-46}$) loci. Of note, 25 of these loci had not been previously identified, and 10 of the lead single-nucleotide variants (SNVs) [RP4-640E24.1, RP11-157N3.1, AC064847.4, ANKAR, LOC105374528, RP11-213H15.1, RP11-445O3.1, SNTB1, F2, EEF2] were monomorphic or rare in non-African populations. F2 (coagulation factor II) gene, for instance, has been linked to renal insufficiency, end-stage renal disease, renal cell carcinoma, and metabolic syndrome. Enrichment analyses in FUMA showed that eGFR-associated variants were significantly overrepresented in genes expressed in the kidney cortex and medulla (Bonferroni correction at $p < 0.00094$). The prioritised genes were linked to renal function, chronic kidney disease, fasting glucose levels, body fat distribution, body mass index, and diastolic blood pressure.

These findings highlight the importance of including African-ancestry populations in genetic studies to identify novel, ancestry-specific loci associated with kidney function, providing new insights into kidney biology and disease risk.

Bridging Genomic Inequities: Chronic Kidney Disease Susceptibility and Pharmacogenomic Variation in Indigenous Australians

Subsession Time: Thursday, October 16 at 11:15am – 11:30am

Authors: Shivashankar Hiriyur Nagaraj (Centre for Genomics and Personalised Health and School of Biomedical Sciences, Queensland University of Technology, Brisbane, QLD, Australia), Simon P Lee (Centre for Genomics and Personalised Health and School of Biomedical Sciences, Queensland University of Technology, Brisbane, QLD, Australia), Vignesh Arunachalam (Centre for Genomics and Personalised Health and School of Biomedical Sciences, Queensland University of Technology, Brisbane, QLD, Australia), Wendy Hoy (Centre of Chronic Disease, Faculty of Medicine, University of Queensland, Brisbane, QLD, Australia), Judith Savage (Royal Melbourne Hospital, The University of Melbourne, Melbourne, VIC, Australia), Brendan J. McMorran (National Centre for Indigenous Genomics, John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia), Andrew J. Mallett (Department of Renal Medicine, Townsville University Hospital, Douglas, QLD, Australia), Pirrawayingi Puruntatameri (Tiwi Genetics Advisory Group, Tiwi Islands)

Abstract: Owing to the injustices of colonization, Indigenous peoples bear a disproportionately high burden of chronic disease but are largely absent from global genomic databases. The health inequities experienced by Indigenous populations are even more dire for Indigenous Australians, who are grossly underrepresented in global databases and genomic studies, have shorter life expectancies than other Indigenous and non-Indigenous populations, exhibit extremely high rates of chronic disease, and experience high rates of polypharmacy (i.e., >5 medications). For example, Indigenous Australians face 6-

and 8-fold higher rates of end-stage renal failure and non-dialysis CKD hospitalizations, respectively, than non-Indigenous Australians. Given the high CKD burden in Tiwi, we conducted the largest sequencing study in this group (n=474), identifying 132 kidney disease-associated variants, of which 16% of SNVs and 46% of SVs were unique to Tiwi. Nearly all individuals carried at least one variant; 44% harbored 12 known pathogenic CKD variants. We also discovered Tiwi-specific variants in kidney-related genes and identified CNVs and SVs in CKD-related genes linked to significant changes in eGFR and ACR, highlighting a genetic predisposition to kidney diseases in the Tiwi population. Although variable drug responses are known to be population specific, most research on pharmacogenes, medication guidelines, and PGx panels is based on Caucasian genetics. Notably, a characterization of the PGx landscape in Tiwi—a genetically distinct founder population of 2,348 with high chronic disease burden—found that all individuals in the cohort (n=500) had at least one actionable genotype, and 77% carried at least three clinically actionable genotypes across 19 pharmacogenes. Furthermore, 41% of the Tiwi cohort were predicted to exhibit impaired CYP2D6 metabolism (vs. 1–2% of Caucasians), and over 50% were predicted to have impaired metabolism of enzymes involved in processing analgesics, statins, anticoagulants, antiretrovirals, antidepressants, and antipsychotics. Future work will explore DNA methylation to elucidate unexplained variability in drug response, with the ultimate goal of reducing health disparities and improving access to personalized medicine for Indigenous Australians.

Uncovering the inconsistencies and potential risks of AI: inaccuracies in nephrogenetic counseling provided by five LLMs

Subsession Time: Thursday, October 16 at 11:30am – 11:45am

Authors: Sarath Babu Krishna Murthy, Dana Liuzzi-Zhang, Kiara Demorla-Ozuna, Sandy Yang, Mariam Elmais, Paavni Suyal, Nishant Shishu, Kelsie Bogyo, Natalie Vina, Hila Milo Rasouly

Abstract: Background: Advances in artificial intelligence, particularly large language models (LLMs), allow the general population to humanize their interactions with LLMs. The high accessibility of LLMs encourages individuals to ask them for medical advice.

Methods: We created six representative scenarios of individuals with various kidney diseases, personal concerns, and race/ethnicity. Using an identical set of questions for each scenario, we simulated a patient encounter with a genetic counselor. We analyzed the responses with five LLMs —Claude 3.5 Sonnet, ChatGPT 4o, Gemini 1.5 Flash, Mistral 7B, and Llama 3.1 70B. The LLMs' responses were analyzed qualitatively for accuracy and comprehensiveness by genetic counselors, quantitatively through response length and readability (Flesch-Kincaid Grade Level), and semi-quantitatively using standardized rubrics to assess content coverage.

Results: No LLM achieved complete accuracy and comprehensive coverage of all essential genetic counseling topics, and only Gemini started its responses by referring the user to a medical professional for advice. Response lengths and readability levels ranged between and within cases and each LLM from 88 to 939 words and from grades 8 to 16, respectively. Consideration of the scenario's race/ethnicity was minimal, whereas genetic testing benefits and clinical and familial information were frequently stated. All LLMs offered generalized educational information about genetic testing and disease in five scenarios,

but none did so for the case of an individual with a heterozygous COL4A4 variant identified through carrier screening. Mistral, Llama, and ChatGPT discussed the benefits and psychological risks of genetic testing for all scenarios, while privacy and confidentiality risks were mostly absent across LLMs. ChatGPT and Claude each discussed support groups and other advocacy resources in 5 scenarios, but only Gemini provided direct resources in 2 scenarios. All LLMs recommended familial testing (in 4-6 scenarios), and most, except for Gemini, asked follow-up questions.

Conclusion: Though prompted to perform as genetic counselors, the LLMs did not ask for additional medical or family history, nor considered alternative genetic etiologies. All provided generalized summaries instead of personalized responses. This study highlights the potential risks of the utilization of LLMs by individuals with kidney diseases for nephrogenetic counseling.

Session 45: Dissecting the Biology of Neurodegenerative Disorders

Location: Ballroom West/Level 3, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 10:45am – 11:45am

Localizing MS GWAS SNPs to pathogenic cell types reveals convergent mechanisms of microglia dysfunction

Subsession Time: Thursday, October 16 at 10:45am – 11:00am

Authors: Xochitl Luna (Whitehead Institute for Biomedical Research, MIT Department of Brain and Cognitive Sciences, co-first author), Michael D. Gallagher (Whitehead Institute for Biomedical Research, co-first author), Wenjuan Du (Whitehead Institute for Biomedical Research, MIT Department of Biology), Kathryn E.A. Hazel (Whitehead Institute for Biomedical Research), Moritz List (Whitehead Institute for Biomedical Research), Orian Stapleton (Whitehead Institute for Biomedical Research), Zeynep Aydin (Whitehead Institute for Biomedical Research), Yiran Cheng (Whitehead Institute for Biomedical Research), Keya Viswanathan (Whitehead Institute for Biomedical Research), Bingbing Yuan (Whitehead Institute for Biomedical Research), Heather R. Keys (Whitehead Institute for Biomedical Research), George W. Bell (Whitehead Institute for Biomedical Research), Richard A. Young (Whitehead Institute for Biomedical Research, MIT Department of Biology), Rudolf Jaenisch (Whitehead Institute for Biomedical Research, MIT Department of Biology, co-corresponding author), Olivia Corradin (Whitehead Institute for Biomedical Research, MIT Department of Biology, co-corresponding author)

Abstract: The pathogenesis of complex traits like multiple sclerosis (MS) involves disruption of many cell types and systems. However, functional studies that aim to characterize the impact of disease-associated SNPs often investigate all SNPs in the same cellular context. MS is both an autoimmune and neurodegenerative disease, in which immune cells and glial cells have been implicated. Here, we integrate a computational approach to predict the pathogenic cell type of individual risk loci with functional CRISPR perturbation studies in iPSC-derived microglia cells (iMGL). This strategy enabled us to focus on the subset of MS GWAS loci that disrupt gene regulation in microglia, and revealed highly convergent biological mechanisms of GWAS loci in these cells.

We previously developed a strategy to pinpoint the cell types impacted by specific disease-linked alleles. This strategy identifies SNPs (termed “outside variants”) that are physically linked to the same gene target as GWAS SNPs and evaluates their impact on disease risk. Using H3K27ac, we identified cell types with a high concordance between enhancer activity and outside variants that alter clinical risk. Applying this method to MS risk alleles, we identified many loci that primarily influence lymphocytes, but also uncovered risk alleles predicted to influence CNS cells, including 15 loci in microglia.

We performed Perturb-seq, pooled CRISPR inhibition followed by single cell RNA-seq, to investigate the impact of 9 of these MS risk loci in iMGL. We identified the target genes and downstream differentially

expressed genes (DEGs) for 5 GWAS loci. Importantly, these DEGs were substantially enriched for MS heritability. DEGs from all 5 target genes show significant overlap, converging on downregulation of cytokinesis and phagocytosis, and upregulation of mitochondrial metabolism. We then compared DEGs to gene expression patterns in MS patient tissue and observed striking similarities, demonstrating that genes dysregulated as the result of GWAS loci perturbation mirrored effects observed in microglia in MS lesions. Microglia in MS gray matter lesions share upregulated DEGs with our dataset that suggest increased mitochondrial metabolic activity, whereas white matter lesions share upregulated DEGs related to immunological activation. Collectively, these results demonstrate how cell type-aware functional studies can reveal new insights into complex trait pathogenesis and elucidate convergent biological mechanisms.

Age at onset GWAS of Parkinson's Disease in the FinnGen study suggests links with genetic risk for earlier age at onset and known prodromal symptoms for Parkinson's

Subsession Time: Thursday, October 16 at 11:00am – 11:15am

Authors: Marika A. Kaakinen (University of Helsinki, Imperial College London), Elisa Lahtela (University of Helsinki), Fedik Rahimov (AbbVie), Rui Jian Chu (University of Turku), FinnGen, Stephanie Loomis (Bristol Myers Squibb), Benjamin Sun (Bristol Myers Squibb), Coro Paisán-Ruiz (Biogen), Anna-Maija Tolppanen (University of Eastern Finland), Filip Scheperjans (University of Helsinki, Helsinki University Central Hospital), Valtteri Kaasinen (University of Turku, Turku University Hospital), Katherine Klinger (Sanofi), Britney Milkovich (AbbVie), Mary Pat Reeve (University of Helsinki, Broad Institute of MIT and Harvard)

Abstract: Genome-wide association studies (GWAS) of Parkinson's Disease (PD) have identified over 130 disease susceptibility loci, highlighting its complex genetic architecture. However, age at onset (AAO) GWAS for PD have provided strong evidence for only two known PD signals at TMEM175 and SNCA, and borderline evidence for APOE. We aimed to shed light to the genetic determinants of PD AAO in 4,381 PD cases in the large Finnish population-based FinnGen study. Given the complex nature of real-world register-based data, we first carefully reviewed the diagnostic evidence for each patient by summarizing and evaluating their data from specialized healthcare, primary care, dispensed prescriptions and medication reimbursements, and next, adjusted the AAO to either the earliest date of diagnosis, dispensing of dopaminergic medication when the use was considered continuous (at least three dispenses within 1.5 years) or medication reimbursement. The AAO varied from 25 to 94 years. We identified three genome-wide significant signals at/near NRXN1 (rs540355435, $\beta=-15.2$, $P=3.0 \times 10^{-8}$, $MAF=0.0018$), TFEC (rs143651829, $\beta=-3.6$, $P=2.1 \times 10^{-8}$, $MAF=0.028$), and ZFYVE26 (rs1286889387, $\beta=-24.4$, $P=4.7 \times 10^{-8}$, $MAF=0.0006$), all predisposing to earlier age at onset. In addition, we replicated at nominal significance the previously reported signals at SNCA ($P=0.001$) and APOE ($P=6.6 \times 10^{-6}$) but not at TMEM175 ($P=0.49$). Variation within 500 kb of the identified three signals has previously been associated with Alzheimer's Disease and its AAO (NRXN1), depression (NRXN1), insomnia (NRXN1), chronotype (TFEC), and spastic paraplegia, also in combination with juvenile-onset levodopa-responsive parkinsonism (ZFYVE26), among other traits. Sensitivity analyses removing PD patients with severe

psychiatric comorbidities (N=229) resulted in slight loss in power but did not alter the main findings ($P < 1.3 \times 10^{-7}$ at all three loci) nor did removing PD patients with comorbid Alzheimer's Disease (N=604) ($P < 2.6 \times 10^{-7}$). In summary, our study suggests the genetic loci influencing AAO in PD are not fully shared with those affecting the risk of PD. Importantly, the identified loci have prior associations with depression and sleep problems, known prodromal features of PD, supporting the hypothesis that these symptoms may not only precede but also reflect genetic contributions to the timing of PD onset.

Large-scale transcriptomic profiling across brain regions in Progressive Supranuclear Palsy

Subsession Time: Thursday, October 16 at 11:15am – 11:30am

Authors: Jennifer Grundman (UCLA, Department of Human Genetics), Tim Chang (UCLA, Program in Neurogenetics), Le Qi (UCLA, Program in Neurogenetics), Ting Zhang (UCLA, Department of Pathology and Laboratory Medicine), Kun Gao (UCLA, Program in Neurogenetics), Riki Kawaguchi (UCLA, Program in Neurogenetics), Lawrence Chen (UCLA, Program in Neurogenetics), Jing Ou (UCLA, Program in Neurogenetics), Ryan Wong (UCLA, College of Letters and Science), Jade Nguyen (UCLA, Program in Neurogenetics), Michael DeTure (Mayo Clinic), Bailey Rawlinson (Mayo Clinic), Clifton L. Dalgard (Uniformed Services University of the Health Sciences, American Genome Center), Huw Morris (Mayo Clinic), Dennis Dickson (Mayo Clinic), Daniel Geschwind (UCLA, Institute of Precision Health, Program in Neurogenetics, Department of Human Genetics)

Abstract: Tau pathology underlies the majority of dementia cases, but the underlying mechanisms of tauopathies are poorly understood, hindering development of effective therapies. Progressive Supranuclear Palsy (PSP) is a prototypical tauopathy characterized by subcortical-to-cortical progression of tau inclusions in neurons and glia. We conducted RNA-sequencing of 200 cases and 52 controls across two brain regions with different vulnerability to PSP — caudate nucleus (CN), and prefrontal cortex (PFC) — to elucidate PSP's molecular underpinnings. In this unprecedented cohort, we discovered widespread differential gene expression, isoform expression, and alternative splicing involving > 2000 genes, which are notably concentrated in CN. With snRNA-seq collected from 25 cases and controls across both regions, we found that differential isoform expression and alternative splicing were enriched in vulnerable cell populations. Using cell-type pathology staging data, we showed that transcriptomic dysregulation in CN correlates with glial pathology. Weighted gene co-expression network analysis showed 32 regional, disease-associated gene co-expression modules with cell type specific pathological changes. We examined genetic risk for PSP across disease-associated modules, discovering that neuronal and oligodendrocytic modules were specifically enriched in genetic risk, consistent with their causal involvement and confirming previous analyses in different data sets. We conducted analysis of hub genes within these modules and found that they are involved in protein processing and degradation pathways, allowing us to nominate several risk genes situated in these modules as therapeutic targets. These findings advance our understanding of the molecular mechanisms driving PSP and elevate promising candidates for therapeutic targeting.

Genomic and epigenomic insights into purkinje and granule neurons in Alzheimer's disease and related dementia using single-nucleus multiome analysis

Subsession Time: Thursday, October 16 at 11:30am – 11:45am

Authors: Xiaoyu Yang (Cleveland Clinic Genome Center, Cleveland Clinic, Cleveland, OH 44195), Yayan Feng (Cleveland Clinic Genome Center, Cleveland Clinic, Cleveland, OH 44195), Margaret E Flanagan (Biggs Institute, University of Texas Health Science Center San Antonio, San Antonio, Texas), Xin Chen (Cleveland Clinic Genome Center, Cleveland Clinic, Cleveland, OH 44195), Borna Bonakdarpour (Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Ken and Ruth Davee Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, IL), Andrew A. Pieper (Department of Neurosciences, Case Western Reserve University, Cleveland, OH 44106), Jeffrey Cummings (Chambers-Grundy Center for Transformative Neuroscience, Department of Brain Health, School of Integrated Health Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA), Feixiong Cheng (Founding Director, Cleveland Clinic Genome Center, Cleveland Clinic, Cleveland, OH 44195),

Abstract: Although the human cerebellum is known to be neuropathologically impaired in Alzheimer's disease (AD) and AD-related dementias (ADRD), the cell type-specific transcriptional and epigenomic changes that contribute to this pathology are not well understood. Here, we report single-nucleus multiome (snRNA-seq and snATAC-seq) analysis of 103,861 nuclei isolated from both cerebellum and frontal cortex of AD/ADRD patients and normal controls. Using peak-to-gene linkage analysis, we identified 431,834 significant linkages between gene expression and cell subtype-specific chromatin accessibility regions enriched for candidate cis-regulatory elements (cCREs). These cCREs were associated with AD/ADRD-specific transcriptomic changes and disease-related gene regulatory networks, especially for RAR Related Orphan Receptor A (RORA) and E74 Like ETS Transcription Factor 1 (ELF1) in cerebellar Purkinje cells and granule cells, respectively. Trajectory analysis of granule cell populations further identified disease-relevant transcription factors, such as RORA, and their regulatory targets. Finally, we pinpointed two likely causal genes, Seizure Related 6 Homolog Like 2 (SEZ6L2) in Purkinje cells and KAT8 Regulatory NSL Complex Subunit 1 (KANSL1) in granule cells, through integrative analysis of cCREs derived from snATAC-seq, genome-wide AD/ADRD loci, and three-dimensional (3D) genome data. Via CRISPRi experiments, we found that perturbation of rs4788201 and rs62056801 significantly inhibited the expression of their target genes, SEZ6L2 and KANSL1, in human iPSC-derived neurons. This cell subtype-specific regulatory landscape in the human cerebellum identified here offers novel genomic and epigenomic insights into the neuropathology and pathobiology of AD/ADRD and other neurological disorders if broadly applied.

Session 46: Functional Dissection of Variants at Scale

Location: Room 205ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 10:45am – 11:45am

Multi-modal functional mapping of non-coding sequences regulating fetal hemoglobin

Subsession Time: Thursday, October 16 at 10:45am – 11:00am

Authors: Basheer F. Becerra (Broad Institute, Massachusetts General Hospital, Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Harvard Medical School), Ting Wu (Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Broad Institute, Harvard Medical School), Zain M. Patel (Broad Institute, Massachusetts General Hospital, Harvard Medical School), Eric Che (Broad Institute, Massachusetts General Hospital, Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Harvard Medical School), Marlina Starrs (Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Broad Institute, Harvard Medical School), Gue-Ho Hwang (Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Broad Institute, Harvard Medical School), Sébastien Levesque (Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Broad Institute, Harvard Medical School), Sabine Studer (Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Broad Institute, Harvard Medical School), Steve Coyne (Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Broad Institute, Harvard Medical School), Jing Zeng (Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Broad Institute, Harvard Medical School), Lucas F. DaSilva (Broad Institute, Massachusetts General Hospital, Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Harvard Medical School), Jonathan D. Rosen (University of North Carolina at Chapel Hill), Jessica L. Bell (University of North Carolina at Chapel Hill), A. Trent Ledford (University of North Carolina at Chapel Hill), Sean Jones (University of North Carolina at Chapel Hill), Justin Delano (Broad Institute, Massachusetts General Hospital, Harvard Medical School), Michael I. Love (University of North Carolina at Chapel Hill), Hyejung Won (University of North Carolina at Chapel Hill), Davide Seruggia (St. Anna Children's Cancer Research Institute, CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences), Luca Pinello (Broad Institute, Massachusetts General Hospital, Harvard Medical School), Daniel E. Bauer (Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Broad Institute, Harvard Medical School)

Abstract: Transcriptional regulation by non-coding elements and their associated DNA-binding transcription factors (TFs) control the developmental transition from fetal to adult hemoglobin – commonly referred to as “hemoglobin switching”. The reactivation of fetal hemoglobin (HbF) has emerged as a promising therapeutic avenue for β -hemoglobinopathies such as sickle cell disease and β -thalassemia. Delineating the cis-regulatory logic of sequences governing HBG1 and HBG2, the genes encoding the γ -globin chains of HbF, is essential for understanding hemoglobin switching and developing

more effective treatments. While population genetics studies have pinpointed single-nucleotide variants (SNVs) linked to HbF induction, this approach is limited to naturally occurring variation, preventing a comprehensive view of causal regulatory elements.

To overcome this limitation, we have performed a systematic comparison of multiple mutagenesis approaches on HbF regulatory sequences in immortalized erythroid progenitors (HUDEP-2) to elucidate the variant-level cis-regulatory logic and corresponding TF binding sites underpinning hemoglobin switching. Specifically, we integrated four complementary approaches: in-situ dense adenine- (ABE8e-SpRY) and dual- (TadDE-SpRY) base-editing mutagenesis, in-situ prime-editing saturation mutagenesis, ex-situ massively parallel reporter assay (MPRA) saturation mutagenesis, and in-silico ChromBPNet saturation mutagenesis. These approaches were performed on the HBG1/2 promoters, BCL11A +55 kb enhancer, and BCL11A +58 kb enhancer given their therapeutic relevance for potent HbF reinduction.

This multi-modal screening strategy identified several novel insights into hemoglobin switching including a BCL11A autoregulatory binding site within its own BCL11A +55 kb enhancer. Additionally, we have identified multiple novel therapeutic editing strategies such as potent TadDE-SpRY base-editing and prime-editing targeting of the BCL11A distal motif in the HBG1/2 promoter, resulting in substantially higher HbF induction compared to gene editing strategies currently pursued in clinical trials.

This comprehensive functional mapping obtained across multiple modalities establishes an example of the strengths, weaknesses, and opportunities to combine in situ, ex situ, and in silico approaches to determine the function of regulatory elements as well as provide an optimized therapeutic roadmap for gene therapies.

A comprehensive MPRA of both trans-ancestral genome-wide and suggestively-associated childhood obesity loci implicates a role for CBLN4 in microglia

Subsession Time: Thursday, October 16 at 11:00am – 11:15am

Authors: Winter S. Bruner (Children's Hospital of Philadelphia), Molly S. Ridler (Children's Hospital of Philadelphia), Amelia C. Tang (Children's Hospital of Philadelphia), Matthew C. Pahl (Children's Hospital of Philadelphia), Alessandra Chesi (University of Pennsylvania), Khanh B. Trang (Children's Hospital of Philadelphia), Struan F.A. Grant (Children's Hospital of Philadelphia, University of Pennsylvania)

Abstract: The rise in pediatric obesity contributes to increased cardiometabolic disease rates. Although our GWAS meta-analyses of childhood obesity for the Early Growth Genetics (EGG) consortium revealed risk loci, such efforts alone cannot pinpoint mechanism. To determine which non-coding variants at these loci drive differential transcription, we utilized massively parallel reporter assay (MPRA).

We selected variants from our childhood obesity genome-wide significant signals, along with any achieving the sub-threshold level of $P < 1 \times 10^{-6}$ in the population-specific datasets or the equivalent in the overall trans-ethnic dataset. Each variant fulfilling these criteria was then subjected to LD expansion ($r^2 > 0.8$). The resulting comprehensive MPRA library consisted of 2,453 variants (4,906 total sequences).

Genomic studies emphasize the brain as central to the pathogenesis of obesity. Similar to other neuropsychiatric traits, glial cells have been increasingly implicated in contributing to the genetic risk of obesity. To explore potential glial involvement in childhood obesity risk, we leveraged our MPRA to determine differential allelic activity in the microglial cell line, HMC3s. Differential regulatory activity was significant ($P_{adj} < 0.05$) for 26 variants with a minimum log fold change of ± 1 . Of these variants, 5 were found to reside in ATAC-seq defined open chromatin in HMC3s and iPSC-derived microglia. Given the largest differential allelic fold change was for the region harboring rs2749796 ($FC = -1.9$), we elected to explore this variant further.

Putative effector gene targets of rs2749796 were nominated by our HMC3 Hi-C data. While rs2749796 is in the topologically associated domain (TAD) harboring CBLN4 and MC3R, the corresponding subTAD is limited to CBLN4. Since the alternative allele decreased expression in the MPRA, CRISPRi was used in HMC3s to repress the immediate non-coding region harboring rs2749796 to determine the impact on the putative effector gene CBLN4. Results indicated downregulation of the CBLN4 gene expression when compared against scrambled and wildtype HMC3 cells. This inactivation is consistent with the decrease expression observed in the MPRA.

Including rs2749796 in the MPRA was due to its high LD with suggestive associated SNPs ($1 \times 10^{-8} > P < 1 \times 10^{-7}$). By incorporating both genome-wide significant and suggestively associated variants, our approach uncovered a previously uncharacterized variant and revealed a novel role for CBLN4 in microglia.

Deciphering causal variant effects in Digestive Disease via Organoid-based allele specific open chromatin and CRISPRi

Subsession Time: Thursday, October 16 at 11:15am – 11:30am

Authors: Luke T. Deary (Department of Molecular and Systems Biology, Department of Biomedical Data Science, Dartmouth Cancer Center), Liyang Yu (Quantitative Biomedical Sciences Program, Geisel School of Medicine), Andrew McCray (Molecular and Cellular Biology Program, Geisel School of Medicine), Sara W. Mayo (Department of Surgery, Geisel School of Medicine), Matthew Z. Wilson (Department of Surgery, Geisel School of Medicine), David T. Breault (Division of Endocrinology, Boston Children's Hospital, Department of Pediatrics, Harvard Medical School), Xiaofeng Wang (Department of Molecular and Systems Biology, Geisel School of Medicine, Dartmouth Cancer Center), Siming Zhao (Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth Cancer Center).

Abstract: Genetics play important roles for many digestive traits and diseases, such as colorectal cancer (CRC) and inflammatory bowel disease (IBD). Genome-wide association studies (GWAS) have identified thousands of associated variants for these traits. However, pinpointing true causal variants and their biological roles remains challenging. Current efforts focusing on expression quantitative trait loci (eQTLs) have explained a substantial fraction of trait-associated loci, however, the majority of the loci remain unresolved, potentially due to the current eQTL list failing to capture context specific variant effect. Here, we leverage allele-specific open chromatin (ASoC) as a novel molQTL readout to identify causal variants in digestive diseases. We harnessed a human colon organoid (hCO) biobank from 24 patients in rural

northern New England, representing all anatomical regions of the large intestine. Under differentiating culture conditions, organoids contain major secretory and absorptive cell types responsible for electrolytic, hormonal, antimicrobial, and regenerative functions essential to colon homeostasis. Using bulk ATAC-seq in differentiated hCOs, we identified over 7000 ASoC variants (FDR < 0.1). 77% of ASoC are not found in the GTEx colon tissue eQTLs. We found ASoCs are highly enriched with CRC and IBD GWAS signals (~ 70 fold compared to baseline). Using ASoCs as functional annotations, we performed annotation-informed fine-mapping and identified ~20% of all loci have ASoC as the likely GWAS causal variant. These variants, located in non-coding regions near genes like CTNNB1, SMAD3, and C1orf106, are implicated in intestinal tissue maintenance. To elucidate their regulatory effects, we employed a high-throughput functional genomics approach, combining CRISPRi with single-cell RNA-seq (Perturb-seq). A MECP2-dCAS9-ZNF10KRAB stable hCO line was transduced with a lentiviral guide RNA library targeting within 100 bp of ASoC loci. This revealed cell-type-specific effects for IBD and CRC variants, linking ASoC loci to differential expression of nearby genes. Integrative analysis uncovered regulatory networks and molecular pathways altered by these variants, including Wnt and TGF- β signaling. Our findings demonstrate that ASoC in hCOs is a powerful tool for resolving causal variants and provide molecular insights into the regulatory mechanisms underlying digestive disease, advancing precision medicine for CRC and IBD.

Multimodal CACNA1C deep mutational scanning to assess the impact of all missense variants on ion channel function

Subsession Time: Thursday, October 16 at 11:30am – 11:45am

Authors: Richard E Dolder III (Vanderbilt University), Sasha Barrera (Vanderbilt University), Marlyssa Lee (Vanderbilt University), Jeremy Smith (Vanderbilt University Medical Center), Jessa Aldridge (Vanderbilt University Medical Center), Sophie Cohen (Vanderbilt University Medical Center), Dan Roden (Vanderbilt University Medical Center), Andrew Glazer (Vanderbilt University Medical Center)

Abstract: CACNA1C encodes the major subunit of the CaV1.2 calcium channel, which mediates calcium influx during cardiac and neuronal action potentials. Variants in this gene are linked to arrhythmias, neurodevelopmental disorders, and birth defects. Identifying individuals with deleterious variants enables clinicians to initiate preventative care, diagnose patients and families, and customize treatment. Unfortunately, most CACNA1C variants cannot be confidently classified, and 89% of missense variants in ClinVar are labeled as Variants of Uncertain Significance (VUS). We report here development of a series of assays that will enable a deep mutational scan (DMS) of thousands of CACNA1C variants in multiplexed experiments. CACNA1C undergoes extensive alternative splicing, so we cloned the two cDNA isoforms that represent major isoforms in cardiomyocytes and neurons. We generated a landing pad Human Embryonic Kidney cell line that expresses the CACNB2 and CACNA2D1 subunits that are required for channel function. To assess calcium influx, we deployed CaMPARI, a calcium indicator that permanently switches from green to red fluorescence at a rate dependent on the intracellular calcium concentration. To validate the approach, cells were transfected with CACNA1C variants and CaMPARI and depolarized to initiate CaV1.2-mediated calcium influx. The assay successfully distinguished cells expressing an allelic series of CACNA1C variants. Compared to wildtype (normalized to 100%), variants

with loss-of-function, partial loss-of-function, and gain-of-function had 8%, 40%, and 134% red fluorescence, respectively. We also observed cellular fitness defects proportional to the extent of channel function. Cells were transfected with a blank control plasmid and either wildtype CaV1.2, two gain-of-function variants, or a loss-of-function variant. After one week, 0-1% of cells were expressing either gain-of-function variant, 11% of cells were expressing the wild-type CaV1.2, and 30% of cells were expressing the loss-of-function variant. This effect will form the basis of a DMS-compatible cellular fitness assay. Additionally, we are implementing a previously established cell-surface epitope tag to assess cell surface abundance. These experiments establish a DMS framework to interrogate multiple functions of all possible CACNA1C missense variants. Completion of this scan will provide functional data for all CACNA1C missense variants helping to solve the VUS problem for this gene.

Session 47: Genetic Findings in African Genetic Ancestry Populations

Location: Room 253ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 10:45am – 11:45am

Proteogenomic analysis of BMI in ~200,000 continental Africans and admixed African identifies 28 loci and enhances cross-ancestry incident obesity prediction using proteomic scores

Subsession Time: Thursday, October 16 at 10:45am – 11:00am

Authors: Yann Ilboudo (Harvard Medical School, Channing Division of Network Medicine), Monica Muti (SAMRC/Wits Developmental Pathways for Health Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Food and Nutrition Security Research Institute), Opeyemi Soremekun (Institute of Translational Genomics, Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Medical Research Council, Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine), Segun Fatumo (Institute of Translational Genomics, Medical Research Council, Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine, Precision Healthcare University Research Institute Queen Mary University of London), Abram Kamiza (University of the Witwatersrand), Palwende R. Boua (Sydney Brenner Institute for Molecular Bioscience (SBIMB)), Tafadzwa Machipisa (University of Pennsylvania), Timothy Olanrewaju (Department of Medicine, University of Ilorin), Lisa Micklesfield (SAMRC/Wits Developmental Pathways for Health Research Unit), Shane A. Norris (SAMRC/Wits Developmental Pathways for Health Research Unit, School of Human Development and Health, University of Southampton), Julia G. Goedecke (Department of Public Health and Clinical Medicine, Umeå University, Biomedical Research and Innovation Platform, South African Medical Research Council, South African Medical Research Council/WITS Developmental Pathways for Health Research Unit (DPHRU)), Bamidele Tayo (Loyola University Chicago), Dwomoa Adu (University of Ghana Medical School), Michele Ramsay (Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand), Andrew P Morris (Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, University of Manchester), Tinashe Chikowore (SAMRC/Wits Developmental Pathways for Health Research Unit, Sydney Brenner Institute for Molecular Bioscience, Harvard Medical School, Channing Division of Network Medicine)

Abstract: Background: Obesity is a growing global health crisis, projected to increase by 115%, from 524 million in 2010 to 1.13 billion by 2030. Yet over 96% of genomic studies have been conducted in individuals of European ancestry, and the proteomic determinants of obesity in African populations remain underexplored. This disparity limits the relevance of existing genetic discoveries and hinders the development of effective, ancestry-informed risk prediction tools and therapeutic targets for Africans.

Methods: We performed the largest proteogenomic analysis of body mass index (BMI) to date in African-ancestry individuals, comprising a meta-analysis of GWAS across 31,786 continental Africans (16,406

from West, 7,963 from East, and 7,417 from Southern Africa) and 169,540 Admixed Africans from MVP, TOPMed, AllofUs and UK Biobank (UKB), totaling 201,326 individuals. An African ancestry-specific BMI polygenic risk score (PRS) was derived in N=193,909 and validated in an independent South African cohort. In a subset of continental Africans with plasma proteomic data, we used machine learning to construct a weighted African-specific proteomic score and evaluated its predictive utility for 10-year incident obesity in the Middle-aged Soweto Cohort (MASC; N=878) and UKB (N=7,959). We also conducted colocalization analysis to identify shared causal variants between BMI-associated protein quantitative trait loci (pQTLs) (UKB; N=931) and our meta-analysis.

Results: We found 1,910 genome-wide significant ($P < 5 \times 10^{-8}$), variants, and identified 87 BMI loci. 28 were novel loci, and 2 were found in regions never reported before in any populations for BMI. We found evidence for colocalization between a pQTL at APOL1 and genetic risk of BMI (posterior probability 4 of (PP4) = 99.5%). Our African-specific PRS improved predictivity of trait variance by 2.8-fold ($R^2 = 4.35\%$, $P = 1.21 \times 10^{-43}$) compared to prior PRS ($R^2 = 1.58\%$, $P = 2.8 \times 10^{-21}$). The African-derived proteomic score predicted 10-year incident for obesity [IMYT1] (BMI > 30) in MASC with HR = 2.41 (95% CI: 2.18–1.65) and Harrell's C-index = 0.73. Predictive performance generalized to Europeans ancestry individuals in the UKB (HR = 2.95, 95% CI: 2.63–3.32; C-index = 0.74).

Conclusion: This comprehensive proteogenomic study highlights the power of African-ancestry datasets to uncover novel genetic loci and validate cross-ancestry proteomic risk scores. Our findings advance the understanding of obesity biology and support the development of cross-ancestry predictive tools for global precision medicine.

Characterizing the genetic and environmental factors that impact vitamin D metabolism in sub-Saharan Africa

Subsession Time: Thursday, October 16 at 11:00am – 11:15am

Authors: Daniel N. Harris (University of Pennsylvania), Keanu Nahmi Natan (University of Pennsylvania), Steven Christopher Jones (University of Pennsylvania), Daniel Hui (University of Pennsylvania), Matthew E.B. Hansen (University of Pennsylvania), Thomas Nyambo (Kairuki University), Sununguko Wata Mpoloka (University of Botswana), Gaonyadiwe George Mokone (University of Botswana), Gurja Belay (Addis Ababa University), Charles Fokunang (The University of Yaoundé I), Alfred K. Njamnshi (Brain Research Africa Initiative (BRAIN), The University of Yaoundé I), Michael F. Holick (Boston University Chobanian & Avedisian School of Medicine), Sarah A. Tishkoff (University of Pennsylvania), NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium

Abstract: The primary source of Vitamin D (VitD) is from the synthesis of VitD3 in skin when ultraviolet-B radiation (UVB) reacts with 7-dehydrocholesterol. It is hypothesized that light skin pigmentation evolved to increase VitD synthesis in regions with low UVB levels. VitD is important for maintaining bone and immune health. However, little is known about the genetic and environmental factors that impact VitD levels in sub-Saharan Africans. We quantified 25-hydroxyvitamin D2 (25(OH)D2), 25(OH)D3 and total 25(OH)D in 896 sub-Saharan Africans with varied genetic ancestries who live in diverse environments. 25(OH)D2 was very low, except in a few Eastern African groups, possibly due to a unique component of

their diet or food supplements. 25(OH)D3 was highly variable and was inversely correlated to skin melanin levels, although with some exceptions. 25(OH)D3 is the lowest in individuals from the Ethiopian highlands where UVB levels are high while also having some of the most lightly pigmented skin in our study, possibly due to cultural factors that impact VitD3 synthesis such as clothing type. 71% of our dataset had total 25(OH)D levels lower than the 2011 endocrine society's recommendation, which calls for additional evaluation regarding if this guideline is appropriate for individuals of African ancestry. GWAS identified a genome-wide significant locus on chromosome 11. Fine mapping identified one credible set, which includes seven regulatory region variants. Six are GTEx eQTLs that are associated with a decrease in DHCR7 gene expression in skin. DHCR7 codes for a protein that converts 7-dehydrocholesterol into a cholesterol that cannot form VitD3. The variant with the most significant GWAS p-value was associated with a 2.2 ng/mL increase in 25(OH)D3, is absent from the GWAS catalog, and is common in sub-Saharan Africans (20%) but rare in non-Africans (0.05%). This variant is at the highest frequency in rainforest hunter-gatherers from Cameroon (54%) and shows evidence of being under local positive selection. These observations suggest a model where regulatory variants reduce DHCR7 gene expression in skin, which therefore increases the capacity of VitD3 synthesis. Our results identify an African-specific haplotype that is associated with an increase in VitD3 and demonstrates the complex relationship between genetic variation, skin pigmentation, the environment and VitD levels. Funded by R35GM134957, X01HL139409, R01AR076241, and ADA Pathway to Stop Diabetes awards.

Dissecting Genetic Contributors to eGFR Bias: Effects of African-Enriched GATM Regulatory Variants on Detailed Renal Phenotypes

Subsession Time: Thursday, October 16 at 11:15am – 11:30am

Authors: Olivia A. Gray (Variant Bio, Settle, WA, USA), Mohd Karim (Variant Bio, Settle, WA, USA), Jean-Tristan Brandenburg (Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa., Strengthening Oncology Services Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.) Rouvick Mariano Gama (Department of Inflammation of Biology, Faculty of Life Sciences and Medicine, School of Immunology and Microbial Sciences, King's College London, London, United Kingdom.), Kate Bramham (Centre for Nephrology, Urology and Transplantation, School of Immunology and Microbial Science, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom.), Robert Kalyesubula (Department of Physiology, Makerere University College of Health Sciences, Kampala, Uganda., Department of Medicine, Makerere University College of Health Sciences, Kampala, Uganda., Section of Nephrology, Yale School of Medicine, New Haven, CT, USA.), Abram Bunya Kamiza (The African Computational Genomic (TACG) Research Group, MRC/UVRI and LSHTM, Entebbe, Uganda.), Melissa Hendershott (Variant Bio, Settle, WA, USA), Anne-Katrin Emde (Variant Bio, Settle, WA, USA), Laura Yerges-Armstrong (Variant Bio, Settle, WA, USA), Stephane Castel (Variant Bio, Settle, WA, USA), Segun Fatumo (NCD Genomics & The African Computational Genomics (TACG) Research Group, MRC/UVRI and LSHTM Uganda Research Unit, Entebbe, Uganda., Precision Healthcare University Research Institute, Queen Mary University of London, United Kingdom.), June Fabian (Wits Donald Gordon Medical Research Institute, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South

Africa., Medical Research Council/Wits University Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.)

Abstract: Serum creatinine is the most widely used biomarker for assessing kidney function and diagnosing kidney disease worldwide. Estimated glomerular filtration rate using creatinine (eGFR_{Cr}) is considered an efficient alternative to directly measuring GFR (mGFR) and is favored in clinical practice. However, most estimation equations have been modelled among individuals of European ancestry, with persistent issues noted in their cross-ancestry performance, particularly amongst African ancestry individuals. Serum creatinine concentration is a highly heritable trait across populations, and genetic variants influencing creatinine metabolism could introduce a systematic bias that is unrelated to kidney function. Here, we present a deep interrogation of one such locus, with strong genetic links to both creatinine synthesis and eGFR. We performed genome wide association studies (GWAS) on kidney biomarkers and global serum metabolites (mGWAS) in an African ancestry cohort of over 900 Tsonga individuals from the Agincourt district of South Africa. In contrast to previous genetic studies, we directly measured GFR using iohexol and compared this to eGFR_{Cr} in the same cohort to better understand if genetic variation reflected a true impact on kidney function. We replicated a well-established eGFR_{Cr} association signal at the glycine amidinotransferase (GATM) gene and linked it, for the first time, to increased creatine metabolites, the precursor of creatinine. Alleles at this locus were associated with increases in both creatine and creatinine - resulting in lower eGFR_{Cr} - yet had no impact on mGFR. Critically, the haplotype associated with increased creatinine synthesis is found at ~50% frequency in both continental African and admixed African ancestry cohorts but is unobserved in European ancestry cohorts. Expanding our analysis to publicly available GWAS we found evidence that these African-enriched alleles were associated with reduced eGFR_{Cr} and potentially CKD stage III misdiagnosis in the MVP cohort. These findings have broad implications for the use and interpretation of eGFR_{Cr} for the estimation of kidney function and diagnosis of CKD in African ancestry populations.

Gene expression associated with impending severe asthma exacerbations in non-Hispanic Black individuals with asthma

Subsession Time: Thursday, October 16 at 11:30am – 11:45am

Authors: Baojun Wu (Center for Individualized and Genomic Medicine Research, Department of Medicine, Henry Ford Health, Detroit, Michigan, USA), Mao Yang (Center for Individualized and Genomic Medicine Research, Department of Medicine, Henry Ford Health, Detroit, Michigan, USA), Donglei Hu (Department of Medicine, University of California San Francisco, San Francisco, California, USA), Samantha Hochstadt (Center for Individualized and Genomic Medicine Research, Department of Medicine, Henry Ford Health, Detroit, Michigan, USA), David E. Lanfear (Center for Individualized and Genomic Medicine Research, Department of Medicine, Henry Ford Health, Detroit, Michigan, USA), Jonathan Witonsky (Department of Medicine, University of California San Francisco, San Francisco, California, USA), Esteban G. Burchard (Department of Medicine, University of California San Francisco, San Francisco, California, USA), Elad Ziv (Department of Medicine, University of California San Francisco,

San Francisco, California, USA), L. Keoki Williams (Center for Individualized and Genomic Medicine Research, Department of Medicine, Henry Ford Health, Detroit, Michigan, USA)

Abstract: Background: Severe asthma exacerbations contribute substantially to asthma-related morbidity. Early prediction of impending exacerbations remains a clinical challenge.

Objective: To identify blood gene expression signatures associated with impending severe asthma exacerbations.

Methods: We conducted a longitudinal cohort study involving 375 asthma patients monitored for over 60 days following enrollment. Peripheral blood samples were collected at enrollment and analyzed using transcriptomic profiling. Severe exacerbations were defined as events resulting in either an asthma-related emergency department (ED) visits or hospitalization within 60 days of enrollment. An analysis of differentially expressed genes (DEGs) was performed comparing samples without and with a severe exacerbation. Pathway enrichment, immune cell counts, and machine learning analyses (i.e., elastic net regularized regression) were used to identify mechanisms and predictive signatures associated with an exacerbation.

Results: We identified 463 genes associated with a subsequent exacerbation. Neutrophil extracellular trap formation was the most prominent pathway significantly associated with events. Patients with severe asthma exacerbations exhibited a significantly higher percentage of circulating neutrophils ($p = 0.023$), which showed a strong correlation with severe asthma exacerbation ($p = 0.007$). A predictive model based on a 6-DEG signature achieved an area under curve (AUC) of 95% in cross-validation. The top predictor within the signature, GPAT2, alone yielded an AUC of 80%. Co-expression analysis and phenome-wide association suggested that GPAT2 may influence helper T-cell (Th1, Th2, and Th17) differentiation, neutrophil migration, and eosinophil percentage.

Conclusion: Blood gene expression profiling identified expression patterns associated with severe asthma exacerbations. These findings may support the development of better predictive tests for identifying and preventing severe asthma exacerbations.

Session 48: In Situ Insight: Uncovering Cellular Regulation with Spatial-Omics

Location: Room 258ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 10:45am – 11:45am

BRIDGE: Inferring 3D Molecular Tissue Structure from Sparse Spatial Omics and Histological Sections

Subsession Time: Thursday, October 16 at 10:45am – 11:00am

Authors: Shihan Liu (Emory University), Jian Hu (Emory University)

Abstract: Background: Spatial omics technologies have become increasingly popular and are used to explore a wide range of biological questions. Most studies extract only a few 2D slices from a single large tissue, introducing selection bias and limiting 3D structural understanding. Increasingly, researchers generate spatial omics data from consecutive sections to reconstruct 3D tissue architecture, which holds promise but remains prohibitively expensive. As a compromise, hematoxylin and eosin (H&E) staining images are often used to assess tissue morphology; they are cost-efficient but lack molecular information. Given the shared 3D continuity between morphology and omics measurement, we hypothesize that it is possible to reconstruct the 3D molecular structure of tissue using sparse 2D spatial omics slices combined with consecutive H&E images from the same specimen. Method: To address this gap, we developed BRIDGE, a machine learning method for 3D reconstruction of molecularly defined tissue structures using sparse spatial omics slices and consecutive H&E-stained sections. BRIDGE has three steps: 1) It identifies fine-grained spatial functional structures based on molecular information from the spatial omics slices. 2) In parallel, it detects coarse histological regions in the consecutive H&E images and models their 3D continuity through a deformation field. 3) It aligns the spatial omics slices with the H&E images and propagates the functional region boundaries to neighboring H&E sections using diffeomorphic deformation, enabling consistent 3D tracking of molecular structures. Result: We applied BRIDGE to multiple spatial omics datasets and it consistently achieved high-quality 3D reconstructions of molecular substructures. In a colorectal cancer sample, BRIDGE reconstructed 3D tertiary lymphoid structures (TLS) from 3 CyCIF-stained and 22 consecutive H&E sections. While TLS appeared separated in individual 2D slices, the 3D reconstruction revealed their connected architecture, suggesting coordinated local immune responses. In a metastatic lymph node sample, BRIDGE reconstructed the 3D immune microenvironment using only 3 spatial transcriptomics and 27 H&E sections, revealing immune infiltration pathways and spatially organized niches that were undetectable in histological images alone. These results demonstrate BRIDGE's ability to uncover 3D biological tissue organization that cannot be revealed in 2D.

Applying the Novel Illumina Spatial Technology in Pig-Kidney to Human Xenotransplant biopsies allows deep profiling of infiltrating human immune cells

Subsession Time: Thursday, October 16 at 11:00am – 11:15am

Authors: Eloi Schmauch*¹, Qian Guo*¹, Kiana Moi*¹, Susanna Vikman¹, Diana Argibay¹, Justin Little¹, Simon Williams¹, David Berglund², Erik Berglund³, Brendan J. Keating^{1,4}, 1. Division of Transplantation, Department of Surgery, New York University Langone Health, New York, USA, 2. Uppsala University, Uppsala, Sweden, 3. Karolinska Institute, Stockholm, Sweden, 4. Institute of Systems Genetics, New York University Langone Health, NYC 10016, USA

Abstract: In the US >100,000 subjects are on transplant waitlists. Xenotransplantation (XTx) has shown promise in addressing organ shortages but there is incomplete understanding of immune responses. We have performed 8 pig-to-human XTx's integrating over 12 different omics per procedure including spatial transcriptomics to identify molecular dynamics associated with clinical trajectories. We applied the Illumina spatial technology, a poly-A capture based wet-lab and analytic pipeline to biopsies from a recent pig kidney to human XTx. Large Day0 & Day28 (explant) xenokidney OCT-biopsies were sectioned from a XTx procedure. Sequenced cDNA was spatially mapped back to H&E images generated for each tissue section. We performed upstream sequence processing in addition to nuclei and cell segmentation on Illumina Connected Analytics (ICA) using the DRAGEN spatial pipeline. We aligned transcripts to a joint human GRCh38 and pig Sus scrofa 11.1 genome and used custom scripts to partition pig kidney and infiltrating human immune cells. The cell segmentation model is trained on a diverse set of challenging tissue images. This allows accurate nuclei detection for accurate mapping of transcripts and species assignment of cells based on genome-specific expression profiles, resulting in single cell resolution counts matrices for both pig and human cells.

We identified >300k and >100k cells in the Day0 & Day28 biopsies, with ~17% identified as human: Median transcript counts = 2,000-3,500 and gene counts = 1,000-1,500 per cell. We identified all major pig kidney cell types (n=15) including resident immune cells. Most of those populations showed distinct expression shifts between Day0 & Day28 biopsies. SFRP2⁺ fibroblasts were enriched in the explant, indicating fibrotic activation. Multiple cell types showed elevated expression of fibrotic-associated genes including COL1A1, SPP1, IL33, FOSB, JUN and complement. Diverse human immune cells infiltrated the pig graft, including macrophages, NK cells, and T cells. The profiling depth enabled identification of macrophage subtypes, such as pro-inflammatory (SPP1+CD36+CD9+) cells, and activated T-cells subpopulations.

We employed this new spatial assay to delineate pig and infiltrating human cells and uncover molecular signatures within the xenograft biopsies, supported by robust cell segmentation through the DRAGEN spatial pipeline. The high number of transcripts/genes provide a robust sensitive assay with precise transcript locations.

Mitochondrial lineage tracing within spatially intact human tissues

Subsession Time: Thursday, October 16 at 11:15am – 11:30am

Authors: Jiazhen Rong (University of Pennsylvania, Graduate Group in Genomics & Computational Biology), Sydney Bracht (University of Pennsylvania, Bioengineering Department), Sydney Shaffer

(University of Pennsylvania, Pathology and Bioengineering Department), Nancy Zhang (University of Pennsylvania, Statistics and Data Science Department)

Abstract: Understanding tissue development and intra-tissue evolution requires the ability to trace cellular lineages in intact tissues coupled with high-plex molecular profiling preserving spatial context. However, current lineage tracing tools are incompatible with spatial omics. Here, we present SUMMIT (Spatially Unveiling Mitochondrial Mutations In Tissues), a spatially-resolved lineage tracing technology that integrates gene expression profiling with mitochondrial mutation-based clone identification. Since SUMMIT relies only on endogenous mutations, it can be applied to human tissues. To address the compositional mixing of cell types within spatial spots, SUMMIT includes a rigorous statistical framework to confidently assign variants to specific cell subpopulations. SUMMIT achieves high power for rare subclones that are spatially localized by pooling information across neighboring spots. We validated SUMMIT using a controlled model in which we mixed two cancer cell lines in a mouse tumor, then demonstrated it on multiple human tissues including Barrett's esophagus, gastric cardia, small bowel, and colorectal cancer. Across these samples, we distinguished between global mutations and mutations marking locally restricted clones. The coupled transcriptomic data allowed us to characterize the cell type composition within each clone and delineate their spatial configuration. This integrated spatial transcriptomic and clonal profiling provides a framework for dissecting how cellular ancestry and spatial context shape tissue heterogeneity.

Subcellular spatial transcriptomics with spatial neighborhood analysis framework reveals immune–stromal crosstalk within the synovium of patients with juvenile idiopathic arthritis

Subsession Time: Thursday, October 16 at 11:30am – 11:45am

Authors: Jun Inamo (University of Colorado School of Medicine), Roselyn Fierkens (University of Colorado School of Medicine), Michael CLay (University of Colorado School of Medicine), Clara Lin (University of Colorado School of Medicine), Kari Hayes (University of Colorado School of Medicine), Nathan Rogers (Children's Hospital Colorado), Heather Leach (Children's Hospital Colorado), Kentaro Yomogida (University of Colorado School of Medicine).

Abstract: Juvenile idiopathic arthritis (JIA) is a chronic inflammatory disorder affecting children, yet its tissue-level cellular architecture remains under-characterized. Here, we present a high-resolution spatial transcriptomic map of JIA synovium, integrating novel computational pipelines, R package for spatial neighborhood analysis (SNA), to dissect cellular composition and spatial neighborhood structure.

We profiled synovial biopsies from nine JIA patients using the 10x Genomics Xenium platform (Prime 5K panel), capturing >350,000 cells across FFPE tissue sections. We implemented a pipeline for preprocessing, normalization, batch correction, and unsupervised cellular clustering. In the SNA framework, spatial neighborhood enrichment was computed via permutation-based Z-scores, and we also developed a single-cell–level colocalization scoring system. Ligand–receptor inference was performed, integrated with custom colocalization metrics to prioritize spatially proximal signaling.

We identified diverse immune and stromal cell populations, including CXCL9+ inflammatory macrophages, THY1+NOTCH3+ sublining fibroblasts, and Tph/Tfh-like CD4 T cells. Spatial neighborhood enrichment uncovered fibroblast–endothelial pairings, validated by co-expression of NOTCH3–JAG1, and macrophage–T cell interactions enriched in CXCL9–CXCR3 signaling. Colocalization scores between T and B/plasma cells successfully identified tertiary lymphoid structures (TLS), which were characterized by canonical chemokine pairs (CXCL13–CXCR5). Comparative integration with RA scRNA-seq data revealed JIA-dominant expansion of THY1+NOTCH3+ sublining fibroblasts and GZMB+ CD8 T cells.

This study introduces a computational framework for integrative spatial transcriptomic analysis at subcellular resolution, revealing disease-specific cellular niches and signaling landscapes in JIA synovium. Our approach enables quantitative modeling of immune–stromal interactions and offers a scalable strategy for decoding tissue microenvironments in complex diseases.

Session 49: Methylation Mechanisms: New Insights into DNA-Methylation in Health and Disease

Location: Room 210C/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 10:45am – 11:45am

★ Epigenome-Wide and Causal Mediation Analyses Identify DNA Methylation Pathways Mediating Longitudinal Glycemic Effects on Diabetic Retinopathy

Subsession Time: Thursday, October 16 at 10:45am – 11:00am

Authors: Brendon Chau (UCLA, Atlanta VA), Peter D. Reaven (Phoenix VA), Aubrey Jensen (UCLA, Phoenix VA), Joe H. Breyear (NIEHS), Todd L. Edwards (Vanderbilt University), Hua Zhou (UCLA, Phoenix VA), Lawrence S. Phillips (Emory, Atlanta VA), Yan V. Sun (Emory, Atlanta VA), Jin J. Zhou (UCLA, Phoenix VA) the Million Veteran Program

Abstract: Cumulative glycemic exposure is a key driver of diabetic retinopathy (DR) in individuals with type 2 diabetes (T2D). DNA methylation (DNAm) may mediate the long-term effect of glycemic exposure on DR through transcriptional regulation. To address this, we conducted the largest to date epigenome-wide association study (EWAS) in T2D patients from the Million Veteran Program (MVP), incorporating causal mediation analyses to identify DNAm sites that mediate the effect of glycemia on incident DR (of any severity).

We analyzed five years of longitudinal HbA1c data preceding blood-based DNAm profiling for 11,599 T2D individuals. Glycemic metrics included: (1) mean HbA1c, (2) coefficient of variation (CV) of HbA1c, and (3) excess HbA1c (area above 6.1%). 768,569 DNAm sites were analyzed via EWAS models adjusted for age, sex, ancestry, T2D duration, and cell type composition, with empirical null correction, and Bonferroni correction for multiple testing ($p < 6.7 \times 10^{-8}$). Gene-level enrichment was examined using hypergeometric testing (FDR < 0.05). Among 8,058 DR-free participants at DNAm profiling, time to DR onset was modeled using Weibull accelerated failure time regression. DNAm sites significant in both EWAS and DR-onset models (FDR < 0.05) were tested for significant mediation effects using bootstrap.

EWAS identified 2,301, 1,170, and 245 DNAm sites significantly associated with mean, excess, and CV HbA1c, respectively. Cg19693031 (TXNIP) was the top shared signal. Novel findings included cg16504526 (KLF9), cg14115749 (GATAD2A), and CV-specific sites such as cg07941916 (CARINH) and cg17475445 (MORC2). DNAm sites associated with mean and excess HbA1c were enriched in glucose and lipid metabolism genes (ABCG1, RARA), while CV highlighted SLC1A5 and STRN4. Among EWAS-significant sites, 231 (mean), 122 (excess), and 117 (CV) were associated with DR onset. Mediation analysis identified 14 (mean), 7 (excess), and 12 (CV) DNAm sites with significant indirect effects. Cg00407100 (C5orf64) showed the strongest mediation for mean and excess HbA1c. Only CV HbA1c showed a significant total DNAm-mediated effect on DR risk (34%, 95% CI: 23%–48%).

This study presents the most comprehensive glyceic EWAS to date, examining three glyceic metrics and is the first to evaluate DNAm as a mediator of glyceic variability's effect on DR. Our findings identify novel glyceia-associated DNAm sites and suggest that glyceic variability contributes to DR through epigenetic pathways.

Widespread non-Mendelian inheritance of DNA methylation patterns in mice

Subsession Time: Thursday, October 16 at 11:00am – 11:15am

Authors: Adam Davidovich (Johns Hopkins University), Danila Cuomo (Texas A&M University), Hang Su (University of North Carolina at Chapel Hill), Sandeep Kambhampati (Johns Hopkins University), Qingqing Gong (Johns Hopkins University), Alexandra Naron (Texas A&M University), Rake! Tryggvadottir (Johns Hopkins University), Leonard McMillan (University of North Carolina at Chapel Hill), David W. Threadgill (Texas A&M University), Kasper D. Hansen (Johns Hopkins University), Andrew P. Feinberg (Johns Hopkins University, Tel Aviv University)

Abstract: Serving as the foundation of modern genetics, Mendel's laws of inheritance shape much of our current understanding of monogenic and complex trait inheritance. However, despite the established potential of DNA methylation to violate Mendelian inheritance through genomic imprinting, comparatively little is known about its intergenerational inheritance. Here, we investigated the intergenerational inheritance of DNA methylation patterns in mice using allele-specific methylation data from whole-genome ONT sequencing of 36 liver and 23 muscle samples from two genetically distinct inbred mouse strains and their heterozygous F1 crosses, alongside targeted ONT sequencing of 19 liver samples from F2 crosses of the same strains.

The use of long-read sequencing enabled a phased analysis of methylation over ~12M CpGs, accounting for ~90% of CpGs outside of regions for which the two strains are identical by descent, i.e. lacking genetic divergence. We identified widespread Mendelian patterns of epigenetic inheritance, primarily composed of 7,087 cis-acting meQTLs (FWER < 0.1), as well as many examples of non-Mendelian forms of epigenetic inheritance. These include 54 examples of emergent epigenetic inheritance patterns, in which at least one allele of the heterozygous offspring exhibits a consistent methylation pattern not found in either parental strain. Emergent patterns appear over genes of high clinical significance, including critical regulators of immune function. Furthermore, we identified the first naturally occurring example of paramutation in a mammalian genome, a phenomenon in which one dominant (paramutant) allele alters the methylation level of the homologous (paramutable) allele in a manner which can be inherited independent of the paramutant allele. This paramutation event, identified over the promoter of a gene involved in the regulation of signal transduction during spermatogenesis, was shown to be stable through the F2 generation. We also discovered six novel imprinted genes, including one of the first imprinted loci discovered on the X chromosome which overlaps a CpG island within a zinc-finger protein associated with X-linked intellectual disability, as well as 305 instances of sex-specific methylation localized to the liver. Overall, we identified 522 autosomal non-Mendelian inheritance patterns, representing ~7% of autosomal epigenetic inheritance patterns. These results have profound implications for the understanding of complex trait heritability.

Differential Bulk and Cell-Type Specific DNA Methylation in Asthma with Severe Exacerbations in Pediatric Populations of Diverse Ethnicities

Subsession Time: Thursday, October 16 at 11:15am – 11:30am

Authors: Javier Perez-Garcia (Stanford University), Mario Martin-Almeida (Universidad de La Laguna), Elena Martin-Gonzalez, (Universidad de La Laguna), Jonathan Witonsky (University of California, San Francisco), Celeste Eng (University of California, San Francisco), Aditya Gorla (University of California Los Angeles), Zeyuan Johnson Chen (University of California Los Angeles), Fabian Lorenzo-Diaz (Universidad de La Laguna), Anne Bozack (Stanford University), Jennifer R. Elhawary (University of California, San Francisco), Donglei Hu (University of California, San Francisco), Scott Huntsman (University of California, San Francisco), Jesús Villar (Instituto de Salud Carlos III, Hospital Universitario Dr. Negrín, St. Michael's Hospital, Universidad del Atlántico Medio), Jose Rodriguez-Santana (Centro de Neumología Pediátrica de San Juan), Luisa N. Borrell (University of New York City), Elinor Rahmani (University of California Los Angeles), Elad Ziv (University of California, San Francisco), Esteban G. Burchard (University of California, San Francisco), Andres Cardenas (Stanford University), Maria Pino-Yanes (Universidad de La Laguna)

Abstract: Background: Asthma is the most prevalent pediatric chronic disease, where exacerbations contribute to the major global burden. It disproportionately affects different populations due to biological and socio-environmental factors, which can be reflected in DNA methylation (DNAm). However, to date, we lack biomarkers with clinical applicability for asthma exacerbations, and the role of DNAm remains understudied.

Aim: To identify DNAm markers associated with asthma with severe exacerbations in children of diverse ethnicities.

Methods: Leukocyte DNAm was profiled in 721 African Americans (SAGE), 571 Puerto Ricans, and 373 Mexican American children (GALA II) with and without asthma using the Illumina EPICv1 array. We conducted an epigenome-wide association study (EWAS) of asthma with exacerbations using regression models adjusted for age, sex, ancestry, cell heterogeneity, and batch effect. We corrected genomic inflation using a Bayesian method and declared significance using a false discovery rate (FDR)<0.05. Additionally, we estimated differentially methylated regions (DMRs) and identified cell-type-specific signals in a cell-type deconvolution EWAS. SAGE was analyzed in the discovery phase, and GALA II in the replication phase. We cross-tissue evaluated the association with asthma in nasal samples from 541 diverse children (Project Viva).

Results: 356 CpGs and 68 DMRs were significantly associated with asthma with severe exacerbations in African Americans (FDR<0.05; $\lambda=1.00$), and 312 CpGs and 55 DMRs replicated in any Latino groups (p<0.05). Top-hits were annotated to genes related to immune and inflammatory processes (e.g., GRAMD4, TPST2, GLCCI1, CCL4, IL18RAP, JAK1, PRF1, TNFRSF4). Among them, 25 were associated with asthma in nasal epithelia (e.g., HDAC4, ITPR1, PIK3R1). We detected 12 cell-type-specific DNAm signals, most of them driven by lymphocytes (e.g., PRTN3, SLAMF1, GPR141) and eosinophils (e.g., GPR55, TPST2), including genes involved in drug response.

Conclusions: We discovered novel associations of DNAm loci for asthma with severe exacerbations in children from diverse ethnicities. We detected bulk and cell-type-specific signals in blood leukocytes involved in immune response, with cross-validated effects in nasal tissue, suggesting potential asthma therapeutic targets.

Funding: NHLBI (5R01HL117004-06, 1R01HL155024-01); NHGRI; NIEHS (R01 ES031259); NIMHD; MCIN/AEI/10.13039/501100011033 (PID2020-116274RB-I00).

Therapeutic epigenome editing for Prader-Willi Syndrome using a novel non-viral, non-nanoparticle delivery platform for CRISPR/dCas9-based gene regulation

Subsession Time: Thursday, October 16 at 11:30am – 11:45am

Authors: Xiaona Lu (Genetics, Yale University), Sung-Eun Wang (Genetics, Yale University), Ying Xie (Neurosurgery, Yale University), Jiangbing Zhou (Neurosurgery, Yale University), Yong-Hui Jiang (Genetics, Yale University)

Abstract: Epigenome editing using CRISPR/dCas9 offers a promising therapeutic avenue for neurogenetic disorders. However, its clinical translation is hindered by challenges such as poor blood-brain barrier penetration, limited cellular uptake, and narrow therapeutic windows. Conventional strategies, including EHMT2/G9a inhibitors to reduce H3K9 methylation or enzymatic DNA demethylation, have shown partial efficacy but are often associated with genome-wide off-target effects, limiting their therapeutic potential. Prader-Willi Syndrome (PWS), a neurodevelopmental disorder caused by paternal deficiency of the 15q11-q13 imprinting domain, presents a compelling therapeutic opportunity due to the presence of an intact but epigenetically silenced maternal allele of SNRPN and SNORD116. Here, we present a novel, non-viral, non-nanoparticle delivery platform, STEP, for targeted delivery of CRISPR/dCas9-based epigenetic editors to reactivate these genes in PWS. As part of the NIH Somatic Cell Genome Editing (SCGE) consortium, we developed STEP-RNP, a chemically modified, non-viral ribonucleoprotein (RNP) delivery platform optimized for tissue and CNS penetration. We used this system to deliver dCas9-JMJD2a, a fusion of dCas9 with the H3K9me2/3 demethylase JMJD2a, along with sgRNA designed to reduce repressive chromatin marks and reactivate gene expression. We applied STEP-dCas9/sgRNA-JMJD2a to neurons and brain organoids differentiated from PWS patient-derived iPSCs, and to 2 PWS mouse models via day1 IT or IV injection: (1) PWSm+/pS-U, with a paternal deletion from Snrpn to Ube3a, and (2) PWSm+/p-LD, a novel model with the entire paternal 15q11-q13 deletion. RT-qPCR and RNA-seq confirmed reactivation of Snrpn and Snord116 from the maternal allele. Importantly, STEP-RNP rescued perinatal lethality, growth failure, and neurodevelopmental delay. Over 80% of PWSm+/pS-U mice and 100% PWSm+/p-LD mice die without treatment, whereas STEP-treated animals showed marked improvements in survival, growth, and neurodevelopment. Accordingly, this treatment led to significant improvement in multiple behavioral domains, including locomotor function, learning and memory. Similar gene reactivation was observed in PWS patient's iPSC-derived neurons and brain organoids. These results demonstrate that STEP enables effective, brain-penetrant delivery of

CRISPR/dCas9 epigenetic editors and represents a promising therapeutic platform for PWS and other neurogenetic disorders.

Session 50: Precision in Chaos: Exploring the Intricacies of Immune Regulation and Autoimmunity

Location: Room 156ABC/Level 1, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 10:45am – 11:45am

A multi-omics resource of B cell activation reveals genetic mechanisms for immune-mediated diseases

Subsession Time: Thursday, October 16 at 10:45am – 11:00am

Authors: Vitor R. C. Aguiar (Boston Children's Hospital, Broad Institute of MIT and Harvard), Marcella E. Franco (Boston Children's Hospital, Broad Institute of MIT and Harvard, Max Delbrück Center for Molecular Medicine in the Helmholtz Association), Nada Abdel Aziz (Boston Children's Hospital, Broad Institute of MIT and Harvard), Daniela Fernandez-Salinas (Boston Children's Hospital, Broad Institute of MIT and Harvard), Marcos Chinas (Boston Children's Hospital, Broad Institute of MIT and Harvard), Mariasilvia Colantuoni (Boston Children's Hospital, Broad Institute of MIT and Harvard), Qian Xiao (Boston Children's Hospital), Nicolaj Hackert (Boston Children's Hospital, Broad Institute of MIT and Harvard, Heidelberg University Hospital), Yifei Liao (Brigham and Women's Hospital), Rodrigo Cervantes-Diaz (Boston Children's Hospital, Broad Institute of MIT and Harvard), Marc Todd (Boston Children's Hospital), Brian Wauford (Boston Children's Hospital), Alex Wactor (Boston Children's Hospital), Vaishali Prahalad (Boston Children's Hospital), Raquel Laza-Briviesca (Boston Children's Hospital), Roxane Darbousset (Boston Children's Hospital), Qiang Wang (Boston Children's Hospital), Scott Jenks (Emory University), Kevin S. Cashman (Emory University), Esther Zumaquero (University of Alabama at Birmingham), Zhu Zhu (Brigham and Women's Hospital), Junning Case (Brigham and Women's Hospital), Paloma Cejas (Dana-Farber Cancer Institute), Miguel Gomez (Dana-Farber Cancer Institute), Hannah Ainsworth (Wake Forest School of Medicine), Miranda Marion (Wake Forest School of Medicine), Mehdi Benamar (Boston Children's Hospital), Pui Lee (Boston Children's Hospital), Lauren Henderson (Boston Children's Hospital), Margaret Chang (Boston Children's Hospital), Kevin Wei (Brigham and Women's Hospital), Henry Long (Dana-Farber Cancer Institute), Carl D. Langefeld (Wake Forest School of Medicine), Benjamin E. Gewurz (Brigham and Women's Hospital), Ignacio Sanz (Emory University), Jeffrey A. Sparks (Brigham and Women's Hospital), Esra Meidan (Boston Children's Hospital), Peter A. Nigrovic (Boston Children's Hospital, Brigham and Women's Hospital), Maria Gutierrez-Arcelus (Boston Children's Hospital, Broad Institute of MIT and Harvard))

Abstract: Most genetic variants that confer risk of complex immune-mediated diseases (IMDs) affect gene regulation in specific cell types. Their target genes and focus cell types are often unknown, partially because some effects are hidden in untested cell states. B cells play central roles in IMDs, including autoimmune, allergic, infectious, and cancer-related diseases. Despite this established importance, B cell activation states are underrepresented in functional genomics studies. In this study, we obtained B cells from 26 healthy female donors and stimulated them in vitro with six activation conditions targeting key

pathways: the B cell receptor (BCR), Toll-like receptor 7 (TLR7), TLR9, CD40, and a cocktail that promotes differentiation into double negative 2 (DN2) IgD⁺ CD27⁺ CD11c⁺ CD21⁺ B cells, a likely pathogenic subset implicated in autoimmunity and infection. We profiled up to 24 B cell activation states and up to 5 control conditions using RNA-seq, single-cell RNA-seq with surface protein markers (CITE-seq), and ATAC-seq. We characterize how IMD-associated genes respond to stimuli and group into distinct functional programs. High-depth RNA-seq data reveals widespread splicing effects during B cell activation. Using single-cell data, we describe stimulus-dependent B cell fates. Chromatin data reveal transcription factors likely involved in B cell activation, and activation-dependent open chromatin regions that are enriched in IMD genetic risk. We experimentally validate a lupus risk variant in a stimulus-specific open chromatin region that regulates TNFSF4 expression, highlighting the relevance of studying B cell activation to elucidate disease association. These data are shared via an interactive browser that can be used to query the dynamics of gene regulation and B cell differentiation during activation by different stimuli, enhancing further investigation of B cells and their role in IMDs: <https://mgalab.shinyapps.io/bcellactivation>.

Multi-Trait GWAS and Fine-Mapping of 14 Traits Identify 17 Ancestry-Specific Immune Loci in Fibrotic and Inflammatory Diseases

Subsession Time: Thursday, October 16 at 11:00am – 11:15am

Authors: Gabrielle Hampton (Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University), Jacklyn Hellwege (Vanderbilt Genetic Institute, Division of Genetic Medicine, Vanderbilt University), Catherine Greene (Vanderbilt Genetic Institute, Division of Genetic Medicine, Vanderbilt University), James Jaworski (Vanderbilt Genetic Institute, Division of Genetic Medicine, Vanderbilt University), Megan J Roy-Puckelwartz (Center for Genetic Medicine, Northwestern University Clinical and Translational Sciences Institute), Yuan Luo (Department of Preventive Medicine, Northwestern University Feinberg School of Medicine), Digna R Velez Edwards (Division of Quantitative and Clinical Sciences, Department of Obstetrics and Gynecology, Department of Biomedical Informatics, Vanderbilt University), Todd Edwards (Department of Biomedical Informatics, Division of Genetic Medicine, Vanderbilt University)

Abstract: Chronic fibroproliferative (FPD) and inflammatory diseases (IFD) differ in prevalence across populations, likely due partly to divergent immune evolution. FPDs (e.g., hypertension, fibroids, keloids) are more common among individuals of African ancestry, while IFDs (e.g., rheumatoid arthritis, multiple sclerosis) are more prevalent in European ancestry populations. These disparities may reflect long-standing selective pressures: African ancestral populations evolved under exposure to extracellular parasites, favoring Th2-biased immune responses that promote tissue repair and fibrosis, whereas European populations faced greater pressure from intracellular pathogens, driving Th1-dominant inflammatory responses. However, the genetic basis underlying these population-specific immune biases remains incompletely understood. We used an ancestry-stratified, multi-layered genomic approach to identify shared and ancestry-specific genetic contributors to immune-related disease. We analyzed 8 FPD and 6 IFD traits in over 45,000 Non-Hispanic Black (NHB) and 500,000 Non-Hispanic White (NHW) individuals from BioVU, eMERGE, and UK Biobank. Methods included GWAS meta-analysis, SNP-based

heritability and genetic correlation analysis, multi-trait GWAS (MTAG), and fine-mapping. In NHBs, we identified a fibrosis-enriched architecture, with 6 shared loci across hypertension, fibroids, keloids, and asthma. Notably, GPHN (rs10713830, $P = 6.2 \times 10^{-18}$; P = genome-wide significance, PIP = 1.0; PIP = posterior inclusion probability) and TNXB (rs565080746, $P = 7.9 \times 10^{-10}$, PIP = 0.98) emerged as high-confidence loci, both involved in extracellular matrix remodeling, supporting a Th2-biased immune profile. In NHWs, IFD traits showed convergent signals at STAT4 (rs7574865, $P = 2.5 \times 10^{-18}$, PIP = 0.99), SH2B3 (rs3184504, $P = 4.1 \times 10^{-10}$, PIP = 0.93), and PSMB9 (rs4148878, $P = 2.0 \times 10^{-18}$, PIP = 0.91), consistent with a Th1-skewed autoimmune profile. Fine-mapping identified 17 ancestry-specific credible sets, including HLA-DQA2, FCGR2B, WNT2B, and MAP2K5, implicating known immune regulators and novel fibrotic genes. These findings reveal striking immunogenetic divergence across ancestries and emphasize the value of multi-trait, ancestry-aware analyses for uncovering disease mechanisms and guiding equitable precision medicine.

T cell receptor repertoire-wide association study identified potential biomarkers for rheumatoid arthritis

Subsession Time: Thursday, October 16 at 11:15am – 11:30am

Authors: Kazuyoshi Ishigaki (Department of Microbiology and Immunology, Keio University School of Medicine; Laboratory for Human Immunogenetics, Riken Center for Integrative Medical Sciences; Keio University Human Biology-Microbiome-Quantum Research Center), Takahiro Nishino (Department of Microbiology and Immunology, Keio University School of Medicine; Laboratory for Human Immunogenetics, Riken Center for Integrative Medical Sciences), Takahiro Arakawa (Laboratory for Human Immunogenetics, Riken Center for Integrative Medical Sciences), Rumi Sasai (Laboratory for Genotyping Development, Riken Center for Integrative Medical Sciences), Mitsuyo Yamaguchi (Laboratory for Genotyping Development, Riken Center for Integrative Medical Sciences), Mikiko Endo (Laboratory for Genotyping Development, Riken Center for Integrative Medical Sciences), Michihiro Kono (Department of Microbiology and Immunology, Keio University School of Medicine; Laboratory for Human Immunogenetics, Riken Center for Integrative Medical Sciences), Bunki Natsumoto (Laboratory for Human Immunogenetics, Riken Center for Integrative Medical Sciences), Kenichiro Asahara (Laboratory for Human Immunogenetics, Riken Center for Integrative Medical Sciences), Haruka Takahashi (Department of Microbiology and Immunology, Keio University School of Medicine; Laboratory for Human Immunogenetics, Riken Center for Integrative Medical Sciences), Hiroaki Hatano (Department of Microbiology and Immunology, Keio University School of Medicine; Laboratory for Human Immunogenetics, Riken Center for Integrative Medical Sciences), Masahiro Nakano (Laboratory for Human Immunogenetics, Riken Center for Integrative Medical Sciences), Koichi Matsuda (Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo), and Yukihide Momozawa (Laboratory for Genotyping Development, Riken Center for Integrative Medical Sciences).

Abstract: Antigen-specific immune responses are orchestrated by the T cell receptor (TCR), which distinguishes self- from non-self-epitopes presented by HLA molecules—whose polymorphisms represent the strongest known genetic risk factors for rheumatoid arthritis (RA). Given the canonical role

of HLA in presenting epitopes to TCRs, HLA-TCR interactions are central to RA pathogenesis. However, previous studies have largely failed to investigate TCR associations with RA, as standard genotyping platforms cannot assess TCR sequences. This limitation arises because TCR is primarily generated by somatic recombination of multiple gene segments during T cell development, rather than by germline-encoded genetic variation. To overcome this limitation, we performed the largest scale targeted sequencing of the TCR region to directly identify RA-specific TCR sequence patterns. We first developed a multiplex PCR-based platform with primers covering all TCR component genes. Using identical aliquots of genomic DNA (gDNA), we benchmarked our platform against that of Adaptive Biotechnologies, a widely used standard for TCR sequencing. Our platform demonstrated comparable sensitivity and accuracy. We next implemented an automated library preparation system for our platform to enable high-throughput processing. This system was applied to gDNA samples from 2,000 RA patients and 2,000 controls from BioBank Japan. Each sample was processed in four technical replicates, resulting in a total of 16,000 libraries. In total, we identified 49,089,481 unique TCR clonotypes, of which 13.4% were observed in multiple donors. As expected, the number of distinct clonotypes was lower in RA patients than in controls (on average, 14,388 vs 19,413 clonotypes; Wilcoxon test $P = 6.1E-107$), suggesting greater clonal expansion in RA. Intriguingly, RA patients were significantly more likely to share clonotypes with other RA patients than with controls (Wilcoxon test $P = 8.8E-69$), and we identified multiple TCR clonotypes significantly enriched in RA patients. In line with these results, clustering analysis based on sharedness TCR clonotypes revealed distinct donor groups, primarily driven by HLA genotypes and RA status. Together, our study establishes a scalable strategy to generate population-level TCR repertoire data from gDNA, enabling TCR repertoire-wide association studies. The RA-enriched TCR clonotypes identified in this study may serve as diagnostic markers of RA and hold potential clinical value.

Post-GWAS analysis in autoimmune-inflammatory diseases: strength and patterns of allele-specific DNA (ASM) vary between disease-relevant cell types

Subsession Time: Thursday, October 16 at 11:30am – 11:45am

Authors: Angelica Castano (Hackensack Meridian Health Center for Discovery & Innovation: HMH-CDI), Carolina Alberca-Doto (HMH-CDI), Tatiana Karakasheva (Children's Hospital of Philadelphia: CHOP), Martha Salas (HMH-CDI), Mariam Tariq (HMH-CDI), Sophia Hartley (HMH-CDI), Isabella V. Peterson (CHOP), Veerin Sirihorachai (HMH-CDI), Amanda Muir (CHOP), Arunjot Singh (CHOP), Kevin Tong (HMH-CDI), Mauricio Rigo (HMH-CDI), Andy Madrid (HMH-CDI), Kathryn Hamilton (CHOP), Benjamin Tycko (HMH-CDI)

Abstract: Background: Mapping of haplotype-dependent allele specific DNA methylation (ASM) can help pinpoint regulatory SNPs (rSNPs) under GWAS peaks. Some ASM differentially methylated regions (DMRs) show strong asymmetry in CpG methylation between alleles in multiple cell types, but most are tissue specific, with ASM strength depending on the cell type. Defining ASM strength in human cell types that are disease targets can give insights to which cell types are able to “read” a given risk allele.

Methods: Starting with a subset of the ASM DMRs that we had identified under GWAS peaks for autoimmune-inflammatory diseases in a previous genome-wide study, we quantitated ASM strength

using targeted bisulfite sequencing applied to purified cell types from multiple individuals. Cell types analyzed were total peripheral blood mononuclear cells (PBMC), CD3+ T cells, CD19+ B cells, CD14+ monocytes, and importantly intestinal epithelial cells (IECs) from patient biopsies, expanded as organoids.

Results: We measured the strength of ASM in 7 DMRs, each tagged by candidate functional rSNPs: rs10058821, rs10051804 in IRGM (Crohn's disease); rs10751776 in RUNX3 (inflammatory bowel disease (IBD) and celiac disease); rs2145623 in PSMA6-NFKBIA (IBD, psoriasis and others); rs9535274 in RCBTB1-ARL11 (multiple sclerosis and systemic lupus erythematosus); rs35615354 in CDC37-PDE4A (IBD and ankylosing spondylitis); rs13406632, rs3116499 in CD28-CTLA4 (biliary cholangitis and rheumatoid arthritis); and rs2542148, rs2847274 in PSMG2-PTPN2 (autoimmune disease pleiotropy). In data from an average of 15 heterozygous individuals per cell type per locus, we observed strong cell type-dependence of ASM strength: the IRGM, PSMA6-NFKBIA, RCBTB1-ARL11, CD28-CTLA4 and CDC37-PDE4A DMRs show strong ASM in immune system cell types but weak or absent ASM in IECs. In contrast, RUNX3 shows strong ASM in IECs with weaker though still substantial ASM in immune cells, while the PSMG2-PTPN2 DMR has strong ASM in both immune cells and IECs.

Conclusions: Measurements of ASM strength across disease target cells can generate new functional genomic hypotheses. Notably, the strong ASM in the RUNX3 DMR in IECs suggests that the rSNP in this region can be "read" not only by immune cells but also by epithelial cells, while the strong ASM seen in the IRGM DMR in immune cells but not in IECs suggests that while this gene can function in both cell types, the rSNP in the ASM DMR is not read by IECs.

Session 51: Resolving Complex Genomic Variation in Rare Disease

Location: Room 210AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 10:45am – 11:45am

Comprehensive characterization and analysis of nuclear mitochondrial DNA segments (NUMTs) from long-read sequencing data

Subsession Time: Thursday, October 16 at 10:45am – 11:00am

Authors: Zechen Chong (University of Alabama at Birmingham), Chia Hsuan Fan (University of Alabama at Birmingham), Kaili Hu (University of Alabama at Birmingham), Yuwei Song (University of Alabama at Birmingham), Zishan Peng (University of Alabama at Birmingham)

Abstract: Nuclear mitochondrial DNA segments (NUMTs), resulting from the insertion of mitochondrial DNA (mtDNA) into the nuclear genome, are a significant source of genomic variation. However, their repetitive and often chimeric nature has made comprehensive detection, especially of long insertions, challenging with short-read sequencing. To overcome this limitation, we developed fnumt, a Python-based computational tool optimized for discovering NUMTs ranging from approximately 20 bp to full-length mitochondrial DNA (~16.5 kb), using PacBio HiFi and Oxford Nanopore Technologies (ONT) long-read sequencing data. The fnumt pipeline clusters reads, aligns them to a mitochondrial reference, generates consensus sequences, and annotates and outputs NUMTs in VCF format. Applying fnumt to 1,365 human genomes sequenced with long-read technologies, we identified 14,926 NUMT insertions across 538 unique genomic loci, with 386 (71.75%) previously unreported. Individuals harbored an average of 10.93 NUMTs, including 3.48 potential de novo insertions. Notably, 3.65% of detected NUMTs exceeded 1 kb in length. NUMTs showed significant enrichment near repetitive elements, particularly LINEs (22.86%) and SINEs (15.24%). Chromosome 11 had recurrent NUMT insertions near OR51D1 in more than 85% of individuals, whereas chromosome 2 contained the most unique NUMT loci. Analysis of flanking sequences identified non-random trinucleotide motifs, indicating preferential genomic contexts for insertion. Evolutionary analysis revealed both recent and ancient insertion events, some dating back as far as 6.4 million years. Functional enrichment analyses linked NUMTs to essential biological processes, including nervous system development, immune regulation, and skeletal morphogenesis. Pathway analyses (KEGG and GO) further implicated NUMTs in mitochondrial dysfunction-associated pathways, such as N-glycan biosynthesis and neurodevelopment. Additionally, cis-eQTL mapping suggested potential regulatory roles for NUMTs, notably influencing immune-related gene expression. fnumt provides a robust and scalable solution for high-resolution NUMT discovery, revealing a more complex landscape of mitochondrial insertions in the human genome than previously appreciated. Our findings underscore the significant contribution of NUMTs to human genetic diversity and suggest their involvement in gene regulation, health, and disease, paving the way for deeper investigations into their functional impact.

Identification of de novo variants from parent-proband duos via long-read sequencing

Subsession Time: Thursday, October 16 at 11:00am – 11:15am

Authors: Leandros Boukas (Harvard Medical School and Boston Children's Hospital, Division of Genetics and Genomics & UCI-GREGoR), Emmanuèle C. Délot (University of California, Irvine, Institute of Clinical and Translational Science & UCI-GREGoR), Georgia Pitsava (University of California, Irvine, Institute of Clinical and Translational Science & UCI-GREGoR), Christine Lambert (PacBio), Cairbre Fanslow (PacBio), Primo Baybayan (PacBio), Sami Belhadj (Ambry Genetics), Bojan Losic (Ambry Genetics), John Harting (Ambry Genetics), Krista Bluske (Ambry Genetics), Jonathan LoTempio (University of California, Irvine & UCI-GREGoR), Huda B. Al-Kouatly (Thomas Jefferson University, Department of Obstetrics/Gynecology), Rachid Karam (Ambry Genetics), William J. Rowell (PacBio), Changrui Xiao (University of California, Irvine, Department of Neurology & UCI-GREGoR), Eric Vilain (University of California, Irvine, Institute of Clinical and Translational Science & UCI-GREGoR) & Seth I. Berger (Ambry Genetics & UCI-GREGoR)

Abstract: De novo status can often upgrade a variant from VUS to pathogenic. However, asserting de novo status currently requires sequencing of the proband and both biological parents. As a result, millions of single-parent families are less likely to receive a genetic diagnosis and benefit from precision treatment and management options.

To address this limitation, we developed duoNovo, which identifies de novo variants from parent-proband duos, i.e. using only one biological parent. duoNovo leverages long-read sequencing followed by haplotype reconstruction and detection of identical-by-descent (IBD) haplotype blocks. duoNovo can determine whether a candidate variant of interest is de novo without knowledge of the missing parent's genotype, by testing whether the variant is present on an IBD haplotype shared by the proband and the sequenced parent.

To evaluate duoNovo, we sequenced 40 trios with the PacBio HiFi technology. To the best of our knowledge, this is one of the largest trio cohorts sequenced with long-read sequencing to date. We applied duoNovo to each of the 80 duos constructed by masking one parent, classifying over 20 million single-nucleotide variants/indels. We assessed duoNovo's performance against ground truth classifications obtained from the full trios (which included over 1900 de novo variants), demonstrating very high precision (average 95%) and low error rate (average 1.15×10^{-6}). Notably, duoNovo had perfect accuracy (100% positive predictive value) when tested on variants absent from gnomAD, which are enriched for rare pathogenic alleles and are thus the most likely to be clinically relevant.

A practical issue is that most single-parent families are single-mother families, while most de novo variants are transmitted from the paternal germline. To address this, we show that unaffected siblings can serve as surrogates for the missing father (since they share 50% of their genome with the father), allowing duoNovo to detect a significant fraction of paternally transmitted de novo variants.

In summary, duoNovo has the potential to transform the diagnostic yield of single-parent genetic testing, and represents an example where long-read sequencing provides clear benefit over short-read

sequencing even for single nucleotide variants. We have implemented duoNovo as an R package, aiming to facilitate its use and adoption by the community.

Improved Transcriptome Diagnostics in the Undiagnosed Diseases Network Using the Telomere-to-Telomere CHM13 Genome

Subsession Time: Thursday, October 16 at 11:15am – 11:30am

Authors: Shamika Ketkar (Molecular and Human Genetics, Baylor College of Medicine), Jill A Mokry (Molecular and Human Genetics, Baylor College of Medicine), Undiagnosed Disease Network, Brendan H Lee (Molecular and Human Genetics, Baylor College of Medicine)

Abstract: Background: The Undiagnosed Diseases Network (UDN) employs comprehensive genomic and transcriptomic strategies to uncover the molecular basis of rare and complex conditions. While genome and exome sequencing resolve some cases, many remain undiagnosed. RNA sequencing (RNA-seq) offers critical insights into gene expression and splicing—especially when DNA testing is inconclusive—but is limited by the quality of the reference genome. The CHM13 telomere-to-telomere assembly (v2.0) improves upon hg19 and hg38, providing more complete coverage and enabling detection of previously inaccessible transcriptomic alterations.

Methods: We analyzed RNA-seq data from 182 individuals (115 probands, 67 relatives) enrolled at the Baylor UDN site. RNA extracted from whole blood and skin fibroblasts was aligned to both hg19 and CHM13. Outliers in gene expression and splicing were identified using the DROP pipeline, which includes OUTRIDER and FRASER. CHM13-specific outliers were annotated using OMIM, loss-of-function intolerance scores, GTEx tissue expression, and pathway data. We also used SpliceAI to assess splicing disruption potential of associated variants.

Results: Alignment to CHM13 revealed 31 unique transcriptomic outlier regions not detectable using hg19, spanning approximately 2.39 megabases or 0.078% of the CHM13 genome. The genes affected are highly relevant to rare disease mechanisms commonly seen in UDN patients, including pathways related to neurodevelopment and neuromuscular function (e.g., ARHGEF10, ARHGEF12, KDM5C, ZNF592), immune dysregulation and inflammatory signaling (e.g., DPP4, DOCK5, UBE4A), splicing and RNA processing (e.g., CWF19L2, RBMS2), and transporter-mediated detoxification (e.g., ABCC1, ABCC11). Precomputed SpliceAI predictions were available for CWF19L2 and ARHGEF10, providing additional support for splicing-disruptive potential in CHM13-unique genes. The majority of candidate loci lacked coverage in existing SpliceAI reference sets, highlighting the need for reannotation using CHM13 coordinates.

Conclusions: Transitioning RNA-seq alignment from hg19 to CHM13 in the UDN cohort uncovered additional clinically relevant outliers, particularly in regions previously unmappable or poorly annotated. These findings support the use of the CHM13 genome to improve transcript-level diagnostic resolution in rare disease investigations and highlight the importance of reference genome selection in clinical genomics.

Characterization of Rare Genomic Structural Variants Across 2,981 Genomes Reveals Significant Involvements in Recessive Conditions

Subsession Time: Thursday, October 16 at 11:30am – 11:45am

Authors: Zirui Dong (The Chinese University of Hong Kong), Keying Li (The Chinese University of Hong Kong), Chi Chun Chan (The Chinese University of Hong Kong), Jia Zheng (The Chinese University of Hong Kong), Mengmeng Shi (The Chinese University of Hong Kong), Wenbin He (Reproductive and Genetic Hospital of CITIC-Xiangya), Matthew Hoi Kin Chau (The Chinese University of Hong Kong), Ye Cao (The Chinese University of Hong Kong), Juan Du (Reproductive and Genetic Hospital of CITIC-Xiangya), Yvonne Ka Yin Kwok (The Chinese University of Hong Kong), Yue-qiu Tan (Reproductive and Genetic Hospital of CITIC-Xiangya), and Kwong Wai Choy (The Chinese University of Hong Kong)

Abstract: Background: Genomic structural variants (SV) underlie genomic diversity and diseases, encompassing various types of deletions/duplications and genomic rearrangements. Current methods experience challenges in precisely identifying and determining SV composition which often lead to incorrect interpretation and/or missing detection of clinically significant variants. Here, we introduce SCARD (Split-read, Chimeric read-pAirs, and Read-depth Difference), an integrated approach optimized for detection and characterization of rare SVs using genome sequencing, applied to investigate the SV spectrum across 2,504 individuals and 477 trios from the 1000 Genomes Project.

Methods: We evaluated the rare SV (with minor allele frequency <1%) detection performance using 16 clinical cases with 18 previously ascertained exonic deletions/duplications. In addition, SCARD characterized the spectrum, inheritance rate, and clinical implications of rare SVs identified from 2,981 genomes. We described clinically significant, loss-of-function (LoF) SVs involving genes associated with recessive disorders, defined by OMIM and/or ClinGen Dosage Sensitivity Map. Finally, to evaluate the accuracy of rare SV calling, we compared our SVs in NA12878 to that of long-read sequencing datasets, and those results from 13 SV detection methods (13callers).

Results: SCARD detected all the 18 exonic deletions/duplications, ranging from 490bp to 24.5kb, outperforming other methods that only achieved 15/18 at best. Among 2,504 genomes analyzed, we identified 427,240 rare simple deletions/duplications and 11,981 structural rearrangements, with 10.7% of deletions/duplications and 43.9% of structural rearrangements being novel to that from 13callers. 96.6% (8,755/9,064) of insertions and 84.1% (1,488/1,769) of inversions were smaller than 100kb. Notably, 50.92% (1,275/2,504) of the subjects carried LoF SVs involving recessive disorders related genes. In 477 trios, the de novo rate for rare SVs was 0.5% (400/77,541). In NA12878, we identified 140 rare simple deletions/duplications ranging from 227bp to 873kb, and four insertions and four inversions, all validated by long-read sequencing datasets, and 18.9% of them were missed detection or misinterpreted by 13callers.

Conclusions: Our study demonstrates that SCARD enables comprehensive, genome-wide identification and precise delineation of rare SVs, particularly for those affecting exons, to reveal a high risk of rare SVs involvement underlying diseases.

Session 52: Autism Spectrum Disorder: From Genetic Variation to Neurobiology

Location: Room 156ABC/Level 1, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 1:30pm – 2:30pm

The role of topologically associating domains in defining the pathogenicity of de novo non-coding variants on autism spectrum disorder risk

Subsession Time: Thursday, October 16 at 1:30pm – 1:45pm

Authors: Shota Mizuno (Riken Center for Brain Science), Takumi Nakamura (Riken Center for Brain Science), Junko Ueda (Riken Center for Brain Science), Kurara Honda (Riken Center for Brain Science), An-a Kazuno (Riken Center for Brain Science), Hirona Yamamoto (Riken Center for Brain Science), Tomonori Hara (Riken Center for Brain Science), Atsushi Takata (Riken Center for Brain Science)

Abstract: Whole-genome sequencing (WGS) studies of autism spectrum disorder (ASD) have demonstrated contributions of non-coding de novo variants (DNVs), especially those affecting gene promoters, to the disease risk. However, a large portion of the enrichment of promoter DNVs in ASD cannot be explained by those immediately upstream of known ASD genes. Recently, a phenomenon called “enhancer release and retargeting” has been described, wherein disruption of a promoter leads to the activation of other genes within the same topologically associating domain (TAD) through enhancer redirection. Based on this model, we analyzed DNVs in genomic regulatory elements, such as promoters/enhancers, within TADs in the human dorsolateral prefrontal cortex using two large-scale ASD WGS datasets from the Simons Simplex Collection and the Simons Foundation Powering Autism Research for Knowledge (5,044 ASD probands, 4,095 unaffected siblings, and their parents). By meta-analyzing the two datasets totaling 19,227 individuals, we found that promoter DNVs located in TADs containing known ASD genes (as defined by the SFARI Gene database), but not immediately downstream of the DNVs, were significantly associated with ASD ($P = 0.00292$, $OR = 1.096$). No such enrichment was observed for promoter DNVs in TADs lacking ASD-associated genes. This finding was replicated in brain- and stem-cell-specific TAD maps, but not in TADs defined from peripheral tissues. A genome-wide search for TADs enriched for promoter DNVs in ASD identified two TADs that surpassed Bonferroni-corrected significance thresholds. Promoters within nominally significant TADs (uncorrected $P < 0.05$), but not those in non-significant TADs, were significantly enriched for ASD common variant heritability, supporting the idea that both rare and common variants in these regions contribute to ASD genetic liability. Functional validation using human induced pluripotent stem cells (hiPSC) carrying ASD TAD promoter DNVs revealed that even single promoter DNVs could affect the expression of multiple genes within the same TAD. This led to widespread dysregulation of genes involved in ASD and neurodevelopment. Collectively, our results demonstrate that incorporating TAD-based gene regulatory architecture enhances our understanding of the role of rare non-coding variants in ASD. Furthermore,

this framework may be applicable to other types of rare non-coding variants beyond promoter DNVs. In our presentation, we also plan to share preliminary data that supports this broader hypothesis.

Rare coding variants identify novel genes and phenotypic context of autism

Subsession Time: Thursday, October 16 at 1:45pm – 2:00pm

Authors: Lu Qiao (Columbia University Irving Medical Center, Department of Pediatrics), Pamela Feliciano (Boston Children's Hospital, Department of Pediatrics), Jacob Hagen (Boston Children's Hospital, Department of Pediatrics), Chao Hou (Columbia University Irving Medical Center, Department of Systems Biology), Xiaoqi Lin (Columbia University Irving Medical Center, Department of Systems Biology), Sarah D. Barns (Boston Children's Hospital, Department of Pediatrics), Alexandra Marsden (Medical University of South Carolina, Development-Behavioral Pediatrics), Angela L. Rachubinski (University of Colorado Anschutz Medical Campus, Department of Pediatrics), Ashley Andrade (Rush University Medical Center, Department of Psychiatry and Behavioral Sciences), Ashley Quinones (Children's Hospital of Philadelphia, Philadelphia, Department of Child and Adolescent Psychiatry and Behavioral Sciences), Ayana Davis (University of Mississippi Medical Center, Department of Pediatrics), Camilla Strathearn (University of Iowa Health Care Medical Center, Department of Psychiatry), Catherine C. Bradley (Medical University of South Carolina, Development-Behavioral Pediatrics), Cheryl Klaiman (Emory University, Department of Pediatrics), Cordelia Robinson Rosenberg (JFK Partners, University of Colorado Anschutz Medical Campus, Department of Pediatrics), Emily Neuhaus (University of Washington, Department of Psychiatry & Behavioral Sciences), Ericka L. Wodka (Johns Hopkins School of Medicine, Kennedy Krieger Institute), Gabrielle Banks (University of Mississippi Medical Center, Department of Pediatrics), Gabriel S. Dichter (UNC-Chapel Hill School of Medicine, Department of Psychiatry and Carolina Institute for Developmental Disabilities), Hallie Brown (Weill Cornell Medicine, Department of Psychiatry), Jacob Michaelson (University of Iowa Health Care Medical Center, Department of Psychiatry), Jaylaan Parker (Vanderbilt University Medical Center, Department of Pediatrics), Kelly Real (UNC-Chapel Hill School of Medicine, Chapel Hill, Carolina Institute for Developmental Disabilities), Marilyn Featherston (Rush University Medical Center, Chicago, Department of Pediatrics), Marilyn Lopez (Children's Specialized Hospital, Autism Department), Opal Ousley (Emory University, Department of Psychiatry), Zachary Warren (Vanderbilt University Medical Center, Department of Pediatrics), Yufeng Shen (Columbia University Irving Medical Center, Department of Systems Biology), Wendy K. Chung (Boston Children's Hospital, Department of Pediatrics)

Abstract: Autism is a highly heritable complex condition often associated with other neurodevelopmental traits, but the role of rare-particularly inherited- variants remains unclear due to limited power and the uncertain impact of most missense variants. In this study, we analyzed exome sequencing data from 65,856 autistic individuals and their close family members (n=181,276 total) from SPARK study and other autism cohorts.

We used new methods (MisFit and GeneBayes) for estimating the selection coefficient of missense and loss of function (LoF) variants. We assessed rare variant burden by selection coefficients. The rate of over-transmission to autistic offspring is highly correlated with the estimated selection coefficient for

both rare missense and LoF variants. Notably, we observed a significantly higher burden of de novo and over-transmitted inherited LoF and missense variants under strong selection in autistic individuals, mainly in simplex families. Inherited damaging missense variants have similar population attributable risk (PAR) to inherited LoF (6.8% vs 7.3%), with 88% of inherited PAR in unknown susceptibility genes. Rare variants contribute more in simplex than multiplex families.

We applied a Bayesian statistical framework for autism gene discovery that accounted for variant type, mode of inheritance and estimated selection coefficient. We identified 273 genes associated with autism at a false discovery rate (FDR) 0.01. Among these 273 genes, a higher proportion of the 170 novel genes were associated exclusively through missense variants compared to the 103 known genes (50% vs 32%, $P=3.9 \times 10^{-3}$). The novel genes driven solely by missense variants were enriched in the VEGFA-VEGFR2 signaling pathway (odds ratio=7.7, $P=5.1 \times 10^{-5}$). Inheritance patterns show opposite PAR trends in novel genes (0.45% de novo vs 1.0% inherited) to known genes (3.0% vs 0.65%). Some novel genes exhibit domain-specific variant enrichment. For example, SORCS1 and SORCS3 are associated largely by variants in the specific domains such as the protein-folding stability of the β -propeller domain in these two related genes ($P<1.9 \times 10^{-3}$). Carriers of rare deleterious variants in autism susceptibility genes were more likely to exhibit cognitive impairment, attention deficit hyperactivity disorder and seizures, but not bipolar, anxiety or depression.

These results provide new insights into the role of rare variants in autism in other neurobehavioral phenotypes, largely in simplex families.

Shared and distinct genetic architectures of autism and neuropsychiatric disorders

Subsession Time: Thursday, October 16 at 2:00pm – 2:15pm

Authors: Jack M Fu (CGM, MPG, MGH), F. Kyle Satterstrom (MPG, ATGU, Stanley Center), Kirsty McWalter (GeneDx), Harrison Brand (CGM, MGH), Robert Kueffner (GeneDx), Michelle Morrow (GeneDx), Eren Shin (CGM), Justin Lim (CGM), Caroline Cusick (Stanley Center), Cal Liao (CGM, MPG, ATGU), Lily Wang (HMS), David J. Cutler (Emory), Kaitlin Samocha (CGM), Elise Robinson (CGM, ATGU, HSPH), Joseph D. Buxbaum (MSSM), Bernie Devlin (University of Pittsburgh), Kathryn Roeder (Carnegie Mellon), Paul Kruszka (GeneDx, University of Virginia), Stephan J. Sanders (University of Oxford, UCSF), Mark J. Daly (MPG, ATGU, Stanley Center), Michael E. Talkowski (CGM, MPG, MGH, Stanley Center)

Abstract: Discovery of genetic factors underlying autism and related neurodevelopmental and neuropsychiatric conditions (NPDs) has progressed rapidly, driven by data aggregation and analytic innovations. Here, we report an analysis of 62,429 affected individuals (38,680 with parental data). By integrating de novo and inherited single-nucleotide variants, indels, and copy number variants via a Bayesian model, we identify 257 genes associated with autism at a false discovery rate (FDR) < 0.001 - a threshold calibrated to be equivalent to exome-wide significance. These genes recapitulate strong enrichment in pathways involved in chromatin remodeling, development, and synaptic communication.

To contextualize these findings in the broader landscape of NPDs, we aggregated our results with large-scale studies of schizophrenia (SCHEMA; 24,248 cases, 97,322 controls), epilepsy (Epi25K; 20,979 cases, 33,444 controls), and bipolar disorder (BD) (BIPEX; 13,933 cases, 14,422 controls). Applying burden heritability regression (HBR) to rare variants, we observed moderate genetic correlation between autism and schizophrenia, epilepsy, and BD (~ 0.2); while schizophrenia and epilepsy exhibited stronger correlation (~ 0.5). BHR also estimates autism to have the greatest rare variant heritability (3.04%), followed by schizophrenia (1.67%) and epilepsy (1.3%). Additionally, 25% of rare variant heritability in autism resides in the 257 autism-associated genes, while the same genes account for $\sim 20\%$ of rare variant heritability in epilepsy, but less than 10% in schizophrenia and BD, suggesting potential shared genetic mechanisms between autism and epilepsy. The autism associated genes are significantly more likely to also be associated with schizophrenia (odds ratio [OR]=10.8, $p=1.5e-14$) and epilepsy (OR=11.6, $p=1.1e-11$), with weaker enrichment in BD (OR=2.8, $p=0.094$). Finally, autism genes in chromatin, development, and synaptic signaling pathways are significantly enriched for genes associated with schizophrenia (OR=1.76, $p=7.6e-2$) and epilepsy (OR=2.22, $p=3.2e-4$).

Our findings highlight a core set of highly penetrant genes with impact across autism and NPDs and reveal disorder-specific genetic architecture differences. Additional efforts to integrate these gene-level discoveries with developmental expression patterns, cell-type specificity, and functional networks are ongoing and will further inform functional mechanisms for rare variant risk across diagnostic boundaries.

Chd8-dependent molecular and functional vulnerabilities in the developing and adult mouse cerebellum

Subsession Time: Thursday, October 16 at 2:15pm – 2:30pm

Authors: Cesar P. Canales (UC Davis Center for Neuroscience), Nicolas Seban (UC Davis Center for Neuroscience), Esteban Merino (UC Davis Center for Neuroscience), Se Jung Jung (UC Davis Center for Neuroscience), Karol Cichewicz (UC Davis Center for Neuroscience), Stephanie Lozano (UC Davis Center for Neuroscience), Darlene Rahbarian (UC Davis Center for Neuroscience), Melissa Correa (UC Davis Center for Neuroscience), Diasynou Fioravante (UC Davis Center for Neuroscience), Alex S. Nord (UC Davis Center for Neuroscience).

Abstract: The cerebellum (CB) is increasingly implicated in higher-order cognitive processes and autism spectrum disorder (ASD) pathogenesis, yet it remains underexplored relative to the forebrain. CHD8, a chromatin remodeler highly associated with ASD, is expressed broadly during development and remains elevated in the adult cerebellum. Chd8 haploinsufficiency in mice is a robust model for ASD-related mechanisms, though most studies have focused on cortical disruption. Here, we explore Chd8-dependent cerebellar pathophysiology across developmental and adult timepoints using complementary genomic, anatomical, functional, and behavioral approaches. To define developmental alterations, we performed single-nucleus RNA-seq and ATAC-seq on cerebellum from Chd8^{+/-} and wild-type (WT) mice at postnatal day 12. We found no gross anatomical differences by immunohistochemistry, but identified significant cell type-specific changes in gene expression and chromatin accessibility. Developmental trajectory analyses indicated transcriptional asynchrony in Purkinje neurons and molecular layer

interneurons, implicating disrupted circuit formation. Functional consequences were supported by patch-clamp recordings and ongoing RNA-FISH validation. To assess adult cerebellar function, we analyzed bulk RNA-seq from adult *Chd8*^{+/-} CB and identified differential expression in synaptic and metabolic genes. Electrophysiological recordings from deep cerebellar nuclei showed elevated firing rates, consistent with hyperexcitability and increased CB output. Conditional deletion of *Chd8* in the adult CB revealed impairments in social novelty preference and regulation of learned fear, highlighting persistent functional vulnerabilities beyond development. Together, these results demonstrate that *Chd8* haploinsufficiency disrupts cerebellar development and adult function through distinct yet converging molecular mechanisms. This suggests that cerebellar circuits remain sensitive to *Chd8* dosage across the lifespan. Our work highlights the CB as a critical and sustained site of vulnerability in CHD8-linked ASD and may offer a broader therapeutic window than previously appreciated.

Session 53: Every Story Has a Beginning and an End: About UTRs and Poly(A) Tails

Location: Room 210AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 1:30pm – 2:30pm

Genome-wide detection of human 5'UTR variants of functional significance in translation

Subsession Time: Thursday, October 16 at 1:30pm – 1:45pm

Authors: Matthieu Chaldebas (The Rockefeller University, Imagine Institute, INSERM), Khoren Ponsin (The Rockefeller University), Jonathan Bohlen (Imagine Institute, INSERM, Ludwig-Maximilians-Universität, German Center for Child and Adolescent Health), Clement Conil (Imagine Institute, INSERM), Peter D. Stenson (Cardiff University), David N. Cooper (Cardiff University), Laurent Abel (The Rockefeller University, Imagine Institute, INSERM), Jean-Laurent Casanova (The Rockefeller University, Imagine Institute, INSERM, Necker Hospital for Sick Children, Howard Hughes Medical Institute), Aurélie Cobat (The Rockefeller University, Imagine Institute, INSERM), Peng Zhang (The Rockefeller University, Imagine Institute, INSERM)

Abstract: The 5' untranslated region (5'UTR) of messenger RNAs (mRNAs) plays a key role in regulating the initiation of protein synthesis, particularly via the Kozak sequence and upstream open reading frames (uORFs). Genetic variants within these regions may affect translation, thereby influencing gene expression and contributing to clinical phenotypes in humans. We developed 5ULTRA (5' UntransLaTed Region Annotation), a computational method for analyzing whole-exome sequencing (WES) and whole-genome sequencing (WGS) data to detect, annotate, and prioritize 5'UTR variants potentially affecting mRNA translation. 5ULTRA specifically identifies single-nucleotide variants (SNVs), indels and splicing variants predicted to affect uORFs by creating or disrupting start/stop codons. It also assesses variants altering Kozak sequence strength for either uORFs or the main coding sequence. 5ULTRA incorporates recent uORF databases and provides relevant user-friendly annotations. 5ULTRA uses a machine-learning score to prioritize candidates that not only has a better predictive performance than CADD for relevant ClinVar variants, but also provides specific mechanistic predictions. The score is strongly correlated with experimentally measured effects of the variant on protein levels, as validated with protein quantitative trait locus (pQTL) data. We applied 5ULTRA to diverse datasets to identify candidate variants in different disease contexts: potential cancer-driving somatic mutations (COSMIC data) predicted to decrease ABI1 protein levels or to alter or increase the levels of NRAS protein; common variants (GWAS Catalog) linked to traits such as multiple sclerosis or lung functions, for which 5ULTRA predicts altered TAGAP or VRTN protein production, respectively; and rare variants in our patient cohort, including one predicted to reduce TNF levels and to predispose to tuberculosis. 5ULTRA is publicly available from <https://hgidsoft.rockefeller.edu/5ULTRA> and <https://github.com/mchaldebas/5ULTRA>.

MicroRNA Perturb-Seq Reveals Genome-wide Functional Targets and Deleterious 3'UTR Variants

Subsession Time: Thursday, October 16 at 1:45pm – 2:00pm

Authors: Eyal Ben-David, Doyeon Kim, Wayne Xianding Deng, Zakaria Louadi, Robin Bombardi, Marcos Nascimento, Jacob Ulirsch, Thy Pham, William J. Greenleaf, Kyle Kai-How Farh

Abstract: MicroRNAs (miRNAs) are essential regulators of post-transcriptional gene expression, yet our understanding of their targets and mechanisms remains incomplete. To obtain a comprehensive catalog of miRNA-transcript target pairs, we performed miRNA perturb-seq with a library of ~650 conserved human miRNAs, profiling ~10 million single cells using Illumina Single Cell 3' RNA Prep, and used nine diverse cell lines to ensure that the vast majority of the human transcriptome was expressed in at least one cell line.

Our design enabled robust coverage, with an average of over 1,000 cells per perturbation, allowing precise quantification of gene expression changes. This approach identified approximately 760,000 significantly downregulated miRNA-target transcript pairs, with targeting specificity determinants that were consistent with established miRNA targeting rules.

Furthermore, we leveraged this dataset to uncover rare 3'UTR variants that disrupt miRNA-mediated gene regulation, linking non-coding genetic variation to functional consequences in human populations. Using the UK Biobank plasma proteome dataset, we validated that variants affecting miRNA target sites lead to measurable changes in protein expression levels. Additionally, analysis of gnomAD dataset revealed that these target site-disrupting variants are under negative selection, highlighting their functional importance. Together, our study provides a comprehensive and high-resolution resource for deciphering functional miRNA targets and understanding the consequences of non-coding 3' UTR genetic variation.

Modulating gene expression in human cells via high-throughput prime editing of regulatory elements

Subsession Time: Thursday, October 16 at 2:00pm – 2:15pm

Authors: Magdalena Armas Reyes (The Francis Crick Institute), Laura Cubitt (The Francis Crick Institute), Michael Herger (The Francis Crick Institute), Gregory Findlay (The Francis Crick Institute).

Abstract: Modulating human gene expression by editing regulatory regions of the genome affords greater flexibility and precision to cell-based therapies. However, edits that induce desired effects on gene expression remain challenging to predict. Here, we present PETRA (Prime Editing of Transcribed Regulatory elements to Alter expression), a multiplexed method that leverages prime editing to quantify the effects of regulatory variants on gene expression.

We demonstrated PETRA in Jurkat cells by scoring 13,935 6-nucleotide insertions in the 5' untranslated regions (UTRs) of 4 genes relevant to T cell function: VAV1, IL2RA, CD28 and OTUD7B. Expression scores

for variants were highly correlated across replicates and validated when tested independently. We identified 57 variants that lead to a significant increase in expression of over 2-fold and 147 variants that significantly reduced expression to below 50% of baseline. Correlation of expression scores to ENFORMER CAGE and DNase predictions were significant for some but not all loci tested.

To investigate mechanisms, we analysed specific sequence features associated with changes in expression. This revealed negative effects on expression due to cryptic splice sites and positive effects attributable to the creation of upstream open reading frames. Analysis of de novo transcription factor binding sites (TFBSs) revealed 108 and 88 factors associated with increased or decreased gene expression, respectively. For instance, MYBL2 binding sites were significantly associated with increased expression of IL2RA. Indeed, combinatorial insertion of MYBL2 binding sites to 3 positions in the IL2RA 5'UTR revealed multiplicative effects leading to more than a 10-fold increase in expression. Conversely, insertion of multiple CTCF motifs decreased expression of IL2RA and VAV1 to less than 2%.

To establish PETRA's utility in primary cells, we next tested 3,279 VAV1 and IL2RA variants in human CD3+ T cells. Expression scores were significantly correlated between Jurkat and primary CD3+ cells (Spearman's $r=0.80$, VAV1) and spanned two orders of magnitude.

In summary, PETRA is a highly scalable method for identifying sequences that can be engineered in human cells to increase or decrease gene expression. We anticipate this approach will reveal ways in which the human genome can be edited for therapeutic purposes while also enabling the generation of large datasets to improve predictive models of gene regulation.

Atlas of cell type-specific genetic impacts of alternative polyadenylation in immune-related diseases

Subsession Time: Thursday, October 16 at 2:15pm – 2:30pm

Authors: Ting Zhang (Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen, China), Hui Chen (Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen, China), Shuxin Chen (Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen, China), Stephen Montgomery (Departments of Pathology and Genetics, Stanford School of Medicine, Stanford, USA), Qin Li (Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA), Lei Li (Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen, China)

Abstract: Background: Genetic variants associated with immune disease risk often act through complex regulatory mechanisms that vary across cell types and contexts. Alternative polyadenylation (APA), a key post-transcriptional process that shapes mRNA stability and translation, has emerged as an important mechanism in disease. However, its cell type-specific genetic regulation remains largely unexplored.

Methods: Using single-cell RNA-sequencing of over 7.27 million peripheral blood mononuclear cells from 2,022 individuals, we developed scGEAR to comprehensively analyze APA phenotypes in population-scale scRNA-seq data and mapped genetic variants affecting cell type-specific APA across 31 immune cell types and diverse stimulation conditions. Colocalization and transcriptome-wide association studies (TWAS)

leveraging genome-wide association study (GWAS) data of 16 immune-related disease were performed to pinpoint shared causal variants and susceptibility genes, while multi-trait colocalization analyses were conducted using protein quantitative trait loci (pQTL) data for APA-related therapeutic target mining.

Results: We identified 10,211 single-cell APA quantitative trait loci (sc-aQTLs) linking 4,484 independent variants to APA changes of 5,448 nearby genes across diverse immune cell types. Notably, APA regulation of the autoimmune disease gene STAT6 operates independently of gene expression changes, highlighting distinct regulatory pathways. Context-dependent aQTL analyses revealed that immune stimuli, rather than sex and ancestry, primarily drive disease regulatory effects. Strikingly, sc-aQTLs exhibited even stronger heritability enrichment for immune-related diseases than expression QTLs. Through colocalization and TWAS, we identified 267 APA-linked putatively causal disease genes, with 80.52% acting through expression-independent regulatory pathways. Further integration of pQTLs reveals 211 previously uncharacterized protein-level regulatory variants and 12 disease causal genes, offering potential therapeutic targets.

Conclusion: Our comprehensive sc-aQTLs atlas transforms our understanding of regulatory mechanisms underlying immune-related diseases by revealing the critical and previously underappreciated role of APA regulation in immune disease susceptibility and immune cells.

Session 54: Function of Complex Variation and Repeats

Location: Room 253ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 1:30pm – 2:30pm

A phenome-wide association study of CNVs genotyped from genome sequencing data in the UK Biobank

Subsession Time: Thursday, October 16 at 1:30pm – 1:45pm

Authors: Andrew Sharp (Icahn School of Medicine at Mount Sinai), Paras Garg (Icahn School of Medicine at Mount Sinai), Bharati Jadhav (Icahn School of Medicine at Mount Sinai), Mariya Shadrina (Icahn School of Medicine at Mount Sinai), Alejandro Martin-Trujillo (Icahn School of Medicine at Mount Sinai)

Abstract: We developed a novel read depth-based approach that allows accurate and highly scalable genotyping of copy number across the entire allele frequency spectrum from whole genome sequencing data, including mosaic, recurrent and multi-allelic CNVs that are difficult to genotype using other methods. We applied this approach to generate CNV maps for every 5kb segment throughout the entire genome in 405,362 unrelated European individuals from the UK Biobank. Using monozygotic twin-pairs, we estimated our genotyping accuracy at ~99.98%, or one genotyping error every ~55 Mb.

We used these CNV maps to perform phenome-wide association studies using 13,215 traits under three different inheritance models: additive, dosage sensitive and recessive. Applying a stringent Bonferroni correction, we identified 4,477 unique CNV:trait pairs, representing 501 CNVs (482 autosomal and 19 X-linked) associated with 1,537 different traits, including many known and novel associations. We also performed PheWAS in individuals of African and South Asian ancestry, replicating many signals observed in Europeans and also identifying several additional associations.

Examples of novel associations included copy number of AMY1 (salivary amylase, which converts starch to sugars in the mouth) with denture use ($p=2.4 \times 10^{-29}$), a negative association between a large coding repeat within MUC1 (mucin 1, which forms the mucus lining of the gut) and duodenal polyps ($p=7.7 \times 10^{-24}$), and a multi-allelic coding CNV within NEB, a muscle sarcomere protein, associated with measures of muscle mass ($p=9.7 \times 10^{-24}$). We found examples of large regions that show relatively common multi-allelic CNVs where rare extreme copy numbers associated with strong effects on cognitive function, suggesting these may underlie some cases of clinically-diagnosed intellectual disability that have been previously ignored due to their multi-allelic nature.

We also identified several rare intergenic CNVs that showed strong effects on traits known to be regulated by nearby genes, suggesting that these CNVs act via cis-dysregulation of known disease genes. For example, carriers of a rare non-coding deletion (MAF ~0.0002) located ~100kb upstream of MC4R, a gene where coding mutations are known to represent the most common cause of monogenic obesity, were an average of ~31 pounds heavier than controls ($p=9.5 \times 10^{-11}$). In some cases, these non-coding CNVs encompassed annotated regulatory elements of the adjacent candidate gene. Using burden tests,

we identified an excess of rare predicted damaging non-coding SNVs within some of these regulatory elements associated with the same traits observed in the CNV carriers, indicating convergence of mutational mechanisms.

Our study provides a detailed map of functional CNVs, including complex loci that are recalcitrant to other methods, providing many novel insights into their role on human traits.

Identification of genetic modifiers of somatic instability across multiple repeat expansion disease loci using in vivo CRISPR–Cas9 genome editing

Subsession Time: Thursday, October 16 at 1:45pm – 2:00pm

Authors: Ricardo Mouro Pinto (Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard) Maheswaran Kesavan (Center for Genomic Medicine, Massachusetts General Hospital, Department of Molecular Medicine, University Laval) Antonia Vitalo (Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School) Anh Nhu (Center for Genomic Medicine, Massachusetts General Hospital) Muzhou Wu (Center for Genomic Medicine, Massachusetts General Hospital) Steven Crescenti (Center for Genomic Medicine, Massachusetts General Hospital) Cameron Douglas (Center for Genomic Medicine, Massachusetts General Hospital) Jessica Ulloa (Center for Genomic Medicine, Massachusetts General Hospital) Neil Doherty (Center for Genomic Medicine, Massachusetts General Hospital) Jennie Roy (Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School) Shota Shibata (Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard)

Abstract: Somatic expansion of disease-causing repeats is a critical driver of disease pathogenesis in Huntington's disease (HD) and other repeat expansion disorders. Targeting the DNA repair pathways that mediate this process presents a promising therapeutic avenue. To accelerate the discovery of genetic modifiers of repeat instability, we developed a scalable in vivo screening platform. This system utilizes systemic AAV-based delivery of sgRNAs to knockout candidate genes in Cas9-expressing mouse models of repeat expansion diseases.

We first established this platform in an HD mouse model (HttQ111), systematically testing over 60 candidate genes. The screen successfully validated known modifiers from the mismatch repair (MMR) pathway (Msh2, Msh3, Mlh1, Mlh3, among others) and identified novel players, including components of the DNA polymerase delta complex (Pold1, Pold3), providing robust in vivo evidence for their role in driving somatic CAG repeat expansion.

We have now extended this powerful approach to investigate the mechanism of somatic instability at different disease-associated repeat loci. Specifically, we tested a panel of candidate genes in a humanized mouse model of Friedreich ataxia (FA), characterized by a GAA repeat expansion in the FXN gene (YG8s.300 and YG8s.800), and in a new knock-in mouse model (f-ATXN1.146Q) of Spinocerebellar Ataxia Type 1 (SCA1), which is another CAG repeat expansion disorder.

Collectively, our cross-disease modifier screening highlights a core set of genes involved in promoting repeat expansion across different disease-associated loci and repeat motifs. However, we also observe evidence for locus-specific effects, even for the same CAG repeat, suggesting both shared and distinct mechanistic features. This in vivo platform provides a robust framework for rapidly validating candidate modifiers, uncovering disease-specific pathways, and prioritizing shared therapeutic targets for a growing class of neurodegenerative diseases.

Computational analysis of tandem repeats for genome-wide identification of novel repeat expansions

Subsession Time: Thursday, October 16 at 2:00pm – 2:15pm

Authors: Egor Dolzhenko (PacBio), Ben Weisburd (Massachusetts General Hospital, The Broad Institute of MIT and Harvard), Matt Danzi (University of Miami Miller School of Medicine), Adam English (Baylor College of Medicine), Georgi Hudjashov (University of Tartu), Tom Mokveld (PacBio), Guilherme De Sena Brandine (PacBio), Zev Kronenberg (PacBio), Aaron Wenger (PacBio), Hope Tanudisastro (Garvan Institute of Medical Research, Murdoch Children's Research Institute), Caitlin Karniski (PacBio), Evan E Eichler (University of Washington School of Medicine, Howard Hughes Medical Institute), Fritz Sedlazeck (Baylor College of Medicine), Stephan Zuchner (University of Miami Miller School of Medicine), Lili Milani (University of Tartu), Heidi L Rehm (Massachusetts General Hospital, Broad Institute of MIT and Harvard), Michael A Eberle (PacBio)

Abstract: Diseases caused by expansions of short tandem repeats (STRs) are estimated to affect 1 in 3000 people. Robust methods have been developed to detect expansions of over 60 known pathogenic repeats from long-read sequencing data. However, because every genome contains hundreds of long and rare repeats whose pathogenicity is difficult to ascertain, the remaining millions of STRs are not routinely assessed in rare disease studies.

We developed computational methods to enable effective genome-wide STR analysis of both single samples and large datasets. Combined, these methods 1) generate a dataset-specific repeat catalog, 2) genotype the repeats across all samples, 3) score the longest pure repeat segments (LPS) within each repeat allele to prioritize potentially-pathogenic expansions, and 4) perform outlier analysis comparing LPS scores in cases to a panel of controls.

The repeat catalog generated in step one includes only polymorphic repeats with reference coordinates that encompass dataset-wide variation. This reduces the catalog to ~250–500k loci and improves allele resolution in complex regions, where Mendelian accuracy rises from 90.86% to 99.31%. Then the LPS scores are calculated for all repeats using an extension to the tandem repeat analysis tool, TRGT, that we recently developed. The variation of LPS scores was recently shown to prioritize potentially pathogenic repeats in population datasets. Here we also show that known pathogenic expansions are among the longest pure STRs in the genome, suggesting that LPS scores can be used to identify novel pathogenic expansions.

We applied our workflow to 1,027 HiFi whole-genome samples from the All of Us program (controls) and 30 samples with known repeat expansions (cases). The workflow produced a repeat catalog containing 450K polymorphic STRs in this dataset. The LPS score analysis prioritized all intronic pathogenic expansions in the top 5 loci genome-wide. Remarkably, all the exonic pathogenic expansions were assigned top ranks when compared to other exonic STRs, confirming the utility of LPS scores. Additional analysis of LPS scores in the All of Us data identified ~40 novel candidate poly-glutamine and 5'-UTR CGG repeats as potentially pathogenic.

Our results show that profiling LPS scores instead of overall allele lengths is a powerful way to search for candidate pathogenic expansions. To make LPS and other STR properties easy to explore, we created a companion online portal (trexplorer.broadinstitute.org).

Going Viral: Targeted Detection and Profiling of Endogenous Retrovirus Transcripts in the Human Genome

Subsession Time: Thursday, October 16 at 2:15pm – 2:30pm

Authors: Isabelle Heifetz Ament (University of Pennsylvania), Zelin Liu, PhD (Children's Hospital of Philadelphia), Yi Xing, PhD (University of Pennsylvania/Children's Hospital of Philadelphia), Lan Lin, PhD (University of Pennsylvania/Children's Hospital of Philadelphia)

Abstract: Endogenous retroviruses (ERVs), remnants of ancient retroviral integrations, comprise a substantial and largely uncharacterized portion of the human genome. To systematically resolve the complexity of the ERV transcriptome, we developed ERV-seq, a novel long-read RNA sequencing method incorporating targeted probe enrichment to robustly capture ERV-derived transcripts across both intragenic and intergenic loci. Leveraging ultra-deep sequencing and full-length reads, ERV-seq achieves unprecedented locus-specific resolution, enabling the identification of over 100,000 novel transcript isoforms, 3,000 previously unannotated ERV loci, and 50,000 ERV-derived exons within annotated gene bodies—vastly surpassing existing detection methods. In addition to enhancing ERV locus annotation, ERV-seq reveals new layers of ERV biology, including post-transcriptional regulation through alternative splicing, and tissue-specific expression dynamics. By providing the most comprehensive and precise catalog of expressed human ERV loci and transcript isoforms to date, ERV-seq establishes a powerful platform for elucidating the functional contributions of ERVs to development, health, disease, and genome evolution.

Session 55: Gene Editing and Gene Replacement for Rare Disease

Location: Room 206AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 1:30pm – 2:30pm

Corrective editing to treat the connective tissue disorder pseudoxanthoma elasticum

Subsession Time: Thursday, October 16 at 1:30pm – 1:45pm

Authors: Lauren Testa (University of Pennsylvania), Hooda Said (Children's Hospital of Philadelphia), Ping Qu (University of Pennsylvania), Dora Obiri-Yeboah (Thomas Jefferson University), Andres Vargas Romero (Thomas Jefferson University), Sara Nouri (Thomas Jefferson University), Mohamad-Gabriel Alameh (Children's Hospital of Philadelphia), Kiran Musunuru (University of Pennsylvania), Qiaoli Li (Thomas Jefferson University), Xiao Wang (University of Pennsylvania)

Abstract: Pseudoxanthoma elasticum (PXE) is a rare, autosomal recessive connective tissue disorder caused by pathogenic variants in *ABCC6*, resulting in ectopic calcification of elastic fibers. The multisystemic manifestations of PXE contribute to significant morbidity and mortality, and there are no targeted therapies.

The c.3490C>T (p.R1164X) and c.3421C>T (p.R1141X) variants are the two most frequent single-nucleotide changes and are potentially correctable with adenine base editing. The *ABCC6* gene encodes a liver-expressed efflux transporter that provides the mineralization inhibitor pyrophosphate (PPi) to the circulation.

We aim to develop adenine base editing therapies for PXE. We screened various adenine base editors (ABEs) and guide RNAs (gRNAs) in human hepatoma cells engineered to contain the R1164X or R1141X variant. We identified ABE8.8 with PXE1 gRNA and SpG-ABE8.8 with PXE2 gRNA as the most efficient for corrective editing.

To assess off-target editing, we performed OligoNucleotide Enrichment and sequencing (ONE-seq) and confirmed several off-target sites by targeted amplicon sequencing. To reduce off-target editing, we tested hybrid gRNAs with RNA replaced by DNA at specific positions of the spacer. We designed and screened hybrid gRNAs (PXE1_hyb1 through 21, PXE2_hyb16 and 17) in R1164X or R1141X hepatoma cells. PXE1_hyb18, with DNA at positions 5-7, and PXE2_hyb16, with DNA at positions 3-5, reduced off-target and bystander editing while improving on-target editing efficiency.

For in vivo studies, we developed a humanized mouse model in which part of the mouse *Abcc6* gene was replaced with the orthologous human sequence containing the R1164X variant. Homozygous R1164X mice exhibited reduced blood PPi and marked muzzle calcification, mimicking PXE. We treated these mice with lipid nanoparticles (LNPs) containing ABE8.8 mRNA and either standard PXE1 gRNA or

PXE1_hyb18 gRNA. PXE1_hyb18 had greater corrective editing and reduced off-target and bystander editing than the standard gRNA in vivo. LNP treatment restored blood PPI and eliminated muzzle calcification in mice on either a normal or pro-calcification diet.

These findings suggest that correcting ABCC6 variants in the liver is sufficient to prevent ectopic calcification at distant sites and support the feasibility of in vivo adenine base editing to permanently correct two of the most frequent pathogenic ABCC6 variants, paving the way for a potential treatment for PXE.

High-Coverage Whole-Genome Sequencing Uncovers Complex CRISPR-Induced Genomic Rearrangements in Human Hematopoietic Stem Cells

Subsession Time: Thursday, October 16 at 1:45pm – 2:00pm

Authors: Nechama Kalter (Bar-Ilan University), Michael Rosenberg (Bar-Ilan University), Gila Lithwick-Yanai (Ultima Genomics), Sarah Pollock (Ultima Genomics), Doron Lipson (Ultima Genomics), Zohar Shipony (Ultima Genomics), Ayal Hendel (Bar-Ilan University)

Abstract: CRISPR-Cas genome editing introduces a range of genomic alterations that must be thoroughly characterized to ensure safe clinical application. These include small insertions and deletions (indels) as well as structural variants (SVs), such as large deletions, inversions, and inter- or intra-chromosomal translocations. Such events can compromise genome integrity and potentially drive clonal expansion or tumorigenesis. Current methods for analyzing CRISPR-induced changes rely on targeted deep sequencing of predefined sites. While effective for detecting small indels, they require prior knowledge of off-target loci and are poorly suited for uncovering large or unexpected rearrangements.

Whole-genome sequencing (WGS) provides a comprehensive, unbiased alternative but has traditionally been limited by cost and depth. To overcome these challenges, we used Ultima Genomics' sequencing platform, enabling 1000× WGS at reduced cost, to profile CRISPR-edited human hematopoietic cells. Editing was performed with three therapeutic gRNAs targeting HBB, HBG, and BCL11A (used in CASGEVY, the first FDA-approved CRISPR therapy). To analyze the full spectrum of genomic outcomes from these edits, we developed a genome-wide pipeline for detecting CRISPR-induced alterations.

High-coverage WGS revealed a broad and complex landscape of CRISPR-induced genomic changes. At on-target sites, we detected events including small indels and long deletions (200 bp to several kb). Notably, 1–4% of reads contained translocations involving the targeted loci. Most of these did not match known off-target sites, suggesting unanticipated breakpoints. We also observed intra-chromosomal SVs. For the HBG gRNA, which induces two cuts, >5% of edited alleles showed ~5 kb inversions—missed by targeted methods. At HBB, we confirmed editing at off-target site OT1, with high-frequency indels and SVs. While both methods captured short deletions (<200 bp), only WGS revealed longer deletions and novel translocations.

These results highlight the diversity of CRISPR-induced alterations and the importance of unbiased, high-resolution tools for safety evaluation. A streamlined, ultra-deep WGS strategy offers comprehensive insight into genome integrity, advancing preclinical assessment of genome editing therapies.

A one and done therapeutic strategy to correct ELP1 splicing defect in familial dysautonomia

Subsession Time: Thursday, October 16 at 2:00pm – 2:15pm

Authors: Krishnakanth Kondabolu (Center for Genomic Medicine, Massachusetts General Hospital Research Institute), Shuqi Yun (Center for Genomic Medicine, Massachusetts General Hospital Research Institute), Anil Chekuri (Center for Genomic Medicine, Massachusetts General Hospital Research Institute), Paolo Pigni (Center for Genomic Medicine, Massachusetts General Hospital Research Institute), Jennifer Art (Center for Molecular Medicine, University of Georgia), Susan A. Slaughaupt (Center for Genomic Medicine, Massachusetts General Hospital Research Institute), Nadja Zeltner (Center for Molecular Medicine, University of Georgia), Benjamin P. Kleinstiver (Center for Genomic Medicine, Massachusetts General Hospital Research Institute), Christiano R. R. Alves (Center for Genomic Medicine, Massachusetts General Hospital Research Institute), Elisabetta Morini (Center for Genomic Medicine, Massachusetts General Hospital Research Institute)

Abstract: Familial Dysautonomia (FD) is a fatal autosomal recessive neuropathy characterized by sensory and autonomic dysfunction. Key clinical features include reduced pain and temperature sensation, cardiovascular instability, progressive vision and balance loss, and significantly shortened life expectancy, with most patients not surviving beyond early adulthood. FD is caused by a T-to-C transition at the 5' splice site of intron 20 in the ELP1 gene, resulting in tissue-specific skipping of exon 20. This leads to a premature termination codon and reduced ELP1 protein levels, primarily in the central and peripheral nervous systems. Notably, all individuals with FD carry at least one copy of this mutation, with 99.5% being homozygous.

In this study, we developed a CRISPR base editing (BE) approach to precisely correct the disease-causing mutation. Through the optimization of Cas9 variants and screening of multiple guide RNAs, we identified a BE strategy that achieved up to 70% on-target editing in HEK293T cells harboring the FD mutation. These editing levels successfully restored inclusion of exon 20 in the ELP1 transcript. To enable in vivo delivery, we engineered a dual intein-split system compatible with packaging into adeno-associated virus (AAV) vectors. These constructs were tested in ELP1 mutant iPSC-derived sympathetic neurons and in a humanized FD mouse model carrying the human ELP1 T-to-C mutation. In both models, the BE approach effectively corrected the splicing defect. In FD sympathetic neurons, splicing correction reduced disease-associated neuronal hyperactivity and restored normal firing patterns. Importantly, comprehensive off-target analyses confirmed the high precision and safety of our base editing strategy.

Together, these findings demonstrate the feasibility of a highly precise, AAV-deliverable base editing therapy for correcting the ELP1 mutation and its downstream molecular consequences, establishing a strong proof-of-concept for a transformative and durable treatment for FD.

Mouse models of, and AAV gene therapy for, cobalamin C (cblC) deficiency

Subsession Time: Thursday, October 16 at 2:15pm – 2:30pm

Authors: Charles P. Venditti (National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland), Giuliana Lopez (National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland), Pauline Hoffman

Stefanos Koutsoukos (National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland), Kelsey Murphy (National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland), Madeline Arnold (National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland), Nathan Achilly (National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland), Gene Elliot (National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland), Victoria Hoffmann (Division of Veterinary Resources, National Institutes of Health, Bethesda, Maryland), Jennifer Sloan (National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland)

Abstract: Combined methylmalonic acidemia and homocystinuria, cobalamin C type (cblC), is the most common inborn error of intracellular cobalamin (vitamin B12) metabolism and is caused by pathogenic variants in MMACHC, an enzyme responsible for the transport and processing of vitamin B12. Early onset disease can be severe, especially for those who are homozygous for the common infantile-onset variant MMACHC c.271dupA

(p.Arg91LysfsTer14). Symptoms of cblC deficiency include growth failure, neurocognitive impairment, anemia, cardiomyopathy, and progressive vision loss due to aggressive macular degeneration that begins early in life. To explore disease pathophysiology and investigate the efficacy of AAV gene addition therapy as a possible treatment for cblC deficiency, we generated and characterized four loss of function Mmachc alleles : Mmachc^{del2} c.163_164delAC p.(Pro56CysfsTer4) , Mmachc^{del3} c.162_164delCAC p.(Ser54_Thr55delinsArg), Mmachc^{del23} c.163_185del p.(Thr55ProfsTer42), and Mmachc^{c.271dupA}, containing the humanized MMACHC variant, c.271dupA, knocked into exon 2 of Mmachc. The Mmachc^{-/-} mutants had reduced viability and recapitulated the clinical and biochemical phenotypes of cblC deficiency including embryonic and neonatal lethality, decreased survival (median 5 days, with complete lethality by 1 month), growth failure, neuropathology (hydrocephalus and thinning of the corpus callosum), elevated plasma methylmalonic acid, increased homocysteine, and decreased methionine. Transgenes designed to express human MMACHC under the control of the CBA or EF1 promoters were pseudoserotyped with AAV9, AAVrh10, or AAV PHP.eB capsids, and systemically delivered on P0-2 at doses of 1 x 10¹⁰ (self-complementary) or 1 x 10¹¹ GC/pup (single strand). All Mmachc^{-/-} mice could be rescued from lethality using systemic AAV gene therapy. The treated mutants had improved growth and metabolic parameters, mitigated neuropathology, and prolonged survival, with many surviving beyond 1 year. Systemic AAV gene therapy therefore holds promise as a potential therapeutic strategy to treat individuals with cblC deficiency, the most common disorder of intracellular vitamin B12 metabolism.

Session 56: Genetic Insights Fueling Drug Discovery and Precision Health

Location: Ballroom West/Level 3, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 1:30pm – 2:30pm

Human genetics evidence supporting ANGPTL4 inhibition to reduce remnant cholesterol and triglycerides informs successful Phase 2a trial outcomes

Subsession Time: Thursday, October 16 at 1:30pm – 1:45pm

Authors: Sahir Rai Bhatnagar (5 Prime Sciences), Tianyuan Lu (Department of Population Health Sciences, Department of Biostatistics and Medical Informatics, University of Wisconsin School of Medicine and Public Health), Urvashi Singh (5 Prime Sciences, McGill University), Matthew Tudball (5 Prime Sciences), Yiheng Chen (5 Prime Sciences), Vincenzo Forgetta (5 Prime Sciences), Brent Richards (5 Prime Sciences, McGill University)

Abstract: Few human genetics-based drug development programs have progressed from GWAS findings for common traits to Phase 2 trials. ANGPTL4 inhibition showed promise for reducing triglycerides, a key contributor to cardiovascular risk, but earlier attempts were halted due to abdominal lymphatic complications in animal models. To evaluate ANGPTL4 as a therapeutic target, we conducted comprehensive human genetics analyses, including GWAS, exome-wide association studies (ExWAS), cis-protein quantitative trait loci (cis-pQTL) Mendelian Randomization (MR), and phenome-wide association studies (PheWAS).

First, our GWAS confirmed that the ANGPTL4 p.E40K loss-of-function variant is associated with reduced triglycerides and remnant cholesterol. These findings were replicated using ExWAS analyses.

Next, we used the p.E40K variant as an instrumental variable in cis-pQTL MR analyses. A one standard deviation (SD) decrease in circulating ANGPTL4 protein levels from deCODE SomaScan assays showed significant causal reductions (in SD units) in remnant cholesterol ($\beta=-0.47$, $P=1 \times 10^{-74}$) and triglycerides ($\beta=-0.74$, $P<10^{-300}$). These results were replicated using UK Biobank (UKB) Olink assays (remnant cholesterol $\beta=-0.42$, $P=1 \times 10^{-74}$; triglycerides $\beta=-0.66$, $P=2 \times 10^{-291}$). All results colocalized with high certainty (PP.H4>99%).

A PheWAS using the p.E40K variant across UKB, FinnGen, and GWAS Catalog, found no associations with abdominal lymphatic complications. Electronic medical record data (FinnGen, provided by Marea Therapeutics) from three homozygous ANGPTL4 p.Cys80ValfsTer12 carriers showed no abdominal or lymphatic clinical pathology. These individuals had reduced triglyceride levels (mean<60 mg/dL) compared to heterozygous carriers (mean=123 mg/dL, N=67) and non-carriers (mean=156 mg/dL, N=363,130) from UKB exomes. Together, this supports that germline ANGPTL4 loss does not confer severe abdominal effects in humans, reinforces its role in plasma triglyceride regulation, and suggests a lower physiological limit to triglyceride reduction.

Marea Therapeutics undertook Phase 1a/2a trials (Cummings et al., 2025, The Lancet). Their ANGPTL4 antibody significantly reduced triglycerides (up to 52.7%) and remnant cholesterol (up to 52.5%) without inflammatory, lymphatic, or adverse metabolic effects.

These findings exemplify how insights from human genetics can rapidly progress to novel therapies and identify side effects that are likely unique to animal models.

Associations between ultra-rare loss-of-function variation and 454 traits in the Mexico City Prospective Study

Subsession Time: Thursday, October 16 at 1:45pm – 2:00pm

Authors: Sam Morris (Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, UK), Eirini Trichia (Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, UK), Diego Aguilar-Ramirez (Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, UK), Michael Turner (Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, UK), Alejandra Vergara-Lope (Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, UK), Elizabeth Barrera (Experimental Research Unit from the Faculty of Medicine, National Autonomous University of Mexico, Mexico City, Mexico), Paulina Baca (Experimental Research Unit from the Faculty of Medicine, National Autonomous University of Mexico, Mexico City, Mexico), Fernando Rivas (Experimental Research Unit from the Faculty of Medicine, National Autonomous University of Mexico, Mexico City, Mexico), Tianshu Liu (Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, UK), Prof Jonathan R Emberson (Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, UK), Prof Jesus Alegre-Díaz (Experimental Research Unit from the Faculty of Medicine, National Autonomous University of Mexico, Mexico City, Mexico),

Prof Pablo Kuri-Morales (Tecnológico de Monterrey, Proyecto OriGen, Monterrey, Mexico), Prof Roberto Tapia-Conyer (Faculty of Medicine, National Autonomous University of Mexico, Mexico City, Mexico), Prof Jaime Berumen (Experimental Research Unit from the Faculty of Medicine, National Autonomous University of Mexico, Mexico City, Mexico), Jason Torres (Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, UK)

Abstract: Ultra-rare loss-of-function (pLoF) variants can exert large phenotypic effects and reveal potential drug targets, yet remain under-characterised in non-European populations. The Mexico City Prospective Study (MCPS) is a large Latin American cohort with exome sequencing data available from 141,046 admixed participants. Previously, we identified over 1.3 million variants not present in UK Biobank, TOPMed, or gnomAD—including 125,344 pLoF and 892,337 missense variants. MCPS therefore provides a unique opportunity to explore the relationship between rare, protein-damaging variants and trait variation. We annotated exome sequencing (mean depth 20x) using the BraVa annotation pipeline (<https://github.com/BRaVa-genetics/variant-annotation>) and retained high-confidence predicted pLoF and predicted-damaging missense/protein-altering (DM/PA) variants with MAF < 0.01%. Gene-based

tests in REGENIEv4 (using the “GENE_P” combination test) were performed across 18,262 protein-coding genes and 454 traits: 78 binary (e.g. type 2 diabetes) and 363 continuous (e.g. height, HbA1c) traits, which included 325 NMR-based metabolite features. We identified 181,675 high-confidence pLoF variants (42% singletons) and 520,998 DM/PA variants. Individuals carried a median of 52 pLoF and 1,179 DM/PA variants. Association tests yielded 192 gene–trait associations, with lipoprotein traits accounting for 32% of associations, driven primarily by pLoF variants in APOB, LDLR, PCSK9, APOA5, and LIPC. PCSK9 pLoF carriers showed -1.07 SD lower LDL-C ($p = 2 \times 10^{-19}$), consistent with the known effect of PCSK9 inhibition. Pathway enrichment analysis identified significant overrepresentation of genes involved in plasma lipoprotein assembly, remodelling, and clearance (Benjamini-Hochberg adjusted p-value $< 1 \times 10^{-20}$). Other key pathways included HDL remodelling and chylomicron processing, implicating central regulators such as APOB, LDLR, and CETP in these metabolic signatures. This study provides the first large-scale atlas of ultra-rare coding variation in Latin America and shows that even single-carrier pLoF alleles can yield meaningful biological insights.

GLP1R and GIPR coding variants reveal differential impacts on metabolic and mental health – Implications for drug response in incretin receptor agonist treatment

Subsession Time: Thursday, October 16 at 2:00pm – 2:15pm

Authors: Sheen Lei (Imperial College London), Zhanna Balkhiarova (Imperial College London, University of Surrey), Giuseppe Deganutti (Coventry University), Christopher A Reynolds (Coventry University), Ben Jones (Imperial College London, Imperial College Healthcare NHS Trust), Inga Prokopenko (University of Surrey)

Abstract: Incretins, GLP-1 and GIP, play a crucial role in glucose metabolism, insulin secretion, and impact mental health. Incretin receptor agonist drugs, such as semaglutide and tirzepatide, target GLP-1R and GIPR, and are widely used to treat type 2 diabetes (T2D) and obesity. However, patients exhibit differential metabolic and mental health responses to drugs, given individual genetic variability.

Objective: to define GLP1R and GIPR coding variant effects on metabolic/mental health, predicting individual drug response, i.e. beneficial, neutral, and adverse, to incretin receptor agonist drugs.

We evaluated 192 GLP1R coding variants and 208 GIPR coding variants from whole-exome sequencing data for 393,764 UK Biobank European individuals. The variant association with four metabolic (body-mass index [BMI], T2D, random glucose and HbA1c) and two mental health phenotypes (major depressive disorder [MDD, ICD-10], depressive symptoms, PHQ-9 questionnaire) was assessed through: single-variant association testing, assuming (log)-additive model of inheritance for binary/quantitative phenotypes in PLINKv.1.9, adjusted for age/sex/six principal genetic components/time since last meal (for RG); variant-specific Z-score effects in hierarchical clustering with pheatmap() in R; SKAT-O set-based association analysis for variant cluster effects using SAIGE-GENE+.

Five distinct clusters of GLP1R variants were identified, exhibiting differential effects on metabolic and mental health phenotypes. Notably, Cluster 1 (21 variants) confers a higher risk of MDD, Cluster 2 (29 variants) leads to better glycemic health, and Cluster 3 (28 variants) may cause adverse drug reactions (ADR) for all four metabolic health phenotypes. Four distinct clusters of GIPR variants were identified, including Cluster 3 (23 variants), leading to higher T2D risk and high HbA1c, and Cluster 4 (7 variants), leading to lower BMI, while conferring risks on glycemic/mental health. We evaluated GLP1R variant effects through functional and molecular dynamics experiments.

Incretin receptor agonist drug target GLP1R and GIPR genes coding variant effects suggest distinct drug response metabolic and mental health profiles, driving individual low response and potential ADRs. We propose to use these predictive pharmacogenetic profiles to define new pre-emptive testing panels as support for drug treatment stratification to improve drug efficacy and prevent adverse effects of incretin receptor agonist drugs.

Hypercholesterolemia polygenic score modulates response to statin therapy

Subsession Time: Thursday, October 16 at 2:15pm – 2:30pm

Authors: Ariana Lichtenstein (Helix), Natalie Telis (Helix), Matthew Levy (Helix), Catherine Hajek (Helix), Amy C. Sturm (The Ohio State University Wexner Medical Center), Chad R. Haldeman-Englert (Cone Health Precision Health), Steven Powell (Sanford Health), Douglas Stoller (University of Nebraska Medical Center), Christopher N. Chapman (Department of Pathology & Laboratory Medicine, St. Luke's University Health Network), C. Anwar A. Chahal (Center for Inherited Cardiovascular Diseases, WellSpan Health), Daniel P. Judge (Division of Cardiology, Medical University of South Carolina), Douglas A. Olson (HealthPartners), Joseph J. Grzymski (Renown Health), Nicole L. Washington (Helix), William Lee (Helix), Alexandre Bolze (Helix), Kelly M. Schiabor Barrett (Helix), Elizabeth T. Cirulli (Helix)

Abstract: Background: Studies have shown that individuals with high CAD PGS have greater reduction in cardio outcomes from statin use than do those with low PGS. Here, we quantify the effect of a hypercholesterolemia PGS on response to statin therapy.

Methods: A retrospective study of real world EMR data from ~200k Helix Research Network participants, enrolled from the general population during routine medical care from >15 health systems across the country. This study utilizes Exome+(R) sequencing; hypercholesterolemia PGS000889; longitudinal LDL levels; and statin start date. We analyzed a median of 9 LDL measurements spanning 8 years and studied percent change in LDL after beginning statins.

Results: We analyzed 12,444 individuals who had baseline LDL above 100 mg/dL and then began statins, dropping their LDL by a median of 31%. We found that highPGS individuals (>90%) had less reduction in their LDL compared to lowPGS (<10%). This effect was most pronounced in those with lower baseline LDL: for those with LDL between 100-130, lowPGS individuals saw a 35% reduction, compared to 20% reduction for highPGS, while for those with LDL >160, both high and low PGS experienced a 34% reduction. In contrast, the effect of statins on reducing subsequent CAD was higher for highPGS: a 1.6x reduction in risk for lowPGS vs. a 4.2x reduction for highPGS with baseline LDL 100-300, and a 4.1x

reduction for lowPGS vs. 7.1x for highPGS with baseline LDL >160. This observed difference was mostly due to untreated highPGS individuals having high risk even without clinically elevated LDL.

Conclusion: Our data shows an association between hypercholesterolemia PGS and CAD risk reduction following statin use. The observed percent LDL reduction appeared lower in highPGS individuals, with the greatest difference in association observed in individuals with relatively low LDL.

Session 57: Genetic Mechanisms of Cancer Risk

Location: Room 210C/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 1:30pm – 2:30pm

Comprehensive characterization of functional pleiotropy at the 5p15.33 multi-cancer risk locus

Subsession Time: Thursday, October 16 at 1:30pm – 1:45pm

Authors: Aidan O'Brien (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Hyunkyung Kong (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Minal Patel (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Katelyn E. Connelly (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Mai Xu (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Irene Collins (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Jun Zhong (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Jason Hoskins (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Harsh Patel (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Thong Luong (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Michelle Ho (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Brenen Papenberg (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), International Lung Cancer Consortium (ILCCO), Bladder Cancer Consortium, Mark M. Iles (Leeds Institute for Data Analytics, School of Medicine, University of Leeds), Matthew H. Law (Statistical Genetics Laboratory, QIMR Berghofer Medical Research Institute),

Maria Teresa Landi (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Melanoma Meta-Analysis Consortium, Rachael Z. Stolzenberg-Solomon (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Brian M. Wolpin (Department of Medical Oncology, Dana Farber Harvard Cancer Center), The Pancreatic Cancer Cohort Consortium (PanScan), Alison P. Klein (Department of Oncology Pathology and Medicine, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine), Donghui Li (Department of Gastrointestinal Medical Oncology, University of Texas MD Anderson Cancer Center), The Pancreatic Cancer Case-Control Consortium (PanC4), Nick Orr (Patrick G Johnston Centre for Cancer Research, Queen's University Belfast), Stephen J. Chanock (Division of Cancer Epidemiology and Genetics, National

Cancer Institute, National Institutes of Health), Sara Lindström (Department of Epidemiology, University of Washington), Marc-Henri Stern (Institut Curie),

Jianxin Shi (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Ludmila Prokunina-Olsson (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Jiyeon Choi (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Kevin M. Brown (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Laufey T. Amundadottir (Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health)

Abstract: 5p15.33 is one of the most replicated cancer GWAS loci, with ten signals reported across multiple cancer types. Notably, alleles associated with risk for one cancer are often protective for others. This locus harbors two compelling target genes: TERT, central to telomerase activity, and CLPTM1L, a growth promoting gene. While fine-mapping has identified multiple independent signals at 5p15.33, functional follow-up has been limited to single-cancer studies, leaving pleiotropic relationships unresolved.

To investigate causal variant–gene relationships across cancers, we fine-mapped 117 candidate causal variants and assessed their regulatory activity using a massively parallel reporter assay (MPRA) screen. In parallel, we performed a tiled CRISPRi proliferation screen using 25,778 sgRNAs across a 150 kb region in two cell lines from each of melanoma, lung, pancreatic, and bladder cancer. Guided by convergent screening hits, we performed targeted CRISPRi validation in lung and pancreatic cancer cells for eight SNPs across five signals. We noted multi-cancer regulation of TERT by regions harboring rs7705526 and rs2736108, CLPTM1L by rs13167280, and rs37005, and both genes by rs2853677 and rs13167280.

Notably, our CRISPRi screen implicated an intronic variable number tandem repeat (VNTR) in CLPTM1L enriched for TEAD transcription factor motifs. Targeted CRISPRi of this region potentially downregulated both TERT and CLPTM1L. PacBio sequencing of 1000G EUR samples and imputation into a pancreatic cancer GWAS (9,040 cases) revealed a strong association between long VNTR alleles and risk ($OR=1.17$, $P=2.21 \times 10^{-12}$). This signal was attenuated by conditioning on the lead SNP for pancreatic cancer (rs31490, $P_{cond}=0.55$), supporting the inclusion of the VNTR as a potentially causal variant. Luciferase assays of VNTR alleles revealed up to 500-fold enhancer activity which was abrogated by inhibition of TEAD-YAP binding, implicating Hippo signaling in VNTR-mediated regulation of TERT and CLPTM1L.

Our findings add to evidence that cancer risk at 5p15.33 may be mediated by VNTRs as well as SNPs. These variants span multiple independent signals and likely influence cis-regulation of TERT and CLPTM1L. Concordant MPRA/CRISPRi data suggests a common allelic function underlying risk across multiple cancers with pleiotropic signals at 5p15.33. Our work lays critical groundwork for further investigation and offers a broadly applicable framework for dissecting complex, pleiotropic GWAS loci.

Genome-wide characterization of clonal hematopoiesis reveals extensive non-coding putative driver mutations

Subsession Time: Thursday, October 16 at 1:45pm – 2:00pm

Authors: Joshua Weinstock (Emory University, Department of Human Genetics), Karen Conneely (Emory University, Department of Human Genetics), Cameron Russell (Emory University, Department of Human Genetics), Janghee Woo (Emory University, Department of Hematology and Medical Oncology), Marios Arvanitis (Division of Cardiology, Department of Medicine, Johns Hopkins University), Mitchell Machiela (National Cancer Institute, Division of Cancer Epidemiology and Genetics)

Abstract: As humans age, we acquire somatic mutations in our blood, leading to distinct groups of genetically identical hematopoietic stem cells (HSC) called clones. Expanded clones have previously been linked to increased risk for hematologic cancer and cardiovascular disease. Despite the prevalence of clonal hematopoiesis (CH), recent unbiased searches for selective sweeps from both single-cell derived colonies as well as population-scale exome sequencing have revealed that the majority of clones have expanded without a known driver mutation. This extensive, unexplained CH motivated our search for novel driver mutations across ~490K blood whole genome sequences from the UK Biobank. We searched across variants with a minor allele count of at least 10 (~147M variants) to discover alleles that are enriched in aged individuals using a generalized linear mixed model. We included age at recruitment, sex, smoking history, and genetic ancestry principal components as covariates. We observed 68 variants where the alternate allele was associated with age at genome-wide significance ($P < 5 \times 10^{-8}$). After excluding hits that were common germline variants in loci known to affect longevity (e.g. APOE), we identified 45 rare variants (MAF < 0.5%) with large effect sizes, consistent with somatic genesis. Among these variants, we recapitulated known CH driver genes (e.g., DNMT3A, ASXL1, SF3B1). Among the 36 novel age-associated variants, we identified 50 carriers of a somatic mutation in the TERT promoter. Although previously reported in pan-cancer analyses, TERT promoter mutations have never been reported in population searches for CH. We also observed a somatic intronic insertion in UGT2B7 in 1,165 carriers. We conducted a phenome-wide association study (PheWAS) among 30 common disease phenotypes to characterize the phenotypic correlates of these mutations. The UGT2B7 insertion was associated with increased odds of myeloproliferative neoplasms (odds-ratio of 2.6, local false sign rate = 1.6×10^{-8}) after adjustment for non-linear age transformations, sex, and smoking history, along with increased risk for several other cancer and non-cancer phenotypes. We then performed a genome-wide association study of the somatic UGT2B7 insertion, finding a common variant in cis at genome-wide significance. Overall, we characterize CH at both increased breadth and resolution and characterize the entire cascade from upstream germline risk haplotypes to downstream clinical correlates.

Detection and quantification of cancer-associated mutagenic processes in normal human somatic cells

Subsession Time: Thursday, October 16 at 2:00pm – 2:15pm

Authors: Safia Mahabub Sauty (Genome Integrity and Structural Biology Lab, NIEHS), Yun-Chung Hsiao (Genome Integrity and Structural Biology Lab, NIEHS), Leszek J. Klimczak (Integrative Bioinformatics Support Group, NIEHS), Dmitry Gordenin (Genome Integrity and Structural Biology Lab, NIEHS)

Abstract: Somatic mutations are inevitable in human genomes and can lead to tumorigenesis. Mutational processes operating in cancer cells are unveiled by several bioinformatic methods, including agnostic mutational signature extraction. Signature extraction methods identified many known and yet unknown sources of somatic mutations in cancer cells, however, our understanding of baseline mutagenesis in normal cells is scarce due to lack of sufficient number of mutations and sequenced genomes required for signature extraction. To circumvent this limitation, we performed knowledge-based, motif-centered analyses on the mutation profiles of 11,987 normal cell genomes across 25 different tissue types. We applied stringent statistical hypothesis derived from mechanistic knowledge to detect statistically significant enrichment and enrichment-adjusted Minimal Estimate of Mutation Load (MEML) with oligonucleotide motifs preferred by known mutagenic mechanisms. We found that several mutational motifs previously identified in human tumors are readily detected in cancer-free tissues. C to T mutations in nCg motif caused by spontaneous deamination of mCpG sites were detected across all 25 tissue types. nCg MEML accounted for approximately 21% of all mutations in C:G-pairs and showed age-dependent (clock-like) accumulation in most tissues. Another ubiquitous motif, T to C substitutions in aTn motif associated with mutagenic epoxides and SN2-reacting electrophiles, was detected in >80% samples and showed clock-like feature in most tissues. Comparison with cancer-free diseased tissues identified a high burden of both nCg and aTn mutational motifs in donors with Xeroderma Pigmentosa, Chronic Obstructive Pulmonary Disease, and several gastrointestinal diseases. We also detected C to T and C to G mutations in tCw motif, diagnostic of cytidine deaminase enzymes APOBEC, in normal bladder, lung, small intestine, and breast samples with preference for APOBEC3A-like mutagenesis. UV-induced C to T mutations in yCn motif were only detected in skin samples where yCn MEML contributed approximately 25% of all mutations in C:G-pairs. Clock-like UV-mutagenesis accumulation was detected in donors with psoriasis. Together, these analyses quantified the mutagenic pathways in different tissues at healthy and non-cancer disease states, identified clock-like motifs that can be used to predict disease progression, and provided new variables for assessing individual predisposition to somatic mutagenesis.

A multiplex assay of variant effect (MAVE) of MSH6 enables accurate, prospective Lynch Syndrome clinical variant interpretation

Subsession Time: Thursday, October 16 at 2:15pm – 2:30pm

Authors: Anthony Scott (Department of Human Genetics & Department of Internal Medicine, University of Michigan), Felicia Hernandez (Ambry Genetics), Ashley PL Marsh (Ambry Genetics), Kristin Kohler (Ambry Genetics), Xiaoyan Jia (Department of Human Genetics, University of Michigan), Sebastian Vishnopolska (Department of Human Genetics, University of Michigan), Sami Belhadj (Ambry Genetics), Matthew Varga (Ambry Genetics), Grace Clark (Department of Human Genetics, University of Michigan), Sajini Jayakody (Department of Human Genetics, University of Michigan), Bala Burugula (Department of Human Genetics, University of Michigan), Rachid Karam (Ambry Genetics), Jacob Kitzman (Department of Human Genetics & Department of Computational Medicine & Bioinformatics, University of Michigan)

Abstract: Pathogenic germline variants in DNA mismatch repair factor genes cause Lynch Syndrome (LS), a dominant, hereditary cancer predisposition syndrome affecting more than 1:300 individuals worldwide. However, the large and growing burden of variants of uncertain significance (VUS)

complicates genetic counseling for this condition. Missense variants make up the majority of VUSs, reflecting both their sheer number and the difficulty of assigning their pathogenicity. Incomplete penetrance and variable age of onset compound this challenge for MSH6, making family history difficult to interpret. To systematically and prospectively provide functional evidence in Lynch Syndrome, we have established and validated a mismatch repair gene multiplex assay of variant effect (MAVE), which we apply here to the key C-terminus of the mismatch repair factor MSH6. Our variant-to-function map accurately recapitulates standing clinical variant interpretations across a large validation set of previously classified benign (n=33) and pathogenic (n=26) variants (ROC AUC = 0.9779). Therefore, under the OddsPath framework, our MSH6 MAVE scores may be considered strong evidence for or against MSH6 missense variant pathogenicity. To apply this resource in a heterogeneous, real-world clinical setting, we intersected our MAVE scores with a large clinical database of genetic test orders. We identified 3,527 individuals who carry an MSH6 missense variant in this region, and looked up their variants in the MAVE results. Individuals carrying a missense variant indicated by the MAVE dataset to be functionally abnormal showed a higher risk for colorectal cancer (hazard ratio [HR]=2.4, 95% CI 1.3-4.6; both sexes) and uterine cancer (HR 3.4, 95% CI 1.42-7.0; females only), as compared to individuals who carried a functionally normal missense variant (p=0.00593 and 0.00458, respectively). These results can be used to guide the VUS resolution. Of the 535 MSH6 missense VUS found in this cohort, 17% were functionally abnormal. Ongoing reclassification of these variants as pathogenic or likely pathogenic will result in new diagnoses, enabling optimal prevention and treatment for those individuals and family members who carry the same variants.

Session 58: Mendelian Disorders of the Epigenetic Machinery: New Cellular and Phenotypic Insights

Location: Room 258ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 1:30pm – 2:30pm

Cell type-specific dysregulation of gene expression due to Chd8 haploinsufficiency during mouse cortical development

Subsession Time: Thursday, October 16 at 1:30pm – 1:45pm

Authors: Marybeth Baumgartner (Yale School of Medicine), Kristina M. Yim (Yale School of Medicine), Martina Krenzer (Mount Sinai School of Medicine, Yale School of Medicine), María F. Rosales Larios (National Autonomous University of Mexico, Yale School of Medicine), Guillermina Hill-Terán (National University of Tucumán, Yale School of Medicine), Timothy Nottoli (Yale Genome Editing Center, Yale School of Medicine), Rebecca A. Muhle (New York State Psychiatric Institute, Columbia University, Yale Child Study Center), James P. Noonan (Yale School of Medicine, Wu Tsai Institute, Yale University)

Abstract: Disruptive variants in the chromodomain helicase CHD8, which acts as a transcriptional regulator during neurodevelopment, are strongly associated with risk for autism spectrum disorder (ASD). CHD8 haploinsufficiency is hypothesized to contribute to ASD by perturbing neurodevelopmental gene expression. However, insight into the cell type-specific transcriptional effects of CHD8 haploinsufficiency remains limited. We used single-cell and single-nucleus RNA-sequencing to globally profile gene expression and identify dysregulated genes in the embryonic and juvenile wild type and Chd8+/- mouse cortex, respectively. Chd8 and other ASD risk-associated genes showed a convergent expression trajectory that was largely conserved between the mouse and human developing cortex, increasing from the progenitor zones to the cortical plate. Genes associated with risk for neurodevelopmental disorders and genes involved in neuron projection development, chromatin remodeling, signaling, and migration were dysregulated in Chd8+/- embryonic day (E) 12.5 radial glia. Genes implicated in synaptic organization and activity were dysregulated in Chd8+/- postnatal day (P) 25 deep- and upper-layer excitatory cortical neurons, suggesting a delay in synaptic maturation or impaired synaptogenesis. Our findings reveal a complex pattern of transcriptional dysregulation in Chd8+/- developing cortex, potentially with distinct biological impacts on progenitors and maturing neurons in the excitatory neuronal lineage.

ASXL1 mutations drive metabolic alterations through mitochondrial dysfunction

Subsession Time: Thursday, October 16 at 1:45pm – 2:00pm

Authors: Michael Sigfrid S. Reyes (Department of Pathology and Laboratory Medicine David Geffen School of Medicine University of California Los Angeles Los Angeles California USA, Department of Human Genetics David Geffen School of Medicine University of California Los Angeles Los Angeles California USA, Department of Computational Medicine University of California Los Angeles Los Angeles

California USA, Biological Models Laboratory Department of Biochemistry and Molecular Biology College of Medicine University of the Philippines Manila Philippines), Isabella Lin (Department of Pathology and Laboratory Medicine David Geffen School of Medicine University of California Los Angeles Los Angeles California USA, Department of Human Genetics David Geffen School of Medicine University of California Los Angeles Los Angeles California USA), Abigail Krall (Department of Biological Chemistry University of California Los Angeles Los Angeles California USA), Aileen Ning (Department of Pathology and Laboratory Medicine David Geffen School of Medicine University of California Los Angeles Los Angeles California USA, Department of Human Genetics David Geffen School of Medicine University of California Los Angeles Los Angeles California USA, Department of Computational Medicine University of California Los Angeles Los Angeles California USA), Suzanne Sarkissian (Department of Pathology and Laboratory Medicine David Geffen School of Medicine University of California Los Angeles Los Angeles California USA, Department of Human Genetics David Geffen School of Medicine University of California Los Angeles Los Angeles California USA, Department of Computational Medicine University of California Los Angeles Los Angeles California USA), Neerja Vashist (Department of Pathology and Laboratory Medicine David Geffen School of Medicine University of California Los Angeles Los Angeles California USA, Department of Human Genetics David Geffen School of Medicine University of California Los Angeles Los Angeles California USA, Department of Computational Medicine University of California Los Angeles Los Angeles California USA), Linsey Stiles (Department of Medicine Endocrinology David Geffen School of Medicine University of California Los Angeles Los Angeles CA USA, Department of Molecular and Medical Pharmacology University of California Los Angeles Los Angeles California USA), Heather Christofk (Department of Biological Chemistry University of California Los Angeles Los Angeles California USA), Paul Mark B. Medina (Biological Models Laboratory Department of Biochemistry and Molecular Biology College of Medicine University of the Philippines Manila Philippines), Valerie A. Arboleda (Department of Pathology and Laboratory Medicine David Geffen School of Medicine University of California Los Angeles Los Angeles California USA, Department of Human Genetics David Geffen School of Medicine University of California Los Angeles Los Angeles California USA, Department of Computational Medicine University of California Los Angeles Los Angeles California USA)

Abstract: ASXL1 (additional sex combs-like 1) modulates the epigenetic landscape, controls many downstream transcriptional pathways, and plays important roles during development and in regulating myeloid differentiation. Germline truncating mutations in ASXL1 cause the rare neurodevelopmental syndrome known as Bohring-Opitz Syndrome (BOS) characterized by intellectual disability, developmental delay, distinctive facial features and posture, and growth failure. Somatic mutations in ASXL1 drive clonal hematopoiesis of indeterminate potential (CHIP) and is associated with worse prognosis in myeloid malignancies. Here, we present findings showing that ASXL1 mutations directly drive metabolic reprogramming through a gain-of-function mechanism. Transcriptomic analysis in cells overexpressing truncated ASXL1 protein showed enrichment of genes related to carbohydrate metabolism. In both patient-derived cell models of BOS and in our overexpression models of mutant ASXL1, we see decreased levels of mitochondrial pyruvate carrier (MPC), a mitochondrial membrane protein that is essential for transporting pyruvate into the mitochondria. We performed Seahorse assay to measure oxygen consumption rate (OCR) and extracellular acidification rate (ECAR), which reflect

mitochondrial activity and glycolysis, respectively. Interestingly, ASXL1 mutations drive decreased reliance on mitochondrial activity. Targeted metabolomic analysis showed upregulation of glycolytic pathways, likely in response to decreased mitochondrial function. The inefficient energy production due to mitochondrial dysfunction could contribute to the growth and developmental delay seen in BOS. This metabolic shift towards aerobic glycolysis, similar to the Warburg effect seen in cancer cells, also suggests a mechanism for how ASXL1 mutations drive CHIP and myeloid malignancies. Overall, the metabolic alterations driven by ASXL1 mutations provide insight into the molecular pathogenesis of BOS and myeloid malignancies and offers a new paradigm of potential therapeutic targets.

Defining the source of motor deficits in KMT5B-linked syndrome

Subsession Time: Thursday, October 16 at 2:00pm – 2:15pm

Authors: Holly A. Feser Stesman (Creighton University, School of Medicine), Jason Hulen (Creighton University, School of Medicine), Jod Hallgen (Creighton University, School of Medicine), Cynthia Watson (Creighton University, Core Facilities)

Abstract: Germline heterozygous disruptive variants in lysine methyl transferase 5B (KMT5B/SUV420H1) have been linked to a syndrome characterized by developmental delays, intellectual disability, and autism spectrum disorder with significant motor delays. Using a *Kmt5b* gene trap mouse to model this human syndrome, we have previously reported early motor reflex deficits, decreased hindlimb skeletal muscle weights, and decreased neuromuscular strength in germline *Kmt5b* heterozygous (HET) compared to wild-type (WT) littermates. HET skeletal muscles displayed increased neuromuscular junction fragmentation and increased numbers of fast-twitch myofibers that were significantly smaller in size compared to WT muscles. To determine the source of these motor phenotypes, floxed *Kmt5b* mice were crossed to a B6.129S1-Mnx1tm4(cre)Tmj/J line expressing cre-recombinase driven by the *Mnx1* promoter for motor neuron-specific *Kmt5b* knockout (“mn”) or a B6.Cg-Tg(ACTA1-cre)79Jme/J line expressing cre-recombinase driven by the ACTA1 promoter for skeletal muscle-specific *Kmt5b* knockout (“sm”). From both lines, conditional WT, HET, and knockout (KO) progeny were subjected to a small battery of early motor reflex, strength, and muscle tests. Independent cohorts of male and female mice were tested in all experiments. Groups were compared using a 2-way ANOVA (variables of genotype and sex) or repeated measures ANOVA with post hoc testing (Šídák's test or similar). While early motor reflexes were not changed in mn mice, surface righting time was significantly slower in smHET compared to smWT female mice ($p=0.0248$; Šídák's test). Weekly measures of strength collected from postnatal day (P) 17-59 showed a significant interaction between age and genotype ($F(6, 177)=2.681$; $p=0.0163$) driven primarily by females in the mn model. While strength was also reduced in smHET and smKO mice, this was not statistically significant. Skeletal muscle and total body weights were not significantly different in the sm model. However, there was a significant effect of genotype ($F(2, 56)=15.26$; $p<0.0001$) on P59 fast-twitch muscle weight driven by large increases in mnHET males and females. MnHET animals also had higher body weights than their mnWT littermates which reached statistical significance in males ($F(2, 59)=3.955$; $p=0.0244$). ScRNA-seq data of the developing embryonic brain provided additional evidence of excessive motor neuron differentiation in HET mice as early E14.5. Our data suggest that the increased myofiber counts reported in the germline *Kmt5b* model are due to specific KMT5B effects on

motor neuron development and/or function rather than skeletal muscle development. Further, our data suggest that the full spectrum of motor deficits identified in germline *Kmt5b* mice are likely due to impacts on multiple systems.

Rare genetic variants in histone modification pathways implicate epigenetic dysregulation in ADHD

Subsession Time: Thursday, October 16 at 2:15pm – 2:30pm

Authors: Ryan Neil Doan (Division of Genetics & Genomics, Boston Children's Hospital, BCH), Paulina E. Gonzalez Tovar (Division of Genetics & Genomics, BCH), Andrada Georgescu (Division of Genetics & Genomics, BCH), Erin O'Halloran (Division of Genetics & Genomics, BCH), Clemence Mbyia (Division of Genetics & Genomics, BCH), Isabella Saunders (Division of Genetics & Genomics, BCH), Daniela Koplin (Division of Genetics & Genomics, BCH), Julieta Bonacina (Division of Genetics & Genomics, BCH), Mia O'Connell (Division of Genetics & Genomics, BCH), Katheryn Hayes (Division of Genetics & Genomics, BCH), Julia Scotellaro (Division of Developmental Medicine, Boston Children's Hospital, DDM, BCH; Division of Genetics & Genomics, BCH), Soleha Patel (Division of Genetics & Genomics, BCH), Stephanie J. Brewster (Division of Developmental Medicine, BCH, DDM; Division of Genetics & Genomics, BCH), Ashish Jain (Research Computing, Information Technology, BCH), Liang Sun (Research Computing, Information Technology, BCH), Shira Rockowitz (Division of Genetics & Genomics, BCH; Research Computing, Information Technology, BCH), William Barbaresi (Division of Developmental Medicine, BCH, DDM), Elizabeth Harstad (Division of Developmental Medicine, BCH, DDM), Anne B. Arnett (Division of Developmental Medicine, BCH, DDM).

Abstract: BackgroundDespite high heritability (~75%), the genetic architecture of attention-deficit/hyperactivity disorder (ADHD) remains poorly understood. While common polygenic risk explains ~28% of phenotypic variance, rare genetic variants—particularly those affecting chromatin regulation—are increasingly implicated. Notably, similar enrichment of rare variants in chromatin-modifying genes has been reported in autism spectrum disorder (ASD), suggesting overlapping molecular mechanisms across neurodevelopmental conditions.

ObjectiveTo investigate the contribution of rare de novo and inherited genetic variants to pediatric ADHD, with a focus on genes involved in chromatin modification.

Design/MethodsWe performed exome and genome sequencing in over 160 families of youth with primary ADHD (i.e., without comorbid ASD, intellectual disability, known genetic syndromes, or congenital anomalies), recruited through the Developmental Medicine Center at Boston Children's Hospital from 2019 to 2024. Families underwent deep phenotyping, including multi-informant clinical assessments, neuropsychological testing, and psychiatric family history interviews. Pathway enrichment analyses were performed, and selected variants were modeled in knock-in/knock-out cell lines and CRISPR-engineered mouse models.

ResultsConsistent with prior findings, 50% (80/160) of families harbored a likely genetic etiology (22% likely pathogenic, 28% VUS), with enrichment for de novo variants in simplex families (OR=7.22,

$p=1.49 \times 10^{-5}$) and rare inherited variants in multiplex families ($OR=1.11$, $p=0.0027$). Pathway analyses revealed significant enrichment for genes involved in chromatin modification—particularly histone methylation and acetylation ($p=2.16 \times 10^{-11}$). Multiple unique variants were identified in EP400, KMT2B, SETD1B, DIP2C, and DOT1L, which have also been implicated in ASD. Functional modeling demonstrated that loss or disruption of these genes leads to altered chromatin accessibility, dysregulated gene expression, and aberrant histone and/or DNA methylation patterns—suggesting a convergent epigenetic mechanism shared across ADHD and ASD.

Conclusions These findings support a substantial role for rare variants in chromatin-modifying genes in pediatric ADHD and highlight shared molecular pathways with ASD, pointing to disrupted epigenetic regulation as a common mechanism underlying neurodevelopmental disorders.

Session 59: Shaking The Evolutionary Tree: Multi-Species Studies to Understand Human Biology

Location: Room 205ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Thursday, October 16 at 1:30pm – 2:30pm

Single Cell Multiomics Across Nine Mammals Reveals Cell Type Specific Regulatory Conservation in the Brain

Subsession Time: Thursday, October 16 at 1:30pm – 1:45pm

Authors: Ashlyn G. Anderson (HudsonAlpha Institute for Biotechnology, University of Alabama at Birmingham), Brianne B. Rogers (HudsonAlpha Institute for Biotechnology), Erin A. Barinaga (HudsonAlpha Institute for Biotechnology), Jacob M. Loupe (HudsonAlpha Institute for Biotechnology), Elisa WaMaina (HudsonAlpha Institute for Biotechnology), S. Quinn Johnston (HudsonAlpha Institute for Biotechnology), Henry L. Limbo (HudsonAlpha Institute for Biotechnology), Lindsay F. Rizzardi (University of Alabama at Birmingham), Richard M. Myers (HudsonAlpha Institute for Biotechnology), Gregory M. Cooper (HudsonAlpha Institute for Biotechnology), J. Nicholas Cochran (HudsonAlpha Institute for Biotechnology)

Abstract: Understanding the gene regulatory mechanisms underlying brain function is crucial for advancing knowledge of the genetic basis of neurologic diseases. Cis-regulatory elements (CREs) play a pivotal role in gene regulation, and their evolutionary conservation can offer valuable insights. Importantly, the function of CREs is affected not only by primary sequence, but also by the cis- and trans-regulatory milieu. For this reason, measuring the function of CREs across cell types and species is likely to yield substantial insights over considering primary sequence conservation alone. In this study, we generated single nucleus multiomic (snRNA- and snATAC-seq) data from post-mortem cortex of nine mammals (human, macaque, marmoset, cat, rabbit, horse, cow, rat, and mouse). We characterized candidate CREs (cCREs) in a cell type-specific manner and developed a multi-dimensional conservation framework that incorporates sequence conservation, chromatin accessibility, and predicted enhancer-gene relationships. We assessed regulatory activity of hundreds of these elements using massively parallel reporter assays (MPRA) in iPSC-derived human neuronal precursor cells (NPCs) and neurons, selecting regions ranging from deeply conserved to human-specific. CRISPR interference (CRISPRi) in neurons further validated the regulatory roles of conserved enhancers, including those linked to neurodevelopmental genes such as FAM181B. Motif enrichment and chromBPnet modeling identified distinct transcription factors associated with conserved versus recently evolved cCREs and highlighted cell type-specific regulatory logic. LDSC analyses revealed that both conserved and human-specific CREs were enriched for neuropsychiatric disease risk loci, whereas enrichment for neurodegenerative traits was largely confined to conserved elements. We prioritized a subset of GWAS variants in MPRA-active CREs to identify disease-associated variants that modify enhancer activity. Our findings define core features of evolutionary constraint in brain CREs and reveal how conserved and recently evolved CREs

contribute to gene regulation and risk for neurologic disease. By integrating data from diverse mammals and using genomic and functional validation techniques, this study provides a deeper understanding of the evolutionary dynamics of gene regulation in the brain. This work permits a better understanding of how conserved regulatory mechanisms influence brain physiological function and disease.

Genetic control of regulatory dispersion across human and chimpanzee cell types

Subsession Time: Thursday, October 16 at 1:45pm – 2:00pm

Authors: Yoav Gilad

Abstract: Gene regulatory mutations are a powerful engine of phenotypic diversity across cell types, individuals, and species. For years, the study of gene regulation has been central to human genetics, yet nearly all efforts have focused on a single summary statistic: the mean expression level of a gene, averaged across a population of cells. Astonishingly, we have paid little attention to another key dimension of gene regulation: the variability in expression levels from cell to cell within the same population. This regulatory dispersion may hold the key to long-standing questions about penetrance, robustness, and tissue-specific vulnerability. One reason this dimension has remained understudied is the widespread assumption that most observed variability is technical noise rather than biologically meaningful. In addition, there has been deep skepticism that variance in gene regulation is itself under genetic control.

We designed a pair of studies to specifically ask whether regulatory variance is genetically controlled, and if so, how. We used single-cell RNA-seq in a novel experimental platform based on heterogeneous differentiated cultures (HDCs), which generate a broad spectrum of human cell types in vitro. We first asked whether regulatory dispersion is structured across cell types. We found that patterns of high and low dispersion are often shared among related cell types and are clearly non-random. Strikingly, genes that are highly expressed in a specific cell type tend to show lower dispersion in that same context. Although this defies statistical expectations, where higher mean is usually associated with higher variance, it suggests biological constraint where precise regulation is essential.

In a second set of experiments, we used a comparative model including human and chimpanzee iPSC lines, a fused human–chimpanzee hybrid line, and mixed-species HDCs. We show that interspecies differences in regulatory dispersion are driven by cis-regulatory divergence, revealing that expression noise itself is subject to evolutionary tuning through local sequence changes.

These findings establish regulatory dispersion as a meaningful, genetically encoded layer of gene regulation, with broad implications for understanding evolution, development, and disease.

Elucidating the metabolic mechanisms of cellular homeostatic resilience

Subsession Time: Thursday, October 16 at 2:00pm – 2:15pm

Authors: Matthew D Pollard (University of Massachusetts Chan Medical School), Diane P Genereux (Broad Institute of MIT and Harvard), L Safak Yilmaz (University of Massachusetts Chan Medical School), AJ Marian Walhout (University of Massachusetts Chan Medical School), Elinor K Karlsson (University of Massachusetts Chan Medical School, Broad Institute of MIT and Harvard)

Abstract: Mammals vary in how narrowly they regulate internal variables such as body temperature, blood oxygen, and blood glucose. Humans are strict regulators, requiring tight homeostatic control to avoid severe medical consequences. By contrast, flexibly homeostatic mammals are robust to conditions that challenge human cells: Camels allow their core temperature to rise to 41°C, fruit- and nectar-consuming bats tolerate elevated blood glucose, and diving mammals deplete their blood oxygen stores. Understanding the molecular basis of cellular resilience will inform methods to reduce human injury from conditions such as diabetes, hyperthermia, and ischemic stroke.

To discover the basis of cellular robustness, we leveraged a panel of primary dermal fibroblasts from 13 mammals, including human, culturing cells from five individuals per species across graded environmental perturbations (temperature: 32, 37, 41°C; glucose: 2.5, 8, 30 mM; hypoxia: 0, 6, 24 h). To quantify how these perturbations rewire metabolic pathway use, we integrated RNA-seq profiles for each individual with genome-scale metabolic models to estimate reaction-level fluxes. We then developed a phylogenetically informed mixed modeling and evolutionary simulation approach to identify differential responses between homeostasis strategies while accounting for statistical non-independence between related species.

We found metabolic responses that are universal across our study species and others that are diverged between flexible and strict regulators. Universal responses highlight deeply conserved programs activated by environmental stress: extreme temperature reduced flux through cholesterol and fatty acid homeostasis reactions, while hypoxia elevated flux through glycolysis reactions. Reactions with divergent flux responses represent candidate targets for improving resilience to environmental stress. For example, cold-flexible species exhibited distinct flux responses in arachidonic acid metabolism reactions, while heat-flexible species showed differences associated with keratan sulfate biosynthesis.

By integrating genome-scale metabolic flux modeling within a phylogenetic framework, we establish a broadly applicable approach to unravel the evolutionary and molecular bases of complex physiological traits. This work provides a powerful resource for future comparative genomic studies aimed at translating evolutionary insights into actionable targets for biomedical discovery.

The human and non-human primate developmental GTEx projects reveal tissue-specific gene expression across developmental stages

Subsession Time: Thursday, October 16 at 2:15pm – 2:30pm

Authors: Laura Domènech Salgado* (Broad Institute of MIT and Harvard, Cambridge MA, USA), Ricardo del Rosario* (Broad Institute of MIT and Harvard, Cambridge MA, USA), Fabien Degalez (Centre for Genomic Regulation, Barcelona, Spain), Irepan Salvador (Barcelona Supercomputing Center, Barcelona, Spain), Winona Oliveros (Barcelona Supercomputing Center, Barcelona, Spain), José Miguel Ramírez

(Barcelona Supercomputing Center, Barcelona, Spain), Edyta Malolepsza (Broad Institute of MIT and Harvard, Cambridge MA, USA), Lisa Anderson (Broad Institute of MIT and Harvard, Cambridge MA, USA), François Aguet (Broad Institute of MIT and Harvard, Cambridge MA, USA), Roderic Guigó (Centre for Genomic Regulation, Barcelona, Spain), Marta Melé, Kristin Ardlie (Broad Institute of MIT and Harvard, Cambridge MA, USA), and the dGTEX Consortium

Abstract: Despite major advances in functional genomics, most studies have focused on adult tissues, leaving early developmental stages largely uncharacterized. Moreover, the conservation of these developmental programs across non-human primates (NHPs) remains poorly understood, limiting our insights into both human evolution and the developmental origins of pediatric and adult onset diseases. Here, we present an analysis of the first data release of the developmental Genotype-Tissue Expression (dGTEX) Consortium, which integrates whole-genome sequencing (WGS) and bulk RNA sequencing (bulk RNA-seq) data from humans and two NHP species to explore tissue-specific gene expression across development and species.

The cohort includes: 364 tissues from 38 human donors spanning birth to adulthood; 384 tissues from 49 rhesus macaques (*Macaca mulatta*); and 318 tissues from 23 common marmosets (*Callithrix jacchus*), sampled at developmentally matched age groups across species.

Since NHP genomes are less well annotated than the human genome, we generated improved NHP annotations by merging Refseq and Ensembl data, and by incorporating lncRNA annotations derived from mapping the human lncRNA sequences from GENCODE v47 onto the primate genomes. These enhanced genome annotations led to the detection of a greater number of genes, and resulted in higher exonic and intragenic read rates, along with a lower intergenic read rate, in the bulk RNA-seq expression quantification.

Consistent with earlier studies, we observed that tissue-specific gene expression patterns are conserved across species. Differential gene expression analysis across developmental stages identified hundreds of genes whose expression vary with age -referred to as developmentally regulated genes, or dev-DEGs-, in humans, macaques and marmosets at specific organs.

Together, these findings lay a foundation for understanding developmental processes across humans and non-human primates and offer insight into the developmental origins of childhood disorders.

Session 69: Advances in Cell-Free DNA Applications from Prenatal Genomics to Cancer Diagnostics

Location: Room 205ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Friday, October 17 at 10:45am – 11:45am

Noninvasive fetal exome sequencing from maternal plasma: Large-scale validation and optimization for prenatal genetic use

Subsession Time: Friday, October 17 at 10:45am – 11:00am

Authors: Ieva Miceikaite (Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard), Harrison Brand (Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard), Christopher W. Whelan (Massachusetts General Hospital, Broad Institute of MIT and Harvard), John Lemanski (Massachusetts General Hospital), Michael Duyzend (Massachusetts General Hospital, Broad Institute of MIT and Harvard, Boston Children's Hospital), Krithika Nathamuni (Broad Institute of MIT and Harvard), Eren Shin (Massachusetts General Hospital), Mythily Ganapathi (Columbia University Irving Medical Center), Jessica Giordano (Columbia University Irving Medical Center), Vaidehi Jobanputra (New York Genome Center, Columbia University Irving Medical Center), Ronald Loosen (New York Genome Center), Monica Salani (Massachusetts General Hospital), Maya Pettes (University of Washington), Amira Yashruti (Brigham & Women's Hospital), Caroline Cusick (Broad Institute of MIT and Harvard), Dayne Filer (University of North Carolina at Chapel Hill), Stephanie Guseh (Brigham & Women's Hospital), Louise Wilkins-Haug (Brigham & Women's Hospital), Ignatia Van den Veyver (Baylor College of Medicine), Neeta L. Vora (University of North Carolina at Chapel Hill), Kathryn J. Gray (University of Washington), Wendy K. Chung (Boston Children's Hospital), Ronald Wapner (Columbia University Irving Medical Center), Michael Talkowski (Massachusetts General Hospital, Harvard Medical School, Broad Institute of MIT and Harvard)

Abstract: Noninvasive prenatal testing assesses fetal genetic risks by analyzing cell-free DNA (cfDNA) in maternal plasma, which contains both maternal and placental (fetal-representative) DNA. While current cfDNA-based platforms have transformed screening for common fetal aneuploidies, their low resolution restricts detection to a narrow set of conditions. Comprehensive prenatal diagnosis still requires invasive procedures such as amniocentesis or chorionic villus sampling, which many pregnant individuals decline due to concerns about procedural risks. This underscores the urgent need for a comprehensive, noninvasive alternative. We have developed a noninvasive fetal exome sequencing (NIFS) method using maternal plasma as a sole input, aiming to provide comprehensive detection of fetal aneuploidies, single nucleotide variants (SNVs), small insertions/deletions, and copy number variants (CNVs) that matches the sensitivity, precision, and interpretive power of standard invasive genome sequencing. In a proof-of-concept study, we previously demonstrated the technical feasibility of NIFS, achieving high concordance with invasive exome data in 11 matched samples for de novo and paternally inherited variants (median SNV sensitivity: 93%, precision: 93%). In a large-scale validation of NIFS against a field standard

prospective study of invasive fetal genome sequencing (PrenatalSEQ) and other cohorts, we have analyzed 563 maternal plasma samples (9–38 weeks' gestation) across a full range of fetal fractions, with matched fetal sequencing data for 83% of cases. Molecular optimizations yielded a mean target coverage of 900x. Novel analytic approaches to dissect fetal variation derived a de novo and paternally inherited variant detection sensitivity of ~99% and precision of ~97% for all coding variation captured by invasive genome sequencing. When comparing directly against pathogenic variants reported in standard clinical care, 97.5% (40/41) were confidently detected by NIFS, including 100% (n = 7) of rare CNVs >1 Mb. Follow-up analyses revealed that one missed diagnosis, a pathogenic NIPBL variant, was mosaic in a postnatal placental biopsy with only 15% allele fraction. These findings suggest that NIFS is a robust and scalable prenatal noninvasive genetic testing platform, capable of accurately detecting a broad spectrum of clinically relevant sequence variants currently only accessible through invasive diagnostic testing.

Performance of prenatal cell-free DNA screening to identify sex chromosome abnormalities: Experience of a commercial laboratory

Subsession Time: Friday, October 17 at 11:00am – 11:15am

Authors: Lisa M. Blazejewski (Quest Diagnostics), Tessa J. Niemchak (Quest Diagnostics)

Abstract: In 2022, the American College of Medical Genetics and Genomics (ACMG) published a Systematic Evidence Review (SER) on the use of prenatal cell-free DNA (cfDNA) screening in general risk pregnancies. The SER data were used to inform their 2023 Practice Guideline, which includes a recommendation that cfDNA screening be offered to patients with singleton gestations to screen for fetal sex chromosome abnormalities (SCAs). The SER meta-analysis found the positive predictive value of cfDNA screening for any SCA to be >50%, but these figures vary by condition, ranging from 29.5% (45,X) to 74.5% (47,XYY). Our aim was to assess the performance of the cfDNA assay developed by our laboratory for the detection of any SCA and compare it with the SER data.

We retrospectively analyzed data from all patients with cfDNA screening results from our laboratory during the study period that indicated increased risk for an SCA. Outcome data were obtained from results of follow-up fetal or neonatal cytogenetic testing (if performed at our institution) or information reported by ordering providers on a standardized outcome form. Data were summarized using descriptive statistics.

During the study period, 3,497 patients received a result indicating increased SCA risk. Outcome data were obtained for 893 patients (25.5%). Cytogenetic testing was performed for 656 patients; screening results were confirmed for 412 (62.8%). When stratified by condition, confirmation rates varied: 145/347 (41.8%) for 45,X; 76/83 (91.6%) for 47,XXX; 150/175 (85.7%) for 47,XXY; 41/51 (80.4%) for 47,XYY.

An additional 197 patients with increased risk for 45,X did not have cytogenetic testing, but the fetuses were suspected to have the condition based on ultrasound findings (e.g. cystic hygroma, fetal hydrops; n=49) or early miscarriage (n=148). If these cases represent true positives, the overall confirmation rate for 45,X is 62.9% (342/543).

For 5 patients, mosaic maternal aneuploidy explained the cfDNA results. For 2 patients, a different aneuploidy than that expected from the cfDNA results was found. For 3 patients with 45,X cfDNA results, the fetuses were found to have mosaic 47,XXX karyotypes, suggesting possible 45,X placental mosaicism. Finally, for 30 patients, providers reported that the patients terminated their pregnancies without confirmatory testing.

These data show robust performance of cfDNA screening to identify SCAs, with confirmation rates higher than those previously reported.

Prenatal cell-free DNA screening detects fetal genetic conditions linked to uniparental disomy

Subsession Time: Friday, October 17 at 11:15am – 11:30am

Authors: Xing Qiu (Zhejiang University), Qiong Luo (Zhejiang University), Luming Sun (Shanghai Jiao Tong University), Dan Zhang (Zhejiang University), He-Feng Huang (Zhejiang University), Chenming Xu (Fudan University), Jinglan Zhang (Zhejiang University)

Abstract: BACKGROUND: Current prenatal cell-free DNA (cfDNA) screening primarily is designed to screen aneuploidies and chromosomal segmental copy number variations (CNVs) but cannot identify genetic conditions in chromosomally balanced euploid fetuses, such as uniparental disomy (UPD). UPD may result in imprinting disorders, including Prader-Willi, Angelman, and Beckwith-Wiedemann syndromes, all associated with substantial clinical morbidity. The estimated prevalence of UPD-related conditions is approximately 1 in 2,000 live births, surpassing that of trisomy 13 (1 in 12,000) and trisomy 18 (1 in 4,000). This study aims to develop a cfDNA-based method for the noninvasive prenatal detection of UPD-mediated conditions, extending beyond traditional screening for aneuploidies and CNVs.

METHODS: High-depth targeted sequencing (average coverage >800x) of maternal plasma cfDNA was performed on selected chromosomal regions to detect common aneuploidies, CNVs, and UPD involving imprinting regions. For copy-number neutral samples, a custom hidden Markov model was applied to infer the parental origin of fetal alleles using genotype likelihoods across informative SNPs. High-risk UPD was defined as absence of biparental inheritance on disomic chromosomes or presence of aneuploidies or CNVs involving imprinting regions on chromosomes 6, 7, 11, 14, 15, or 20. Orthogonal validation was performed using parental peripheral blood and fetal genomic DNA from amniocytes or placental tissue, followed by SNP microarray and methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA).

RESULTS: A total of 493 retrospective maternal plasma samples were analyzed from a cohort enriched for UPD, all with available diagnostic confirmation. The mean gestational age at sampling was 17.8 weeks, and the average fetal fraction was 13.4%. Prenatal cfDNA analysis identified 472 true-negative and 10 true-positive cases, including 3 aneuploidies, 2 uniparental isodisomies, and 5 uniparental heterodisomies. Two discordant results were observed: one false-negative caused by low-level mosaic UPD, confirmed through five biopsies from a single placenta, and one false-positive associated with parental consanguinity. The assay achieved a clinical sensitivity of 91.0% (95% CI: 57.1–99.5%) and

specificity of 99.8% (95% CI: 98.6–100%). An additional 9 cases exhibited extensive homozygosity in non-imprinting regions, all verified by fetal genomic DNA analysis.

CONCLUSIONS: Combined chromosomal dosage and SNP-based biparental inheritance analysis using high-depth cfDNA sequencing enables accurate, noninvasive prenatal detection of UPD, warranting further studies to establish its clinical validity and utility for population-based screening and improved management of UPD-related conditions.

Genome-Wide Variations of End Motif in Cell-Free DNA Fragments Distinguish Immunotherapy Responders from Non-Responders in Head and Neck Cancer: A Multi-Institute Prospective Study

Subsession Time: Friday, October 17 at 10:45am – 11:45am

Authors: Ravi Bandaru (Northwestern University), Hailu Fu(Northwestern University), Haizi Zheng(Cincinnati Children’s Hospital Medical Center), Li Wang(Northwestern University), Shuchi Gulati(University of Cincinnati College of Medicine), Mingxiang Teng(Moffitt Cancer Center), Bin Zhang(Cincinnati Children’s Hospital Medical Center), David A. Hildeman(Cincinnati Children’s Hospital Medical Center), Trisha Wise-Draper(University of Cincinnati College of Medicine), Yaping Liu(Northwestern University)

Abstract: Reliable, minimally invasive biomarkers for predicting immunotherapy response in head and neck squamous cell carcinoma (HNSCC) remain an unmet clinical need. Here, in a prospective, multi-institutional phase II clinical trial (NCT02641093), we performed whole-genome sequencing of 185 plasma cell-free DNA (cfDNA) samples collected longitudinally from 68 patients with locally advanced, surgically resectable HNSCC undergoing neoadjuvant and adjuvant pembrolizumab treatment. We developed the regional motif diversity score (rMDS), a novel fragmentomic metric quantifying the entropy of cfDNA 5’ end motifs across genomic regions. Remarkably, unsupervised analysis revealed that rMDS robustly distinguished immunotherapy responders from non-responders, outperforming established cfDNA fragmentomic metrics and copy number alterations, while demonstrating independence from technical confounders. Longitudinal analysis revealed dynamic rMDS changes in genomic regions enriched for immune-, hemoglobin-, and keratinization-related genes—hallmarks of squamous cell carcinoma—reflecting the interplay between tumor and peripheral immunity during the immunotherapy treatment. Interestingly, the regions with the most dynamic rMDS changes were highly enriched in telomere-proximal loci, suggesting a novel link between telomere biology and cfDNA fragmentation. A machine learning classifier based on rMDS achieved robust predictive performance across multiple validation settings (AUC 0.92-0.94), with the highest accuracy at post-treatment timepoints and superior to PD-L1 expression and tumor fraction. Predicted responders demonstrated trends toward improved disease-free survival (DFS), with significant differences between high-risk non-responders and intermediate-risk responders (DFS: $p = 0.002$; overall survival: $p = 0.011$), underscoring the clinical utility of rMDS-based stratification. These findings position rMDS as a biologically meaningful and clinically actionable biomarker for immunotherapy response in HNSCC, supporting its integration into future risk assessment frameworks and broader cancer care.

Session 70: Contextualizing Somatic Variation: From Clonal Expansion to Molecular Mechanisms

Location: Room 258ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Friday, October 17 at 10:45am – 11:45am

★ Landscape of somatic mutations across genomes, tissues, and individuals

Subsession Time: Friday, October 17 at 10:45am – 11:00am

Authors: Huixin Xu (Lewis Sigler Institute for Integrative Genomics, Princeton University), Rob Bierman (Lewis Sigler Institute for Integrative Genomics, Princeton University), Dayna Akey (Lewis Sigler Institute for Integrative Genomics, Princeton University), Joshua M. Akey (Lewis Sigler Institute for Integrative Genomics, Princeton University)

Abstract: A human zygote undergoes trillions of cell divisions to form a fully differentiated multicellular organism. This complex developmental process provides many opportunities for somatic mutations to arise. Even in terminally differentiated cells, somatic mosaicism can continue to accumulate through non-replicative processes, such as the spontaneous deamination of methylated cytosines at CpG dinucleotides. Thus, the burden of somatic variation in humans is expected to be substantial. In addition, somatic mutations contribute to cancer, aging, neurodegeneration, and a growing number of other diseases. Despite its biological and clinical significance, our understanding of somatic mosaicism in healthy individuals remains limited. Key challenges include the lack of comprehensive datasets that profile multiple tissues from the same individuals and the absence of statistical models that account for the complex structure of such data. In this study, we analyze somatic mutations using a high-coverage (~160X) whole-genome sequencing dataset of 300 tissue samples spanning seven tissues across 55 GTEx donors. Leveraging a novel probabilistic framework, we identified a comprehensive catalog of somatic variants to estimate their burden and mutational spectra across tissues and individuals. We show that the burden of deleterious somatic mosaicism varies widely across individuals and tissues and that age is significantly related to the burden of somatic mosaicism, but only in a subset of tissues. Interestingly, we find that the rate of somatic mutations in non-coding DNA is significantly higher ($p = 10^{-9}$) than the rate of somatic mutations in coding DNA in all tissues except sun-exposed skin. We show that UV-light-induced somatic mutations in coding regions drive the higher burden of coding somatic variants in sun-exposed skin. We also observe that the mutational spectrum of somatic variants is significantly different between coding and noncoding regions, and explore how chromatin structure contributes to this difference. Finally, we also identify somatic mosaicism within mitochondrial DNA (mtDNA) and report widespread heteroplasmy across tissues. In summary, our data provides a high-resolution compendium of somatic variants across genomes, tissues, and individuals, and offers novel biological insights into mutational mechanisms.

Clonal Hematopoiesis of Indeterminate Potential (CHIP) Mutations Are Enriched in Alzheimer's Disease Patients with APOE $\epsilon 3/\epsilon 3$ Genotype

Subsession Time: Friday, October 17 at 11:00am – 11:15am

Authors: Jaejoon Choi (Boston Children's Hospital, Harvard Medical School), Kyung Sun Park (Kyung Hee University Medical Center), Yann Le Guen (Stanford University), Jong-Ho Park (Seoul National University Bundang Hospital), Zinan Zhou (Boston Children's Hospital, Harvard Medical School), Liz Enyenihi (Boston Children's Hospital, Harvard Medical School), Ila Rosen (Boston Children's Hospital, Harvard Medical School), Yoon-Ho Choi (Samsung Medical Center), Christopher A. Walsh (Boston Children's Hospital, Harvard Medical School), Jong-Won Kim (Samsun Medical Center), August Yue Huang (Boston Children's Hospital, Harvard Medical School), Eunjung Alice Lee (Boston Children's Hospital, Harvard Medical School)

Abstract: Alzheimer's Disease (AD) is a neurodegenerative disorder, with advanced age and the APOE $\epsilon 4$ allele as its strongest risk factors. Somatic mutations accumulate with age, and our studies have identified an increased somatic mutation burden in AD brains, particularly in genes implicated in clonal hematopoiesis of indeterminate potential (CHIP). However, findings in AD blood have been inconsistent, making it challenging to understand the relevance of CHIP mutations in AD pathogenesis.

Here, we analyzed 315 peripheral blood samples from individuals with and without AD. Using molecularly barcoded deep (~400X) gene panel sequencing, we identified significant enrichment of somatic single nucleotide variants (sSNVs) and somatic indels (sIndels) in CHIP genes within the blood of AD patients compared to age-matched controls. Interestingly, a higher frequency of CHIP mutations in AD blood compared to controls was observed only in individuals with the APOE $\epsilon 3/\epsilon 3$ genotype, whereas those with the $\epsilon 3/\epsilon 4$ genotype showed no such increase. Additionally, CHIP mutations in AD blood samples accumulated at an accelerated rate and exhibited a higher dN/dS ratio, indicating evidence of stronger positive selection.

We further validated these findings in a larger, independent dataset from the Alzheimer's Disease Sequencing Project (ADSP), which consists of the blood whole-genome sequencing of ~30,000 AD and control individuals. Consistent with our observations from panel sequencing, AD blood samples exhibited a significantly higher burden of CHIP mutations compared to controls, especially in non-APOE $\epsilon 4$ genotypes.

Our findings suggest that somatic CHIP mutations in blood are common in normal aging but are further enriched in AD. This aligns with previous findings of the enrichment of CHIP mutations in AD microglia, suggesting a potential crosstalk between blood- and brain-resident macrophages. Our study provides valuable insights into the clonal dynamics of blood in AD pathogenesis and identifies potential avenues for developing new diagnostic and therapeutic approaches.

Multi-ethnic Transcriptomic Analysis of Clonal Hematopoiesis Reveals Driver-Specific Pathways and Causal Associations

Subsession Time: Friday, October 17 at 11:15am – 11:30am

Authors: Linke Li* (Massachusetts General Hospital), Seyedeh Zekavat* (Massachusetts Eye and Ear Hospital), Xinyu Wang (Harvard T.H. Chan School of Public Health), Buu Truong (Harvard T.H. Chan School

of Public Health), Alexander Bick (Vanderbilt University Medical Center), Jerome Rotter (The Lundquist Institute), Stephen Rich (University of Virginia), Yongmei Liu (Duke University), Kristin Ardlie (Broad Institute of MIT and Harvard), Francois Aguet (Broad Institute of MIT and Harvard), Kent Taylor (The Lundquist Institute), Xiaohui Li (The Lundquist Institute), Wendy Post (Johns Hopkins University School of Medicine), Ramachandran Vasan (University of Texas Health San Antonio), Alex Reiner (University of Washington), Pinkal Desai (Weill Cornell Medicine), Eric Whitsel (University of North Carolina), Charles Kooperberg (Fred Hutch Cancer Center), Daniel Levy (National Heart, Lung, and Blood Institute), JoAnn Manson (Brigham and Women's Hospital), Siddhartha Jaiswal (Stanford University School of Medicine), Joanne Murabito (Boston University), Karen Conneely (Emory University School of Medicine), Pradeep Natarajan* (Massachusetts General Hospital), Zhi Yu* (Massachusetts General Hospital)

Abstract: Introduction Clonal hematopoiesis of indeterminate potential (CHIP), the age-related clonal expansion of hematopoietic stem cells with somatic mutations, is a major risk factor for diverse pathologies, yet its underlying molecular mechanisms that contribute to these risks are not well defined. Hence, this study investigates the blood transcriptomic signatures of CHIP.

Methods We analyzed RNA-Seq data from 4,028 participants across 4 NHLBI TOPMed cohorts: Multi-Ethnic Study of Atherosclerosis, Framingham Heart Study, Genetic Epidemiology of Chronic Obstructive Pulmonary Disease, and Women's Health Initiative. Cohort-specific linear mixed-effects models were used to examine associations between blood transcript levels and CHIP (composite and its common drivers DNMT3A, TET2, ASXL1), followed by a random-effect meta-analysis, pathway enrichment analyses, and two-sample Mendelian randomization (MR) using eQTLGen instruments and CHIP GWAS on significant associations.

Results Among participants (mean age 62 years, 69% female, 71% White), 5.1% had CHIP. We identified 319 significant (FDR <0.05) transcript-CHIP associations, driven mainly by TET2 (149; 74% downregulated) and ASXL1 (88; 77% upregulated). Our analysis revealed driver-specific transcriptomic signatures, with findings for TET2 and ASXL1 aligning with previous proteomic studies. Specifically, TET2 were associated with upregulated immune activation transcripts (e.g., IL1R1; β SD=0.28, FDR=0.01); ASXL1 were strongly linked to cell adhesion transcripts (e.g., FN1; β SD=1.06, FDR<0.001); and DNMT3A showed associations with cellular transport transcripts (e.g., KRT18; β SD=0.27, FDR<0.001). Pathway analysis (p<0.05) revealed 17-fold enriched immune response pathway TET2 mutations, 33-fold enriched downregulated extracellular matrix organization for ASXL1, and 50-fold enriched vesicle trafficking for DNMT3A. MR analysis provided causal evidence for several protective associations (p<0.001), linking genetically predicted increases in CDK18 expression (a regulator of genomic stability) to a reduced risk of ASXL1-CHIP (β MR=-2.34); and separately, increased ALDH2 expression (critical for HSC function) to a reduced risk of TET2-CHIP (β MR=-0.33).

Conclusion Our multi-cohort, multi-ethnic transcriptomic analyses with causal inference revealed driver-specific transcriptomic signatures of CHIP implicating genomic stability and immune pathways, shedding light on the potential molecular mechanisms linking CHIP and diseases.

★ Population scale single-cell multi-omic analysis of clonal haematopoiesis

Subsession Time: Friday, October 17 at 11:30am – 11:45am

Authors: Zhen Qiao (Garvan Institute of Medical Research, University of New South Wales), Michael Geaghan (Garvan Institute of Medical Research), Yves Fontaine (Garvan Institute of Medical Research), Hao Lawrence Huang (Garvan Institute of Medical Research), Oscar Dong (Garvan Institute of Medical Research), Peter C. Allen (University of New South Wales, Garvan Institute of Medical Research), Jianan Fan (University of New South Wales, Garvan Institute of Medical Research), Angli Xue (University of New South Wales, Garvan Institute of Medical Research), Hope A. Tanudisastro (University of New South Wales, Centre for Population Genomics, Garvan Institute of Medical Research, Murdoch Children's Research Institute, University of Sydney), Anna S.E. Cuomo (University of New South Wales, Garvan Institute of Medical Research, Centre for Population Genomics, Murdoch Children's Research Institute, UNSW Cellular Genomics Futures Institute), Blake Bowen (University of New South Wales, Garvan Institute of Medical Research), Albert Henry (University of New South Wales, Garvan Institute of Medical Research), Anne Senabouth (University of New South Wales, Garvan Institute of Medical Research), Giulio Genovese (Broad Institute of MIT and Harvard, Stanley Center for Psychiatric Research), Zhili Zheng (Broad Institute of MIT and Harvard, Massachusetts General Hospital), Shannon Ji (Garvan Institute of Medical Research), Michael Harper (University of New South Wales, Centre for Population Genomics, Garvan Institute of Medical Research, Murdoch Children's Research Institute), Etienne Masle-Farquhar (University of New South Wales, Garvan Institute of Medical Research), Eleanor Spenceley (University of New South Wales, Garvan Institute of Medical Research), Eszter Sagi-Zsigmond (Garvan Institute of Medical Research), Caitlin Bartie (Garvan Institute of Medical Research), Agota Tuzesi (Garvan Institute of Medical Research), Julia Forkgen (Garvan Institute of Medical Research, University of New South Wales), Michael Silk (University of New South Wales, Centre for Population Genomics, Garvan Institute of Medical Research, Murdoch Children's Research Institute), Matthew Hobbs (Garvan Institute of Medical Research), Joe Coptly (Garvan Institute of Medical Research), Matt A Field (Garvan Institute of Medical Research, James Cook University), Andrew D. Calcino (James Cook University), Caitlin Uren (University of New South Wales, Centre for Population Genomics, Garvan Institute of Medical Research, Murdoch Children's Research Institute), Hannah R. Nicholas (Centre for Population Genomics, Garvan Institute of Medical Research, Murdoch Children's Research Institute), Kristof Wing (Menzies Institute for Medical Research, University of Tasmania, Royal Hobart Hospital, Centre for Eye Research Australia, University of Melbourne), Owen Tang (Charles Perkins Centre, The University of Sydney, Kolling Institute of Medical Research, Royal North Shore Hospital), Michael Gray (Charles Perkins Centre, The University of Sydney, Kolling Institute of Medical Research, Royal North Shore Hospital), Katrina M. de Lange (Centre for Population Genomics, Garvan Institute of Medical Research, Murdoch Children's Research Institute), Alex W. Hewitt (Menzies Institute for Medical Research, University of Tasmania, Royal Hobart Hospital, Centre for Eye Research Australia, University of Melbourne), Gemma A. Figtree (University of Sydney, Charles Perkins Centre, The University of Sydney, Kolling Institute of Medical Research, Royal North Shore Hospital), Daniel G. MacArthur (University of New South Wales, Centre for Population Genomics, Garvan Institute of Medical Research, Murdoch Children's Research Institute), Joseph E. Powell (University of New South Wales, Garvan Institute of Medical Research, UNSW Cellular Genomics Futures Institute), Owen M. Siggs (Garvan Institute of Medical Research, University of New South Wales)

Abstract: Clonal haematopoiesis (CH) is an age-related process characterised by the accumulation of somatic mutations in haematopoietic stem cells that drive clonal expansion. CH has been implicated in a wide variety of adverse clinical outcomes, including increased risks of haematological malignancies and cardiovascular disease, although the underlying mechanisms remain incompletely understood.

As part of the TenK10K Phase 1 project, we leveraged comprehensive multiomic data from over 2,000 individuals and 5 million peripheral blood mononuclear cells (PBMCs), including both healthy controls and individuals with cardiovascular disease, to elucidate the biological impact of CH. By integrating sample-matched whole-genome sequencing (WGS) and single-cell RNA sequencing (scRNA-seq) data, along with single-cell ATAC-seq (scATAC-seq) and genotyping array data for a subset of over 1,000 individuals, we systematically identified diverse forms of CH. These included somatic single nucleotide variants and short indels in known CH driver genes, megabase-scale mosaic chromosomal alterations (mCAs), mosaic loss of sex chromosomes, as well as individuals with high somatic mutational burden lacking known CH drivers or mCAs. We found that the boundaries and cell fractions of WGS-derived mCA events were highly concordant with those inferred from genotyping arrays, with WGS offering improved sensitivity for detecting smaller or lower-frequency events.

To identify cell types associated with the genetic causes of CH, we applied trait–cell type mapping methods that integrate summary statistics from large-scale GWAS of CH with single-cell transcriptomic and chromatin accessibility profiles. Additionally, using summary statistics from cell-type–specific expression quantitative trait loci (eQTL) and chromatin accessibility QTLs (caQTL), we identified CH-associated genes and pathways in specific cell types through colocalisation and causal inference methods. These analyses offer novel insights into the cellular contexts and regulatory mechanisms underlying genetic predisposition to CH.

We further leveraged matched scRNA-seq and scATAC-seq data to assess mosaic loss of the Y chromosome (LOY), the most prevalent subtype of CH, at single-cell resolution in 974 male individuals. We consistently observed that LOY is enriched in monocytes, regulatory T cells, natural killer cells, and dendritic cells. Through detailed characterisation of its lineage distribution, we found that LOY may disrupt immune cell differentiation. For example, we observed a gradual decline in LOY frequency along the B cell trajectory, from naive to intermediate and memory B cells, indicating impaired maturation within the B cell compartment. Further analyses of cell states and differential gene expression revealed LOY-associated functional dysregulation in immune cells, offering mechanistic insights into its contribution to disease risk.

In summary, we present a population-scale multiomic resource for studying somatic mutations and demonstrate its utility in advancing our understanding of CH biology and its potential contribution in disease.

Session 71: From Classroom to Community: Shaping the Future of Genomics Education

Location: Room 206AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Friday, October 17 at 10:45am – 11:45am

PaRtnErIng to build understanding oF gEnomics Responsibly: the PREFER CHW research education program

Subsession Time: Friday, October 17 at 10:45am – 11:00am

Authors: Marie Smith, MS (Clemson), Caitlin Allen, PhD (Wake Forest), Greg Green (Center for Community Health Alignment- University of South Carolina), Paige Menking (Ponderosa Health), Paula Ramos (Emory), Kelsey Jaeger (Wake Forest), Susan Mayfield-Johnson (USM), Honey Estrada (NCCHWA), Charkarra Anderson-Lewis (USM), Ashley Evans (SCSU), Yue Guan (Emory), Laurie Hoffma (TNCHWA), Sarah Covington-Kolb (USC), Tamika Williams (NCCHWA), Cassandra Wright (NCCHWA)

Abstract: Introduction:

Community Health Workers (CHWs) have seen unprecedented growth in recent years. While CHWs are from the communities they serve and regularly interact with diverse populations, they have yet to receive appropriate training in genomics. To address this need, we have developed the PREFER CHW Genomics Research Education Program.

Methods:

The PREFER CHW program was developed in collaboration with 5 partner sites across the SE United States: The Center for Community Health Alignment, NC CHW Association, SC State University, Southern MS University, and the TN CHW Association. The PREFER CHW curriculum, centered around 10 key genomics competencies, and was refined through a 3-round Delphi process involving input from key stakeholders. It has been implemented for the first 3 cohorts, with the goal to train a total of 150 CHWs. Partner sites accompanied by genetics content experts have been trained to deliver the curriculum and are equipped to continue offering and evaluating the program. Participant surveys are conducted before, immediately after and 3 months following the course to assess course impact, self-efficacy, and knowledge retention. Ongoing evaluation of the program will be conducted qualitatively through trainer and CHW participant interviews.

Results:

Curriculum Development: The 1st round of the Delphi process included 20 participants, 86% of which were CHWs. Modules with the lowest agreement (<75%) were Genetics 101 and CHW Roles. The 2nd round of the Delphi process revealed key themes including emphasis on the CHW role and real-word application and simplifying technical content. The 3rd and final round revealed gaps in the application of adult learning theory and the accompanying workbook. The final curriculum includes 10 modules.

Implementation: The first 3 cohorts of the PREFER training program were initiated in early June 2025 with 88 total participants. Most participants were from MS (34%) or TN (38%), female (83%), black (72%), and non-Hispanic/non-Latino (82%). Motivators for interest in the course included a desire to learn about genetics, improve as CHWs, improve healthcare for community members and learn more about health research. Themes for overall course goals included learning about genetics, gaining new skills to better serve communities, and improving their ability to share the benefits of research within their community.

Conclusion:

PREFER CHW offers a sustainable solution to training CHWs in genomics competencies, helping to enhance the capacity of the entry-level genomics research workforce and improve the diversity of the genomics research field.

Psychometric Validation of the Education and Assessment of Genetic Literacy (EAGL) Scale

Subsession Time: Friday, October 17 at 11:00am – 11:15am

Authors: Chris Gunter (National Human Genome Research Institute), Lily S. Barna (National Human Genome Research Institute), Yi Liao (Huntsman Cancer Institute, Department of Communication, University of Utah), Gabriela M. Ramírez-Renta (University of Colorado Anschutz Medical School), Kimberly Kaphingst (Huntsman Cancer Institute, Department of Communication, University of Utah)

Abstract: Genetic literacy, defined as “the sufficient knowledge and understanding of genetic principles for individuals to make decisions that sustain personal well-being and effective participation in social decisions on genetic issues,” is an integral measure to examining society’s interaction with genetics. Widely-used genetic literacy measures lack knowledge comprehension, a crucial part of this definition, as well as psychometric validation. To address these issues, we validated the Education and Assessment of Genetic Literacy scale (EAGL) in a sample of 2708 US participants, using both exploratory and confirmatory factor analysis (EFA and CFA).

In addition to standard subjective and objective knowledge subscales, our measure’s distinct knowledge comprehension subscale focused on autism as an example of a complex condition with environmental and genetic components. Strong scores across those questions confirmed our participants’ ability to understand this model when presented with sufficient information. Regression analyses found a statistically significant interaction when looking at education and connection to autism (identifying yourself or a family member as being autistic) in relation to knowledge comprehension ($F=3.11$, $p=0.008$). Separately, those in our sample with a connection to autism scored higher on the unrelated subjective knowledge section ($F=17.96$, $p<0.001$), but not in the other subscales, concurring with previous demonstrations of a subjective-objective knowledge gap in science literacy.

We explored geographic location as one potential factor in genetic literacy and found that metropolitan vs non-metropolitan status had no significant main effects on overall levels. There was a statistically significant interaction effect between education and metro status for objective knowledge only ($F=2.22$,

$p=0.050$), suggesting potential subtle educational differences that may exist between geographic areas. The most missed question regardless of any demographic variable centered on polygenic traits, indicating an immediate target for future education. After the validation process, we have two multi-domain measures available for wide use: the multi-faceted EAGL-long, which has previously been tested in thousands of participants, or the validated three-factor EAGL-short. The new measure continues to capture three subscales crucial for accurate representation of genetic literacy, the construct needed to understand and productively implement genetic applications.

Learning Competencies and Pedagogical Considerations for the Future Public Health Genetics Workforce

Subsession Time: Friday, October 17 at 11:15am – 11:30am

Authors: Alison E Fohner (University of Washington), Diane Xue (University of Pennsylvania), Joon-Ho Yu (University of Washington)

Abstract: Background. The Public Health Genetics (PHG) workforce serves crucial roles in genomic discoveries and their ethical and equitable translation into the public sphere including public health programs, clinical genetics, and implementation research. As both genomic science and translation advance, PHG training programs must evolve to meet the changing needs of the PHG landscape, yet the field's interdisciplinary nature poses challenges to curricula development and pedagogical practice.

Methods. Using convenience sampling, we invited 114 PHG-related professionals to participate in focus group discussions related to the skills, knowledge, and training needed for a career in PHG based on their experiences in the field. Additional questions solicited perspectives on curriculum structure, analytic skills, and unneeded content. Discussion transcripts were analyzed for thematic content and subsequently used to develop learning competencies and pedagogical recommendations.

Results. Focus groups included 55 professionals (48% recruitment rate) across academia (80%), government (9%), and industry (<5%). Participants were 71% female and represented bioethics (55%), policy (45%), genetic epidemiology (42%), medical genetics (40%), statistical genetics (35%), and law (27%). Discussions led to 17 learning competencies across four learning goals: 1) Foundational knowledge in genetics and public health (e.g. principles of population genetics, history of unethical genomic practices); 2) Data and analytical proficiency across PHG sub-domains (e.g. using qualitative and quantitative datasets); 3) Application of interdisciplinary methods to research design and practice (e.g. developing research questions at the intersection of genomics and health services; and 4) Cross-disciplinary collaboration (e.g. cultural and disciplinary humility in communication). Informants recommended that PHG programs develop pedagogical priorities including strong mentorship, broad core requirements plus specialization tracks, and hands-on, cross-disciplinary projects. Informants highlighted the need for conceptual knowledge and critical thinking, rather than technical proficiency in specific methods.

Conclusions. Current experts in PHG offered guidance for training programs that will best meet the evolving needs of the PHG field. PHG training programs should emphasize broad foundational knowledge and team science skills to facilitate cross-disciplinary communication and meet emerging challenges.

How Epigenetics is Communicated on YouTube: A Content Analysis of Science Communication Strategies

Subsession Time: Friday, October 17 at 11:30am – 11:45am

Authors: Aantaki Rasa (Washington University School of Medicine in St. Louis), Jada G. Hamilton (Memorial Sloan Kettering Cancer Center), Irania Santaliz Moreno (Washington University School of Medicine in St. Louis), Erika A. Waters (Washington University School of Medicine in St. Louis)

Abstract: Background: Effective communication of epigenetic information to lay audiences is critical as epigenetic therapeutics, direct-to-consumer products, and disease-related applications become more prevalent. While social media platforms like YouTube are increasingly used to disseminate epigenetic information, the sources and strategies behind this dissemination remain understudied.

Purpose: This study examines how epigenetic information is communicated on YouTube.

Method: We conducted a content analysis of 294 YouTube videos published between 2008-2023, applying an a priori codebook based on science communication best practices. Videos were transcribed for analysis and the transcripts were independently coded by two coders on NVivo. Disagreements (if any) were resolved by weekly discussions. Additionally, we assessed channel affiliations, narrator demographics, visual strategies, and rhetorical techniques (e.g., metaphors, scientific examples) of the videos. Readability of transcripts of the videos was evaluated using the Flesch-Kincaid grade score.

Results: Most videos (54.4%) were hosted by online-only platforms (e.g., Veritasium, SciShow), followed by (27.6%) online-only personal accounts (e.g., McDowell, Raw Maraby). Fewer videos were hosted by platforms affiliated with governmental agencies like the NIH (0.7%), educational institutions like Johns Hopkins (3.1%), or healthcare industries like Blue Cross Blue Shield (3.7%). The videos frequently featured male-presenting (34.7%), white-presenting (44.9%) narrators compared to female presenting (22.1%), non-white-presenting (13.3%) narrators. Multimodal visual strategies dominate, including animated graphics with voiceovers (37.4%) and on-screen presenters (40.1%) compared to slideshows with voice over (16%), conversation between multiple people (16%), and graphics or text with no voice over (4.4%). Common communication techniques included metaphors/analogies (30.6%), scientific examples (27.9%), and storytelling (22.1%). However, 21.8% of the videos used undefined jargon, and 92.8% failed to acknowledge scientific uncertainty. Readability averaged 10.7 (above the NIH-recommended 8th grade level).

Conclusions: Epigenetics communication on YouTube often omits discussion of scientific uncertainties and exceeds recommended readability levels. These findings highlight gaps between current practices and evidence-based guidelines, underscoring the need for clearer, more accessible messaging.

Session 72: Functional Consequences of Transcriptome Variation Across Tissues and Populations

Location: Room 253ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Friday, October 17 at 10:45am – 11:45am

Multi-modal dissection of coronary artery disease GWAS loci reveals transcript splicing-mediated risk mechanisms in vascular smooth muscle cells

Subsession Time: Friday, October 17 at 10:45am – 11:00am

Authors: Noah Perry (Civelek Lab, UVA), Jennifer Korchak (Sheynkman Lab, UVA), Mayank Murali (Sheynkman Lab, UVA), Faruk Erin (Civelek Lab, UVA), Ben Jordan (Sheynkman Lab, UVA), Nicolas Barbera (Civelek Lab, UVA), Gloria Sheynkman (University of Virginia, Department of Molecular Physiology and Biological Physics), Mete Civelek (University of Virginia, Department of Genome Sciences)

Abstract: Smooth muscle cells (SMCs) play a pivotal role in atherosclerosis, the underlying cause of coronary artery disease (CAD), through their ability to adopt diverse phenotypic states that influence both plaque stability and progression. Regulatory mechanisms guiding SMC phenotypes remain incompletely understood, especially at the post-transcriptional level. Alternative splicing (AS), a major contributor to transcriptomic and proteomic diversity, is a critical yet underexplored mechanism linking non-coding genetic variation to disease risk. We implemented a targeted prioritization pipeline to identify splice-altering variants that colocalize with CAD GWAS loci and show in silico splice disruption. We identified 23 novel SMC sGenes colocalized with CAD GWAS loci and combined these with previous results to study 169 colocalized sGenes. We found 1,555 variants that directly share both splicing and CAD genetic signals. We prioritized 115 of these variants to be causal of affecting splicing machinery according to their location within known splice protein motifs. Candidates were filtered using SMC-specific long-read (LR) RNA sequencing to confirm associated junctions. LR sequencing identified over 190,000 isoforms, nearly 80,000 of them novel. We identified nine CAD loci with GWAS SNPs at splice sites predicted to affect isoforms found by LR sequencing. Isoform-specific PCR validated genotype-dependent splicing in SMCs for DCLRE1B, ITGA1, PARP12, and WIP1. Protein modeling showed these AS events disrupt key domains. Exon 10 skipping in PARP12 caused loss of the catalytic domain; intron retention in ITGA1 produced a truncated isoform lacking transmembrane and cytoplasmic regions; exon 2 skipping in DCLRE1B removed the DNA repair domain. Overexpression of PARP12 isoforms in VSMCs revealed isoform-specific effects on NF- κ B signaling. Gene set enrichment identified TNF α signaling via NF- κ B as the most enriched difference. ELISA of phosphorylated p65 confirmed increased NF- κ B activation in canonical versus exon 10 skipping isoform-overexpressing SMCs. We linked rs2286196 at the PARP12 locus to AS regulation through predicted SRp40 binding at an exon splice enhancer. The CAD risk allele (G) promotes exon 10 inclusion and upregulates inflammatory signaling. These findings establish AS as a mediator of CAD risk in VSMCs and demonstrate the value of isoform-resolved approaches for interpreting GWAS loci.

Isoform-Level Transcriptomic Signatures of Glycemic Traits

Subsession Time: Friday, October 17 at 11:00am – 11:15am

Authors: Jennifer Kang (Framingham Heart Study, NHLBI), Tianxiao Huan (Framingham Heart Study, NHLBI), Roby Joehanes (Framingham Heart Study, NHLBI), Daniel Levy (Framingham Heart Study, NHLBI)

Abstract: Introduction: Alternative splicing produces diverse RNA isoforms that may be a direct cause or consequence of disease. The role of splice variation in the pathogenesis of dysglycemia and type 2 diabetes (T2D) is not well understood. To address this, we identified isoform-level expression signatures of four glycemic traits.

Methods: Whole blood RNA-seq was analyzed from two cohort studies: the Framingham Heart Study (FHS, discovery cohort, N=3,469, 55% women) and the Women's Health Initiative (WHI, replication cohort, N=1,699, women only). The glycemic traits were fasting glucose, fasting insulin, the Homeostatic Model Assessment for Insulin Resistance (HOMA-IR), and glycosylated hemoglobin (HbA1c). To identify isoform-level associations, we applied linear mixed models using isoform ratios (ratio of isoform to gene-level expression), adjusting for multiple covariates (age, age², sex, body mass index, white blood cell proportions, batch effects, surrogate variables, and family structure). Isoform ratio is defined as the TMM value of a given isoform divided by the sum of TMM values for all isoforms of the corresponding gene. Putative causality of the isoforms were evaluated using two-sample Mendelian randomization (MR). We conducted gene ontology analysis to uncover enriched biological pathways.

Results: In the FHS discovery cohort, there were 54 isoforms (based on ratio) associated with glucose, 2,430 associated with insulin, 2,607 associated with HOMA-IR, and 235 associated with HbA1c. Among these, replication was achieved for 1 glucose-associated isoform, 52 insulin-associated isoforms, and 60 HOMA-IR-associated isoforms. In discovery analysis limited to women (N=1,929), 9 isoforms were significant for glucose, 38 for insulin, and 42 for HOMA-IR with replication rates of 22%, 61%, and 54% respectively. Gene ontology analysis showed significant enrichment for genes involved in protein-complex assembly, apoptosis, and immune function. MR identified 14 of the replicated genes as causal for the corresponding glycemic trait (PMR<0.05). Of these, PSMB8-201, which encodes an immunoproteasome, was associated with insulin and was the top isoform from MR analysis.

Conclusion: Our findings highlight isoform associations with glycemic traits and advance our understanding of splicing-dependent pathways involved in T2D, which can inform the development of more precise diagnostic tools and targeted therapeutics.

Allele-specific transcriptome variation identification at single-cell resolution in human retina enabled by single-cell long-read RNA sequencing

Subsession Time: Friday, October 17 at 11:15am – 11:30am

Authors: Meng Wang (University of California Irvine, Department of Ophthalmology Research, Center for Translational Vision Research, Gavin Herbert Eye Institute), Yumei Li (University of California Irvine, Department of Ophthalmology Research, Center for Translational Vision Research, Gavin Herbert Eye

Institute), Jun Wang (Baylor College of Medicine, Department of Molecular and Human Genetics), Rui Luo (University of California Irvine, Department of Ophthalmology Research, Center for Translational Vision Research, Gavin Herbert Eye Institute), Rui Chen (University of California Irvine, Department of Ophthalmology Research, Center for Translational Vision Research, Gavin Herbert Eye Institute, Baylor College of Medicine, Department of Molecular and Human Genetics, Human Genome Sequencing Center)

Abstract: Background: Regulation of gene expression and transcript structure contributes to transcriptomic diversity and plays a critical role in human disease. Systematically characterizing how genetic variation influences gene regulation in the corresponding cell context is fundamental to understanding complex traits. The combination of long-read sequencing with single-cell transcriptomics technologies enables the exploration of genetic effects on both gene expression and transcript architecture at single-cell resolution.

Methods: To investigate this, we performed transcriptome profiling of over 500,000 cells derived from 89 healthy human retina samples collected from 77 donors across age groups from diverse ancestry groups, utilizing both Illumina short-read and Oxford Nanopore Technologies (ONT) long-read sequencing. Paired with whole genome sequencing data from each donor, allele-specific expression (ASE) and allele-specific transcript structure (ASTE) across major retinal cell types are systematically characterized. The list of ASE identified from long and short read data is compared. ASE and ASTE are further annotated with previously reported eQTLs, transcript ratio QTLs (trQTLs), and genome-wide association study (GWAS).

Results: We identified more than 50,000 novel transcripts associated with annotated genes, revealing cell class-specific patterns of transcript expression and isoform usage. Then, by integrating genomic and transcriptomic data, we pinpointed regulatory variants with distinct effects on gene regulation across retinal cell classes. Finally, we linked GWAS loci to enriched cell types, candidate causal variants and genes, and uncovered cell class-specific regulatory mechanisms underlying complex traits.

Conclusion: Utilizing optimized technology and method, we systematically characterized allele-specific expression and transcript structure at single-cell resolution in human retina, uncovering novel insights into gene expression and transcript alterations driven by both common and rare genetic variants. The use of long-read sequencing enabled a more detailed view of these events than previously possible. Overall, genetic influences on gene regulation exhibit strong context specificity. Our findings further indicate that hierarchical interactions among transcription factors play a pivotal role in mediating the cell class-specific regulatory effects of genetic variants.

Multi-omics in 10,000 diverse individuals expands mapping genetic mechanisms in health and disease

Subsession Time: Friday, October 17 at 11:30am – 11:45am

Authors: Evin M. Padhi (Stanford University), Marianne DeGorter (Stanford University), Elizabeth Kiernan (Broad Institute), Jonathan Nguyen (Stanford University), Wail Baalawi (Broad Institute), Alex Miller (Stanford University), Yulia Mostovoy (Broad Institute), Alex Yenkin (Broad Institute), Lee Lichtenstein

(Broad Institute), All of Us Research Program Multi-omics Working Group, Namrata Gupta (Broad Institute), Tara Dutch (NIH), Konrad J. Karczewski (Broad Institute), Niall Lennon (Broad Institute), Stephen B. Montgomery (Stanford University)

Abstract: Multi-omics in population biobanks provides new opportunities to map the impact of genetic variation on molecular phenotypes and inform the mechanisms of health traits and diseases. We generate and analyze multi-omics from 10,000 ancestrally diverse, whole genome-sequenced participants of the All of Us Research Program. These data consist of whole blood transcriptomes using the Watchmaker RNA-seq protocol, detecting both coding and non-coding transcripts, paired with plasma proteome profiling via an expanded Olink Explore HT platform to quantify 5,419 proteins, as well long-read genome sequencing and associated methylation profiling.

The large sample size and ancestrally diverse composition of the All Of Us multi-omics cohort enable new opportunities to study common and rare variants impact on gene expression, splicing and protein levels. We perform deep cis-QTL mapping ($MAF > 0.1\%$) and identify 70,949 significant ($FDR < .05$) QTL loci across -omes, with 80% of these signals having reproducible effects detected in at least 3 out of 5 populations studied. Fine-mapping analysis with SuSiE identified over 700,000 variants across all QTLs and more than 35,000 variants with a $PIP > 0.9$.

Among fine-mapped pQTL variants, we observe comparable numbers of common ($>5\%$ MAF), low-frequency ($1-5\%$ MAF), and rare ($1-0.1\%$ MAF) variants across PIP deciles, demonstrating that rare variants are important drivers of QTL signals that are missed in many studies. We find that fine-mapped variants are significantly enriched for ENCODE regulatory annotations and coding variant effects similarly across allele frequencies. However, when examining pathogenicity scores from REVEL, we find that likely-causal fine-mapped ($PIP > .9$) rare and low-frequency variants are significantly enriched for pathogenic scores ($OR = 6.1, p = 0.00067$), suggesting these rare QTL variants may play a role in disease.

We are conducting analyses to integrate transcriptomic QTLs, structural variant and methylation data from All of Us long-read sequencing, and performing analyses integrating GWAS data from All of Us and external datasets to evaluate the degree that rare fine-mapped variants provide expanded discovery power for trait-relevant genes. Collectively, we have characterized molecular QTLs from All of Us, data imminently available through the Researcher Workbench, and expanded our identification of rare variants driving molecular phenotypes to human traits and diseases.

Session 73: Grey Matters: The Science of Aging

Location: Room 210C/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Friday, October 17 at 10:45am – 11:45am

Linking single-cell transcriptomic and genomic changes in the aging human brain

Subsession Time: Friday, October 17 at 10:45am – 11:00am

Authors: Tianxiong Yu (Umass Chan Medical School), Ailsa M. Jeffries (Umass Chan Medical School), Jennifer S. Ziegenfuss (Umass Chan Medical School), Allie K. Tolles (Umass Chan Medical School), Christina E. Baer (Umass Chan Medical School), Cesar Bautista Sotelo (Umass Chan Medical School), Yerin Kim (Umass Chan Medical School), Zhiping Weng (Umass Chan Medical School), Michael A. Lodato (Umass Chan Medical School)

Abstract: Aging disrupts diverse biological processes across tissues, often driven by the gradual accumulation of stochastic damage in individual cells. While bulk transcriptome studies have mapped age-associated gene expression changes in the human brain, the cell-type-specific transcriptomic and genomic alterations that underlie non-pathological aging remain poorly understood.

Here, we integrated single-nucleus RNA sequencing, single-cell whole-genome sequencing, and spatial transcriptomics to identify transcriptomic and genomic changes in the prefrontal cortex of the 19 human brains across life span, from infancy to centenarian. Analyzing the transcriptome of 712,798 nuclei and cells, we identified infant-specific cell clusters marked by elevated expression of neurodevelopmental genes. Across cell types, aging was associated with a consistent downregulation of core homeostatic genes involved in ribosomal function, intracellular transport, and metabolism. Conversely, neuron-specific gene expression remained largely stable over time. Using single-cell genome sequencing of 100 neurons, we identified a somatic mutational signature shaped by gene length and expression level. Notably, short, highly expressed homeostatic genes accumulate somatic mutations faster during life—correlating with their reduced expression in aged neurons.

Together, our study reveals coordinated transcriptomic and genomic changes of aging in the human brain, offering new insights into the cellular mechanisms that drive functional decline during aging.

Epigenetic mosaicism acquired in human embryogenesis and aging

Subsession Time: Friday, October 17 at 11:00am – 11:15am

Authors: Young Seok Ju (Korea Advanced Institute of Science and Technology)

Abstract: During early human embryogenesis, epigenetic reprogramming occurs in parallel with cell fate specification, orchestrating the formation of the human body. However, ethical constraints limit direct investigation of human embryos, leaving our understanding of epigenetic heterogeneity, particularly mosaicism within the same cell type after embryogenesis, largely incomplete. This knowledge gap poses challenges to assessing the epigenetic accuracy in embryonic cellular differentiation. In this study, we

investigate developmental and age-related epigenetic changes at single-embryonic-lineage resolution by integrating whole-genome, DNA methylome, and transcriptome sequencing from 500 clones derived from single cells of the colon and blood in 15 individuals. Using spontaneous somatic mutations accumulated during embryogenesis, we reconstructed developmental phylogenies of these clones and uncovered several key insights. Notably, two colorectal epithelial cells from the same individual differed by ~1 million CpG methylation marks at birth, with these differences stably maintained throughout their somatic lineages. These early embryonic epigenetic mosaicisms (EEEMs) were not random; instead, they recurred at ~10,000 loci across individuals, frequently overlapping with gene promoters. Methylation levels at these loci correlated with gene expression, suggesting functional roles in early development and aging. Strikingly, over 70% of promoter demethylation events at active LINE-1 elements occurred during early embryogenesis, implicating imperfect methylation reprogramming as a major contributor to retrotransposon activation in colorectal epithelium. During aging, CpG methylation changes accumulated at an estimated rate of ~13,000 sites per year, further compounding epigenetic mosaicism in older cells. To our knowledge, this is the first systematic characterization of EEEMs and age-associated methylation drift, offering a comprehensive view of epigenetic mosaicism across human lifetime.

Epigenome-Wide Analysis of Aging in the Largest Nanopore Methylation Cohort of the Emirati Genome Program

Subsession Time: Friday, October 17 at 11:15am – 11:30am

Authors: Judith Arres Reche (M42, Emirati Genome Program), Abdelrahman Saad (M42, Emirati Genome Program), Daniel Sanchez (M42, Emirati Genome Program), Ahmed Al Awadhi (M42, Emirati Genome Program), Santosh Elavalli (M42, Emirati Genome Program), Senchao Zhang (M42, Emirati Genome Program), Sneha Leela (M42, Emirati Genome Program), Luis F Paulin (Baylor College of Medicine), Yilei Fu (Baylor College of Medicine), Haiguo Wu (M42, Emirati Genome Program), Hanil Sajad (M42, Emirati Genome Program), Joseph Mafofo (M42, Emirati Genome Program), Paul Jones (M42, Emirati Genome Program), Albarah El-Khani (M42, Emirati Genome Program), Fahed Al Marzooqi (M42, Emirati Genome Program), Asma Al Mannaei (DOH), Mohamed Al Ameri (DOH), Wadha Mohammed Abdulrahman (DOH), Sahar Ahmed Al Marzooqi (DOH), Tiago Magalhaes (M42, Emirati Genome Program), Fritz Sedlazeck (Baylor College of Medicine), Javier Quilez (M42, Emirati Genome Program)

Abstract: The Emirati Genome Program (EGP) is conducting ~30x whole-genome sequencing (WGS) of ~1 million Emirati nationals. With over 100,000 genomes sequenced using Oxford Nanopore Technologies (ONT), EGP represents the largest human genome-wide methylation dataset to date. Unlike conventional array- or bisulfite-based profiling, which targets limited CpG sites, EGP leverages native DNA sequencing to produce genome-wide methylation maps at single-base resolution. This unique resource enables large-scale discovery of epigenetic markers with biological and clinical relevance.

In this work, we aimed to (1) define quality control (QC) metrics for WGS-based methylation from ONT data; and (2) identify age-associated CpG sites using an epigenome-wide association study (eWAS) approach.

We analyzed ~10,000 blood-derived genomes from EGP participants sequenced with ONT R9 chemistry, calling CpG methylation levels with modkit and profiling ~29 million CpG sites across the genome. To evaluate per-sample quality, we established a QC framework by defining key performance metrics, including the percentage of average methylation site coverage relative to genomic coverage, the distribution of CpG sites across binned methylation levels (0–25%, 25–75%, 75–100%), and the proportion of sites meeting minimum coverage cutoffs (score ≥ 10). We then performed a large-scale eWAS to assess the association between methylation and chronological age using non-parametric models. The analysis included QC-pass CpGs with sufficient coverage and variability across the cohort.

The eWAS revealed thousands of CpG sites significantly associated with age, reflecting consistent and biologically meaningful methylation changes across individuals. Among them, we highlight two genes comprising CpG sites with a significant (adjusted $p < 4.36 \times 10^{-44}$) increase in methylation with age: CILP2 (chr19:19540366 and chr19:19540317) and PPP1R13L (chr19:45396817 and chr19:45396577). Significantly, PPP1R13L has been associated with cognitive aging in previous studies, supporting the functional relevance of this locus in age-related epigenetic regulation.

In summary, in this study we established systematic QC criteria for WGS-derived methylation data from ONT, demonstrating the high quality of EGP datasets. These data enabled robust, population-scale epigenome-wide analysis, uncovering strong age-associated methylation signals and positioning EGP as a key resource for advancing epigenetic research.

Non-linear molecular QTLs in the human brain reveal context-dependent effects on aging and neurodegenerative disorders

Subsession Time: Friday, October 17 at 11:30am – 11:45am

Authors: Anjing Liu (Center for Statistical Genetics, The Gertrude H. Sergievsky Center, Columbia University), Ruixi Li (Center for Statistical Genetics, The Gertrude H. Sergievsky Center, Columbia University), Zining Qi (Center for Statistical Genetics, The Gertrude H. Sergievsky Center, Columbia University, Department of Human Genetics, The University of Chicago), Xuewei Cao (Center for Statistical Genetics, The Gertrude H. Sergievsky Center, Columbia University, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center),

Roulan Jiang (Department of Statistics, Colorado State University), Kushal K. Dey (Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, Gerstner Sloan Kettering Graduate School of Biomedical Sciences), Philip L. De Jager (Center for Translational & Computational Neuroimmunology, Columbia University, Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, Department of Neurology, Columbia University), David Bennett (Rush Alzheimer's Disease Center and Department of Neurological Sciences, Rush University Medical Center), The FunGen-AD Consortium, Tianying Wang (Department of Statistics, Colorado State University), Gao Wang (Center for Statistical Genetics, The Gertrude H. Sergievsky Center, Columbia University, Department of Neurology, Columbia University)

Abstract: Most molecular QTL (xQTL) studies rely on linear models (lm-xQTL), while non-linear xQTL (nl-xQTL) applications like interaction QTL (iQTL) are limited because they require predefined factors, lack statistical power and fine-mapping capabilities for putative causal variants. Variance QTL (vQTL) methods can capture hidden interactions but remain underpowered. In this study, we conducted quantile QTL (qQTL) analysis to detect distributional heterogeneity in gene expression and protein abundance across human brain tissues, identifying context-dependent regulatory effects without requiring prior knowledge of interacting factors. We created a nl-xQTL atlas from >2,300 brain donors across 22 molecular contexts, including vQTL and iQTL with APOE ϵ 4 status, age, sex, and cell-type proportions to benchmark power and utility of qQTL. For disease applications, we developed qSuSiE for fine-mapping qQTL signals within candidate GWAS loci, and created quantile-based transcriptome-wide association study (qTWAS) for gene prioritization in Alzheimer's disease (AD), Parkinson's disease, and aging-related traits.

Among 2,986,002 brain qQTLs identified, ~30% are not detected as lm-xQTL. qQTL effectively captured known interaction effects, recovered varying proportions of iQTL across brain regions/cells (e.g. 100% of sex-iQTLs in oligodendrocytes, 68.42% in inhibitory neurons), and overlapped with 94.5% of the 29,270 vQTLs. Compared to lm-xQTL, qQTL in several bulk tissues show stronger enrichment in chromatin accessibility, chromatin loops, and cell-specific enhancer regions, indicating they capture epistasis and cell-cell interactions. qQTL also explains additional GWAS heritability for brain-related traits beyond lm-xQTL. When applied to AD, qSuSiE, our new quantile fine-mapping method, identified a 95% credible set in TMEM106B showing significant qQTL effects ($q\text{-value}=2.24\times 10^{-5}$) in excitatory neurons, which we further dissected into a combination of independent context-dependent effects, including interactions with APOE ϵ 4, cell fraction, and trans-eQTL association. qTWAS uncovered 34 novel AD-related genes beyond standard TWAS, most prominently IGFL4 in posterior cingulate cortex (qTWAS $p=4.34\times 10^{-11}$ vs. TWAS $p=0.466$).

In sum, qQTL reveals regulatory effects missed by standard methods and improves disease gene discovery. We recommend augmenting standard xQTL and TWAS with qQTL to prioritize functional variants and qTWAS for broader insights into complex disease pathogenesis.

Session 74: Pan-Genome Perspectives on Cardiovascular Diseases

Location: Room 156ABC/Level 1, Thomas M. Menino Convention and Exhibition Center

Session Time: Friday, October 17 at 10:45am – 11:45am

Uncovering Hundreds of New Loci from a Multi-Ancestry Study of Blood Pressure Traits in Over 2.5M Individuals

Subsession Time: Friday, October 17 at 10:45am – 11:00am

Authors: Hannah M. Seagle (Vanderbilt University Genetics Institute, Division of Genetic Medicine, Division of Epidemiology, Department of Medicine, Vanderbilt University, Joseph Maxwell Cleland Atlanta VA Medical Center), Jeewoo Kim (Vanderbilt University Genetics Institute, Division of Genetic Medicine, Division of Quantitative Science, Department of Obstetrics and Gynecology, Vanderbilt Medical Scientist Training Program, Vanderbilt University Medical Center, Vanderbilt University School of Medicine), Joseph H. Breeyear (Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, National Institutes of Health), Catherine A. Greene (Vanderbilt University Genetics Institute, Division of Quantitative Science, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center), John B. Buse (Division of Endocrinology, University of North Carolina School of Medicine), Lanna Caruth (Department of Pathology and Laboratory Medicine, Institute of Biomedical Informatics, University of Pennsylvania), Lindsay Guare (Department of Pathology and Laboratory Medicine, Genomics and Computational Biology Graduate Program, University of Pennsylvania), Navya S. Josyula (Department of Population Health Services, Geisinger Health), Qianqian Liang (Department of Population Health Services, Geisinger Health), Chelsea Okeh (Department of Pathology and Laboratory Medicine, University of Pennsylvania), Hannah Poisner (Vanderbilt University Genetics Institute, Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center), Yash Pershad (Vanderbilt University Genetics Institute, Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center), Serena Houghton (VA Boston Healthcare System), Yuk-Lam Anne Ho (Massachusetts Veterans Epidemiology Research Information Center), Xiaofeng Zhu (Department of Population and Quantitative and Health Sciences, Case Western Reserve University), Alexander G. Bick (Vanderbilt University Genetics Institute, Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center), Digna R. Velez Edwards (Division of Genetic Medicine, Department of Medicine, Division of Quantitative Science, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center), J. Michael Gaziano (VA Boston Healthcare System), Alison A. Motsinger-Reif (Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, National Institutes of Health), Helen R. Warren (William Harvey Research Institute, Queen Mary University of London, NIHR Leicester Biomedical Research Centre), Patricia B. Munroe ((William Harvey Research Institute, Queen Mary University of London, NIHR Leicester Biomedical Research Centre), Shefali S. Verma (Department of Pathology and Laboratory Medicine, Genomics and Computational Biology Graduate Program, University of Pennsylvania), Anne E. Justice (Department of

Population Health Services, Geisinger Health), Peter W.F. Wilson (Joseph Maxwell Cleland Atlanta VA Medical Center, Emory Clinical Cardiovascular Research Institute), Yan V. Sun (Joseph Maxwell Cleland Atlanta VA Medical Center, Department of Epidemiology, Emory University Rollins School of Public Health), Jacklyn N. Hellwege (Vanderbilt University Genetics Institute, Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Joseph Maxwell Cleland Atlanta VA Medical Center, VA Tennessee Valley Healthcare System (626)), Todd L. Edwards (Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, VA Tennessee Valley Healthcare System (626))

Abstract: Elevated blood pressure (BP) is a leading contributor to morbidity and mortality, with notable ancestral and racial differences in BP traits in the United States and globally. To date, most BP genome-wide association studies (GWAS) are predominantly European (EUR) ancestry and have not included the X chromosome, with the current largest BP GWAS being EUR only. We combined newly conducted GWAS with publicly available summary statistics to study three BP traits: systolic [SBP], diastolic [DBP], and pulse pressure [PP]. BP traits were regressed onto additively encoded genotypes, adjusted for age, age-squared, sex, body mass index, and the first 10 principal components. We performed fixed-effects inverse-variance-weighted meta-analyses including over 2.5M individuals (16 cohorts/biobanks; 35% non-EUR; 50% female), within and across ancestry groups. Multi-ancestry meta-analyses of autosomes identified a total of 3,204 unique significant independent loci via GCTA software ($P < 5 \times 10^{-8}$; minor allele frequency > 0.01), with 1,351 for SBP, 1,206 for DBP, and 1,201 for PP. Among these, we detected 629 (SBP), 532 (DBP), and 540 (PP) novel independent signals, which corresponded to 1,521 unique novel loci across BP traits. Ancestry-stratified analyses detected 43 Admixed American-specific loci, 135 specific to African, and 343 specific to the Asian population across all three BP traits. Sex-stratified analyses across all BP traits detected 274 independent loci in females and 403 for males, of which 52 and 181 were unique to females and males, respectively. Chromosome X meta-analyses identified 8 independent significant loci for SBP, 80 for DBP, and 8 for PP. The summary statistics from the multi-ancestry and stratified meta-analysis were used to evaluate genetically predicted gene expression (GPGE) associations using S-PrediXcan. For all three BP traits across cardiovascular tissues, we identified a total of 3,835 significant (Bonferroni $P < 9.08 \times 10^{-8}$) gene-tissue pairs. We observed significant GPGE with SBP at 496 genes (32 novel), DBP at 608 genes (40 novel), and PP at 499 genes (166 novel). This expansion of associations is made possible through the first-ever chromosome X S-PrediXcan analysis for BP traits and unprecedented ancestral heterogeneity and sample size. This study including > 2.5 M individuals builds a strong foundation for precision medicine opportunities among traditionally understudied populations.

Cigarette smoking modifies genetic associations with serum lipids in cross-population meta-analyses of 920,966 individuals

Subsession Time: Friday, October 17 at 11:00am – 11:15am

Authors: Amy R. Bentley (National Human Genome Research Institute), Michael Brown (University of Texas Health Science Center at Houston School of Public Health), Thomas Winkler (University of Regensburg), Karen Schwander (Washington University School of Medicine), Paul S. de Vries (University of Texas Health Science Center at Houston School of Public Health), Raymond Noordam (Leiden

University Medical Center), Heming Wang (Brigham and Women's Hospital, Harvard Medical School), Nora Franceschini (University of North Carolina), Charles N. Rotimi (National Human Genome Research Institute), the CHARGE Gene-Lifestyle Interactions Working Group

Abstract: Previous meta-analyses have shown that smoking status can modify genetic associations with serum lipids and that evaluating smoking interactions can help identify novel lipids loci. However, these analyses were well-powered to detect only some types of interactions and the need for larger sample sizes was acknowledged.

We conducted a genome-wide gene-smoking interaction study on serum lipids (high-density lipoprotein cholesterol [HDL], low-density lipoprotein cholesterol [LDL], and triglycerides [TG]) in 919,188 individuals in populations defined by European (n=672,889), East Asian (n=144,836), African (n=48,031), Hispanic (n=29,815), Middle Eastern (n=13,128), and South Asian ancestry (n=12,267). Smoking was represented by current smoking status, cigarettes/day, and packyears. We evaluated both population-specific and cross-population meta-analyses. Statistical significance was determined using 3 approaches: 1) $p < 5 \times 10^{-9}$ for a 1df test of the interaction, 2) $p < 5 \times 10^{-9}$ for a 2df joint test of the main and interaction effects and variant 500 kb from known lipids loci, 3) $p < 1 \times 10^{-5}$ for a main effect and a Bonferroni-corrected $p_{1df} < 0.05$ [2-step].

In total, we identified 35 significant loci ($N_{1df}=5$, $N_{2df}=28$, $N_{2-step}=2$). These associations were largely from cross-population meta-analyses, with variants common across populations. We identified interactions with known lipids loci, including APOA5, GCKR, and PCSK9. For instance, rs12117661-G (PCSK9) was associated with 1.1 mg/dl lower LDL among current smokers ($p_{1df}=2.0 \times 10^{-11}$). Among the novel lipids loci, we identified rs3943782-G in CADM2, previously associated with smoking behaviors and smoking-BMI interactions, was associated with reduced HDL among current smokers ($p_{2df}=3.7 \times 10^{-10}$). Interactions with variants in a region on chromosome 12 were observed among East Asians (MAF ~ 0.2 , while rare or absent in other populations). For example, in this region, rs144504271-A (HECTD4) was associated with 0.099 lower lnTG ($p_{1df}=8.6 \times 10^{-17}$) and 5.6 mg/dl higher LDL ($p_{1df}=3.9 \times 10^{-15}$) in current smokers. Notably, this region includes ALDH2, which breaks down aldehydes (a component of cigarette smoke) as part of a major oxidative pathway.

In summary, these meta-analyses identified gene-smoking interactions on serum lipids. The identification of interactions with known lipids loci and with smoking-relevant loci provides promising insights for further exploration of biological mechanisms and potential clinical relevance.

Genetic underpinnings of stroke in Hispanic/Latino populations

Subsession Time: Friday, October 17 at 11:15am – 11:30am

Authors: Yao Tu (Department of Biobehavioral Health, Pennsylvania State University), Zhe Wang (Department of Epidemiology, University of Alabama at Birmingham), Scott Ratliff (Department of Epidemiology, University of Michigan), Jie Yao (The Lundquist Institute, Harbor-UCLA Medical Center), Xiuqing Guo (The Lundquist Institute, Harbor-UCLA Medical Center), Jennifer Smith (Department of Epidemiology, University of Michigan), Kari E. North (Border Health Research Center, UTHealth Houston)

School of Public Health–Brownsville campus), Misa Graff (Department of Epidemiology, University of North Carolina at Chapel Hill), Lindsay Fernandez-Rhodes (Department of Biobehavioral Health, Pennsylvania State University)

Abstract: Hispanic/Latino (HL) populations are the largest ethnic minority in the United States (US). However, they are highly underrepresented in genetic studies for cardiovascular diseases, especially stroke. We have assembled the largest HL cardiovascular study to date, including electronic health record (EHR)-based biobanks and cohorts. Our preliminary genome-wide association studies (GWAS) for total stroke included 63,082 HL adults (3036 cases and 62% female) from 7 datasets. GWAS results were adjusted by sex, age, and 10 principal components (PCs). The following meta-analysis identified a potentially novel locus (rs17070134, a highly conserved gene PRICKLE2, with missense mutations involved in neurodevelopmental disorders) that reached genome-wide significance (GWS; $p < 5 \times 10^{-8}$). In a female-specific analysis of 40,227 HL adults, three loci (rs76295001 and rs1166044176, known stroke loci; rs77378396) reached GWS in females (1745 cases). The novel locus rs77378396 on chromosome 9, is located within PCSK5, a gene previously associated with high-density lipoprotein (HDL) levels and heart defects. Among 23,855 HL males (1291 cases), another 4 loci reached GWS (rs457903 and rs139559202, known loci; rs118016885 and rs387693). The two known loci (rs457903 at MACROD2 and rs139559202 at CHL1) are known ischemic stroke loci in Asian populations, yet, the index SNP at the rs139559202 locus is common in African populations and $MAF < 0.1\%$ in Asian populations. rs118016885 at SNX31 and rs387693 at FCGR2A are both potentially novel loci. While no variants in FCGR2A have been previously associated with stroke, a case-control study has identified its protein expression level as significantly associated with the presence of atherosclerotic plaques, a known risk factor for stroke.

Our preliminary data have identified 4 established and 4 novel stroke-associated loci in HL populations. Also, sex-stratified GWAS results suggest the need for further studies on sex differences in stroke genetic architecture. Currently, ~200k GWAS results from other studies in the HISLA-CVD consortium are being added. Future meta-analyses will also consider stroke subtypes, including ischemic and hemorrhagic stroke, to further the specificity of our findings on the genetic underpinnings of stroke in HL populations.

Global Evaluation of Congenital Heart Disease-Associated Non-Coding Variant Effect on Cardiac Transcription Factor Regulation

Subsession Time: Friday, October 17 at 11:30am – 11:45am

Authors: Edwin G. Peña-Martínez (University of Puerto Rico-Río Piedras), Shreya Sharma (Indian Institute of Technology Roorkee), Joshua G. Medina-Feliciano (University of Puerto Rico-Río Piedras), Elise Root (Yale University), Lois Parks (Cincinnati Children's Hospital Medical Center), Marissa Granitto (Cincinnati Children's Hospital Medical Center), Diego A. Pomales-Matos (University of Puerto Rico-Río Piedras), Jean L. Messon-Bird (University of Puerto Rico-Río Piedras), Adriana C. Barreiro-Rosario (University of Puerto Rico-Río Piedras), Leandro Sanabria-Alberto (University of Puerto Rico-Río Piedras), Alejandro Rivera-Madera (University of Puerto Rico-Cayey), Jessica M. Rodríguez-Ríos (University of Puerto Rico-Río Piedras), Rosalba Velazquez-Roig (University of Puerto Rico-Río Piedras), Juan A. Figueroa-Rosado (University of Puerto Rico-Mayagüez), Mackenzie Noon (Yale University), Omer A. Donmez (Cincinnati Children's Hospital Medical Center)

Children's Hospital Medical Center), Carmy Forney (Cincinnati Children's Hospital Medical Center), Hayley K. Hesse (Cincinnati Children's Hospital Medical Center), Katelyn A. Dunn (Cincinnati Children's Hospital Medical Center), Xiaoting Chen (Cincinnati Children's Hospital Medical Center), Matthew R. Hass (Cincinnati Children's Hospital Medical Center), Lucinda P. Lawson (Cincinnati Children's Hospital Medical Center), Matthew T. Weirauch (Cincinnati Children's Hospital Medical Center), Leah C. Kottyan (Cincinnati Children's Hospital Medical Center), Steven K. Reilly (Yale University), Devesh Bhimsaria (Indian Institute of Technology Roorkee), and José A. Rodríguez-Martínez (University of Puerto Rico-Río Piedras)

Abstract: Congenital heart diseases (CHDs) are the most common birth defect, affecting 1 in every 100 liveborn infants. Genome-wide association studies (GWAS) have mapped thousands of CHD-associated variants within non-coding regions of the genome. Non-coding variants can alter gene expression regulatory mechanisms, such as transcription factor (TF) binding, potentially contributing to human diseases. However, functionally validating each of these disease-associated variants remains challenging. In this work, we systematically evaluated all reported CHD-risk variants for genotype-dependent TF-DNA binding activity and gene expression to identify potential causal variants.

Towards this goal, we developed SNP Bind-n-Seq, a high-throughput gel shift assay, to evaluate >3,000 CHD-risk variants for allele-specific binding for the cardiac TFs NKX2-5, GATA4, and TBX5. In doing so, we quantified binding affinity (Kd) of every possible allele of 3,232 CHD-risk variants (12,928 alleles) for all three cardiac TFs. Through this approach, we identified 170 CHD-risk variants with differential allelic binding for NKX2-5 (54), GATA4 (58), or TBX5 (62). Since TF-DNA binding directly controls gene expression, understanding these allele-dependent biochemical interactions can provide insight into the genetic mechanisms behind CHD-risk variants. Next, we complemented our binding affinity measurements with a direct identification of CHD-risk variants with allele-dependent expression through a massively parallel reporter assay (MPRA). We identified 187 CHD-risk variants that exhibited an allelic skew in transcriptional activity (>25% change in expression between alleles), of which 13% occurred in cardiac cis-regulatory elements active during heart development.

By combining our findings from SNP Bind-n-Seq and MPRA, we identified five variants (rs2137643, rs7303642, rs2896074, rs839154, and rs7032149) with both differential allelic binding and transcriptional activity. Additionally, we identified four genes (MGAT4C, CNOT6L, VARS2, and DDR1) previously implicated in cardiovascular diseases in cardiac eQTL with multiple CHD-risk variants with allelic binding or expression. Altogether, our findings provide a systematic genome-wide assessment of CHD-risk variants with genotype-dependent regulatory activity. Our approach is scalable for dissecting the regulatory mechanisms of non-coding variants and understanding the genetic etiology of any complex human disease.

Session 75: The Utility of AI in Clinical Genomics Workflows

Location: Ballroom West/Level 3, Thomas M. Menino Convention and Exhibition Center

Session Time: Friday, October 17 at 10:45am – 11:45am

Automated Phenotyping Using NLP Rivals Human Expert Curation and Enables Scalable Gene Prioritization in Clinical Diagnostic Workflows

Subsession Time: Friday, October 17 at 10:45am – 11:00am

Authors: Kai Lee Yap (Department of Pathology & Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine), Andy Drackley (Department of Pathology & Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine), Anthony F Wong (Quantitative Science Pillar, Stanley Manne Children's Research Institute, Ann & Robert H. Lurie Children's Hospital of Chicago), Alexander Ing (Department of Pathology & Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine), Jorge Pineda (Quantitative Science Pillar, Stanley Manne Children's Research Institute, Ann & Robert H. Lurie Children's Hospital of Chicago), Sachleen K Tuteja (Weinberg College of Arts and Sciences, Northwestern University), Pamela Rathbun (Department of Pathology & Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine), Patrick D McMullen (Department of Pathology & Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine)

Abstract: Phenotype-driven variant prioritization is central in clinical diagnostic pipelines, requiring accurate mapping of clinical features to structured vocabularies like the Human Phenotype Ontology (HPO) before deriving gene associations. Manual phenotyping by expert human reviewers from unstructured notes is time-consuming and subject to inter-reviewer variability, posing challenges in high-throughput diagnostic settings. To address this, we evaluated natural language processing (NLP) as a scalable solution for HPO term extraction and gene prioritization in a clinical diagnostics laboratory pipeline.

We collated 100 cases with confirmed monogenic diagnoses from clinical exome sequencing. To reflect the context of clinical notes available at the time of genetic test initiation, only notes documented prior to the return results were reviewed independently by two laboratory genetic counselors and an NLP-based tool for the extraction of HPO terms. We developed methodologies to measure concordance between human- and NLP-derived terms using Jaccard similarity coefficients and an ontology-aware distance metric calculation. We also developed HPOverlap, an open-source, interactive application that generates customizable gene lists from HPO terms, and allows users to adjust phenotypic specificity through an ontology-based overlap model. The NLP-derived HPO terms showed strong concordance with human-curated terms and comparable performance in isolating the diagnostic gene. If we define a gene list size of <800 as optimal for the purpose of variant prioritization during genomic analysis, the gene lists generated by HPOverlap included the causative gene in >75% of cases for both NLP and human

generated HPO terms. Furthermore, our HPOverlap tool allows for the incorporation of most up-to-date HPO annotation database. This ensures that the most current gene-phenotype associations are integrated into our diagnostic workflows, maximizing possible diagnostic yield of genetic testing.

Overall, automated phenotyping using NLP can democratize access to deep phenotyping, without dependency on highly trained genetics personnel. By automating steps in the analysis pipeline, laboratories can redirect expert effort toward variant interpretation of complex cases. While short-term human oversight remains important, our findings show that NLP-based HPO extraction approximates human expert review in both content and performance when paired with our HPOverlap application.

Artificial intelligence cannot replace human intelligence: publicly available generative AI is not suitable for variant classification

Subsession Time: Friday, October 17 at 11:00am – 11:15am

Authors: Thomas W Laver (University of Exeter), Preeah Sangha (University of Exeter), Oguzhan Kalyon (University of Exeter), James Fasham (University of Exeter, Royal Devon University Healthcare NHS Foundation Trust), Adam Gunning (Royal Devon University Healthcare NHS Foundation Trust)

Abstract: A clinical scientist receives a genetic testing result for a patient. They are not an expert on this gene and are not sure if the variant reported in their patient is pathogenic. They put the variant into a publicly available generative AI tool and ask it to classify the variant. The AI confidently reports that it is pathogenic and gives evidence to support the classification. Satisfied with this answer, the clinical scientist reports the variant. This leads to an incorrect genetic diagnosis and treatment for their patient.

We aimed to assess how well publicly available generative AI tools performed at classifying the pathogenicity of Mendelian disease variants. We used a set of 276 pathogenic and 202 benign variants, which were manually classified by a clinical scientist in a diagnostic setting in an accredited laboratory and reviewed by a second registered clinical scientist. We asked 8 widely used generative AI tools to classify each variant according to the ACMG variant interpretation guidelines.

The generative AI tools correctly classified 7.7-51.9% of variants: ChatGPT (GPT-4o - 7.7%, o4-mini - 32.6%, o3 - 32.4%), Grok (3) -12.6%, Claude (Sonnet 4) -12.8%, copilot - 24.1%, deepseek - 38.7%, Gemini (2.5 flash) - 51.9%. These results compare poorly to meta-predictor tools, which can be considered a form of classification AI. For example, REVEL (thresholds: benign ≤ 0.29 ; pathogenic ≥ 0.644) correctly classified 65.7%. A further problem with generative AI is that the outcome can vary so classifications are not consistent or reproducible, for example 12.3% of ChatGPT GPT4o classifications changed when we repeated the analysis. For a subset of 162 pathogenic variants, we also assessed the ACMG codes that can be applied from just publicly available data. The different AI tools only used the correct codes for 0.6-13% of variants.

Current publicly available generative AI cannot accurately classify Mendelian variants. Generative AI could have a role as a data collection tool to support clinical scientists, but this requires AI models specifically developed for this task. This is in line with OpenAI's usage policy that AI should not provide

tailored medical advice without review by qualified professionals. Current generative AI cannot replace expert clinical scientists and should not be used to classify Mendelian variants.

The Community Texome Project: Advancing Rare Disease Diagnosis Through Genomic Medicine, AI, and Functional Modeling in Texas communities

Subsession Time: Friday, October 17 at 11:15am – 11:30am

Authors: Vanessa A Gomez (Baylor College of Medicine, Jan & Dan Duncan Neurological Research Institute), Blake Vuocolo (Baylor College of Medicine, Jan & Dan Duncan Neurological Research Institute), Mingxi Deng (Baylor College of Medicine, Jan & Dan Duncan Neurological Research Institute), Ryan German (Baylor College of Medicine, Jan & Dan Duncan Neurological Research Institute), Michal Tyrlik (Baylor College of Medicine, Jan & Dan Duncan Neurological Research Institute), Carlos Bacino (Baylor College of Medicine, Texas Children's Hospital, Baylor Genetics Laboratories), Chaya Nautiyal Murali (Baylor College of Medicine, Texas Children's Hospital), Seema Lalani (Baylor College of Medicine, Texas Children's Hospital), Stephanie Baskin (Baylor College of Medicine, Department of Pediatrics), Elizabeth Roeder (Baylor College of Medicine, Department of Pediatrics), Rebecca Okashah Littlejohn (Baylor College of Medicine, Department of Pediatrics), Pengfei Liu (Baylor College of Medicine, Baylor Genetics Laboratories), Zhandong Liu (Baylor College of Medicine, Jan & Dan Duncan Neurological Research Institute), Seon Young Kim (Baylor College of Medicine, Jan & Dan Duncan Neurological Research Institute), Shinya Yamamoto (Baylor College of Medicine, Jan & Dan Duncan Neurological Research Institute, Department of Neuroscience), Oguz Kanca (Baylor College of Medicine, Jan & Dan Duncan Neurological Research Institute), Hugo J Bellen (Baylor College of Medicine, Jan & Dan Duncan Neurological Research Institute, Department of Neuroscience), Michael F Wangler (Baylor College of Medicine, Jan & Dan Duncan Neurological Research Institute)

Abstract: Next-generation sequencing has proven to be an effective clinical tool in resolving diagnostic odysseys for the last decade, yet many individuals have economic or geographic limitations that prevent them from accessing genetic testing. The Texome Project is an NIH-funded research study that offers genomic sequencing to patients with undiagnosed, rare diseases who have experienced challenges in obtaining genetic sequencing clinically. Individuals who meet these criteria receive a genetic evaluation from a medical provider, CAP/CLIA genome (GS) or exome sequencing (ES), and genetic counseling. To date, there have been 353 applications reviewed by the Texome team, and 296 have been accepted into the study (84%). Individuals in our cohort reside in multiple regions of Texas, and most patients are either publicly insured or not insured (61.7%) and live below 200% of the federal poverty line (59%). To date, 172 individuals have undergone exome sequencing, with an etiologic diagnosis confirmed in 58 (33%) of cases, 22 of which (~38%) had changes to their medical management, including disease-specific treatments, referrals to new specialists, surveillance protocols, and cascade testing. Most notably, 28 out of 69 (41%) enrolled adult patients received a genetic diagnosis, a majority of whom experienced symptom onset at birth or early childhood and were able to end decades-long diagnostic odysseys. For the cases that remain unsolved, we utilize Artificial Intelligence through a tool we developed called AI-MARRVEL, allowing prioritization of variants from raw sequencing files. In one example, AI-MARRVEL identified a novel gene candidate for which we have subsequently established a cohort through

matchmaking. We have also developed *Drosophila* models for functional studies of novel genes nominated by re-analysis of genomic sequencing data. We use genetic strategies to study variants in flies, including humanization of these candidate genes in the fly to test the variant function and provide evidence for pathogenicity. In summary, the Texome project's utilization of advanced genomic sequencing, AI reanalysis, and functional modeling drives new knowledge of genetic etiologies of disease in populations with financial barriers to clinical genomic sequencing.

Benchmarking genAI tools for literature retrieval and summarization for genomic variant interpretation

Subsession Time: Friday, October 17 at 11:30am – 11:45am

Authors: Andrea Gazzo (Memorial Sloan Kettering Cancer Center), Matteo Biancospino (Johnson & Johnson Medical Affairs Department), Lorenzo Cuollo (Columbia Center for Genetic Errors of Immunity, Department of Pediatrics, Columbia University), Flavia Dei Zotti (Department of Pathology & Cell Biology, Columbia University), Emanuela Ferraro (European Institute of Oncology IRCCS), Antonio Marra (European Institute of Oncology IRCCS), Enrico Tartarotti (Ca' Foncello Treviso Regional Hospital, Department of Women's and Children's Health), Silvia Berardelli (University of Pavia)

Abstract: Background:

Generative AI is increasingly used to gather detailed information across different fields, including academic research. In genomics, where the accuracy of the extracted information is critical, its applicability has not yet been evaluated. This study investigates the reliability of genAI in extracting genomic variant-related information from scientific literature by comparing the performance of five different tools. Each tool's output was blindly evaluated by six reviewers on a dataset of selected genomic variants.

Methods:

Six reviewers, three germline and three somatic experts, were each assigned 20 variants (20 germline / 20 somatic), selected to represent diverse pathogenicity, literature support, and disease contexts. For each variant, summary reports were generated using five different tools. When supported by the tool, we used a curated prompt reflecting a clinician's need for a literature-based overview of the variant. The benchmark includes ChatGPT, MistralAI, VarChat, Perplexity, and ScholarAI. ChatGPT and MistralAI are general-purpose tools enabled to access websites and biobanks, while VarChat, Perplexity, and ScholarAI can access only scientific literature. Reviewers scored the summaries (1 to 5) using five metrics: Summarization Accuracy & Completeness, Hallucination Rate, Readability & Usability, Literature Coverage & Relevance, and Time Efficiency.

Results:

Preliminary results show that VarChat performs best for both germline and somatic variants, with overall mean scores of 4.27 (SD = 0.76) and 4.4 (SD = 0.55), respectively, demonstrating lower hallucinations and higher accuracy. ChatGPT reached 3.9 (SD = 0.72) on somatic variants and 3.69 (SD = 1.57) on germline

variants. ScholarAI (mean = 3.04, SD = 0.74) and Perplexity (mean = 3.15, SD = 0.75) showed lower performance in both scenarios. For variants with limited literature availability, the overall quality of summaries tends to decrease. In such cases, general-purpose tools with broader access to different sources, i.e. based not only on peer-reviewed literature, may perform better.

Conclusions:

While general-purpose models are not yet able to produce reliable summary reports on genomic variants, especially because they are affected by hallucinations and limited domain understanding, specialized tools demonstrate higher accuracy and relevance. With proper validation and domain-specific tuning, generative AI can become a valuable asset to support clinical genomics research.

Session 76: Unraveling Disease Mechanisms Through Multi-Omics: From Genetics to Microbiome and Chromatin Regulation

Location: Room 210AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Friday, October 17 at 10:45am – 11:45am

Microbiome and host genetic effects on gene regulation in Inflammatory Bowel Disease

Subsession Time: Friday, October 17 at 10:45am – 11:00am

Authors: Shreya Nirmalan (Wayne State University, University of Chicago), Julong Wei (Wayne State University), Sabrina Arif (University of Chicago), Sambhawa Priya (University of Chicago), Ran Blekhman (University of Chicago), Roger Pique-Regi (Wayne State University), Francesca Luca (University of Chicago)

Abstract: Inflammatory Bowel Disease (IBD), a chronic inflammatory disease of the digestive tract, is associated with altered gut microbiome composition and host genetic risk variants located within genes involved in microbial response pathways. Both host genetic variants and the gut microbiome regulate host gene expression in the colon; however, it remains unclear whether interactions between host genetics and the microbiome (GxM) regulate host gene expression in the gut, and what role GxM plays in IBD. Identifying GxM interactions that modify genetic risk for IBD is important for elucidating its pathogenesis and developing treatments. Here, we used the data from the Inflammatory Bowel Disease Multi'omics Database for 86 individuals (64 patients with IBD and 22 controls) consisting of host genotype, host gene expression, and mucosal gut microbiome (16S rRNA) data from rectum and ileum biopsies to perform interaction expression Quantitative Trait Locus (eQTL) mapping and identify GxM interactions in IBD. We identified 3,777 host genes with eQTLs in the rectum and 3,694 host genes with eQTLs in the ileum (FDR = 10%). We performed fine-mapping and used the fine-mapped eQTLs to test for GxM interactions. We found 36 GxM interactions for 31 host genes with 22 microbial taxa in the rectum and 30 GxM interactions in the ileum for 15 host genes and 20 taxa (FDR = 10%). Taxa with GxM interactions clustered into two distinct groups that have opposing interaction effects and reflect different functions of microbes in the gut. i.e, butyrate producers versus sulfate reducers. Using Transcriptome Wide Association Studies (TWAS) and colocalization analyses, we found that 23 SNPs with GxM interactions regulate the expression of host genes putatively causal for IBD, Ulcerative Colitis (UC) or Crohn's Disease (CD) (INTACT FDR = 10%). For example, increased expression of the gene NFKB1, which encodes the transcription factor NF- κ B, is associated with increased risk of UC. We found that increased abundance of the genus Holdemanella, a short-chain fatty acid producing taxon, is associated with decreased expression of NFKB1 in individuals with the alternative allele at rs1020760, potentially decreasing UC risk. Our results show the first evidence of genetic effects on gene expression that are

modulated by microbiome composition in IBD, and provide insight into how IBD risk could be reduced by targeting specific microbial taxa contingent on host genotype.

Multi-omics integration nominates disease-relevant cell types and novel effector genes for Crohn's disease and ulcerative colitis

Subsession Time: Friday, October 17 at 11:00am – 11:15am

Authors: Laura Fachal (Wellcome Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK), Rui Zhang (Massachusetts General Hospital, Boston, MA, USA), Kyle Gettler (Icahn School of Medicine at Mount Sinai, New York, NY, USA), on behalf of the International IBD Genetics Consortium.

Abstract: Genome-wide association studies have identified 370 signals associated with inflammatory bowel disease (IBD) and its two main subtypes, Crohn's disease (CD) and ulcerative colitis (UC). Revealing new association signals and the genes dysregulated by these variants is essential for advancing our understanding of IBD. However, pinpointing the causal genes remains challenging, as most associated variants are non-coding.

We conducted a GWAS meta-analysis of 67 cohorts spanning European, East Asian, and South Asian populations. The dataset included 125,992 IBD patients (59,734 CD; 57,565 UC) and >1.2 million population controls. Fine-mapping (FM) was performed to identify the most likely causal variants.

To assess the cellular contexts of these variants, we performed heritability enrichment analyses using bulk and single-cell (sc) ATAC-seq data from five studies. Gene prioritisation was achieved by colocalising cis expression QTL (eQTL) data from over 30 relevant tissues and cell types. We generated disease-specific protein QTL (pQTL) data, and integrated it, along with two large pQTL studies, with our GWAS using Mendelian randomisation (MR).

We identified 290 novel, independent, disease-associated signals. Combined with all known loci, these explain >77% of the heritability attributed to common variants. While many signals were shared between CD and UC, 38% were subtype-specific.

IBD-associated variants showed significant enrichment in regulatory elements of blood and tissue-resident T cells, macrophages, and B cells. Notably, UC variants showed greater enrichment in regulatory regions of intestinal epithelial cells compared to CD.

Using FM, colocalisation, and MR, we prioritized 626 candidate genes. This set was enriched for known receptor–ligand pairs (e.g. SERPINE1–ITGAV and VTN–PLAUR). Thirteen genes, including CARD11, SHARPIN, and IL10RB, are implicated in monogenic IBD. Multiple lines of evidence supported the molecular mechanisms linking variants to disease risk. For example, three novel risk loci were associated with increased expression of PDCD1, SELL, and MMP10, respectively. MR analyses confirmed that serum levels of PD-1 and L-selectin (encoded by PDCD1 and SELL) are causally associated with CD susceptibility, while MMP10 is linked to IBD.

This study advances our understanding of the cellular contexts through which genetic variants influence IBD susceptibility and provides a robust set of candidate genes for future functional and therapeutic research.

Integrating GWAS and CRISPR screens reveals context-specific T cell cytokine regulation in inflammatory bowel disease

Subsession Time: Friday, October 17 at 11:15am – 11:30am

Authors: Ruize Liu, Rui Zhang, Ruifei Zhu, Hailiang Huang (Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital; Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard)

Abstract: Inflammatory Bowel Disease (IBD), affecting ~1% of the US population, is a complex autoimmune disorder. While GWAS identified >300 IBD risk loci, their molecular mechanisms and cell-type-specific roles are largely unclear. Studies have suggested dysregulation of cytokine production, e.g., IL-2 and IFN- γ , in T cells as a potential pathogenesis mechanism. In this study, we integrate large-scale IBD genetic study (N=59,957, PMC5289481) with the latest genome-wide CRISPR screens to reveal context-specific (e.g., CD4+, CD8+, Tregs, Teffs) T cell regulators in IBD pathogenesis.

We performed LDSC analysis on CRISPRa/i screens-identified gene sets regulating IL-2 and IFN- γ in T cells (PMC9307090). We found significant enrichment of IBD heritability in genes regulating IFN- γ production in CD8+ T cells and IL-2 production in CD4+ T cells ($P < 1e-5$), indicating these cytokine regulators are key contributors to IBD genetic risk. Integrating with fine-mapped IBD common variants (PIP > 0.5) implicated PTPN22 and IL2RA as potential modulators of cytokine production. Further analysis incorporating rare loss-of-function (LoF) variants (MAF < 0.01) from IBD exome sequencing (ongoing unpublished study with 86,213 cases and 478,363 controls) identified three IBD candidate genes potentially influencing the production of cytokines: ARHGDIB, RELA, and FUBP1.

We then integrated our genetic finding with a CRISPR KO screen (PMC11754113) in both regulatory T cells (Tregs, high IL-2 α expression) and effector T cells (Teffs, upregulate IL-2 α upon stimulation) to further investigate the context-specific regulators in IBD pathogenesis. We found that IBD genetic risk was significantly enriched only in genes regulating IL-2 α in stimulated Teff cells ($P < 0.003$), not in Tregs or resting Teffs. This strongly suggests IL-2 α regulation in activated Teffs is a critical IBD pathogenic mechanism.

Notably, using Schizophrenia GWAS data as a negative control, no such significant enrichments were found.

In summary, our findings link IBD genetic risk to regulators of IL-2 (CD4+ T cells) and IFN- γ (CD8+ T cells). Crucially, genetic risk factors appear to predominantly impact IBD by affecting IL-2 α regulation within stimulated Teffs. This context-specific discovery significantly advances IBD mechanistic understanding and suggests that therapeutic interventions specifically targeting these T cell states could be beneficial.

Resolving genome regulatory complexity with new multiomic long-read sequencing tools for chromatin state, DNA methylation, and genetic variation

Subsession Time: Friday, October 17 at 11:30am – 11:45am

Authors: Andrea L. Johnstone (EpiCypher, Inc), James T. Anderson (EpiCypher, Inc), Connor P. Frasier (EpiCypher, Inc), Sabrina R. Hunt (EpiCypher, Inc), Paul W. Hook (Department of Biomedical Engineering, Johns Hopkins University), Zu-Wen Sun (EpiCypher, Inc), Martis W. Cowles (EpiCypher, Inc), Bryan J. Venters (EpiCypher, Inc), Andrew Stergachis (Division of Medical Genetics, Department of Medicine, University of Washington), Winston Timp (Department of Biomedical Engineering, Johns Hopkins University), and Michael-Christopher Keogh (EpiCypher, Inc)

Abstract: Many disease-associated genetic variants reside in non-coding and structurally complex regions of the human genome where they influence transcriptional regulation by modulating chromatin accessibility, transcription factor binding, and epigenetic modifications, including DNA methylation (DNAm) and histone post-translational modifications (PTMs). Resolving these regulatory mechanisms is critical for translating genetic discoveries into clinical insights. However, commonly used short-read sequencing approaches lack multiomic resolution and cannot capture haplotype-specific regulatory features or fully profile repeat-rich loci, segmental duplications, and sites of structural variation.

To overcome these limitations, we present two multiomic long-read sequencing (LRS) methods that preserve native relationships between genetic and epigenetic features on individual DNA molecules, generating long reads (>10 kb) that enable variant-to-function insights across challenging regions of the genome. Fiber-seq employs an N6-adenine methyltransferase (6mA MTase; Hia5) to profile chromatin accessibility while retaining endogenous DNA methylation (5mC). We demonstrate that Fiber-seq is highly sensitive, enabling detection of transcription factor binding footprints alongside broadly accessible chromatin. Fiber-seq has provided diagnostic insights in previously unsolved genetic disease cases involving cis-regulatory haplotypes within structurally complex loci (PMID: 40166185). CUTANA-LRS complements this approach by using a tethered 6mA MTase (pAG-M.EcoGII) to label DNA beneath antibody-targeted histone PTMs or chromatin-associated proteins. CUTANA-LRS is ideal for mapping chromatin proteins that do not have a distinct binding motif or interact indirectly with DNA, while Fiber-seq provides an ideal approach for unbiased analysis of chromatin accessibility and inferred transcription factor footprinting.

Through validation against orthogonal short-read epigenomic assays (e.g. CUT&RUN) and application across multiple human cell types, we demonstrate that both methods enable robust parallel profiling of chromatin features, DNA methylation, and haplotype-resolved genetic variation. Multiomic LRS thus provides a powerful new tool for advancing human genetics, epigenomics, and clinical genomics by supporting interpretation of variants of uncertain significance (VUS), elucidating gene regulatory mechanisms, and identifying novel biomarkers and therapeutic targets.

Session 84: All That and a Bag of Variants: Functional Genomics of Brain Disorders

Location: Room 210AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Saturday, October 18 at 8:30am – 9:30am

Optimally matching healthy and disease-specific adult brains onto the EN-TE_x personal epigenome resource

Subsession Time: Saturday, October 18 at 8:30am – 8:45am

Authors: Alexander Popescu (Department of Molecular Biophysics and Biochemistry, Yale University; Program in Computational Biology and Biomedical Informatics, Yale University), Matthew Jensen (Department of Molecular Biophysics and Biochemistry, Yale University; Program in Computational Biology and Biomedical Informatics, Yale University), Marcus Lee (Department of Molecular Biophysics and Biochemistry, Yale University; Program in Computational Biology and Biomedical Informatics, Yale University), Eric Nguyen (Department of Molecular Biophysics and Biochemistry, Yale University; Program in Computational Biology and Biomedical Informatics, Yale University), Jiaqi Li (Department of Molecular Biophysics and Biochemistry, Yale University; Program in Computational Biology and Biomedical Informatics, Yale University), Timur Galeev (Department of Molecular Biophysics and Biochemistry, Yale University; Program in Computational Biology and Biomedical Informatics, Yale University), Mark Gerstein (Department of Molecular Biophysics and Biochemistry, Yale University; Program in Computational Biology and Biomedical Informatics, Yale University)

Abstract: The ability to identify allele-specific (AS) events at both bulk and single-cell levels is essential for understanding how genomic variants influence tissue- and cell-type-specific gene regulation. However, discovering these AS events requires mapping to diploid personal genomes rather than a haploid reference genome. Our objective is to map and compare AS expression, binding, and chromatin accessibility in the human brain across healthy and disease states, with a particular emphasis on the impact of both single nucleotide variants (SNVs) and structural variants (SVs) toward disease phenotypes in neuropsychiatric disorders. First, to investigate AS events in the healthy brain, we integrated matched functional bulk RNA-seq and ChIP-seq datasets from 12 brain tissues in four matched control individuals from the Genotype-Tissue Expression (GTEx) project with the non-brain-specific EN-TE_x resource, which mapped ~15 functional genomics assays across ~30 tissues to personal diploid genomes. Using personal genomes incorporating roughly five million variants per individual, we identified over 25,000 AS expression and binding events, a subset of which were uniquely observed in brain tissues. As the effects of SVs toward regulatory elements are not as well characterized as for SNVs, we overlapped ~200,000 ChIP-seq peaks with SV calls for each haplotype. Peaks disrupted by SVs on one haplotype, which we term “haplo-novel” peaks, were located within 50kbp of several thousand genes on average per individual. We further analyzed bulk and single-nucleus RNA-seq, ATAC-seq, and Multiome data from the prefrontal cortex of eight healthy individuals in the PsychENCODE Consortium, matched by age and sex

to the EN-TEX samples. Using a similar personal-genome mapping approach, we identified thousands of AS loci with significant expression or epigenetic imbalances between the two parental haplotypes in 27 brain cell types. A subset of these loci showed cell-type specificity compared to bulk datasets. Finally, to exemplify the relevance of this work for disease genomics, we compared these cell-type-specific AS loci to those from individuals with schizophrenia (n=4) or autism spectrum disorder (n=4) and identified a subset of AS loci that are active in a disease-specific manner across relevant cell types. Overall, this work deepens our understanding of the functional impact of genetic variants toward brain-specific gene regulation in both healthy individuals and individuals with neuropsychiatric disorders.

★ Genetic atlas of the human brain revealed by a 2-Million-Voxel GWAS at millimeter resolution

Subsession Time: Saturday, October 18 at 8:45am – 9:00am

Authors: Xingzhong Zhao(Department of Bioinformatics and Systems Medicine, McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, Texas, United States of America), Ziqian Xie(Department of Bioinformatics and Systems Medicine, McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, Texas, United States of America), Han Chen(Human Genetics Center, Department of Epidemiology, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, United States of America), Degui Zhi (Department of Bioinformatics and Systems Medicine, McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, Texas, United States of America),

Abstract: Understanding how genetic variations shape the human brain structure remains a central challenge in neuroscience and genomics. Traditional brain imaging genetics often relies on atlases with predefined regions of interest (ROIs), limiting the detection of effects beyond the ROI-boundaries. Voxel-level GWAS (vGWAS) was attempted, but it used lower-resolution brain imaging data and did not produce many significant associations due to limited sample sizes. Here, we present the first 2mm-resolution brain imaging genetic atlas consisting of tests for all single nucleotide polymorphism (SNP)-voxel pairs for the 2,208,500 voxels of the brain template and the 8.9 million SNPs by voxel-wise genome-wide association study (vGWAS) using T1-weighted magnetic resonance imaging (MRI) data from the UK Biobank (N = 33,417). Leveraging our custom-developed GPU-accelerated tool Torch-GWAS, we were able to perform these large-scale analyses within five days. We identified over 15.9 million genome-wide significant associations ($P < 2.26 \times 10^{-15}$) between 66,226 SNPs (v-snps) and 621,317 voxels (s-voxels). Some significant SNPs (e.g., rs273944, rs1677498) achieved an extremely significant level ($P = 4.48 \times 10^{-251}$). In addition, we identified 11,849 lead SNPs ($LD\ r^2 < 0.1$) and mapped these to 1,406 genes. These significant genes are enriched in neurobiological pathways, including synapse organization (Bonferroni corrected $P = 5.77 \times 10^{-4}$, Hypergeometric Test) and projection development (Bonferroni corrected $P = 1.44 \times 10^{-3}$, Hypergeometric Test). Our voxel-level brain imaging genetics atlas can reveal the spatially specific genetic influences(e.g., SNP rs4971088, previously found white matter related, shows more significant GWAS p-value on WM than other tissue), structural symmetry (e.g., the number of significant SNP overlaps was significantly higher on counterpart ROIs between hemispheres (e.g., left-

and right-hippocampus than on asymmetric brain regions, $P = 2.63e-9$, Wilcoxon test), and links to neurological disorders. As a use case, our atlas for a SNP can serve as a reference t-map for the interpretation of the SNP. For instance, the voxel-level t-map of our atlas for a known epilepsy risk SNP rs28498976 (nearest gene LINC02497) was significantly overlapping with ENIGMA epilepsy-related functional maps (Bonferroni corrected $P = 1.11e-5$, Spearman correlation test). This resource offers new opportunities for studying the genetic architecture of brain structure and its relevance to disease.

★ Uncovering molecular mechanisms of neuropsychiatric risk via functional genomics and AlphaFold 3

Subsession Time: Saturday, October 18 at 9:00am – 9:15am

Authors: Sherif Gerges (Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard), Nikolaj Catois Straarup (Department of Molecular Biology and Genetics, Aarhus University), Mohamed El-Brolosy (Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology), F. Kyle Satterstrom (Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard), Nolan Kamitaki (Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard), Emi Ling (Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard), Sabina Berretta (Department of Psychiatry, Harvard Medical School), Jonathan S. Weissman (Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology), Jen Q. Pan (Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard), Charlott Stock (Department of Molecular Biology and Genetics, Aarhus University), Poul Nissen (Department of Molecular Biology and Genetics, Aarhus University), Steve McCarroll (Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard), Hilary Finucane (Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard), Mark Daly (Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard)

Abstract: While genetic studies have identified numerous risk loci for neuropsychiatric disorders, the biological mechanisms they perturb remain unclear. We generated single-nucleus RNA sequencing (snRNA-seq) data from multiple human brain regions to weight gene co-expression by polygenic heritability, revealing disease-relevant pathways. We then leveraged the largest exome datasets from the Autism Sequencing Consortium (presented at ASHG 2024 -38,680 probands, 9,567 siblings) and schizophrenia (SCHEMA consortium-24,248 cases, 97,322 controls) to show that both rare and common variants converge on Ca^{2+} homeostasis as a key etiological process in neuropsychiatric disorders.

Within this pathway, we identified ATP2B2—a P-type ATPase responsible for Ca^{2+} extrusion – as consistently downregulated in the prefrontal cortex in donors with schizophrenia compared to controls, both synaptic proteomes (34 cases, 35 controls, $P=3 \times 10^{-3}$) and snRNA-seq (94 cases, 97 controls, $P=6.6 \times 10^{-7}$). This reduction is specific to excitatory neurons, pointing to a cell type-specific loss-of-function (LoF) mechanism linking ATP2B2 to schizophrenia risk.

ATP2B2 shows a notable excess of missense variants linked to schizophrenia and autism. To examine how these variants disrupt Ca^{2+} homeostasis, we developed a 3D enrichment method using AlphaFold 3 to test for excess case variants within 15Å spherical neighborhoods around each residue. We found variants

significantly clustered around the Ca^{2+} binding site ($P=2.06 \times 10^{-6}$) and ATP:Mg^{2+} coordination region ($P=2.01 \times 10^{-4}$). Suggesting two distinct mechanisms of ATP2B2 perturbation. In the Ca^{2+} binding neighborhood, substitution of the Ca^{2+} -binding residue E457 with lysine (E457K) introduces a charge-reversal, suggesting disrupted binding as a focal mechanism of pathogenic variation in ATP2B2. We used AlphaFold 3 to simulate ATP2B2 and Ca^{2+} with and without E457K, and found it markedly reduced Ca^{2+} contact probabilities relative to wildtype, supporting an LoF effect.

We validated E457K's impact in two orthogonal assays. The variant abolished ATPase activity in recombinant ATP2B2 and in a cellular context it impaired Ca^{2+} extrusion in HEK293 cells using a GCaMP6s-based imaging assay - both consistent with a LoF mechanism aligned with the direction of genetic risk.

Together, our findings link genetic risk to neuronal dysfunction via a critical calcium signaling axis, offering mechanistic insight into the pathogenesis of neuropsychiatric disorders.

SingleBrain: A Meta-Analysis of Single-Nucleus eQTLs Linking Genetic Risk to Brain Disorders

Subsession Time: Saturday, October 18 at 9:15am – 9:30am

Authors: Beomjin Jang(Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA), Kailash BP(Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA), Alex Tokolyi(New York Genome Center, New York, NY, USA), Winston H. Cuddleston(Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA), Ashvin Ravi(Biological and Medical Informatics PhD Program, University of California, San Francisco, San Francisco, CA), Sang-Hyuk Jung(Department of Medical Informatics, Kangwon National University College of Medicine, Chuncheon, Republic of Korea), Tatsuhiko Naito(Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA), Beomsu Kim(Department of Digital Health, SAIHST, Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea), Min Seo Kim(Medical and Population Genetics and Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, USA), Minyoung Cho(Department of Digital Health, SAIHST, Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea), Mi-So Park(Department of Digital Health, SAIHST, Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea), Mikaela Rosen(Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA), Joel Blanchard(Nash Family Department of Neuroscience & Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA), Jack Humphrey(Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA), David A Knowles(New York Genome Center, New York, NY, USA), Hong-Hee Won(Department of Digital Health, SAIHST, Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea), Towfique Raj(Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA)

Abstract: Introduction: Non-coding genetic variants identified by GWAS influence brain disorders through cell-type-specific gene regulation, but their functional effects remain largely unresolved. While

eQTL studies in bulk brain tissues have provided insight, single-nucleus RNA-seq (snRNA-seq) enables eQTL discovery at cell-type resolution. However, statistical power has been limited by sample size and cell counts.

Methods: We conducted a meta-analysis of cis-eQTLs and trans-eQTLs by integrating snRNA-seq from mostly prefrontal cortex and genotype data from four major cohorts (Fujita, Mathys, Gabitto, Bryois), totaling 5.8 million nuclei from 983 European-ancestry donors. Using linear mixed models, we mapped cell type and cell state eQTLs across seven major brain cell types and 28 subtypes. We performed colocalization, Mendelian randomization (MR), and fine-mapping to link eQTLs to neurodegenerative and neuropsychiatric GWAS loci.

Results: "SingleBrain" revealed over 11K eGenes in excitatory neurons and 1.8K cell-type-specific eQTLs. Compared to prior datasets, eQTL discovery improved 3–10X. Colocalization and MR identified 143 GWAS loci linked to eQTLs across major disorders. AD risk was strongly mediated by microglial eQTLs, while SCZ, BPD, and PD showed enrichment in neurons. Subtype-level analysis revealed microglia subpopulation-specific eQTLs for AD and PD, including FGF20 (PD, MG3) and SORL1 (AD, MG1). Fine-mapping reduced credible variants and identified many causal variants within cell-type-specific enhancers, many of which were validated in independent microglia-specific MPRA study.

Conclusion: "SingleBrain" is the largest snRNA-seq eQTL meta-analysis in the human brain to date. It provides a high-resolution map of regulatory variants across brain cell types and states, linking non-coding variants to cell-type-specific gene regulation in brain disease. This resource prioritizes causal genes and variants for future functional validation and therapeutic targeting.

Session 85: Better Together: Multi-Ancestry Approaches for Gene Discovery

Location: Room 253ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Saturday, October 18 at 8:30am – 9:30am

Multi-ancestry genome-wide and transcriptome-wide association analyses identify new risk loci and genes for Inflammatory Bowel Disease

Subsession Time: Saturday, October 18 at 8:30am – 8:45am

Authors: Linshuoshuo Lyu (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA), Qing Li (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA), Chao Li (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA), Wanqing Wen (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA), Ran Tao (Department of Biostatistics, Vanderbilt University Medical Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA), Ghadeer K. Dawwas (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA), Xiao-Ou Shu (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA), Wei Zheng (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA), Zhijun Yin (Department of Biomedical Informatics, Vanderbilt University Medical Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA), Ken S Lau (Department of Cell and Developmental Biology, Vanderbilt University Medical Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA), Xingyi Guo (Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, and Vanderbilt-Ingram Cancer Center, Department of Biomedical Informatics, Vanderbilt University Medical Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA)

Abstract: Background: Genome-wide association studies (GWASs) have identified approximately 400 loci associated with inflammatory bowel disease (IBD), while the biological mechanisms and causal genes remain largely unsolved.

Methods: We conducted a GWAS meta-analysis of 49,022 IBD cases (including 20,093 Crohn disease (CD) and 28,485 Ulcerative colitis (UC)) and 1,293,489 controls of European ancestry using data from previous GWAS, Million Veteran Program, FinnGen, and UK Biobank, following by subtype-stratified analyses. We further conducted a multi-ancestry analysis by incorporating additional 14,393 cases of IBD and 15,456

controls of East Asians. To identify putative causal genes, we applied our developed sTF- TWAS to conduct multi-ancestry TWAS, alternative splicing (AS-WAS), and alternative polyadenylation (APA-WAS) using RNA-seq data from 364 East Asian and 707 European normal colon tissues. We further leveraged scRNA-seq reference data from the COLON MAP to deconvolute the bulk transcriptomes and performed multi-ancestry single cell-TWAS (scTWAS). We next conducted x-QTLs (eQTLs, AS-QTLs, APA-QTLs) and colocalization analyses to identify putative target genes for GWAS loci. High-confidence genes were selected using FOCUS fine-mapping (PIP > 0.5) and colocalization (PP.H4 > 0.5). Druggable targets were annotated using drug- protein database (e.g., ChEMBL), focusing on drugs that has been approved or at phase II/III clinical trials.

Results: We identified 90 novel IBD risk loci (either CD, UC, or IBD), including 59 from the European-only analysis and an additional 31 from the multi-ancestry meta-analysis at $p < 5 \times 10^{-8}$ threshold. Integrative analyses across TWAS, APA-WAS, AS-WAS, scTWAS, and GWAS loci identified 613 high-confidence IBD risk genes, including 477 novel genes located within 46 newly discovered GWAS loci. Of these, 94 genes are linked to 386 drugs. Notably, we identified key genes involved in established IBD-related pathways, including cytokine signaling (e.g., IL23R), innate immunity (e.g., ATG16L1), adaptive immunity (HLA-DRA), MAPK and NF- κ B signaling (e.g., NFKB1), lipid and fatty acid metabolism (FADS1), and chromatin regulation and epigenetics (e.g., EZH2).

Conclusion: Our study identified large number of novel genes using integrative, multi-ancestry, and cell type-resolved framework. These findings have the potential to broaden the genetic and molecular landscape of IBD towards improving precision medicine in IBD.

★ Largest multi-ancestry genetic study of 99 infectious disease traits identifies hundreds of novel disease-associated loci across the whole genome

Subsession Time: Saturday, October 18 at 8:45am – 9:00am

Authors: Abhijith Biji (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029), Sinead Cullina (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029), Eimear E. Kenny (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029), Samira Asgari (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029)

Abstract: Background: Infectious diseases remain a major concern, amplified by population growth and climate change. While genetic variation affects disease risk and severity, few loci have been reliably identified. We aim to address this gap.

Methods: We performed ancestry-stratified GWAS (MAF0.5%) with SAIGE across 99 infectious disease phecodes in All of Us, BioMe, and UK Biobank (N=813,000), with 33.6%, 62%, and 2% non-European individuals respectively. The results were meta-analysed by inverse-variance weighting.

Results: We identified 217 genome-wide significant loci (177 unique loci) associated with 80 phecodes, without systematic inflation (λ_{GC} median[IQR]=1.03[1.01-1.04]), and 94% of variants passed the heterogeneity test. 88% of our variants or those in LD ($r^2 > 0.6$, 1Mbp window) were unreported in GWAS

catalog, 13 variants had infectious disease associations, and only 7 were reported for the same disease. HLA locus showed the strongest and most number of associations (14 variants, lowest $p=3.3e-28$). Beyond HLA, 6 loci had multiple infectious disease associations, indicating pleiotropy. ANNOVAR showed most variants to be intergenic (47.5%) or intronic (27.6%). 33 variants were significant eQTL or in LD with one ($r^2>0.6$, 1Mbp window) in GTEx.

We used ClinVar and CADD to predict pathogenicity. rs113993960 (CADD=19.2), homozygous pathogenic for cystic fibrosis (CF), increased risk of respiratory infections like mycobacteria, pseudomonas, aspergillosis, and drug-resistant Staphylococcus aureus infections (lowest $p=6.8e-24$). rs334 (CADD=13.1), homozygous pathogenic for sickle-cell anaemia (SCA), was associated with sepsis and E.coli risk (lowest $p=4.9e-9$). Thus, contrary to common wisdom, CF and SCA carriers (>13 million in the US) are at a higher risk of infectious diseases. We further queried for signals of selection. The strongest selection coefficient (2.6%, $p=2.1e-22$) was obtained for rs9267123, a novel risk variant for of plantar warts (OR [CI]=1.4[1.2-1.6], $p=5.6e-9$) in the HLA region and is more prevalent in individuals of European ancestry compared to others. This variant is highly pleiotropic with many cancers, blood traits, and autoimmune diseases.

Conclusion: This is the largest, most diverse GWAS of infectious diseases to date. Our results underscore the influence of host genetics beyond classical immune loci on infection risk, highlight new opportunities to study their genetic architecture, and can inform vaccine and drug discovery.

Multi-ancestry genome-wide association study of endometriosis and its clinical manifestations in ~1.4 million women: translating gene discovery into pathogenic mechanisms and therapeutic targets

Subsession Time: Saturday, October 18 at 9:00am – 9:15am

Authors: Dora Koller (Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA; Department of Genetics, Microbiology, and Statistics, Faculty of Biology, University of Barcelona, Barcelona, Spain), Jun He (Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA; Veterans Affairs Connecticut Healthcare Center, West Haven, CT, USA), Solveig Lokhammer (Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA; Department of Clinical Science, University of Bergen, Bergen, Norway; Dr. Einar Martens Research Group for Biological Psychiatry, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway), Selena Aranda (Department of Genetics, Microbiology, and Statistics, Faculty of Biology, University of Barcelona, Barcelona, Spain), Dan Qiu (Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA; Veterans Affairs Connecticut Healthcare Center, West Haven, CT, USA), David Davtian (Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA; Veterans Affairs Connecticut Healthcare Center, West Haven, CT, USA), Qianyu Chen (Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA; University of Connecticut, Storrs, CT, USA), Ziang Xu (Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA), Zhongzheng Mao (Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA), Eleni Friligkou (Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA), Sefayet Karaca (Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA; Faculty of

Health Science, Aksaray University, Aksaray, Turkey), Bru Cormand (Department of Genetics, Microbiology, and Statistics, Faculty of Biology, University of Barcelona, Barcelona, Spain; Biomedical Network Research Centre on Rare Diseases (CIBERER), Instituto de Salud Carlos III, Madrid, Spain; Institute of Biomedicine of the University of Barcelona (IBUB), Barcelona, Catalonia Spain; Sant Joan de Déu Research Institute (IR-SJD), Esplugues de Llobregat, Catalonia, Spain), Idhaliz Flores (Department of Basic Sciences, Ponce Health Sciences University, Ponce, PR, USA; Department of Obstetrics & Gynecology, Ponce Health Sciences University, Ponce, PR, USA), Signe Altmae (Department of Biochemistry and Molecular Biology I, Faculty of Sciences, University of Granada, Granada, Spain; Biosanitary Research Institute, Granada, Spain; Division of Obstetrics and Gynecology, Department of Clinical Science, Intervention and Technology, Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden; Department of Obstetrics and Gynaecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia), Marina Mitjans (Department of Genetics, Microbiology, and Statistics, Faculty of Biology, University of Barcelona, Barcelona, Spain; Institute of Biomedicine of the University of Barcelona (IBUB), Barcelona, Catalonia Spain; Sant Joan de Déu Research Institute (IR-SJD), Esplugues de Llobregat, Catalonia, Spain; Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain; Brenda Cabrera-Mendoza (Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA; Veterans Affairs Connecticut Healthcare Center, West Haven, CT, USA), Renato Polimanti^{1,3,18,19,20} (Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA; Veterans Affairs Connecticut Healthcare Center, West Haven, CT, USA; Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA; Wu Tsai Institute, Yale University, New Haven, CT, USA; Department of Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT, USA)

Abstract: Endometriosis is a systemic inflammatory disease that affects approximately 10% of women. Regardless of an active research in the field, the etiology of endometriosis remains unclear. While genetic factors are believed to contribute to its development, large-scale genomic studies investigating this relationship are still limited. Here, we conducted a multi-ancestry genome-wide association study of endometriosis and its clinical manifestations in ~1.4 million women, including 105,869 cases. Overall, we identified a total of 80 genome-wide significant associations (42 of them novel). Fine-mapping and colocalization analyses uncovered causal loci with respect to >50 endometriosis-related associations. Additionally, transcriptome-wide association analyses identified 54 genes associated with endometriosis through tissue-specific regulation related to uterus, vagina, ovary, breast mammary gland, and colon. Blood-based methylation-wide summary-based Mendelian randomization identified >500 CpG sites associated with endometriosis. Plasma-based proteome-wide association identified nine loci associated with endometriosis. Several loci converged across these omics-wide analyses. For instance, the strongest proteome-wide association was RSPO3, a key modulator of the Wnt signaling pathway, which includes WNT4 identified in the cross-ancestry, European-descent, and East-Asian-descent analyses. Drug-repurposing analyses of endometriosis-associated loci identified potential therapeutic targets, including drugs for breast cancer (neratinib, toremifene, and abemaciclib) and treatments for preterm-birth (17-hydroxyprogesterone caproate) and for endometriosis (norgestimate). Beyond loci discovery, we also investigated polygenic contribution to endometriosis clinical manifestations. For example, we observed that endometriosis polygenic risk interacts with fibromyalgia, infertility, menorrhagia, and depression in

increasing endometriosis risk. Genetic correlation, latent causal variable, and Mendelian randomization analyses identified endometriosis pleiotropy with women's health-related conditions, including excessive menstruation, menstrual-cycle related pain, infertility, and polyp of female genital tract. In conclusion, the present study greatly advances the understanding of endometriosis pathogenesis, characterizing how its systemic nature is linked to the broad range of symptoms reported by affected women.

Next generation multi-population GWAS of Coronary Artery Disease

Subsession Time: Saturday, October 18 at 9:15am – 9:30am

Authors: Satoshi Koyama (Broad Institute of MIT and Harvard, Cambridge, MA, USA; Massachusetts General Hospital, Boston, MA, USA), Anuj Goel (University of Oxford, Oxford, Oxfordshire, United Kingdom), Shoa Clarke (Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA; Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA; VA Palo Alto Health Care System, Palo Alto, CA, USA), Pik Fang Kho (Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA; VA Palo Alto Health Care System, Palo Alto, CA, USA; Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA, USA), Kruthika Iyer (Gladstone Institutes, San Francisco, CA, USA; University of California, San Francisco, San Francisco, CA, USA), Federico Murgia (Nuffield Department of Population Health, Big Data Institute, University of Oxford, Oxford, Oxfordshire, United Kingdom), Pradeep Natarajan (Broad Institute of MIT and Harvard, Cambridge, MA, USA; Massachusetts General Hospital, Boston, MA, USA; Harvard Medical School, Boston, MA), Catherine Tcheandjieu (Gladstone Institutes, San Francisco, CA, USA; University of California, San Francisco, San Francisco, CA, USA; VA Palo Alto Health Care System, Palo Alto, CA, USA), Jemma Hopewell (Nuffield Department of Population Health, Big Data Institute, University of Oxford, Oxford, Oxfordshire, United Kingdom), Themistocles Assimes (Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA; VA Palo Alto Health Care System, Palo Alto, CA, USA; Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA, USA) on behalf of the CARDIoGRAMplusC4D Consortium

Abstract: Coronary artery disease (CAD) accounts for roughly one in six deaths worldwide and remains a leading cause of mortality. Although traditional risk factors are well characterized, they do not fully explain individual susceptibility. To elucidate its molecular mechanisms, large-scale genome-wide association studies (GWAS) have been conducted and identified approximately 300 risk loci. However, the high heritability of CAD suggests that much of its genetic architecture remains unexplained.

On behalf of the CARDIoGRAMplusC4D Consortium, we assembled the largest multi-population GWAS resource to address this gap, analyzing genotype and phenotype data from over 4 million participants including nearly three quarters of a million cases. We performed population-specific fixed-effect meta-analyses in African-like (4.81%), Admixed-American-like (2.39%), East Asian-like (18.9%), European-like (71.2%), South Asian-like (2.08%), and Middle Eastern-like (0.19%) participants, followed by a multi-population meta-analysis. In total, 522 independent loci exceeded the genome-wide significance threshold ($P < 5 \times 10^{-8}$) and 233 of which have not been previously reported. Finemapping identified 96

variants with over 80% posterior probability of inclusion. The putative causal missense variants implicate genes central to lipid metabolism (PCSK9, APOB), vascular integrity (COL4A2), and inflammation (IL6R). SNP heritability was strongly enriched within active regulatory elements of arterial, cardiac, adipose and liver tissues.

Using these expanded data of genetic association for CAD, we built an updated polygenic score (PGS) and tested it in a held-out UK Biobank validation cohort of 336,112 individuals with 20,753 incident CAD events. The new PGS explained an additional 7.5% of phenotypic variance—a 39% relative improvement over previous scores. When combined with the QRISK3 clinical algorithm, the integrated model achieved an area under the receiver operating characteristic curve of 0.805 (versus 0.764 for QRISK3 alone; $P < 1 \times 10^{-300}$). Nonetheless, we observed marked fluctuations in predictive accuracy across not only population groups but also individuals, underscoring the need for methods to achieve suitable calibration.

In summary, our study meaningfully augments contemporary clinical and biological resources for CAD research, identifying >200 additional susceptibility loci, refining the resolution of causal variants and biological pathways, and establishing a clinically relevant PGS.

Session 86: Genomic Causes and Consequences in Drug Treatment

Location: Room 205ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Saturday, October 18 at 8:30am – 9:30am

Genome-wide association study for Glucocorticoid-induced Ocular Hypertension

Subsession Time: Saturday, October 18 at 8:30am – 8:45am

Authors: Jyoti Lama (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Renee Liu (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Alicia Huerta-Chagoya (Program in Metabolism and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, Center for Genomic Medicine and Diabetes Unit, Endocrine Division, Department of Medicine, Massachusetts General Hospital, Boston, MA), Ashley Li (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Katie Huynh (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Lynn Stanwyck (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Samuel Han (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Yan Zhao (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Weilin Chan (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Liyin Chen (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), George N. Papaliodis (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Lucy Q. Shen (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Christine Elkins (Retina Health Center, Fort Myers, Florida), Irma Benavides (Retina Health Center, Fort Myers, Florida), Antonio Cutino (Alimera Sciences Inc., Alpharetta, GA), Da Meng (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Janine Yunfan Yang (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Gayatri Susarla (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Janey L. Wiggs (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Alexander Eaton (Retina Health Center, Fort Myers, Florida), Ayellet V Segrè (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA), Lucia Sobrin (Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA)

Abstract: A major complication of Glucocorticoids (GC), a widely used anti-inflammatory therapy used to treat many ocular disorders, is intraocular pressure (IOP) elevation that can lead to glaucomatous optic nerve damage and vision loss. This study aimed to identify common and rare variants associated with GC-induced IOP change using genome-wide association study (GWAS) and whole exome sequencing (WES) analyses, respectively. Participants for the discovery cohorts were taken from the Fluocinolone Acetonide in Diabetic Macular Edema (FAME) trials and were exposed to GC primarily by intravitreal

injection. Replication was performed in an independent cohort from Mass Eye and Ear and Retina Health Center (MEE/RHC). The primary outcome tested was the maximum difference between IOP before GC treatment and any of the IOP measurements within 6 months after treatment. Common variants were genotyped using Illumina's Global Screening Assay followed by imputation to TOPMed and rare variants were detected with WES. GWAS and rare variant gene burden and SKAT-O analyses were applied to all ancestries and the European samples separately, adjusting for covariates. Meta-analysis was performed between the discovery and replication cohorts using METAL. To prioritize causal genes and pathways, we performed gene-level and pathway analyses using MAGMA and expression/splicing quantitative trait loci (eQTL/sQTL) colocalization using eCAVIAR. 530 FAME and 588 MEE/RHC participants were included in the GWAS. For the primary outcome, two novel variants in the UBE2E3 locus, rs13425173 (MAF=0.095, beta=-2.4mmHg) and rs104022 (MAF=0.093, Beta=-2.3mmHg), reached genome-wide significance ($P<5E-08$) for all ancestries in the discovery cohort and meta-analysis, respectively. The UBE2E3 discovery cohort variant significantly colocalized with eQTLs in 6 GTEx tissues. In the gene-level analysis, UBE2E3 also demonstrated subthreshold significance ($P=6E-06$). In the rare variant analyses, MSTO1 passed Bonferroni correction ($P<1.8E-06$) for the primary outcome using all ancestries in the discovery cohort, and ZNF248 and LDHAP5 passed replication Bonferroni cutoff in the meta-analysis. Several subthreshold genes from the GWAS and WES analyses were found to be associated with IOP in previously published GWAS. We have identified novel common variants, as well as for the first time, rare variants associated with GC-induced ocular hypertension that implicate different pathways, such as ubiquitination.

High-throughput Identification and Genetic Analysis of Adverse Drug Effects in Individuals of African Ancestry

Subsession Time: Saturday, October 18 at 8:45am – 9:00am

Authors: Mojgan Babanejad (Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA), Alyson Dickson (Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA), Yi Xin (Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA), Vivian K. Kawai (Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA, Department of Medicine, University of Alabama at Birmingham, Birmingham, IL, USA), Ran Tao (Department of Statistics, Vanderbilt University Medical Center, Nashville, TN, USA), Wei-Qi Wei (Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA), Monika Grabowska (Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA), Bingshan Li (Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA), Lan Jiang (Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA), Nikhil K. Khankari (Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA), Matthew Krantz (Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA), Xue Zhong (Department of Medicine, Vanderbilt Genetics Institute, Nashville, TN, USA), Eric Kerchberger (Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA), Christopher Guardo (Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA), Adriana Hung

(Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA), Gul Karakoc (Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA), Benjamin Sun (Bristol Myers Squibb, Informatics and Predictive Sciences, Research, Cambridge, MA, USA), Xiuwen Zheng (Abbvie pharmaceuticals, North Chicago, IL, USA), Aparna Vasanthakumar (Abbvie pharmaceuticals, North Chicago, IL, USA), Nizar Smaoui (Abbvie pharmaceuticals, North Chicago, IL, USA), Britney Milkovich (Abbvie pharmaceuticals, North Chicago, IL, USA), Jamie Wenke (Abbvie pharmaceuticals, North Chicago, IL, USA), Megan Shuey (Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA), C. Michael Stein (Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA), QiPing Feng (Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA)

Abstract: Background:

Adverse drug events (ADEs) are a common cause of preventable morbidity and mortality, and genetics are known contributors to ADEs. Rates and risk factors for ADEs vary among ancestral groups, yet most research has focused on individuals of European ancestry. We assessed genetic associations with ADEs in individuals of majority African ancestry using data from the Alliance for Genomic Discovery, which has generated whole-genome sequencing data for individuals in Vanderbilt's electronic health record (EHR)-linked BioVU biobank.

Methods:

The cohort included individuals with 1) >50% African ancestry, 2) genetic data that passed quality control, and 3) a problem list in their EHRs (to ensure adequate utilization). We used high-throughput methods to detect cases of ADEs for 11 drugs/drug classes (penicillin, CYP2D6-mediated opioids, sulfonamides, angiotensin converting enzyme inhibitors, morphine, statins, aspirin, cephalexin, erythromycin, levofloxacin, and meperidine) from the structured allergy/problem list entries. For each drug/drug class, controls were those individuals in the cohort without mention of that specific drug/drug class in their problem list. We used Regenie to conduct genome-wide association studies, adjusted for age, sex, EHR length, and 10 principal components. We included variants with a missing rate < 0.02 and minor allele count ≥ 100; $p < 5 \times 10^{-8}$ was significant.

Results:

The cohort included 25,130 individuals; 8,888 (35.4%) experienced ≥ 1 ADE. The highest rates of ADEs were for penicillin (14.1%) and CYP2D6-mediated opioids (10.7%). The frequencies were consistent with the previous limited research in African ancestry populations but differed from reported rates for European descent populations. We identified 7 novel significant loci for ADEs of penicillin, CYP2D6-mediated opioids, morphine, and sulfonamides. A variant (rs73407039) in MMP26 was associated with morphine ADEs (OR=1.93, $p=1.9 \times 10^{-10}$); however, it is in high linkage disequilibrium with variants in the nearby HBB region. Adjustment for HBB rs334 homozygous status attenuated this association ($p=0.025$), suggesting that the MMP26 signal was driven by patients with sickle cell anemia.

Conclusion:

We detected novel genetic risk loci in a less studied population. These findings highlight the value of high-throughput methods for ADE identification and the importance of conducting pharmacogenomic analyses across groups for ancestry-informed approaches to ADE prediction and prevention.

Genetic control of gene expression in whole blood of childhood cancer survivors

Subsession Time: Saturday, October 18 at 9:00am – 9:15am

Authors: Jennifer N French (St. Jude Children's Research Hospital), Kateryna Petrykey (St. Jude Children's Research Hospital), Achal Neupane (St. Jude Children's Research Hospital), Guang Yang (St. Jude Children's Research Hospital), John Easton (St. Jude Children's Research Hospital), Heather Mulder (St. Jude Children's Research Hospital), Evadnie Rampersaud (St. Jude Children's Research Hospital), Gang Wu (St. Jude Children's Research Hospital), Kirsten K Ness (St. Jude Children's Research Hospital), Melissa M Hudson (St. Jude Children's Research Hospital), Gregory T Armstrong (St. Jude Children's Research Hospital), Yutaka Yasui (St. Jude Children's Research Hospital), Yadav Sapkota (St. Jude Children's Research Hospital)

Abstract: Background: While GWAS have shown that childhood cancer survivors are at increased risk of chronic health conditions (CHCs) due to cancer treatment exposures, identifying causative variants can be challenging. Investigating how genetic variants regulate expression of nearby genes can prioritize GWAS findings and highlight mechanisms for increased risk of CHC in survivors. Expression quantitative trait loci (eQTLs) have been identified in the general population but have not been examined in childhood cancer survivors.

Methods: We conducted cis-eQTL analyses of common genetic variants (MAF>5%) within 1Mb of each gene, utilizing whole-genome and RNA-sequencing of whole blood samples from 5-year childhood cancer survivors of European (EUR; N=208) and African (AFR; N=102) ancestry participating in the St. Jude Lifetime Cohort Study. Ancestry-specific analysis was performed using linear regression, adjusting for sex, age at primary cancer diagnosis and sample collection, exposure to radiation and chemotherapy, the first 5 genotype-based principal components and the first 5 expression-based PEER factors. eQTLs with FDR <0.05 were considered statistically significant. Significant eQTLs were compared to those in whole blood from general population samples in GTEx v10.

Results: We identified 376,347 eQTLs for 6,442 eGenes (genes with ≥ 1 eQTL) among EUR survivors, of which 205,836 (54.7%) eQTLs and 3,932 (61.0%) eGenes overlapped with GTEx. In AFR survivors, we found 49,267 eQTLs and 2,131 eGenes, with 19,381 eQTLs and 1,077 eGenes overlapped with EUR. EUR survivor-specific eGenes not present in GTEx were significantly enriched for 9 Gene Ontology (GO) terms related to metabolic regulation and transcription. AFR survivor-specific eGenes not present in EUR were enriched for 6 GO terms, including regulation of mRNA and RNA stability. In EUR, 7 eQTLs significantly interacted with exposures to both platinum and alkylating agent. These treatment-specific eQTLs regulated expression of 4 genes: PTEN (tumor suppressor and lipid phosphatase), UBR7 (ubiquitin ligase), TAF1C (crucial for RNA polymerase), and SET (inhibits histone acetylation).

Conclusion: To our knowledge, this is the first eQTL analysis conducted among childhood cancer survivors, revealing both shared and unique genetic regulation of gene expression in whole blood samples compared to the general population. These eQTLs will be publicly available as a resource to enhance genetic studies of CHCs in survivors.

Genomic and transcriptomic insights into antipsychotic-induced changes in total cholesterol in a multi-ancestry cohort of 58,000 veterans

Subsession Time: Saturday, October 18 at 9:15am – 9:30am

Authors: Hamed Kazemi (University of Arizona), John Drake (University of Arizona), Silviu Bacanu (Virginia Commonwealth University), Julie Kreyenbuhl (Baltimore VA Medical Center), Stephen Marder (University of California Los Angeles), Philip Harvey (University of Miami), Khushbu Agrawal (Pacific Northwest National Laboratory), Alejandro Zuniga (Pacific Northwest National Laboratory), Sutanay Choudhury (Pacific Northwest National Laboratory), Hamed Abbaszadegan (Phoenix VA Medical Center), Benjamin McMahon (Los Alamos National Laboratory), Vladimir Vladimirov (University of Arizona), Ayman Fanous (University of Arizona, Phoenix VA Medical Center)

Abstract: Background: Antipsychotics are a cornerstone of the pharmacotherapy of schizophrenia and are commonly used in many psychiatric illnesses. However, they are known to induce metabolic adverse effects, including hypercholesterolemia, which increases the risk of cardiovascular disease. To date, there have been no adequately powered GWASs or transcriptomic studies of this phenotype.

Methods: We conducted a GWAS to explore the genetic architecture underlying changes in total cholesterol (TC) following antipsychotic treatment. Our cohort included 58,689 patients of European (EUR), African-American (AFR), and Hispanic (HIS) ancestry. Total cholesterol change (Δ TC) was defined as the difference between the highest TC levels before and after the initiation of antipsychotic treatment. Sex, age, and first 20 principal components (PCs) were included as covariates to control for population stratification. Transcriptome-wide association studies (TWAS) were performed to prioritize gene expression-trait associations.

Results: We identified genome-wide significant loci across ancestral groups including genes implicated in TC changes, lipid metabolism, cholesterol biosynthesis, and atherosclerosis. In the AFR cohort, key associations were observed on chromosomes 6 (EDN1, PHACTR1), 1 (CCN1), 2 (CYP27A1, PRKAG3, WNT6), 22 (PNPLA3), 9 (BAAT, PTPRD), 15 (KLF13, CRT3), 8 (TPD52, KIF13B), 5 (NDST1), and 10 (PPP2R2D). In the EUR cohort, significant signals were detected on chromosomes 8 (RNF139, TPD52), 7 (RELN), 4 (UCHL1), 6 (MIR206, IL17A, IL17F), 2 (CLASP1, MARCO, DBI), 9 (VPS13B), and 10 (SFTPD). Among HIS individuals, only two genome-wide significant variants were found on chromosome 4 (CXCL10, CXCL11) and 2 (NLRC4). Notably, TPD52 showed association with TC in both AFR and EUR groups. TWAS results further supported the involvement of several genes in cholesterol and lipid-related biological processes. In HIS, the expression of FABP5 (chr 8) was significantly associated with Δ TC in whole blood and visceral adipose tissue ($p = 1.5 \times 10^{-5}$); and in AFR, SLC44A4 (chr 6) in liver and subcutaneous adipose tissue ($p = 5 \times 10^{-6}$).

Conclusion: This study highlights the genetic contributions to changes in TC levels in response to antipsychotic treatment across diverse ancestries. Our findings improve the understanding of lipid dysregulation in patients with psychiatric disorders and provide a foundation for further research into personalized approaches to managing metabolic side effects.

Session 87: Prenatal Genetics and Reproductive Health

Location: Room 206AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Saturday, October 18 at 8:30am – 9:30am

Anticipating Futures: Stakeholder Perspectives on Ethical Governance of Prenatal Gene Editing

Subsession Time: Saturday, October 18 at 8:30am – 8:45am

Authors: Rosario Isasi¹, Megan A. Allyse³, Yvonne Bombard⁴, Roel Feys¹, Heidi C. Howard⁴, McKenna Horstmann³, Assata S. Kokayi², Marsha Michie³, Kiran Musunuru⁵, Preya Agam², Kirsten A. Riggan⁶, Sabina Rubeck³, Karen M. Meagher², Jane Q. Yap², Kelly E. Ormond⁶,¹University of Miami Miller School of Medicine, Miami, FL, USA, ²Mayo Clinic, Rochester, MN, USA, ³Case Western Reserve University, Cleveland, OH, USA, ⁴University of Toronto, Toronto, ON, Canada, ⁵University of Pennsylvania, Philadelphia, PA, USA, ⁶ETH Zurich, Zurich, Switzerland

Abstract: As prenatal gene editing (PGE) approaches clinical application, questions of ethical legitimacy, social acceptability, and regulatory readiness become increasingly urgent. Equally pressing is the need to ensure that such advances are grounded in robust, ethically accountable science. The Patient-Supported Approaches to Gene Editing (PASAGE) study investigates how diverse stakeholders—including clinicians, scientists, ethicists, and policymakers—understand and shape the governance of emerging reproductive technologies.

Using a mixed-methods design, we conducted 51 semi-structured interviews with international professionals involved in policymaking domains relevant to gene editing and prenatal interventions. Participants reflected on regulatory feasibility, professional ethics, and responsible innovation, as well as the institutional and normative commitments underpinning governance structures. Recurring themes included the tension between individual ethical convictions and institutional norms, as well as concerns about the perceived trustworthiness and capacity of governance institutions. Interview data were triangulated with international policy documents to assess alignment and adequacy.

Findings reveal broad support for polycentric, value-based governance models that integrate legal, professional, and community-based oversight. Participants emphasized that governance must be multi-level, adaptive, and participatory—guided by principles of transparency, accountability, and ethical pluralism. Professionals viewed themselves not as passive stakeholders, but as ethical agents contributing domain-specific insight to responsible innovation. This presentation will present key findings from the stakeholder interviews conducted and introduce a valued-based governance framework for gene editing and PGE.

Utility of Genomic Autopsy for Infant Mortality

Subsession Time: Saturday, October 18 at 8:45am – 9:00am

Authors: Gwendolyn Strickland (Boston Children's Hospital), Monica H Wojcik (Boston Children's Hospital), Casie A Genetti (Boston Children's Hospital), Kathleen Garvey (Boston Children's Hospital), Pankaj B Agrawal (Boston Children's Hospital), Richard Goldstein (Boston Children's Hospital), Heidi L Rehm (Massachusetts General Hospital), Anne O'Donnell-Luria (Boston Children's Hospital), Alan H Beggs (Boston Children's Hospital), Ingrid A Holm (Boston Children's Hospital)

Abstract: Background: Postmortem genomic sequencing, or “genomic autopsy”, has high diagnostic yield for fetal loss or stillbirth. However, incorporation of a genomic autopsy into the investigation of infant deaths is infrequent, in part due to lack of awareness of the diagnostic potential and clinical impact. Consequently, the genomic landscape of infant mortality remains poorly understood, and many families are left without the answers a genomic autopsy could provide.

Objective: To use exome or genome sequencing (ES/GS) to identify Mendelian disorders leading to death in infancy in order to define diagnostic, clinical, and psychosocial utility.

Methods: For a cohort of infants who died prior to one year of age, we obtained detailed phenotype information from medical records, collected DNA samples from the infant, parents (if available), and affected siblings (if pertinent), performed ES/GS, and returned diagnostic results to the families. When available, autopsy findings were reviewed. Parents were surveyed upon enrollment and one year later to evaluate clinical and psychosocial utility. A multivariable logistic regression model was constructed on the outcome of diagnosed status.

Results: 101 infants were enrolled from 99 families, including 80 trios, 3 quads, 8 duos, and 8 proband-only cases. Most decedents were male (56, 55%), born preterm (< 37 weeks, 55, 54%) and had at least one major congenital anomaly (64, 63%). Seventy (69%) deaths were neonatal and 49 (49%) were within the first week of life. Sequenced families received ES (54, 60%) or GS (40, 45%); 3 families had long read GS, and 6 families had more than one of these. Five families had RNA sequencing. Traditional autopsy data were available for 42 sequenced infants (46%).

Overall, 28/89 sequenced families (32%) were considered diagnosed or likely diagnosed, with candidate causal variants in found in an additional 17 (19%). Age at death, sex, prematurity, presence of major congenital anomalies, lack of parental data, and lack of traditional autopsy data were not significantly associated with odds of diagnosis.

Preliminary parent-reported utility data at enrollment reflect high interest in a diagnosis: 25/31 (81%) reported that a diagnosis is very/extremely important, particularly for understanding the infant's health problems and pregnancy planning. On follow-up surveys post-ES/GS with 14 parents (4 diagnosed families, 10 undiagnosed), 6 (43%) met criteria for prolonged grief disorder, all parents whose infants remained undiagnosed.

Conclusion: Genomic sequencing had a high diagnostic yield in an infant mortality cohort, illuminating novel disease-gene relationships, with high parent-perceived importance.

Prospective comparative evaluation of optical genome mapping, mate-pair genome sequencing, and chromosomal microarray analysis in prenatal genetic diagnosis of fetuses with ultrasound anomalies

Subsession Time: Saturday, October 18 at 9:00am – 9:15am

Authors: Matthew Hoi Kin Chau (Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong Hub of Paediatric Excellence, The Chinese University of Hong Kong), Jicheng Qian (Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong Hub of Paediatric Excellence, The Chinese University of Hong Kong), Mingyang Yu (Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong Hub of Paediatric Excellence, The Chinese University of Hong Kong), Cheuk Yan Chung (Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong Hub of Paediatric Excellence, The Chinese University of Hong Kong), Ka Yin Yvonne Kwok (Prince of Wales Hospital, Hong Kong), Ye Cao (Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong Hub of Paediatric Excellence, The Chinese University of Hong Kong), Kwong Wai Choy (Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong Hub of Paediatric Excellence, The Chinese University of Hong Kong), Zirui Dong (Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong Hub of Paediatric Excellence, The Chinese University of Hong Kong)

Abstract: Introduction: Chromosomal microarray analysis (CMA) is widely used in prenatal genetic diagnosis for detecting clinically significant copy-number variants (CNVs). Recently, optical genome mapping (OGM) and mate-pair sequencing have emerged as promising technologies capable of detecting a broader spectrum of structural variants, including balanced and unbalanced rearrangements. OGM and mate-pair sequencing inform structural orientation of copy-number gains and losses. Comparison of these platforms is essential to inform clinical implementation and testing strategies.

Methods: This is a prospective evaluation of OGM, CMA, and mate-pair sequencing for molecular cytogenetic investigation of 223 pregnancies with fetal ultrasound anomalies. CMA was performed on a custom 60k array comparative genomic hybridization + single-nucleotide polymorphism platform. Mate-pair sequencing was performed by an in-house method; paired-end sequencing of large-insert size libraries to 4X sequencing depth. OGM was performed on the Bionano Saphyr platform with at least 80x

effective coverage. All platforms detect chromosomal imbalances. Mate-pair sequencing, and OGM will also detect balanced structural rearrangements.

Results: CMA yielded positive findings in 13 cases (5.6%), OGM in 13 cases (5.6%), and mate-pair sequencing in 15 cases (6.4%). Mate-pair sequencing detected all variants identified by CMA and OGM. Two cases with Alpha-thalassemia Hemoglobin Barts were detected by mate-pair sequencing, but not by CMA and OGM. OGM and mate-pair sequencing informed unbalanced translocation in one case. OGM and mate-pair sequencing enabled reclassification of 4 variants of uncertain significance (detected by CMA). The copy-number gains, with genes at breakpoint junctions, were determined as forward tandem duplications, which did not disrupt those genes.

Conclusion: OGM and mate-pair sequencing demonstrated broader detection capabilities for structural variants compared to CMA, assisting in the resolution of simple and complex rearrangements. Their ability to determine the structural orientation of copy-number variants enabled more accurate classification and interpretation. OGM and mate-pair sequencing may serve as complementary approaches to CMA in prenatal genetic diagnostic testing.

Lifecourse Genetic Risk of Cardiovascular Disease in Women with Reproductive Health Conditions

Subsession Time: Saturday, October 18 at 9:15am – 9:30am

Authors: Tess Cherlin (University of Pennsylvania), Stephanie Mohammed (University of Pennsylvania), Lindsay Guare (University of Pennsylvania), Lannawill Caruth (University of Pennsylvania), Aude Ikuzwe Sindikubwabo (University of Pennsylvania), Ananya Rajagopalan (University of Pennsylvania), Regeneron Genetics Center (Regeneron, Inc.), Penn Medicine Biobank (University of Pennsylvania), Shefali Setia-Verma (University of Pennsylvania)

Abstract: Research and medical care for cardiovascular disease (CVD) has traditionally been one-size-fits all, leading to sex-based disparities when it comes to CVD prevention, diagnosis, and treatment. Recent efforts are being made uncover the etiology of CVD in women and reproductive health conditions (RHC) have been shown to significantly increase the risk for CVD outcomes later in life suggesting a culmination of factors contribute to CVD. While comorbidities between these conditions can be shared, it is still not clear what is driving risk. Genetic correlation studies from the FinnGen Ritsey database show genetic links between many RHC and cardiovascular diseases. We hypothesize that genetic factors may be underlying the increased time to CVD outcome for women with RHCs. We leveraged three biobanks to address this hypothesis: Penn Medicine Biobank, eMERGE network, and the NIH All of Us biobank. All biobanks contain electronic health record (EHR) diagnoses and whole genome sequencing data for over 22,165 and 32,905, and 245,601 women assigned female at birth respectively. Phenotypes for composite CVD and women's health conditions (adverse pregnancy outcomes, endometriosis, gestational diabetes, gestational hypertension, PCOS, pre-eclampsia, uterine fibroids) were defined by ICD codes. We did time-to-event genome-wide association (TTE GWAS) analysis with SAIGE GATE to identify variants associated with disease trajectories. Two different endpoints were used: 1) birth --> RHC and 2) RHC --> first composite CVD event. The analysis was stratified by genetically inferred ancestry and summary

statistics were meta-analyzed across biobanks using METAL software. We ran up to six ancestry-stratified RHC TTE GWAS for timepoint 1 and up to four for ancestry-stratified timepoint 2 across the three biobank datasets. In the PMBB cohort we identified 116 significant ($p < 5e-08$) loci for timepoint 1 and 46 significant ($p < 5e-08$) loci for timepoint 2 across all RCHs. In eMERGE cohort we identified rs2084686 ($7.62e-08$) as the most significant loci in gestational diabetes time to CVD. The variant lies upstream of the gene MYT1L. The most significant association for PCOS time-to-CVD meta-analysis was rs58875218 ($1.18e-05$), is located downstream of KCNV2 a gene important for calcium homeostasis. These findings provide valuable insights into the genetic basis of composite CVD disease risk for women with reproductive health conditions. Future directions include replication and meta-analyses with other large datasets like the GBMI consortium. Given the high prevalence of CVD in women, longitudinal studies are needed to uncover the molecular, genetic, and pharmacogenomic pathways underlying disease progression in women.

Session 88: Using Disease Models to Assess Variant Affect

Location: Room 210C/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Saturday, October 18 at 8:30am – 9:30am

Exploring the Utility of A Knockout Mouse Resource for Insights into Human Disease

Subsession Time: Saturday, October 18 at 8:30am – 8:45am

Authors: Sofia Martin (NHGRI Division of Genome Sciences), Pilar Cacheiro (Queen Mary University of London) Damian Smedley (Queen Mary University of London) Colin Fletcher (NHGRI Division of Genome Sciences), Oleg Mirochnitchenko (NIH Office of the Director)

Abstract: The Knockout Mouse Phenotyping Program (KOMP2) is a trans-NIH effort across 18 institutes and centers to knockout and characterize all protein-coding genes in the mouse genome. Because humans and mice share many orthologous genes, the mouse serves as a key model organism in elucidating the roles they play in health and disease. Knocking out the activity of a gene via targeted mutagenesis allows researchers to understand its function and phenotypic information across biological systems and processes. In a global collaborative effort with the International Mouse Phenotyping Consortium (IMPC), KOMP2 has created a comprehensive and public null mutant mouse resource. 21 KOMP2 and IMPC centers generate their own data with standardized phenotyping protocols, made available through a multitude of free, open-access resources. Mouse lines and embryonic stem (ES) cells are also available to the larger scientific community to promote further research efforts involving the knockout mouse.

With 9,073 phenotyped genes, 111,485 phenotype calls reported, and 6,425 associated publications, enhancing the utility of this comprehensive resource for human disease research will remain significant as it nears completion. KOMP2 approaches this challenge in multiple ways, including the implementation of an automated pipeline to connect phenotyping parameters to human diseases, enhancing the identification of candidate genes, and creating strains with coding and non-coding variants implicated in human disorders. Through these current endeavors and future opportunities, the knockout mouse resource will advance disease discovery and yield valuable insights for improving human health outcomes.

Digenic variants that affect an actin-mitochondria-glutamate pathway cause epilepsy

Subsession Time: Saturday, October 18 at 8:45am – 9:00am

Authors: Shenzhao Lu (Baylor College of Medicine, Department of Molecular and Human Genetics), Mengqi Ma (Baylor College of Medicine, Department of Molecular and Human Genetics), Shabab

Hannan (Baylor College of Medicine, Department of Neurology), Lindsey D. Goodman (Baylor College of Medicine, Department of Molecular and Human Genetics), Mingxi Deng (Baylor College of Medicine, Department of Molecular and Human Genetics), Hu Chen (Baylor College of Medicine, Department of Pediatrics), Sandeep Kumar Dubey (Baylor College of Medicine, Department of Neurology), Wen-Wen Lin (Baylor College of Medicine, Department of Molecular and Human Genetics), Xueyang Pan (Baylor College of Medicine, Department of Molecular and Human Genetics), Zhijian Yu (Baylor College of Medicine, Department of Pediatrics), Debdeep Dutta (Baylor College of Medicine, Department of Molecular and Human Genetics), Jill A. Rosenfeld (Baylor College of Medicine, Department of Molecular and Human Genetics, Baylor Genetics), Xi Luo (Baylor College of Medicine, Department of Molecular and Human Genetics, Baylor Genetics), Zhandong Liu (Baylor College of Medicine, Department of Pediatrics), Joshua M. Shulman (Baylor College of Medicine, Department of Neurology), and Hugo J. Bellen (Baylor College of Medicine, Department of Molecular and Human Genetics)

Abstract: Epilepsy is one of the most prevalent neurological disorders with a complex etiology. Genetic studies have identified over 1,000 epilepsy-associated genes. However, more than 50% of individuals with epilepsy who undergo genetic sequencing fail to receive a genetic diagnosis. One of the main factors limiting our understanding of the molecular mechanisms underlying epilepsy is the involvement of multiple genetic risk factors. Although ion channels and synaptic proteins that directly regulate neuronal activity have been shown to be at the root of epilepsy, the underlying mechanism remains undefined for most of the remaining epilepsy-associated genes, including 27 genes involved in the regulation of actin dynamics. Here, by studying epilepsy-associated genes that regulate actin dynamics, we identified an actin-mitochondria-glutamate pathway associated with epilepsy. We demonstrate that reduced actin polymerization levels promote Dynamin-Related Protein 1 (DRP1)-mediated mitochondrial fission, which elevates reactive oxygen species (ROS) levels, leading to increased glutamatergic transmission and seizures. Reducing mitochondrial fission with the DRP1 inhibitor Mdivi-1, or suppressing ROS with N-acetyl-L-cysteine amide, significantly alleviates the seizure phenotypes. Interestingly, we observe that *Drosophila* trans-heterozygous for loss-of-function (LoF) in genes impacting this pathway exhibit obvious seizure-like phenotypes. Together, these findings suggest that seizures may arise from a combination of two genetic variants affecting this pathway. Furthermore, we assembled an epilepsy cohort of 200 individuals from Baylor Genetics who were not diagnosed based on sequencing. We analyzed their sequencing data by ranking the top risk variants using AI-MARRVEL, an AI-powered tool to prioritize possible disease-causing variants. This led to the identification of 17 individuals with epilepsy who carry two heterozygous LoF variants in genes known to regulate either actin polymerization or mitochondrial ROS. We evaluated the effect of the genetic variants in eight individuals by generating trans-heterozygous mutant flies and found that five exhibited synergistic or additive effects leading to seizures. Overall, this novel “actin-mitochondria-glutamate pathway” may advance the identification of multigenic etiologies in individuals with seizures and highlight novel therapeutic targets.

Investigating cilia-autophagy crosstalk in skeletal development and FGFR3-mediated pathogenesis of achondroplasia

Subsession Time: Saturday, October 18 at 9:00am – 9:15am

Authors: Presenting Author - Neha Quadri (Ph.D. Scholar, Department of Medical Genetics, Kasturba Medical College, Manipal)

Co-Authors - Hitesh Shah (Professor & HOD, Department of Orthopaedics, Kasturba Medical College, Manipal), Priyanka Upadhyai* (Associate Professor, Department of Medical Genetics, Kasturba Medical College, Manipal)

Abstract: Background: Primary cilia (PC) are conserved, immotile sensory organelles that transduce a variety of signals and are essential for both embryonic and postnatal development. They interact with cellular processes such as autophagy, a mechanism for the degradation and recycling of cellular proteins and organelles, in a cell type-specific manner. Although emerging studies suggest a possible crosstalk between PC and autophagy, the nature of their interaction during chondrocyte differentiation remains unclear. A few existing studies have reported abnormalities in autophagy and PC independently in chondrodysplasias, such as achondroplasia (ACH).

Keywords: primary cilia, autophagy, chondrocyte differentiation, FGFR3, achondroplasia, ERK1/2

Research Questions:

We aimed to examine the nature and molecular mechanisms underlying the crosstalk between PC and autophagy during chondrocyte differentiation using ATDC5 prechondrocytes.

We also investigated how FGFR3 overactivation disrupts the PC–autophagy axis in ACH using patient-derived chondrocytes, revealing key molecular defects that contribute to disease pathogenesis

Methodology: Primary cultures of chondrocytes were established from a pediatric ACH subject. Autophagy was chemically modulated, and primary cilium (PC) ablation was achieved by lentiviral knockdown (KD) of Ift52, an essential subunit of the intraflagellar transport machinery required for ciliogenesis. Ciliary structure and autophagic activity were assessed by immunofluorescence, while gene expression and protein levels were evaluated using RT-qPCR and immunoblotting

Results: A mutual positive regulatory role was revealed between PC and autophagy in chondrocyte differentiation

ATDC5 cell line : Ift52 KD, a critical component of the intraflagellar transport machinery, suppressed starvation-induced autophagy flux. Pharmacological inhibition of autophagy using chloroquine impaired both basal and induced ciliogenesis, whereas activation of autophagy with trehalose enhanced ciliation. Disruption of either autophagy or ciliary function resulted in diminished expression of chondrogenic markers, highlighting their cooperative role in promoting chondrogenesis

ACH patient-derived chondrocytes : We observed severely impaired autophagy, defective ciliogenesis, and reduced chondrocyte differentiation capacity. Augmenting autophagy rescued ciliation in ACH chondrocytes. Mechanistically FGFR3 overactivation in ACH cells led to hyperphosphorylation of ERK1/2, but not STAT1, implicating ERK1/2 in the disrupted crosstalk. Ongoing studies aim to further delineate the FGFR3–ERK1/2–autophagy–ciliogenesis signaling axis and its contribution to the pathogenesis of ACH

Conclusion:

This study sheds light on the cilia–autophagy axis in chondrocyte differentiation and how it is disrupted in ACH. It advances our current knowledge of bone development, ACH pathomechanisms and highlights potential avenues for therapeutic interventions. These insights will have broader implications for the treatment of other skeletal ciliopathies.

ZIC1 is a transcription factor that intersects Idiopathic Hypogonadotropic Hypogonadism and Craniosynostosis

Subsession Time: Saturday, October 18 at 9:15am – 9:30am

Authors: Hannah McGinty (Stanley Manne Children's Research Institute, Ann & Robert H. Lurie Children's Hospital of Chicago), Sheraz Khan (Stanley Manne Children's Research Institute, Ann & Robert H. Lurie Children's Hospital of Chicago), Elizabeth LeClair (Stanley Manne Children's Research Institute, Ann & Robert H. Lurie Children's Hospital of Chicago), Ravikumar Balasubramanian (Harvard Reproductive Endocrine Sciences Center and the Reproductive Endocrine Unit of the Department of Medicine, Massachusetts General Hospital), Erica E. Davis (Stanley Manne Children's Research Institute, Ann & Robert H. Lurie Children's Hospital of Chicago, Department of Pediatrics and Cell and Developmental Biology, Feinberg School of Medicine, Northwestern University)

Abstract: Background: Pleiotropy is a phenomenon in which one gene affects multiple distinct phenotypic features. Our ongoing work has highlighted how a subset of causal disease genes overlaps two unique phenotypes: Idiopathic Hypogonadotropic Hypogonadism (IHH) and Craniosynostosis (CS). IHH is a rare reproductive disorder caused by secretory deficits of gonadotropin releasing hormone (GnRH) neurons and/or defective embryonic migration. IHH is caused by pathogenic variants in >70 genes and individuals present with sexual infantilism, absent puberty, and infertility. CS, caused by pathogenic variants in >120 genes, impacts craniofacial development and causes one or more sutures of the skull to fuse prematurely. Published data on CS and our unpublished IHH cohort data suggest that ZIC1 is an overlapping gene that causes both disorders. ZIC1 is a zinc finger protein transcription factor that impacts neural development.

Hypothesis: ZIC1 is required for establishment of the GnRH axis, and heterozygous loss of function variants cause IHH.

Methods: We analyzed exome sequencing (ES) data from 1,489 IHH probands and performed cohort-based collapsed gene burden testing for non-benign rare variants. We then established genetically stable mutant and transient knockdown zebrafish models. We use in vivo complementation in zebrafish to test pathogenicity of human ZIC1 missense variants. For all zebrafish assays, we perform molecular characterization and genotype/phenotype correlation on wild type or tg(gnrh3:gfp) reporter lines.

Results: Fourteen CS genes were detected with nominally significant rare variant enrichment in the IHH cohort, including ZIC1. We identified four unrelated families with rare ZIC1 variants that segregate with

IHH but without overt CS. We isolated two independent *zic1* frameshift mutant zebrafish lines: a 5 base-pair (bp) insertion and a 5 bp deletion targeting exon 1. We observe a dose-dependent significant increase in *zic1* transcript in mutant larvae of each line. Brightfield imaging of 5 bp deletion mutants and morpholino knockdown larvae showed significantly smaller head size versus wild type siblings.

Conclusion: Our genetics data suggest that heterozygous rare variants in *ZIC1* cause IHH. Our work will involve assessment of *zic1* zebrafish models to measure differences in *gnrh3*-GFP+ area and rescue assays will test pathogenicity of IHH-associated missense variants. *zic1* zebrafish models will be used to better understand pleiotropy in IHH and CS.

Session 89: Biobank-Scale Genomic Data Resources of Diverse Populations

Location: Room 210AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Saturday, October 18 at 10:00am – 11:00am

Large-scale metabolomic profiling enables genome-wide and phenome-wide association studies in the Million Veteran Program

Subsession Time: Saturday, October 18 at 10:00am – 10:15am

Authors: Christian Gauthier (Center for Data and Computational Sciences, Booz Allen Hamilton), Michael Francis (CDACS, Booz Allen Hamilton), Nalvi Duro (CDACS, Booz Allen Hamilton), Rebecca Chance (CDACS), Million Veteran Program (MVP), Patrick Schreiner (CDACS, Booz Allen Hamilton), Bryan R. Gorman (CDACS, Booz Allen Hamilton), Saiju Pyarajan (CDACS)

Abstract: Background: Modern advances in high-throughput metabolomics platforms have enabled the efficient and comprehensive profiling of small-molecule metabolites in large-scale human cohorts. However, integration of this data with genomic and medical health records, particularly in underrepresented ancestries, remains limited. We present a new metabolomics profiling dataset ($n = 57,325$) generated from blood plasma samples of Million Veteran Program (MVP) participants, processed by the Metabolon Discovery HD4 platform, to characterize metabolite abundance and evaluate genetic and phenotypic associations.

Methods: Peak data from 1995 metabolites across 57,325 MVP plasma samples was measured using Metabolon's Discovery platform. The metabolite intensities were normalized and transformed using rank-based inverse normal transformation (RINT). Differential abundance analyses were conducted to validate the data against published results. We performed association studies with phenotypic and genotypic data, stratifying on European ($n = 41,492$), admixed African ($n = 11,081$), and Hispanic ($n = 3,842$) ancestries. Cotinine, docosahexaenoic acid (DHA), and (non-fasting) glucose were selected as exemplar biomarkers to evaluate against known associations.

Results: Differential analysis identified biomarker signatures that coincided with published results (e.g. elevated levels of sulfate compounds and steroids) with the most significant values present in nicotine-metabolizing compounds such as cotinine, hydroxycotinine, and o-cresol sulfate. PheWAS results showed the strongest associations of cotinine abundance with tobacco use disorders, DHA abundance with mixed hyperlipidemia and hyperglyceridemia, and elevated levels of glucose were closely associated with type 1 and 2 diabetes. GWAS replicated known loci ($P < 5 \times 10^{-8}$) for cotinine at CYP2A6, ADH1C, and PPP1R3B, for DHA at FADS1/FADS2, CDKR, TRIB1, and ALDH1A2, and for glucose at CDKN2B, TCF7L2, and CCDN2. We also identified a putatively novel glucose locus at SLC22A11, whose gene product is a urate re-uptake transporter that may have downstream consequences on insulin resistance.

Conclusion: We present one of the broadest and most diverse single-cohort metabolomic profiling datasets to-date. Our association analyses highlight the power of integrating high-throughput metabolomic profiling with the rich genetic and phenotypic resources of the MVP to advance our understanding of human metabolism and uncover novel associations.

Our Future Health: The World's Largest Cohort Study is Open for Genetic Research

Subsession Time: Saturday, October 18 at 10:15am – 10:30am

Authors: Michael B. Cook (Our Future Health, Imperial College London, University of Oxford), Ashley Adjetey (Our Future Health), Ryan Arathimos (Our Future Health), Marko Balabanovic (Our Future Health), Benjamin J. Cairns (Our Future Health), Ben Elsworth (Our Future Health), Kate Evans (Our Future Health), Eva Gradovich (Our Future Health), Cosima Gretton (Our Future Health), David J. Hunter (Our Future Health, University of Oxford), Kamil Lipinski (Our Future Health), Fiona Maleady-Crowe (Our Future Health), Fiona Reddington (Our Future Health), Tim Sprosen (Our Future Health), Raghbi Ali (Our Future Health, University of Cambridge)

Abstract: Our Future Health is a new UK prospective cohort study with an ambition to recruit 5 million adult participants. Our Mission is to help everyone live longer, healthier lives through the discovery and testing of more effective approaches to prevention, earlier detection and treatment of diseases. We opened recruitment in October 2022 and, by May 2025, 2.4 million participants had consented, 1.8 million had completed the baseline questionnaire, and 1.4 million had completed an in-person appointment to donate health metrics and blood and have been linked to electronic health records. Genotyping is conducted using a 700,000-variant Illumina custom array that is optimised for multi-ethnic imputation and has up-to-date disease-, phenotype-, blood type-, and pharmacogenomic-associated variants.

With over 1 million participants already genotyped, Our Future Health is providing the world's largest and most diverse genetic cohort that will catalyse discovery and translation research. A forthcoming imputed dataset will expand the possibilities even further. The large scale of this study is already facilitating GWAS, PRS and other study types on many common and rare conditions, with strong statistical power for complex interactions, causal factors, algorithmic prediction, and AI/ML methods using exposures that have the requisite recency and relevance for the disease etiology of today.

Many promising diagnostic and prognostic biomarkers fail at the stage of prospective evaluation. Our Future Health will solve this by being a translational catalyst by facilitating access to selected stored biosamples. Our Future Health will also accelerate recruitment to external studies and trials with participants selectively invited based on demographics, phenotypes, and disease risks.

The cohort includes over 600,000 18–40-year-olds, facilitating research into diseases that disproportionately affect younger people including obesity, mental health disorders and addiction, and 900,000 over 60s, enabling research into vascular dementia, Alzheimers and aging.

Our Future Health is available for you to use today. To learn more, please visit <https://research.ourfuturehealth.org.uk>

The All of Us Research Program data release 2026 (CDR v9): Powering genomic research through All of Us

Subsession Time: Saturday, October 18 at 10:30am – 10:45am

Authors: Tara Dutka (All of Us Research Program, National Institutes of Health), Anjene. Musick, (All of Us Research Program, National Institutes of Health), Melissa Basford (Vanderbilt University Medical Center), Eric Boerwinkle (University of Texas Health Sciences Center at Houston), Mine Cicek (Mayo Clinic, Rochester), Kim Doheny (Center for Inherited Disease Research, Department of Genetic Medicine, Johns Hopkins University School of Medicine), Evan Eichler (University of Washington), Stacey Gabriel (Broad Institute), Kiran Garimella (Broad Institute), Richard Gibbs (Baylor College of Medicine), Paul Harris (Vanderbilt University Medical Center), David Glazer (Verily), Gail Jarvik (University of Washington), Chris Kachulis (Broad Institute), Konrad Karczewski (Broad Institute), Nial Lennon (Broad Institute), Stephan Montgomery (Stanford University), Donna Muzny (Baylor College of Medicine), Heidi Rehm (Center for Genomic Medicine, Massachusetts General Hospital), Dan Roden (Vanderbilt University Medical Center), Michael Schatz (Johns Hopkins University), Fritz Sedlazeck (Baylor College of Medicine), Michael Talkowski (Broad Institute), Winston Timp (Johns Hopkins University), Scott Topper (Color Health), Lee Lichtenstein (Broad Institute), The All of Us Research Program Genomics Investigators, All of Us Research Program Long Read Working Group, All of Us Research Program Multi-omics Working Group

Abstract: The NIH's All of Us Research Program aims to enroll ~1 million participants who reflect the broad spectrum of backgrounds in the United States. The program is well-positioned to drive precision medicine by combining data from surveys, electronic health records, physical measurements, genomics, public data linkages and digital health technologies. Here we describe the program's next data release, currently in QC and planned for release in early 2026, which increases the size and comprehensiveness of the -omics data available from participants. There are >500k arrays and short-read whole genome sequences (srWGS) and >13k long-read WGS (lrWGS). Proteomes (Olink Explore HT) and transcriptomes (whole blood - Watchmaker protocol) will be available for ~10k participants who have lrWGS, as well as pQTLs, eQTLs, and sQTLs. Genomic data were generated at one of four All of Us Genome Centers (GC) using equivalent library construction protocols, sequencers, software, and software configuration. The Broad GC performed additional QC, structural variant (SV) discovery, joint calling, and data curation. The release will provide data for common analysis use cases, including pharmacogenomics, inferred continental ancestries, genetic admixture estimates, local-ancestry, phasing, HLA haplotype calls, results of principal components analysis, variant annotations, relatedness and maximal set of unrelated individuals, and multiple variant file formats (e.g., plinkbed, bgen, pgen, Hail, VDS), and commonly used subsets of data (e.g., common variants, exomes, ClinVar variants, and challenging medically relevant genes). The lrWGS data were generated at 8x-30x using both Oxford Nanopore (ONT) and PacBio (PB) across five genome centers. This includes >600 paired lrWGS (same participant on both ONT & PB) with genome variant discovery with alignment-based analyses against two references, GRCh38 and T2T-CHM13, as well as de novo assembly-based. There will be methylation profiles from most lrWGS.

Summary-level data will be publicly available on the All of Us data browser, including aggregated genetic variant counts. Additional genomic tools released in 2025 include an “All by All” browser mapping associations between genotypes and phenotypes and an imputation service based jointly from All of Us and AnVIL data. The All of Us Research Program is enabling novel health research and medical breakthroughs due to the unprecedented scale and breadth of participants and data sources.

Uncovering genetic etiologies in 30,000 rare disease genomes from a nationwide Chinese cohort

Subsession Time: Saturday, October 18 at 10:45am – 11:00am

Authors: Fengxiao Bu (West China Hospital of Sichuan University, Institute of Rare Diseases), Chao Wang (West China Hospital of Sichuan University, Institute of Rare Diseases), Qian Zhang (West China Hospital of Sichuan University, Institute of Rare Diseases), Ting Tang (West China Hospital of Sichuan University, Institute of Rare Diseases), Jia Geng (West China Hospital of Sichuan University, Institute of Rare Diseases), Fandi Ai (West China Hospital of Sichuan University, Institute of Rare Diseases), Yang Wu (West China Hospital of Sichuan University, Institute of Rare Diseases), Guanglin He (West China Hospital of Sichuan University, Institute of Rare Diseases), Yu Lu (West China Hospital of Sichuan University, Institute of Rare Diseases), Jing Cheng (West China Hospital of Sichuan University, Institute of Rare Diseases), and Huijun Yuan (West China Hospital of Sichuan University, Institute of Rare Diseases)

Abstract: While whole genome sequencing (WGS) has significantly advanced the genetic diagnosis of rare diseases, large-scale nationwide cohorts utilizing WGS (e.g. the UK 100,000 Genomes Project) remain limited, particularly in non-European populations. To address this gap, the Genome Sequencing of Rare Diseases (GSRD) project was initiated in July 2020, with the aim of enrolling 100,000 individuals affected by rare diseases across China. Through population-scale genome sequencing, the GSRD project seeks to elucidate the molecular underpinnings and pathogenic mechanisms of rare diseases, thereby advancing precision diagnostics.

As of May 2025, WGS has been completed for 49,138 individuals (30,194 cases and 18,944 controls), with an average coverage depth of 46x. The cohort spans 339 rare disease entities across 15 clinical specialties. It exhibits broad national representation, with participants’ birthplaces covering 315 of China’s 333 prefecture regions and including individuals from 49 of 55 officially recognized ethnic minority groups. Population stratification analysis revealed nine distinct genetic clusters, highlighting the extensive genetic diversity captured within the cohort.

Comprehensive genome-wide variation detection was performed. SNV and indel calling were accelerated using a GPU-optimized pipeline, enabling completion of secondary analysis in approximately 20 minutes per sample. To support population-specific genomic analysis, a haplotype reference panel was constructed. Compared with existing panels, this new reference achieved a 72.56% increase in the number of imputed variants, with an imputation precision of 0.983 and sensitivity of 0.980. Substantial gains in genotype concordance, particularly for rare variants, were observed across diverse Chinese subpopulations.

To date, genetic diagnoses have been established for 4,295 rare disease cases, yielding a diagnostic rate of 27.33% (n = 1,174). Importantly, WGS contributed an incremental diagnostic yield of 13.5% (n = 159) through the detection of complex variants—including SVs, STRs, deep intronic variants, and MEIs. For example, a pathogenic noncoding SV disrupting enhancers located 10 kb upstream of PAX3 was identified as the cause of Waardenburg syndrome.

These findings underscore the transformative potential of nationwide WGS initiatives in improving diagnostic outcomes, characterizing population-specific variation, and enabling equitable genomic medicine in underrepresented populations.

Session 90: Cardiometabolic Traits: Insights from Multi-Omics Approaches

Location: Room 210C/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Saturday, October 18 at 10:00am – 11:00am

Large scale metabolomics study in 69,671 South Asians identifies novel rare variant associations with lipids

Subsession Time: Saturday, October 18 at 10:00am – 10:15am

Authors: Moeen Riaz (Regeneron Genetics Center, Tarrytown, NY USA), Joshua Backman (Regeneron Genetics Center, Tarrytown, NY USA), George Hindi (Regeneron Genetics Center, Tarrytown, NY USA), Olukayode A. Sosina (Regeneron Genetics Center, Tarrytown, NY USA), Sameer Malhotra (Regeneron Genetics Center, Tarrytown, NY USA), Regeneron Genetics Center, the BELIEVE study group, Sohel Choudhury (National Heart Foundation Hospital & Research Institute, Dhaka, Bangladesh), Fazal Khan (iccdr,b, Dhaka, Bangladesh), Md Khalequzzama (Bangladesh Medical University (BMU), Dhaka, Bangladesh), Wing Tung Fung (University of Cambridge, Cambridge, UK), Scott Ritchie (University of Cambridge, Cambridge, UK), Adam S. Butterworth (University of Cambridge, Cambridge, UK), John Danesh (University of Cambridge, Cambridge, UK), Emanuele Di Angelantonio (University of Cambridge, Cambridge, UK), Jonathan Marchini (Regeneron Genetics Center, Tarrytown, NY USA), Aris Baras (Regeneron Genetics Center, Tarrytown, NY USA), Gonçalo R. Abecasis (Regeneron Genetics Center, Tarrytown, NY USA), Luca A. Lotta (Regeneron Genetics Center, Tarrytown, NY USA), Alan R. Shuldiner (Regeneron Genetics Center, Tarrytown, NY USA)

Abstract: Blood biomarkers are essential for understanding metabolic and physiological processes related to disease risk assessment, treatment, and prevention. Genome-wide association studies (GWAS) coupled with metabolic profiling have identified numerous loci associated with circulating metabolic traits; however, these studies have predominantly focused on European-ancestry populations. Our study represents the first large-scale genetic association study of metabolic biomarkers in South Asians (SA). Utilizing Nuclear Magnetic Resonance technology on the Nightingale Health's metabolic biomarker profiling platform, we quantified and analyzed 326 blood biomarkers, including lipoproteins, lipids, fatty acids, amino acids, ketones, and glycolysis metabolites in 69,671 South Asians from the Bangladesh Longitudinal Investigation of Emerging Vascular Events (BELIEVE) study. Genome-wide genotype and exome sequencing with mean coverage of 48X was conducted using the Illumina NovaSeq 6000 platform and imputation was performed. Followed by, Exome-genome wide association analyses for single variants and gene aggregates under an additive model, adjusting for age, sex, and genetic principal components using REGENIE. We found South Asian enriched variants associated with lipids traits in known and novel genes including, Angiopoietin-like 3 (ANGPTL3), Apolipoprotein B (APOB), Low-Density Lipoprotein Receptor (LDLR), and Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9). The predicted loss of function (pLOF) burden mask of ANGPTL3 was associated with 171 lipids and lipoprotein traits

such as free cholesterol in chylomicrons and extremely large VLDL ($p=3.20e-42$, effect=-0.51 standard deviation (SD)). The APOB pLOF mask was associated with 240 biomarkers, with the most significant association being free cholesterol in very small VLDL ($p=1.90e-48$, effect=-1.4 SD). Two novel associations were found with rare missense variants, p.Asp367His (PCSK9) and p.His583Tyr (LDLR), which were associated with many lipid biomarkers such as APOB ($p=9.1e-49$, effect=0.49 SD) and higher total cholesterol ($p=2.0e-09$, effect=0.63 SD, respectively). The frequency of p.Asp367His is 0.004 in SA compared to $1.6e-6$ in Europeans. Variant p.His583Tyr is found in SA (MAF 0.0006) and East Asians (0.0004) only. Our study highlights the power of studying diverse ancestries at scale to identify ancestry-specific associations.

Connecting proteomics and genomics to identify sex-specific plasma biomarkers for calcific aortic valve stenosis

Subsession Time: Saturday, October 18 at 10:15am – 10:30am

Authors: Pardis Zamani (Institut universitaire de cardiologie et de pneumologie de Québec-Université Laval (IUCPQ-Ulaval)), Ursula Houessou (IUCPQ-Ulaval), Hasanga Manik Purage (IUCPQ-Ulaval), Manel Dahmene (IUCPQ-Ulaval), Zhonglin Li (IUCPQ-Ulaval), Aida Eslami (IUCPQ-Ulaval), Nathalie Gaudreault (IUCPQ-Ulaval), Patrick Mathieu (IUCPQ-Ulaval), Yohan Bossé (IUCPQ-Ulaval), Benoît Arsenault (IUCPQ-Ulaval), Sébastien Thériault (IUCPQ-Ulaval)

Abstract: Introduction: Calcific aortic valve stenosis (CAVS) is a progressive disease characterized by the calcification and narrowing of the aortic valve, leading to significant morbidity and mortality. Early detection remains a major clinical challenge as current diagnosis relies on late-stage echocardiographic changes. The identification of valve-specific biomarkers could improve risk stratification and guide the development of novel therapies.

Objective: This study aims to (1) identify plasma protein biomarkers for incident CAVS, (2) investigate sex-specific associations, and (3) establish causal relationships using genetic approaches to prioritize molecular targets.

Methods: We assessed the association between 2923 unique plasma proteins measured using the Olink Explore platform and CAVS incidence in 53,014 UK Biobank participants, including 487 incident CAVS cases over a median follow-up of 13 years. Multivariable Cox proportional hazards models adjusted for cardiovascular risk factors were used to evaluate associations. A stratified analysis was performed to investigate the association in men and women separately. For causal inference, we used protein quantitative trait loci (pQTL) to perform Mendelian Randomization (MR) analyses. We then verified the aortic valve specificity of the proteins using our transcriptomic dataset of 500 human aortic valves. Using expression quantitative trait loci (eQTL)-based MR, we evaluated the causal role of aortic valve gene expression.

Results: Our fully adjusted model identified 55 proteins significantly associated with CAVS incidence, with GDF15 showing the strongest association (HR=2.25; 95% CI=2.02-2.52, P value= 3.79×10^{-46}). In stratified analyses, the associations were different in each sex: 10 proteins showed significant association

with CAVS incidence in women, whereas 16 proteins were identified in men. pQTL-MR suggested a potential causal role for 7 proteins in CAVS incidence. Transcriptomic integration showed aortic valve-enriched expression for 10 candidate proteins, with LTBP2 and B2M demonstrating significant associations in aortic valve eQTL-MR analyses.

Conclusion: This study integrating proteomics, causal inference, and tissue-specific validation identified promising sex-specific biomarkers for CAVS and provides potential targets for prevention and treatment.

Proteomic discovery in an inception cohort of acute myocardial infarction survivors

Subsession Time: Saturday, October 18 at 10:30am – 10:45am

Authors: Colleen Sitlani (University of Washington), Jennifer Brody (University of Washington), Kerri Wiggins (University of Washington), Joshua Bis (University of Washington), Laura Harrington (Kaiser Permanente Washington Health Research Institute), Robert Gerszten (Beth Israel Deaconess Medical Center), Bruce Psaty (University of Washington), James Floyd (University of Washington)

Abstract: Background. While much is known about the etiology and incidence of coronary heart disease (CHD) from human genetics, far less is known about the recurrent disease in patients with established CHD. We used affinity-based proteomics profiling to discover new biomarkers and drug targets for high-risk CHD patients.

Methods. In the Heart and Vascular Health (HVH) Study, we used the Olink Explore platform to measure 2,941 plasma proteins in a cohort of 1,528 survivors of acute myocardial infarction (AMI). We used Cox proportional hazards models to estimate hazard ratios (HR) for the longitudinal protein associations with the risk of recurrent CHD (recurrent AMI or CHD death) during up to 20+ years of follow up. We replicated statistically significant results ($p\text{-value} < 0.05/2941$) in UK Biobank participants with prevalent MI at time of proteomics measurement ($n=1,285$), who were followed for recurrent hospitalized AMI ($n=346$). To evaluate whether protein associations with recurrent CHD risk have causal support, two-sample Mendelian randomization (MR) analyses were conducted using results from the largest available GWAS of CHD and Olink pQTLs. Proteins with FDR $q\text{-values} < 0.05$ were considered to have statistically significant evidence of a causal effect on CHD.

Results. Among 1,528 HVH participants (mean age 67), 626 experienced recurrent CHD. After adjustment for CHD risk factors, 127 proteins were significantly associated with recurrent CHD, of which 57 (45%) replicated in the UK Biobank. In MR analyses, 13/115 (11%) proteins had evidence of a causal role in the development of CHD, including eight that have not been identified in previous large-scale GWAS of CHD (CHCHD10, GGT5, RARRES2, REG1B, CSF1, HAVCR1, CX3CL1, EFNA1). Higher levels of all MR-significant proteins correlated with higher recurrent CHD risk. Ten of the MR-significant proteins had causal effects that were concordant with the observational association, including CHCHD10 (observational HR 1.26, MR HR 1.22). GGT5 (observational HR 1.27, MR HR 0.95) and two other proteins had discordant associations, which may suggest a counter-regulatory mechanism in response to biological processes that lead to CHD.

Conclusions. We used proteomic profiling and large-scale genomic data to identify, replicate, and validate casual protein associations with recurrent CHD, which should be prioritized for further investigation as therapeutic targets for secondary prevention of CHD.

Adipose single nucleus transcriptomic and epigenomic sequencing discovers 192 colocalized regulatory variants for eight cardiometabolic disease traits

Subsession Time: Saturday, October 18 at 10:45am – 11:00am

Authors: Seung Hyuk T. Lee (UCLA), Asha Kar (UCLA), Kyla Z. Gelev (UCLA), Zeyuan Johnson Chen (UCLA), Sankha Subhra Das (UCLA), Sini Heinonen (University of Helsinki), Tuure Saarinen (Helsinki University Hospital, University of Helsinki), Maija Vaittinen (University of Eastern Finland), Dorota Kaminska (University of Eastern Finland, UCLA), Hilkka Peltoniemi (Eira Hospital), Chongyuan Luo (UCLA), Anne Juuti (Helsinki University Hospital, University of Helsinki), Ville Männistö (University of Eastern Finland, Kuopio University Hospital), Jussi Pihlajamäki (University of Eastern Finland, Kuopio University Hospital), Minna U. Kaikkonen (University of Eastern Finland), Kirsi H. Pietiläinen (University of Helsinki, Helsinki University Hospital), Brunilda Balliu (UCLA), Päivi Pajukanta (UCLA)

Abstract: Previous colocalization analyses of cardiometabolic diseases (CMDs) using bulk subcutaneous adipose tissue (SAT) are confounded by cellular heterogeneity, limiting resolution of gene-trait associations and masking causal mechanisms. To overcome this, we performed cell-type level cis-eQTL mapping in single nucleus RNA-seq data of 127 SAT samples from Finns with obesity. We integrated these results with SAT cell-type level cellular compartments from epigenomic data generated using snm3C-seq. We identified 166,168 lead cis-eQTL variants for 1,684 eGenes, with 62-76% of the eGenes residing in active (A) cellular compartments of the cell-type. Using stratified LD score regression in the UK Biobank, we found that only A-compartment cis regions were enriched for CMD heritability; B compartments were depleted, consistent with the presence of active promoters and distal enhancers in A compartments. We then performed Bayesian colocalization between SAT cell-type eQTLs and GWAS loci from eight CMD traits: body mass index (BMI), coronary artery disease (CAD), triglycerides (TGs), type 2 diabetes (T2D), waist-to-hip ratio adjusted for BMI (WHRadjBMI), Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD), C-reactive protein (CRP), and alanine transaminase (ALT). We identified 192 colocalized eQTL–GWAS pairs for 129 eGenes. Of these, 19 BMI, 7 CAD, 46 TGs, 31 T2D, and 13 WHRadjBMI colocalizations were not detected in bulk SAT. For traits not previously tested in adipose (e.g., MASLD, CRP, ALT), we discovered 6, 16, and 13 colocalizations, respectively. As an example, T2D variants colocalized with 23 eGenes in adipocytes, 8 in adipose stem cells, 8 in myeloid, 4 in endothelial, and 2 in lymphoid cells. In adipocytes alone, 87% of T2D colocalizations were missed in bulk data, including rs6565259 near SH2B1, involved in adipocyte differentiation. For known T2D loci (e.g., CMIP, MARK3, SETD9), we pinpointed their cell types of action. Finally, 81% of colocalized T2D variants localized to adipocyte A compartments, underscoring their regulatory relevance. Together, our results uncover 129 CMD-associated target genes and assign regulatory effects to distinct SAT cell types. This work reveals cell-type-specific regulatory architecture underlying CMDs and emphasizes the power of single-nucleus functional genomics.

Session 91: End Results—Telomere Resolution and Involvement in Genomic Instability

Location: Room 205ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Saturday, October 18 at 10:00am – 11:00am

Telomere dysfunction promotes remote recombination events between telomeres, centromeres and ribosomal DNA that are catalyzed by long-range chromatin interactions

Subsession Time: Saturday, October 18 at 10:00am – 10:15am

Authors: Yi-An Chen, T. Rhyker Ranallo-Benavidez, Noelle Fukushima, Yue Hao, Ogechukwu Mbegbu, Floris P. Barthel

Abstract: Telomere dysfunction (TD) arises when the DNA end protection afforded by telomeres goes awry and is thought to be an important driver of genomic instability. To better understand how TD drives disease, we modeled TD using two independent systems. First, we transduced normal human astrocyte (NHA) cells with retroviruses containing HPV-18 E6/E7 in order to disrupt cell cycle checkpoints. NHA-E6E7 cells were expanded over long periods of time to progressively shorten telomeres and induce TD. Second, we transduced NHA cells with a dominant negative TRF2 telomere-binding protein to directly disrupt end protection.

Applying long- and short-read DNA sequencing to cells from both models and utilizing our KaryoScope approach to call repetitive elements we identified centromeric satellites and ribosomal DNA (rDNA) juxtaposed to telomeric DNA in TD+ but not control cells, suggesting that TD promotes recombination between these otherwise remote regions. Applying KaryoScope to telomerase+ HeLa S3 cells revealed no events whereas alternative lengthening of telomeres (ALT) positive U2-OS cells exhibited an even higher recombination burden than TD+ NHA cells. This finding is consistent with the notion that ALT+ tumors adapt to ongoing TD whereas telomerase+ tumors, having undergone TD during tumorigenesis, negatively select for further TD via the reactivation of telomerase. In agreement with these findings, we applied KaryoScope to analyze telomeric long reads from an ALT+ astrocytoma and telomerase+ glioblastoma and found abundant aberrant recombination events in astrocytoma but none in glioblastoma.

We next applied our Telomere-C method to test whether recombination between distal repeats was catalyzed by spatial proximity. Indeed, telomeres were found to strongly interact with rDNA elements in all tested lines. Surprisingly, U2-OS cells showed a significant reduction in interactions between telomeres and alpha-satellites, possibly reflecting their juxtaposition.

Our findings reveal that TD promotes telomeric recombination events involving centromeres and rDNA catalyzed by long-range telomeric chromatin interactions. TD-induced distal recombination events are selectively disadvantageous, leading most cancers to activate telomerase to avoid further TD. Conversely,

in ALT+ tumors, these events are tolerated and persist. This mechanistic insight significantly advances our understanding of genomic instability in disease, pointing towards novel diagnostic markers and therapeutic strategies.

A genomic foundation model for ALT detection reveals new therapeutic vulnerabilities in pediatric cancer lacking actionable drug targets

Subsession Time: Saturday, October 18 at 10:15am – 10:30am

Authors: Declan Bennett (St. Jude Children's Research Hospital), Monica Wierdl (St. Jude Children's Research Hospital), JD Johnson (St. Jude Children's Research Hospital), Yuan Feng (St. Jude Children's Research Hospital), Xiaotu Ma (St. Jude Children's Research Hospital), Lily Guenther (St. Jude Children's Research Hospital), Paul Gleeher (St. Jude Children's Research Hospital)

Abstract: Approximately 15–20% of cancers maintain telomeres through the alternative lengthening of telomeres (ALT) pathway, a recombination-based mechanism associated with poor clinical outcomes relative to telomerase-positive tumors. ALT is frequently driven by loss-of-function mutations in ATRX or DAXX, yet it is not mutually exclusive with TERT expression, reflecting the complex spectrum of telomere maintenance mechanisms in cancer. Despite its clinical relevance, gold-standard assays for ALT detection are low-throughput and not scalable to large cohorts, leaving the dependencies of ALT-positive cancers poorly characterized.

To address this, we developed a whole-genome sequencing (WGS)-based framework for inferring ALT status. Using telomere-relevant features extracted from over 1,000 cancer cell lines in the Depmap resource, we trained a foundation model to predict ALT activity. Our method provides the first scalable map of ALT across a densely CRISPR and drug-screened cell line panel to existing tumor-normal matched ALT detection algorithms and enables high-throughput ALT classification across diverse cancer types.

We integrated our ALT predictions with genome-wide CRISPR loss-of-function screening data to identify dependencies unique to ALT-positive cancers. Notably, we found SMARCAL1, whose helicase domain is paralogous to that of ATRX, to be a key dependency in osteosarcoma, neuroblastoma, and soft tissue sarcomas—tumors often characterized by ATRX lof and poor prognosis, and currently lack actionable drug targets. We also identified FANCM, another gene with a paralogous helicase domain, as a top dependency, suggesting that ALT-positive cancers are especially reliant on replication stress response pathways.

Our work provides a scalable method to infer ALT from WGS data and systematically connect ALT status to cancer vulnerabilities. Incorporating ALT predictions into functional genomics platforms such as Depmap will enable precision oncology efforts to uncover novel therapeutic targets in historically hard-to-treat cancers, with few clear therapeutic targets. This approach also lays the groundwork for clinical integration of ALT biomarkers in patient stratification and drug development pipelines.

FLT3 regulates telomere maintenance and suppresses senescence in acute myeloid leukemia

Subsession Time: Saturday, October 18 at 10:30am – 10:45am

Authors: Pei-Yi Chen (Hualien Tzu Chi Hospital, Tzu Chi University), Wan-Yu Yang (Tzu Chi Senior High School Affiliated with Tzu Chi University), Wei-Chen Wang (Tzu Chi University), Hui-Yu Jiang (Hualien Tzu Chi Hospital), Jui-Hung Yen (Tzu Chi University)

Abstract: Acute myeloid leukemia (AML) is the most common type of acute leukemia in adults and is characterized by rapid disease progression. Approximately 30% of AML cases harbor mutations in the FMS-like tyrosine kinase 3 (FLT3) gene, which is associated with poor clinical outcomes. While telomere dynamics are recognized as critical contributors to AML pathogenesis, the regulatory influence of FLT3 on telomere maintenance remains poorly understood. To investigate this relationship, we analyzed 942 AML samples from the OHSU cohort using the cBioPortal cancer genomics database (Cancer Cell, 2022). FLT3 mRNA expression showed a significant positive correlation with major telomere maintenance genes, including TERT, RTEL1, PARN, POT1, MYC, TERC, and TERF2. To examine the functional role of FLT3 in telomere regulation, we established a cytarabine-resistant AML cell line (MV4-11-CR) that exhibits elevated FLT3 expression. We generated stable FLT3-knockdown AML cell clones using shRNA and assessed the effects on senescence and telomere length. Knockdown of FLT3 led to a significant increase in senescent cell populations. Telomere shortening was confirmed in FLT3-deficient clones using quantitative PCR and interphase fluorescence in situ hybridization (FISH). Transcriptomic profiling through RNA sequencing and Gene Ontology-based gene set enrichment analysis (GSEA) revealed that FLT3 knockdown suppressed cell cycle progression and mitotic signaling, inhibiting three key telomere maintenance pathways. RT-qPCR further validated these findings, demonstrating reduced telomere-associated gene expression, including TERT, MYC, POT1, and SMG6, in FLT3-deficient clones. Ongoing work aims to elucidate the molecular mechanisms underlying these observations. In summary, our results suggest that FLT3 may protect against senescence by regulating the expression of telomere maintenance genes. This study uncovers a previously unrecognized role for FLT3 in telomere biology, offering new insights into AML pathogenesis and potential therapeutic strategies.

Determinants of chromosome-specific telomere lengths estimated from long reads among 2,573 All of Us participants

Subsession Time: Saturday, October 18 at 10:45am – 11:00am

Authors: Niyati Jain (The University of Chicago), Jiajun Luo (The University of Chicago), Yuqing Yang (The University of Chicago), Briseis Aschebrook-Kilfoy (The University of Chicago), Lin S. Chen (The University of Chicago), Brandon L. Pierce (The University of Chicago)

Abstract: Telomeres are DNA-protein structures that protect chromosome ends. The DNA component of telomeres shortens as cells divide, and telomere length (TL) is an established biomarker of aging and disease risk. Most previous studies of TL in humans have analyzed average TL across all chromosomes; thus, our knowledge of variability in TL across individual chromosome arms remains limited. The availability of long-read whole-genome sequencing (lrWGS) data has enabled the development of computational methods to measure TL in a chromosome-specific (cs) fashion. In this study, we estimated csTLs for >2,500 individuals by applying the Telogator2 software to lrWGS data from the All of Us (AoU)

population cohort. We characterized variability in csTL attributable to individuals, chromosome arms, participant characteristics, and technical factors using random effects models. After accounting for batch effects, 25.7% of the variation in csTL was explained by biological factors. We found that csTL varies by chromosome arm, which accounts for 9.1% of the variance in csTL. Specific arms observed to be consistently shorter (e.g., 17p, 12p, 20q) or longer (e.g., 4q, 12q, 3p) across individuals were highly consistent with those identified in prior studies, highlighting the potential for chromosome-specific mechanisms of TL regulation. Substantial variability in csTL was attributable to individuals (8.9%), independent of age, supporting the hypothesis that individuals are endowed with an inherently short or long TL in early development, which is maintained throughout life. While age showed a clear negative association with csTL across all arms (3.7% of variance explained), the strength of the association varied by arm. Additional variability in csTL was explained by ancestry, sex, smoking, and BMI. Leveraging phenotypic data from the AoU cohort, we were able to examine disease associations that incorporate features of the csTL distribution. Our results demonstrate that the length of an individual's shortest telomere is more strongly associated with increased cardiovascular disease risk than their average TL across all chromosomes. Our work identified key lrWGS quality metrics that impact csTL estimation for AoU lrWGS data, providing a framework to guide future studies in AoU and beyond. Overall, this study highlights the utility of lrWGS data for csTL profiling and emphasizes the need for larger cohort studies of csTL to advance our understanding of TL's role in aging and disease.

Session 92: Pathogenic Mechanisms in Neurological Disease: Gene to Network and Everything In Between

Location: Room 206AB/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Saturday, October 18 at 10:00am – 11:00am

Dissecting MECP2 cis-regulatory architecture with massively parallel reporter assays to assess its contribution to male-biased autism

Subsession Time: Saturday, October 18 at 10:00am – 10:15am

Authors: Rebecca Meyer-Schuman (Department of Human and Molecular Genetics, Baylor College of Medicine, and the Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston TX 77030, USA), Fisher Cherry (Department of Human and Molecular Genetics, Baylor College of Medicine, and the Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston TX 77030, USA), Yang Sui (Department of Genome Sciences, University of Washington School of Medicine, Seattle WA 98195, USA), Athanasios Papastathopoulos-Katsaros (Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, and Department of Pediatrics, Baylor College of Medicine, Houston TX 77030, USA), Yi Zhong (Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, and Department of Pediatrics, Baylor College of Medicine, Houston TX 77030, USA), Kelsey Hennick (Department of Neurological Surgery, University of California, San Francisco CA 94143, USA), Druha Karunakaran (Center for Human Genetics and Genomics, New York University Grossman School of Medicine, New York NY 10016, USA), Hanna Berk-Rauch (Center for Human Genetics and Genomics, New York University Grossman School of Medicine, New York NY 10016, USA), Zhandong Liu (Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, and Department of Pediatrics, Baylor College of Medicine, Houston TX 77030, USA), Aravinda Chakravarti (Center for Human Genetics and Genomics, New York University Grossman School of Medicine, New York NY 10016, USA), Tomasz Nowakowski (Department of Neurological Surgery, University of California, San Francisco CA 94143, USA), Evan Eichler (Department of Genome Sciences, University of Washington School of Medicine, and Howard Hughes Medical Institute, University of Washington, Seattle WA 98195, USA), Huda Zoghbi (Department of Human and Molecular Genetics and Department of Pediatrics, Baylor College of Medicine, Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, and Howard Hughes Medical Institute, Houston TX 77030, USA).

Abstract: Loss-of-function mutations in the X-chromosome gene MECP2 cause the neurodevelopmental disorder Rett Syndrome, while MECP2 duplications or triplications also cause profound neurological disease. However, even mild hypo-morphic mutations in MECP2 can cause autism, intellectual disability, and neuropsychiatric phenotypes. These milder mutations mainly impact males, while female carriers are protected by a mosaic wild-type allele and are mostly unaffected. This dosage sensitivity is reflected in male mice, where a 30% decrease or 50% increase in *Mecp2* expression causes autism-like social and behavioral defects. Based on this observation, we propose that mutations modestly changing MECP2

expression levels—such as non-coding mutations in MECP2 cis-regulatory regions (CREs)— contribute to the increased incidence of autism in boys, who are nearly four times as likely to be diagnosed with autism as girls.

To understand the MECP2 cis-regulatory landscape and its role in neurodevelopmental disorders, we performed a massively parallel reporter assay (MPRA) in human iPSC-derived neurons (iNeurons). We included MECP2's proximal promoter, a known enhancer, a known repressor, and two novel candidate CREs. We tested the effect of 37 ultra-rare, maternally inherited variants found in male ASD probands from >8,000 sequenced families, along with hundreds of synthetic transcription factor binding site mutations introduced by shuffling predicted motifs.

We identified multiple transcription factor binding hotspots in four of the five tested elements and developed a computational pipeline to prioritize likely motifs and transcription factors for further validation. We also found that multiple variants from male autism probands altered MPRA activity. Consistent with their functional relevance, these variants often overlapped the identified transcription factor binding hotspots. We validated a subset of these variants using luciferase reporter assays, then selected two variants—one in the MECP2 promoter and one in a novel repressor element—to edit into the endogenous MECP2 locus using CRISPR-Cas9. We are currently testing for effects on MECP2 RNA and protein expression in iNeurons. This study provides new insights into MECP2 gene regulation and is the first to report noncoding MECP2 mutations with functional impact in boys with autism spectrum disorder.

CYP2D6 contributes to cognitive dysfunction and promotes excessive synapse loss

Subsession Time: Saturday, October 18 at 10:15am – 10:30am

Authors: Liang Ma (University of Texas Health Science Center at San Antonio)

Abstract: The Psychiatric Genomic Consortium has identified more than a hundred loci associated with schizophrenia. Recently, our integration of genome-wide data has highlighted CYP2D6 as one of the leading risk genes for this disorder. However, the regulatory mechanisms of CYP2D6 and its functions in the brain remain largely unexplored. Here, we leveraged multi-omics data from over a thousand human postmortem brains collected independently from two consortiums. Our analysis revealed that schizophrenia risk alleles are consistently associated with CYP2D6 upregulation in both European and African American populations. By integrating CYP2D6 gene expression with DNA methylation and H3K9ac ChIP-seq data generated from human postmortem brains, we prioritized five SNPs located within active chromatin regions that significantly interacted with four CpG sites. Additionally, we observed a strong correlation between the DNA methylation levels and CYP2D6 gene abundance. SNP-motif analysis suggested that schizophrenia risk alleles may influence the binding affinity of brain-related transcription factors, such as ZIC2. Furthermore, we observed a significant effect of the CYP2D6 genotype on brain activation during a working memory task, as measured by fMRI in 409 unrelated individuals. Next, we employed virus-mediated overexpression of CYP2D6 in the medial prefrontal cortex (mPFC) to investigate the effects of this schizophrenia-susceptible gene on cognitive, emotional, and social

behaviors. Our findings revealed that CYP2D6 impairs working memory, object recognition memory, and spatial memory, while having no significant impact on social abilities, anxiety-like behaviors, or depression-like behaviors. In mouse brain tissue, we observed that CYP2D6 adversely affects synaptic and mitochondrial ultrastructure. Transcriptomic analysis further indicated that CYP2D6 disrupts genes involved in synaptic signaling pathways. Additionally, using both mouse brain tissue and a human cell line, we found that CYP2D6 induces excessive ROS, which may contribute to the observed behavioral, cellular, and molecular alterations. In summary, we deepened our understanding of the CYP2D6 regulatory network in brain, its impact on cognition and synapse through inducing oxidative stress. Those findings offer therapeutic implications for schizophrenia.

Synergistic Loss of Regulatory and Protein-Coding Genes Drives Neuronal and Mitochondrial Dysregulation in Prader-Willi Syndrome

Subsession Time: Saturday, October 18 at 10:30am – 10:45am

Authors: Rachita Yadav (Massachusetts General Hospital, Harvard Medical School, Broad Institute), Derek JC Tai (Massachusetts General Hospital, Harvard Medical School, Broad Institute), Celine EF de Esch (Massachusetts General Hospital, Harvard Medical School, Broad Institute), Xander Nuttle (Massachusetts General Hospital, Harvard Medical School, Broad Institute), Yating Liu (Massachusetts General Hospital), Jaewon Shin (Massachusetts General Hospital), Nicholas D Burt (Massachusetts General Hospital), Serkan Erdin (Massachusetts General Hospital, Broad Institute), Bimal Jana (Massachusetts General Hospital, Harvard Medical School, Broad Institute), Alex Yenkin (Harvard University, Massachusetts General Hospital, Broad Institute), Riya Bhavsar (Massachusetts General Hospital), John Lemanski (Massachusetts General Hospital), Dadi Gao (Massachusetts General Hospital, Harvard Medical School, Broad Institute), James F. Gusella (Harvard University, Massachusetts General Hospital, Harvard Medical School, Broad Institute), Michael E. Talkowski (Massachusetts General Hospital, Harvard Medical School, Broad Institute)

Abstract: Prader-Willi Syndrome (PWS) is a complex neurodevelopmental disorder caused by loss of paternally inherited genes within the imprinted 15q11.2–q13 region. 5–6 Mb deletions (Type I/II) are the most common cause, affecting five paternally expressed protein-coding genes, six C/D box small nucleolar RNAs (SNORDs), and several long non-coding RNAs (IPW, SPA1, SPA2). To dissect the genetic drivers of PWS, we generated an allelic series of CRISPR-edited human induced pluripotent stem cells (hiPSCs) with deletions across the PWS locus. This included isogenic hiPSC lines and matched controls, covering Type I/II deletions, a critical region deletion (CRD; SNORD116, IPW, SNORD109A), and individual gene deletions (SNORD116, IPW, SNORD109A, SPA1, MAGEL2), across male and female backgrounds. These were differentiated into Ngn2-induced excitatory neurons (iGLUTs, n=810) and hypothalamic organoids (HOs). Transcriptomic (mRNA, small RNA) and bioenergetic profiling via Seahorse assays were performed to assess molecular and metabolic changes. We developed a bootstrapping-based differential expression pipeline to enhance detection power and identified 767 PWS signature genes in Type I/II iGLUTs at >60% directional replication across the 100,000 iterations. The co-expression modules of these PWS signature genes are involved in transcription regulation, chromatin remodeling, GPCR ligand binding, ribosomes, and synapses, and form a distinct mitochondrion module with functions including

the citric acid TCA cycle, aerobic respiration, and oxidative phosphorylation. The scRNAseq profiling of PWS HOs revealed that the hypothalamic neuron cluster also shows mitochondrial electron chain dysregulation. Seahorse assays showed elevated mitochondrial stress and increased basal respiration in both Type I/II iGLUTs and HOs. Regulatory analysis revealed that up to 68% of DEGs in PWS-associated pathways are enriched for targets of 15q11.2–q13 SNORDs, indicating extensive post-transcriptional control by non-coding RNAs. The distinct transcriptional dysregulation observed across sub-regional and individual gene deletions suggests a synergistic model, where the combined loss of regulatory RNAs, protein-coding genes, and long non-coding RNAs disrupts interconnected molecular networks. These findings establish a functional map of multigenic contribution that underlies the complexity of PWS and uncover key alterations in neuronal and mitochondrial functions linked to disease pathology.

Transcriptomic Profiling of Extracellular Vesicles from Autism Patient-Derived Forebrain Organoids Reveals Novel Signatures

Subsession Time: Saturday, October 18 at 10:45am – 11:00am

Authors: Phillip Smit (Department of Medical Sciences, Frank H. Netter SoM, Quinnipiac University, North Haven, CT, USA), Isidora Stankovic (Center for Neurogenetics, Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, Cornell University, New York, NY, USA), Dilek Colak (Center for Neurogenetics, Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, Cornell University, New York, NY, USA, Gale and Ira Drukier Institute for Children's Health, Weill Cornell Medicine, Cornell University, New York, NY, USA), Paul Wolujewicz (Department of Medical Sciences, Frank H. Netter SoM, Quinnipiac University, North Haven, CT, USA; Center for Neurogenetics, Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, Cornell University, New York, NY, USA; Department of Biomedical Sciences, School of Health Sciences, Quinnipiac University, Hamden, CT, USA)

Abstract: Autism Spectrum Disorder (ASD) affects 1 percent of the world's population with an increased prevalence of 178 percent since 2000. Although altered synaptic function putatively accounts for many of the abnormalities seen in ASD, the specific molecular mechanisms underlying this disorder remain poorly defined. A growing body of evidence suggests that extracellular vesicles (EVs), specifically exosomes, play a critical role in cellular communication within the brain. While they have been implicated in various types of diseases from cancer to neurodegeneration, their involvement in ASD remains largely unexplored. In this study, we utilized patient-derived cortical organoid models to characterize EVs secreted by human three-dimensional (3D) tissue and defined their cargo with tailored RNA sequencing analyses. We observed significant differences both in coding and non-coding RNA content between ASD and control samples. In fact, we found that 58 coding RNAs were upregulated and 55 coding RNAs were downregulated in ASD EVs compared to CTRLs. These genes are grouped into clinically significant pathways such as polyols and phosphoinositide metabolism. Disruptions in phosphoinositide signaling could affect neuronal connectivity and synaptic function, both of which are central to the pathophysiology of ASD. In addition to coding genes, we also identified a range of upregulated noncoding RNA elements, particularly miRNAs, in ASD-derived EVs. Specifically, exosomal miRNAs were shown to play a pivotal role in regulating gene expression by modulating mRNA stability and translation. For example, miRNA370, which was previously shown to increase permeability of the

BBB is altered in ASD EVs. Indeed, altered BBB permeability and associated inflammation have been observed previously in ASD. Our study also identified novel miRNAs in the context of EV cargo such as miRNA6724-1, miRNA6724-2, miRNA6724-3, and miRNA6724-4. Future directions may involve studies exploring the potential of miRNA expression in ASD EVs as diagnostic markers for this disorder, offering a non-invasive approach to early diagnosis and monitoring of neurodevelopmental conditions.

Session 93: The Metabolic Matrix: Untangling Complex Conditions

Location: Room 253ABC/Level 2, Thomas M. Menino Convention and Exhibition Center

Session Time: Saturday, October 18 at 10:00am – 11:00am

Molecular Risk Scores Enhance Type 2 Diabetes Prediction Across Clinical Contexts

Subsession Time: Saturday, October 18 at 10:00am – 10:15am

Authors: Magdalena Sevilla-Gonzalez (MGH), Paul Hanson (MGH), Alicia Huerta (Broad Institute), Maheak Vora (Broad Institute), Elizabeth W. Karlson (MGH), Jose C. Florez (MGH), Chirag J. Patel (HMS), Josep M. Mercader (MGH), Aaron Leong (MGH)

Abstract: Background:

Real-world integration of genetic and molecular biomarkers for disease prediction remains limited. We evaluated the predictive utility of a metabolomic risk score (MRS) and a polygenic risk score (PRS) for incident type 2 diabetes (T2D) in a medically and ancestrally diverse healthcare system.

Methods:

We analyzed 36,352 patients free of T2D at the time of metabolite measurements in the Mass General Brigham Biobank (mean follow-up: 6.2 years). Cox models evaluated the association of a 33-metabolite risk score—constructed using LASSO-derived weights based on prior associations with T2D (Nightingale platform)—and a genome-wide PRS with incident T2D in four nested models adjusted for increasing amounts of clinical information: Model 1: age, sex, and genetic ancestry; Model 2: Model 1 + BMI, systolic blood pressure, and family history of T2D; Model 3: Model 2 + random glucose; Model 4: Framingham T2D clinical risk score and random glucose.

Results:

Both PRS and MRS were independently associated with incident T2D in all models. In Model 4, MRS showed the largest effect (HR 1.88; 95% CI: 1.72–2.05). PRS effect estimates remained stable in all models, but MRS effect estimates increased when adjusting for more clinical variables. The highest discriminative performance was observed in Model 3 with the combined PRS+MRS model (c-index = 0.794).

Interaction analyses (MRS x clinical variable) revealed that the MRS effect was modified by the most predictive clinical variable in each model: age (Model 1, P interaction = 5.0×10^{-7}), BMI (Model 2, P interaction = 1.6×10^{-14}), random glucose (Model 3, P interaction = 6.8×10^{-5}), and the clinical risk score (Model 4, P interaction = 0.002). Time-varying coefficient (TVC) analyses showed stable MRS effects over

time in Models 1–3 ($P_{TVC} > 0.05$), but the effect attenuated in Model 4 ($P_{TVC} = 0.001$) over time likely due to the adjustment for clinical variables proximal to disease onset.

In Model 4, absolute 5-year T2D risk ranged from under 2% in the lowest joint PRS and MRS tertiles to over 15% in the highest tertiles, highlighting the added stratification power of these molecular scores beyond a traditional clinical risk score.

Conclusion:

Metabolomic and polygenic scores are independent, complementary predictors of T2D risk. Their predictive utility across clinical contexts supports their integration in situations that vary by clinical information availability for the development of targeted preventive strategies.

Male-specific genetic regulation across germline and somatic variation on the Y chromosome contributes to type 2 diabetes

Subsession Time: Saturday, October 18 at 10:15am – 10:30am

Authors: Go Sato (The University of Tokyo, Osaka University, RIKEN IMS), Yuji Yamamoto (Osaka University), Kyuto Sonehara (The University of Tokyo), Ryunosuke Saiki (Kyoto University), Takafumi Ojima (The University of Tokyo, Tohoku University), Masahiro Kanai (Massachusetts General Hospital), Aoxing Liu (Massachusetts General Hospital), Ryuya Eda (Osaka University, RIKEN IMS), Yuya Shirai (Osaka University), Shinichi Namba (The University of Tokyo), Ho Namkoong (Keio University), Atsushi Kumanogoh (Osaka University), Shigeki Nakagome (Trinity College Dublin), Gen Tamiya (Tohoku University), Mark J. Daly (Massachusetts General Hospital), James F. Wilson (The University of Edinburgh), Masayuki Yamamoto (Tohoku University), Koichi Matsuda (The University of Tokyo), Seishi Ogawa (Kyoto University), Toshimasa Yamauchi (The University of Tokyo), Takashi Kadowaki (Toranomon Hospital), Yukinori Okada (The University of Tokyo, Osaka University, RIKEN IMS)

Abstract: Our understanding of the biological role of the Y chromosome remains limited. Here, we systematically profile germline Y haplogroups and somatic loss of the Y chromosome (LOY) for 122,683 East Asian males (BioBank Japan: BBJ) and 181,472 European males (UK Biobank: UKB). Y haplogroup compositions and frequency spectra are population-specific and diverse across geographic regions within Japan, characterized by the Japanese-specific haplogroup D. Long-read WGS reveals an ancient hunter-gatherer Jomon-derived Alu insertion that characterizes haplogroup D. A genome-wide association study with 156,863 males from multiple East Asian cohorts identifies eight novel loci associated with LOY. A phenome-wide scan uncovers male-specific genetic regulation of complex traits, such as pleiotropic effects of haplogroup D involving height ($\beta = -0.050$, $P = 1.5 \times 10^{-19}$) and type 2 diabetes (T2D; odds ratio [OR] = 0.94, $P = 2.2 \times 10^{-5}$). LOY is associated with increased T2D risk in East Asians but with reduced risk in Europeans. These opposite associations are further validated using independent cohorts in East Asians ($nT2D = 32,976$ from BBJ and Tohoku Medical Megabank, OR = 1.15, $P = 6.2 \times 10^{-8}$) and Europeans ($nT2D = 26,595$ from UKB and FinnGen, OR = 0.84, $P = 3.9 \times 10^{-18}$). This disparity may arise from ethnic differences in the role of obesity in T2D pathogenesis. Using BBJ's longitudinal disease follow-up data, we show that the presence of LOY prospectively increases future T2D onset risk in East Asians (HR = 1.28,

P = 0.043). Incorporating germline and somatic variation on the sex chromosomes enhances the polygenic prediction of T2D risk for both sexes. To gain biological insights into the Y chromosome associations with T2D, we leverage multi-omics resources, including metabolomics, proteomics, and single-cell transcriptomics. Individuals with detectable LOY in blood cells tend to exhibit LOY across various body tissues. Single-cell RNA sequencing enables accurate mapping of Y chromosome-deficient cells in peripheral blood, revealing their enrichment in monocytes across populations. We also found that LOY cells in non-blood tissues are enriched in specific cell types relevant to disease pathogenesis. In the pancreas, beta cells lacking the Y chromosome exhibit transcriptional alterations associated with impaired beta cell function. Our study demonstrates male-specific genetic regulation of T2D in East Asians, providing insights to improve diabetes management.

Signatures of positive selection in Pacific-ancestry genomes reveal adaptive signals relevant to metabolic diseases

Subsession Time: Saturday, October 18 at 10:30am – 10:45am

Authors: Connor C. Littlefield (University of Utah, Department of Human Genetics), Jose M. Lazaro-Guevara (McGill University, Department of Medical Genetics), Devorah Stucki (University of Utah, Department of Internal Medicine), Michael Lansford (University of Utah, Department of Human Genetics), Melissa H. Pezzolesi (University of Utah, Department of Internal Medicine - Division of Nephrology and Hypertension), Emma J. Taylor (University of Utah, Department of Internal Medicine - Division of Nephrology and Hypertension), Justina P. Tavana (Brigham Young University, Department of Biology), William L. Holland (University of Utah, Department of Nutrition and Integrative Physiology), Kalani L. Raphael (University of Utah, Department of Internal Medicine - Division of Nephrology and Hypertension), Marcus G. Pezzolesi (University of Utah, Department of Internal Medicine - Division of Nephrology and Hypertension, Department of Human Genetics, Department of Nutrition and Integrative Physiology)

Abstract: Individuals with Pacific ancestry have been underrepresented in human genomic research. To better understand the genetic underpinnings of the elevated burden of metabolic diseases—including Type 2 diabetes, chronic kidney disease, and cardiovascular disorders—we investigated recent positive selection in this understudied population. We collected whole blood and generated high-coverage whole-genome sequencing data (30×) from 450 individuals with Pacific ancestry. Sequencing reads were aligned to a recently developed Pacific ancestry pangenome reference using Giraffe, projected to GRCh38, and processed with DeepVariant for variant calling. This approach improved variant call sensitivity and precision genome-wide, particularly in repetitive and segmentally duplicated regions, enhancing downstream selection analyses.

Principal component analysis, incorporating 1000 Genomes reference samples with our Pacific ancestry cohort, revealed low levels of European and East Asian admixture. We performed local ancestry inference with East Asian, European, and Melanesian reference panels to map continental ancestry across the genome. This enabled ancestry-specific selection scans by masking local ancestry segments

and identifying regions of local ancestry deviation (LAD). Additional XP-EHH scans against East Asian and European samples further refined population-specific signals of adaptation.

To detect recent positive selection, we computed iHS and nSL statistics and identified 98 genes with strong haplotypic signals ($|\text{standardized iHS}| > 4$). The strongest signal was located in intronic variants within the MCTP1 gene ($|\text{iHS}| > 7$), which encodes a calcium-binding transmembrane protein involved in synaptic vesicle trafficking and neuronal function. Variants in MCTP1 have been associated with Type 2 diabetes and Alzheimer's disease, suggesting a possible role in metabolic and neurological phenotypes in Pacific populations. Gene ontology analysis of variants with $|\text{standardized iHS}| > 3$ revealed enrichment for genes involved in cardiac muscle contraction and atrioventricular node signaling ($\text{FDR} = 1.72 \times 10^{-4}$). These pathways may contribute to the increased prevalence of cardiovascular disease observed in Pacific ancestry groups.

Together, this study highlights the utility of a pangenome-informed framework and ancestry-aware selection analysis in uncovering adaptive loci relevant to metabolic health in Pacific ancestry populations.

Phenome-wide Genetic Colocalization and Mendelian Randomization of eQTLs and pQTLs against complex traits in up to 2.5 million participants across multiple Biobanks

Subsession Time: Saturday, October 18 at 10:45am – 11:00am

Authors: João Fadista (Novo Nordisk), Hector Climente Gonzalez (Novo Nordisk), Neil Robertson (Novo Nordisk), Ciara Vangjeli (Novo Nordisk), Dmitry Shungin (Novo Nordisk), Nils Koelling (Novo Nordisk)

Abstract: Unraveling the intricate relationships between genetic variants and diseases is fundamental to drug target discovery and advancing precision medicine. This work presents a comprehensive atlas that delineates gene-disease associations through Bayesian genetic colocalization (coloc) and Mendelian randomization (MR) analyses. We systematically assess how genetic variants influencing gene expression (eQTLs) and protein levels (pQTLs) correlate with complex traits and disease phenotypes.

Our exposure analyses utilize multiple eQTL resources, including the Genotype-Tissue Expression (GTEx) database and published meta-analyses for eQTLs across diverse tissues: liver ($N = 1,937$), adipose tissue ($N = 2,344$), skeletal muscle ($N = 1,256$), kidney ($N = 667$), arteries ($N = 934$), pancreatic islets ($N = 420$), and blood ($N = 31,684$). Similarly, our pQTL analyses draw from large datasets, including those generated using the Olink platform (UK Biobank $N = 49K$, Epic-Norfolk $N = 1K$) and the Somalogic platform (deCODE $N = 35K$, Fenland $N = 10K$, ARIC $N = 7K$). Our outcomes are derived from prominent biobanks, including the UK Biobank (500K participants), FinnGen (500K participants), and the Million Veteran Program (MVP; 600K participants), alongside studies from disease-specific consortia featured in the Genome-Wide Association Studies (GWAS) catalog.

Based on 2,061,393 coloc pairs, our findings reveal significant associations between eQTLs and pQTLs across various disease contexts and quantitative traits, leading to the identification of 3,918 genes and

456 proteins causally implicated, with MR p-value $< 2.5 \times 10^{-6}$ and coloc posterior probability > 0.7 . We unveil multiple novel causal pairs and replicate established associations, including PCSK9 in atherosclerotic cardiovascular disease (ASCVD), ANGPTL3 in lipid metabolism, GDF15 in hyperemesis gravidarum, APOL1 in kidney disease, and GLP1R in diabetes and obesity. We also identify widespread pleiotropy, with 6,639 genes associated with two or more traits (mean: 15 traits; range: 1–407), with GCKR, BRAP and APOE being the most pleiotropic.

Our discoveries have two primary implications: (1) they propose potential novel therapeutic targets amenable to intervention while minimizing adverse effects; and (2) they underscore the necessity of integrating diverse genetic data into therapeutic target discovery, thereby enriching drug development pipelines.

