ASHG 2025 Lightning Talk Abstracts

As of September 30, 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 77: Beyond Sequencing: Genetic Architecture of Complex Traits

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

Session Time: Friday, October 17 at 1:30pm – 2:30pm

Proteomic Signatures as Biomarkers of Atherosclerosis Burden

Subsession Time: Friday, October 17 at 1:35pm – 1:40pm

Authors: Lanyue Zhang (Institute for Stroke and Dementia Research (ISD), LMU University Hospital, LMU Munich, Munich, Germany),

Murad Omarov (Institute for Stroke and Dementia Research (ISD), LMU University Hospital, LMU Munich, Munich, Germany),

LingLing Xu (Institute for Stroke and Dementia Research (ISD), LMU University Hospital, LMU Munich, Munich, Germany),

Barnali Das (Urological Diseases Research Center, Boston Children's Hospital, Boston, MA, USA),

Hong Luo (Institute of Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health (GmbH), Neuherberg, Germany),

Stefanie M. Hauck (Metabolomics and Proteomics Core, Helmholtz Zentrum München, German Research Centre for Environmental Health (GmbH), Neuherberg, Germany),

Agnese Petrera (Metabolomics and Proteomics Core, Helmholtz Zentrum München, German Research Centre for Environmental Health (GmbH), Neuherberg, Germany),

Zhi Yu (Clinical and Translational Epidemiology Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; Broad Institute of MIT and Harvard, Cambridge, MA, USA),

Sascha N. Goonewardena (Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA; Division of Cardiovascular Medicine, VA Ann Arbor Health System, Ann Arbor, MI, USA),

Eleftheria Zeggini (Institute of Translational Genomics, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany; Technical University of Munich, School of Medicine and Health, Munich, Germany),

Annette Peters (Institute for Medical Information Processing, Biometry, and Epidemiology, Faculty of Medicine, Ludwig-Maximilians-Universität, Munich, Germany; German Center for Diabetes Research (DZD), Partner München-Neuherberg, Neuherberg, Germany; German Centre for Cardiovascular Research (DZHK), Munich, Germany),

Martin Dichgans (Institute for Stroke and Dementia Research (ISD), LMU University Hospital, LMU Munich, Munich, Germany; Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; German Centre for Cardiovascular Research (DZHK), Munich, Germany; German Centre for Neurodegenerative Diseases (DZNE), Munich, Germany),

Venkatesh L. Murthy (Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA),

Barbara Thorand (Institute of Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health (GmbH), Neuherberg, Germany; German Center for Diabetes Research (DZD), Partner München-Neuherberg, Neuherberg, Germany; Institute for Medical Information Processing, Biometry, and Epidemiology, Faculty of Medicine, Ludwig-Maximilians-Universität, Munich, Germany),

Marios K. Georgakis (Institute for Stroke and Dementia Research (ISD), LMU University Hospital, LMU Munich, Munich, Germany; Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; Program in Medical and Population Genetics and Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, USA)

Abstract: Atherosclerosis progresses silently over decades before manifesting clinically as myocardial infarction or stroke. Currently, no circulating biomarker reliably quantifies atherosclerotic burden beyond imaging techniques. Here, we sought to define plasma proteomic signatures that systematically capture the burden of atherosclerosis. Using CatBoost machine learning applied to plasma proteomes (Olink Explore 3072; 2,920 proteins) from 44,788 UK Biobank participants, we derived four distinct proteomic signatures based on biologically-informed protein sets: the whole proteome (WholeProteome; n=2,920), proteins associated with genetic predisposition to atherosclerosis (Genetic; n=402), those implicated in atherogenesis (Mechanistic; n=680), and proteins enriched in arterial tissue (Arterial; n=248). These signatures robustly discriminated individuals with known atherosclerotic disease from propensity score-matched controls (ROC-AUC up to 0.91, 95% CI: 0.89–0.93). Among 41,200 individuals without baseline atherosclerosis, all four signatures were strongly associated with future major adverse cardiovascular events over a median follow-up of 13.7 years (HR per SD increase in WholeProteome signature: 1.70, 95% CI: 1.64–1.77), providing significant improvements in risk discrimination (ΔC-index: +0.036; p <0.0001) and reclassification (Net Reclassification Index: 0.085–0.135 at a 10% risk threshold) beyond SCORE2. Signature levels correlated with the number of clinically affected vascular beds and carotid ultrasound-measured plaque burden, and successfully predicted future cardiovascular events in external KORA S4 (n=1,361) and KORA-Age1 (n=796) cohorts. Longitudinal analyses using linear mixed-effects models revealed signature trajectories with steeper annual increases among individuals with higher baseline risk factors or subsequent MACE events, and consistent associations with polygenic risk score for coronary artery disease, reflecting their capacity to capture genetically-influenced atherosclerotic burden. These findings demonstrate that proteomic signatures effectively capture atherosclerotic burden and improve cardiovascular risk prediction in asymptomatic individuals. Plasma proteomics may serve as a scalable and accessible alternative to imaging for identifying subclinical atherosclerosis, thereby supporting prevention strategies for cardiovascular disease.

Beyond heritability: Multimodal AI integrating imaging and genetics enables population scale precision coronary artery disease risk prediction

Subsession Time: Friday, October 17 at 1:40pm – 1:45pm

Authors: Devansh Pandey (Department of Integrative Biology, The University of Texas at Austin), Liaoyi Xu (Department of Integrative Biology, The University of Texas at Austin), Eucharist Kun (Department of Integrative Biology, The University of Texas at Austin), Joyce Y. Wang (Department of Integrative Biology, The University of Texas at Austin), Chenfei Li (Department of Integrative Biology, The University of Texas at Austin), Julie C. DiCarlo (Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin), Alaa Melek (Department of Biomedical Engineering, The University of Texas at Austin), Edward Castillo (Department of Biomedical Engineering, The University of Texas at Austin), Jagat Narula (University of Texas Health Sciences Center, Houston.), Charles A. Taylor (Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin), Vagheesh M. Narasimhan (Department of Integrative Biology, The University of Texas at Austin)

Abstract: Coronary artery disease (CAD) remains the leading cause of death worldwide, yet it is also highly amenable to treatment and prevention. Identifying asymptomatic individuals at highest risk remains a key gap in prevention. Polygenic risk scores (PRS) have recently emerged as powerful tools to stratify individuals by inherited risk; however, genetic prediction is fundamentally constrained by heritability and does not account for environmental or acquired factors. Non-invasive imaging captures environmental exposure and subclinical disease burden. To unify these complementary views of risk, we developed a multimodal AI model that integrates genome-wide PRS, metabolic biomarkers, demographics, with imaging to predict 10-year incident CAD. Using data from over 60,000 UK Biobank participants with ~4,000 CAD events after imaging, we fine-tuned foundational vision models on multiple views of cardiac, liver, and pancreas MRI, as well as DXA scans. Modality specific imaging embeddings were compressed using principal component analysis and integrated with an external multi ancestry PRS (trained on >1.5M individuals), metabolic and ECG traits, and baseline variables in a unified time to event model. In examining the imaging data we found that embeddings derived from imaging models outperformed hand-crafted image derived phenotypes (AUC: 0.794 vs. 0.666). Among imaging modalities, only cardiac long-axis and aortic distensibility MRI contributed substantial independent value; after adjusting for baseline traits - liver, pancreas, and DXA features added minimal predictive power. PRS alone explained pseudo R^2 of 0.08, while the full multimodal model reached 0.45, with imaging contributing nearly three times the incremental variance explained compared to genetics. Commonality analysis confirmed that genetic and imaging signals were largely orthogonal, though imaging captured some aspects of genetic risk. A hierarchical risk stratification framework combining clinical, genetic, and imaging data identified a subgroup with a 10-fold increased incident CAD risk relative to the low-risk baseline, and a 5-fold increase compared to individuals with high clinical and genetic risk alone. These findings show that while genetic risk provides a static baseline of inherited susceptibility, deep learning on clinical imaging adds dynamic markers of disease progression. Multimodal modeling thus enables a clinically practical framework for precision cardiovascular screening at population scale.

An Integrated Multi-Trait PRS and Proteomic Approach Identifies Biomarkers and Therapeutic Targets in Chronic Obstructive Pulmonary Disease

Subsession Time: Friday, October 17 at 1:45pm – 1:50pm

Authors: Chengyue Zhang (Brigham & Women's Hosp., Boston, MA), Iain R. Konigsberg (Univ. of Colorado Anschutz Medical Campus, Aurora, CO), Yixuan He (University of Texas Health, Houston, TX), Jingzhou Zhang (Brigham & Women's Hosp., Boston, MA), Tinashe Chikowore (Brigham & Women's Hosp., Boston, MA), William B. Feldman (Brigham and Women's Hosp., Boston, MA), Xiaowei Hu (University of Virginia, Charlottesville, VA), Yi Ding (Dana Farber Cancer Institute, Bostn, MA), Bogdan Pasaniuc (University of Pennsylvania, Philadelphia, PA), Diana Chang (Genentech, South San Francisco, CA), Qingwen Chen (Brigham & Women's Hosp., Boston, MA), Jessica A. Lasky-Su (Brigham & Women's Hosp., Boston, MA), Julian Hecker (Brigham & Women's Hosp., Boston, MA), Martin D. Tobin (Univ. of Leicester, Leicester, United Kingdom), Jing Chen (Univ. of Leicester, Leicester, United Kingdom), Sean Kalra (Brigham & Women's Hosp., Boston, MA), Katherine A. Pratte (National Jewish Health, Denver, CO), Hae Kyung Im (University of Chicago, Chicago, IL), Emily S. Wan (Veterans Affairs Boston Healthcare System, West Roxbury, MA), Ani Manichaikul (University of Virginia, Charlottesville, VA), Edwin K. Silverman (Brigham & Women's Hosp., Boston, MA), Russell P. Bowler (Cleveland Clinic Lerner School of Medicine, Cleveland, OH), Leslie A. Lange (University of Colorado Anschutz Medical Campus, Aurora, CO), Victor E. Ortega (Mayo Clinic, Scottsdale, AZ), Alicia R. Martin (Massachusetts General Hospital, Boston, MA), Michael H. Cho (Brigham & Women's Hosp., Boston, MA), Matthew R. Moll (Brigham & Women's Hosp., Boston, MA)

Abstract: Background: Multi-trait polygenic risk score (PRS) analyses improve prediction and reveal traits with shared genetic architecture, especially in heterogeneous diseases. However, they often fall short in identifying shared mechanisms and linking them to druggable targets, which is critical for drug repurposing. A two-step approach using elastic net models to identify key traits, followed by genetically predicted protein analysis validated with measured protein levels, can address this gap. We applied this approach to chronic obstructive pulmonary disease (COPD) and COPD exacerbations, major global causes of morbidity and mortality.

Method: We selected 25 traits comorbid with COPD based on literature and clinician input, then trained a multi-trait PRS (PRSmulti) using PRSmix+ to predict COPD in COPDGene non-Hispanic white participants (n=6,647). Associations of PRSmulti and individual PRSs with COPD status and exacerbations were tested using logistic and negative binomial regression, respectively. Validation was conducted in COPDGene African American (n=2,466), ECLIPSE (n=1,858), MassGeneral Brigham Biobank (n=15,152), and All of Us (n=118,566), adjusting for confounders and genetic similarity. Results were meta-analyzed. We next used GWAS summary statistics from traits contributing to PRSmulti and applied S-PrediXcan with pQTL-based protein expression models. Proteins with significantly predicted expression changes were further tested for association with COPD exacerbations using measured plasma protein levels in COPDGene SomaScan (n=5,173) and UK Biobank Olink (n=5,012). Druggable targets were identified using OpenTargets.

Results: PRSmix+ selected 7 PRSs (FEV1/FVC, FEV1, smoking, idiopathic pulmonary fibrosis, BMI, C-reactive protein, and a deep learning-based spirometry trait) to calculate PRSmulti. PRSmulti was associated with COPD (meta-analysis random effects [RE] OR 1.58, 95% CI: 1.28–1.94) and annualized exacerbations (RE β = 0.32, 95% CI: –0.16 to 0.81), with stronger effects in smoking-enriched cohorts. S-PrediXcan identified 73 proteins linked to these PRSs, validated by measured levels; 28 were associated with ≥2 traits, and 25 have clinical trials involving 46 drugs.

Conclusion: Combining elastic net-based multi-trait PRS modeling with genetically predicted protein expression, validated by measured protein levels, revealed shared disease mechanisms and identified promising biomarkers and druggable targets for COPD and COPD exacerbations.

Genetic architecture of fat and lean mass across body compartments in men and women

Subsession Time: Friday, October 17 at 1:50pm – 1:55pm

Authors: Alice Williamson (Computational Medicine Berlin Institute of Health at Charité-Universitätsmedizin Berlin, Friede Springer Cardiovascular Prevention Center at Charité Charité University Medicine Berlin, Precision Healthcare University Research Institute (PHURI) Queen Mary University of London, MRC Epidemiology Unit Institute of Metabolic Science University of Cambridge), on behalf of the Body Composition Genomics Consortium

Abstract: Genetic analyses of overall adiposity and fat distribution have primarily focused on simple measures of body size (e.g., body mass index; BMI) or derived indices of fat distribution (e.g., BMI-adjusted waist-hip ratio; WHR). However, these do not adequately reflect the full spectrum of body size and composition of fat and lean mass across body compartments. Dual X-ray absorptiometry scans (DEXA) enable such assessment but have not been performed at sufficient scale for GWAS analyses to date.

We examined the influence of rare and common genetic variation on fat and lean mass composition across body compartments using genome-wide association (GWAS) and whole exome sequencing (WES). We conducted fixed-effect genome-wide association study meta-analysis for 31 DEXA-derived regional fat and lean mass traits, including 74,554 individuals of European ancestry across 7 studies considering variants with a MAF > 0.01%. Sex-stratified GWAS analyses allowed identification of sex-dimorphic loci. Using predicted DEXA measures from bioimpedance in ~450,000 individuals in UK Biobank, we assessed the rare variant contribution to well predicted segmental body composition traits through WES analyses, including single variant and gene burden analyses.

In GWAS analyses we identified 679 independent signals (P<5x10-8) after fine mapping across traits. Of these, 384 were specific to fat mass compartments, with limited overlap with signals for commonly used measures such as BMI (8%) or BMI adjusted WHR (48%), despite their much larger sample sizes. We identified seven clusters of genetic variants associated with fat phenotypes including those reflective of metabolically unhealthy body composition e.g., high central and low peripheral fat. We identified 47 loci that exhibited differences in genetic effects between sexes on large body compartments, including

android and leg fat distribution. Deep molecular and clinical phenotyping identified subsets of loci with distinct metabolic effects on human health. Finally, rare variant gene burden analyses of predicted measures identified a total of 114 unique significantly associated genes which could be classified into 7 distinct phenotypic clusters.

This is the first large-scale study to identify the complex genetic architecture of both fat and lean mass compartments across the rare and common allelic spectrum. This study demonstrates the value in going beyond simple indices to investigate the mechanisms influencing body composition.

HLA Allele Frequencies and Phenome-Wide Associations in a Diverse Cohort from the All of Us Research Program

Subsession Time: Friday, October 17 at 1:55pm – 2:00pm

Authors: Kwangmi Ahn (NIH), Adam Burkholder (NIH), Tam C Tran (NIH), John S House (NIH), Joseph H Breeyear (NIH), Cristina Justice (NIH), Josh C Denny (NIH) and Alison A Motsinger-Reif (NIH)

Abstract: The human leukocyte antigen (HLA) region is highly polymorphic and plays key roles in immune response, infection, autoimmunity, and transplant compatibility. Its genetic complexity—marked by dense variation, strong linkage disequilibrium, and population-specific patterns—makes large-scale analysis challenging. Existing HLA reference data are mostly derived from European populations, limiting applicability to other ancestries. To address this gap, we characterized HLA variation and conducted ancestry-stratified phenome-wide association studies (PheWAS) in 267,742 participants with genetically inferred ancestries: 163,489 European (EUR), 53,237 African (AFR), and 51,016 Admixed American (AMR) in the All of Us Research Program, a diverse, population-scale cohort. HLA alleles were typed at the 8-digit resolution from whole genome sequencing data using Kourami, a graph-based method designed for accurate resolution of complex HLA regions. We identified 4,780 variants across 20 HLA genes. We found 1,425 alleles with significant frequency differences across ancestries (Bonferroni-adjusted), with the greatest diversity at the DQB1, DRB1, and DQA1 loci. For example, allele A*01:01:01G was more frequent in EUR (18.3%) than AFR (4.4%) or AMR (7.4%), while C*16:01:01G was enriched in AFR (9.5%), underscoring strong population stratification. Ancestry-stratified PheWAS using 114 common HLA alleles (with at least 0.1% allele frequency) and 3,430 phecodes identified multiple variants with robust associations to autoimmune and endocrine diseases. Notably, B*08:01:01G was linked to autoimmune disease and celiac disease in EUR and AMR groups, while DRB*103:01:01G was consistently associated with type 1 diabetes across all ancestries. We also identified ancestry-specific associations, particularly in EUR, involving variants such as A*02:01:01G linked to contact dermatitis and hypercholesterolemia, and A*01:01:01G associated with celiac disease—demonstrating that pan-ethnic analyses may miss key signals. Many HLA variants were pleiotropic: DRB1*03:01:01G and B0*8:01:01G were each linked to over 25 phenotypes, reinforcing the broad relevance of these alleles to complex disease. This study underscores the importance of ancestrally diverse populations in genetic research, especially for complex regions like HLA. Our findings advance equitable discovery and support improved health outcomes in underrepresented groups.

Defining the Molecular Effect of the Autoimmune-Risk Allele on Surface CD40 Expression in B Cells

Subsession Time: Friday, October 17 at 2:00pm – 2:05pm

Authors: Yoshihiko Tomofuji (Brigham and Women's hospital, Broad Institute, Harvard Medical School), Zepeng Mu (Brigham and Women's hospital, Broad Institute, Harvard Medical School), Cassidy Liu (Brigham and Women's hospital, Broad Institute, Harvard Medical School), Vidya Jayanthi (Brigham and Women's hospital, Broad Institute, Harvard Medical School), Hafsa Mire (Brigham and Women's hospital, Broad Institute, Harvard Medical School), Yu Zhao (Brigham and Women's hospital, Broad Institute, Harvard Medical School), Laurie Rumker (Brigham and Women's hospital, Broad Institute, Harvard Medical School), Soumya Raychaudhuri (Brigham and Women's hospital, Broad Institute, Harvard Medical School)

Abstract: GWAS has revealed the genetics of autoimmunity. While polygenic analyses emphasize T cells, they may miss key mechanisms involving other cell types. For example, CD40, a key B cell gene involved in memory B cell differentiation, proliferation, and class switching—and a therapeutic target (e.g., Bleselumab)—shows GWAS associations across autoimmune diseases (e.g., rheumatoid arthritis [RA], multiple sclerosis, and Grave's disease), aligning with the success of B cell-targeted therapies. Thus, functional dissection of individual loci is needed. Although protein QTL (pQTL) is a promising strategy, most pQTL studies rely on bulk tissues, lacking cell-type and cellular-compartment resolution. Also, QTL analysis alone cannot pinpoint causal variants or mechanisms. Even with CRISPR editing, heterogeneous editing outcomes reduce power, necessitating single-cell level genotyping to disentangle heterogeneity. To address these limitations, we combined (1) cell-type-resolved surface pQTL (spQTL) mapping with CITE-seq, and (2) CRISPR Prime Editing with CRAFT-seq, a single-cell multiomics method we previously developed that jointly profiles genotype, transcriptome, and surface proteins.

Using CITE-seq data of 375,094 PBMCs from 96 donors, we identified 23 spQTLs at 11 loci across 7 cell types (P<5×10^{-8}). A spQTL for CD40 on B cells (βlinear=0.63, P=1.6×10^{-15}) colocalized with autoimmune GWAS signals, including RA (coloc PP4=0.99), but showed no significant eQTL effect (βlinear=–0.01, P=0.88). Cell-state abundance QTL analysis showed that the RA-risk and CD40-increasing allele increased memory and decreased naïve B cell state abundance (P=0.02), could be consistent with the role of CD40 in memory B cell differentiation.

To define the function of the variant, we used Prime Editing to introduce the disease variant into a B cell line. Despite low editing efficiency (<25%), CRAFT-seq enabled direct evaluation of the genotype–mRNA/protein association at single-cell level while controlling for environmental factors (e.g., cytokines in culture medium). We found that rs1883832, located in the translation initiation enhancer, significantly altered CD40 surface protein level (βnb=0.13, P=1.8×10^{-8}), without affecting mRNA expression (βnb=–0.16, P=0.14), suggesting a causal effect potentially acting at the translation level.

In summary, our study showed that CITE-seq-based spQTL mapping and CRAFT-seq enables base-pair- and cell-type-resolved functional dissection of disease variants.

MSH3 is a modifier of repeat somatic expansion, somatic contraction and age-at-onset in X-linked dystonia parkinsonism (XDP)

Subsession Time: Friday, October 17 at 2:05pm – 2:10pm

Authors: Alan Mejia Maza (MGH-Harvard), Madison Hincher (MGH), Kevin Correia (MGH), Tammy Gillis (MGH), Ayumi Nishiyama (MGH), Ellen B. Penney (MGH-Harvard), Aloysius Domingo (MGH-Harvard), Rachita Yadav (MGH-Harvard), Nutan Sharma (MGH-Harvard), D. Cristopher Bragg (MGH-Harvard), Michael E. Talkowski (MGH-Harvard), Marcy E. MacDonald (MGH-Harvard), Jong-Min Lee (MGH-Harvard), Laurie J. Ozelius (MGH-Harvard), Vanessa Chantal Wheeler (MGH-Harvard)

Abstract: X-linked dystonia parkinsonism (XDP) is a rare neurodegenerative disease caused by the insertion of a SINE-VNTR-Alu (SVA) retrotransposon in the TATA-box binding protein associated factor-1 (TAF1) gene. Within the SVA, the length of the polymorphic CCCTCT repeat inversely correlates with XDP age-at-onset (AAO), placing XDP among a larger class of neurodegenerative diseases caused by expanded microsatellite repeat tracts. We have previously shown that the XDP CCCTCT repeat exhibits somatic expansion that is tissue-specific and repeat length-dependent, consistent with somatic expansion as a mechanism underlying the CCCTCT length-dependent AAO. Here, we aimed to identify genetic modifiers of CCCTCT instability and to provide a mechanistic link between somatic instability and disease. We quantified metrics of both repeat expansion and contraction in blood and in postmortem brain tissues from XDP males. Using exome-wide association analyses (ExWAS), complemented with directed Miseq sequencing of a coding MSH3 repeat variant, we show that common genetic variation in MSH3 associates with blood somatic expansion, somatic contraction and AAO. MSH3 variants associated with less expansion or more contraction are associated with a later AAO, and vice versa. We further identify rare MSH3 coding variants associated with low expansion. Significantly, MSH3 repeat alleles also modify expansion and contraction in brain tissues, paralleling modifier effects in the blood. This study provides the first direct evidence that MSH3 is a modifier of both repeat expansion and contraction in XDP, providing key evidence that MSH3's role(s) in CCCTCT repeat instability underlies its impact on disease. Our findings strongly indicate that therapeutic strategies targeting MSH3, predicted to both slow CCCTCT expansion and promote CCCTCT contraction, will have a disease-modifying effect.

Targeting glutamatergic pathways: genetic insights into comorbid neurodevelopmental disorders

Subsession Time: Friday, October 17 at 2:10pm – 2:15pm

Authors: Joseph T. Glessner (CHOP), Munir E. Khan (CHOP), Xiao Chang (CHOP), Yichuan Liu (CHOP), Shahram Torkamandi (CHOP), Deborah Abrams (CHOP), F. George Otieno (CHOP), Jihoon Kim (CHOP), Yeshwanth Mahesh (CHOP), Maria Lemma (CHOP), Frank Mentch (CHOP), Jin Li (Tianjin Medical University), Charlly Kao (CHOP), Michael E. March (CHOP), Huiqi Qu (CHOP), John Connolly (CHOP), Hakon Hakonarson (CHOP)

Abstract: Copy number variants (CNVs) affecting metabotropic glutamate receptor (GRM) genes have been previously linked to neurodevelopmental disorders (NDDs), but the broader role of GRM-centered

networks in attention-deficit/hyperactivity disorder (ADHD), particularly in comorbid presentations, remains underexplored. Here, we investigated CNV burden in GRM genes and their molecular interaction networks in a large, ancestrally diverse pediatric cohort enriched for ADHD and comorbid NDD phenotypes.

We analyzed high-density genotyping data from 72,726 participants in the Center for Applied Genomics at Children's Hospital of Philadelphia, including 12,472 individuals diagnosed with ADHD, of whom 7,967 had at least one additional NDD diagnosis. CNVs were detected using PennCNV and annotated for overlap with GRM genes (GRM1, GRM5, GRM7, GRM8) and their protein-protein interaction (PPI) networks curated from STRING-db. MultiOMICS integration was performed, leveraging transcriptomic and proteomic annotations alongside whole genome sequencing (WGS) data to validate CNV calls and functional context. Artificial Intelligence (AI)-based pipelines enabled automated access and curation of raw data, followed by rigorous normalization and phenotype matching.

Significant CNV enrichment (FDR < 0.05) was observed in 27 GRM-interacting genes among ADHD cases with comorbid NDDs, but not in ADHD-only individuals. These genes include NRXN1, DLG2, SHANK3, and SYNGAP1, key regulators of glutamatergic signaling and neurodevelopment. Cross validation using independent ADHD subsets confirmed CNV enrichment patterns in GRM network genes, emphasizing reproducibility.

Our findings reveal a distinct CNV burden signature in ADHD with comorbid NDDs, implicating GRM interaction networks as a key molecular substrate underlying phenotypic complexity. These results support the use of AI-enhanced MultiOMICS tools for discovery of precision medicine targets in genetically stratified NDD populations.

Uncovering Ancestry-Specific Genetic Architecture in Admixed Individuals of African Descent with a Scalable Tractor GWAS Workflow

Subsession Time: Friday, October 17 at 2:15pm – 2:20pm

Authors: Nirav N. Shah (Department of Molecular and Human Genetics, Baylor College of Medicine), Adam X. Maihofer (Department of Psychiatry, School of Medicine, University of California San Diego), Taotao Tan (Department of Molecular and Human Genetics, Baylor College of Medicine; Department of Integrative Physiology, Baylor College of Medicine), Jessica Honorato-Mauer (Department of Molecular and Human Genetics, Baylor College of Medicine), Helen Lin (Department of Molecular and Human Genetics, Baylor College of Medicine), Clement C. Zai (Tanenbaum Centre for Pharmacogenetics, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health; Department of Psychiatry, University of Toronto), PGC-PTSD ancestry working group, Konrad Karczewski (Program in Medical and Population Genetics, Broad Institute of MIT and Harvard; Analytic and Translational Genetics Unit, Massachusetts General Hospital; Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard), Marcos Santoro (Disciplina de Biologia Molecular, Departamento de Bioquímica, Universidade Federal de São Paulo), Caroline M. Nievergelt (Department of Psychiatry, School of Medicine, University of California San Diego), Elizabeth G. Atkinson (Department of Molecular and Human Genetics, Baylor College of Medicine)

Abstract: Genome-wide association studies (GWAS) have often excluded admixed individuals due to concerns over spurious associations from population structure, reducing the generalizability of findings, as seen in the poor accuracy of polygenic risk scores Tractor GWAS addresses this by using local ancestries in its model, offering well-controlled type I error, increased power to detect and generation of ancestry-specific summary statistics. However, it requires three key upstream steps: phasing, local ancestry inference, and ancestry-aware regression, making its implementation complex.

We present a scalable, modular, and portable Nextflow-based Tractor GWAS workflow to simplify and standardize ancestry-aware GWAS. To demonstrate its applicability, we applied the workflow to 3,568 traits in 6,245 UK Biobank participants of African admixed ancestry, requiring ~11,400 CPU-hours. Our large-scale implementation not only demonstrated the scalability, but also replicated canonical loci for key traits, and uncovered novel ancestry-specific associations not previously reported.

For instance, in analyzing Apolipoprotein B (APOB) biomarker levels, we identified a signal in APOE, a key locus, shared across African (AFR) and European (EUR) ancestral haplotypes with nearly identical effect sizes. In contrast, we identified variants in PCSK9 and PMFBP, both previously linked with LDL and total cholesterol but not APOB, to be associated with APOB levels in AFR but not in EUR. This was due to the lead SNPs' absence in European haplotypes, highlighting Tractor's ability to detect population-specific variants that conventional GWAS would miss, thus helping us better characterize the genetic architecture of diseases.

A key question in the field has been the degree of similarity of effect sizes across ancestries. To answer this, we curated ~700 largely independent phenotypes from the original 3,568 spanning EHR trait types. We identified shared and unique significant hits across ancestries and evaluated the concordance in ancestry specific effect sizes for lead SNPs, finding widespread effect size heterogeneity. This suggests that gene discovery methods deriving power from marginal effect size heterogeneity are likely to benefit at many loci genome-wide, even when loci underlying the traits have similar causal effects.

This modular Nextflow workflow democratizes Tractor, enabling researchers to harness admixed genomes at scale and illuminate ancestry-enriched biology missed by standard GWAS

Improved glaucoma polygenic risk score enables clinical utility and risk prediction across major ancestries

Subsession Time: Friday, October 17 at 2:20pm – 2:25pm

Authors: Stuart MacGregor (QIMR Berghofer), Guiyan Ni (QIMR Berghofer, Seonix Bio), Mathias Seviiri (QIMR Berghofer, Seonix Bio), Antonia Kolovos (Flinders University), Ngoc Quynh Le (QIMR Berghofer), Puya Gharahkhani (QIMR Berghofer), Priyanka Nandakumar (23andMe), 23andMe Research Team, Alex W Hewitt (University of Tasmania), Nick Haan (Seonix Bio), Owen Siggs (Garvan Institute), Jamie E Craig (Flinders University).

Abstract: Glaucoma is one of the most heritable complex diseases and is the leading cause of irreversible blindness. Glaucoma is characterised by progressive vision loss, with the disease largely preventable through timely treatment. However, early detection is difficult and 50% of patients remain undiagnosed.

There is an urgent need to develop better risk assessment tools to personalise screening and improve treatment strategies.

In collaboration with 23andMe, Inc., we collated genome-wide association study data on 296,757 open angle glaucoma (OAG) cases and 6,240,939 controls, drawn from several major ancestry groups. We supplemented these with glaucoma related traits (eye pressure, optic disc parameters, ocular hypertension; N=320,296) and applied a multitrait model, enabling construction of a glaucoma polygenic risk score (PRS). We evaluated PRS performance for glaucoma risk in independent case-control samples from each ancestry group. We then evaluated the ability of the PRS to predict a range of clinical outcomes.

Our novel PRS showed marked improvements in OAG risk prediction performance in held out testing sets. In the Australian ANZRAG European ancestry data those in the top 10% PRS had 10 fold increased risk (OR=10.0) relative to the remainder (compared to 4.2 for our previously published PRS). Similar performance was seen in the US NEIGHBOR data (OR=9.1 for top 10% versus remainder in European ancestry individuals; OR 7.5 in Latinos). In other ancestry groups, the novel PRS had similar performance to that seen in Europeans using our previous PRS (OR =4.0 and OR=3.8 for top 10% versus the remainder in South Asians and African Americans, respectively). Our PRS effectively stratified an Australian cohort for age at onset with those in the top 10% PRS developing glaucoma 25 years earlier than those in the bottom 10%. The PRS also predicted the need for treatment escalation in early glaucoma, both prevalent and incident surgery, and the probability of OAG in first-degree relatives.

Our PRS efficiently identifies individuals at very high-risk of developing glaucoma as well as accelerated disease progression and requirement for trabeculectomy. Risk profiling with the enhanced PRS will enable earlier screening and timely treatment of high-risk individuals, with reduced screening and monitoring costs in those at low risk.

Session 78: From Bedside to Bench: Delineating Novel Genomic Etiologies of Mendelian disorders

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

Session Time: Friday, October 17 at 1:30pm – 2:30pm

KAT14 Deficiency as a Novel Genetic Cause of Diamond-Blackfan Anemia

Subsession Time: Friday, October 17 at 1:35pm – 1:40pm

Authors: Shiyu Luo (University of Miami Miller School of Medicine and Holtz Children's Hospital, Boston Children's Hospital and Harvard Medical School), Haoyue Sheng (University of Miami Miller School of Medicine and Holtz Children's Hospital), Anjali Pawar (UC Davis Health), Klaus Schmitz-Abe (University of Miami Miller School of Medicine and Holtz Children's Hospital, Boston Children's Hospital and Harvard Medical School), Qifei Li (University of Miami Miller School of Medicine and Holtz Children's Hospital, Boston Children's Hospital and Harvard Medical School), Lamrot Tulu (University of California Davis), Madesh Ramesh (University of Miami Miller School of Medicine and Holtz Children's Hospital), Victtoria Ardellini (University of Miami Miller School of Medicine and Holtz Children's Hospital), Staci D. Arnold (Aflac Cancer and Blood Disorders Center at Children's Healthcare of Atlanta, Emory University), Bojana B. Pencheva (Aflac Cancer and Blood Disorders Center at Children's Healthcare of Atlanta, Emory University), Casie A. Genetti (Boston Children's Hospital and Harvard Medical School), Alan H. Beggs (Boston Children's Hospital and Harvard Medical School), Suma Shankar (University of California Davis), Pankaj B. Agrawal (University of Miami Miller School of Medicine and Holtz Children's Hospital, Boston Children's Hospital and Harvard Medical School)

Abstract: Diamond-Blackfan anemia (DBA) is a rare congenital disorder characterized by pure red cell aplasia, macrocytic anemia, and variable congenital malformations, with approximately 30% of affected individuals exhibiting growth deficiency. While most DBA cases are caused by heterozygous variants in ribosomal protein genes, 30–40% remain unexplained. Here, we report the genetic and mouse modeling findings on a promising DBA candidate gene encoding lysine acetyltransferase 14 (KAT14), also known as cysteine-rich protein 2-binding protein (CSRP2BP). KAT14 is ubiquitously expressed in human tissues and functions as a component of the ADA2A-containing complex, which possesses histone acetyltransferase activity. In this study, we identified four individuals from three unrelated families carrying rare recessive variants of KAT14, all of whom presented with DBA. The siblings from the same family were identified as compound heterozygotes for a novel missense (NM_001392073.1: c.1783C>T, p.R595W) and a nonsense (c.1696C>T, p.R566X) variant. The third and fourth patients from different families were homozygous for the R595W variant and a 2-bp nucleotide deletion (c.198_199delGT, p.Ser67LeufsTer11), resulting in a frameshift and premature stop codon, respectively. The Kat14 hypomorphic (Kat14tm1a) mice were also characterized using the International Mouse Phenotyping Consortium (IMPC) pipeline. Homozygous mutants exhibit extensive multisystem abnormalities, including severe ocular defects, significant growth retardation, craniofacial and skeletal abnormalities. Notably, these mice displayed decreased hemoglobin content and variable degrees of anemia, mirroring the key hematopoietic

features of DBA. This hypomorph model is associated with incomplete penetrance and pre-weaning lethality in different genetic backgrounds, underscoring the essential role of Kat14 in development and survival. Collectively, these findings identify KAT14 as a critical regulator of mammalian development and hematopoiesis, supporting its candidacy as a novel genetic contributor to DBA. The identification of rare, recessive KAT14 variants in multiple DBA patients, together with the recapitulation of multi-system, growth, and hematologic phenotypes in the Kat14 mouse model, provides compelling evidence for KAT14's role in DBA pathogenesis and establishes a valuable platform for further mechanistic and therapeutic studies.

Detection of Allele-Specific DNA Methylation of X chromosome in a Female Patient With a Truncating MED12 Variant Using Long-Read Sequencing

Subsession Time: Friday, October 17 at 1:40pm – 1:45pm

Authors: China Nagano (Department of Pediatrics, Kobe University Graduate School of Medicine), Naoya Morisada (Department of Clinical Genetics, Hyogo Prefectural Kobe Children's Hospital), Syuhei Aoyama (Department of Pediatrics, Kobe University Graduate School of Medicine), Yuka Kimura (Department of Pediatrics, Kobe University Graduate School of Medicine), Yuta Inoki (Department of Pediatrics, Kobe University Graduate School of Medicine), Nana Sakakibara (Department of Pediatrics, Kobe University Graduate School of Medicine), Tomoko Horinouchi (Department of Pediatrics, Kobe University Graduate School of Medicine), Tomohiko Yamamura (Department of Pediatrics, Kobe University Graduate School of Medicine), Shingo Ishimori (Department of Pediatrics, Kobe University Graduate School of Medicine), Kandai Nozu (Department of Pediatrics, Kobe University Graduate School of Medicine)

Abstract: Background

Hardikar syndrome (OMIM:301068) is a rare X-linked dominant disorder with multiple congenital anomalies, including cleft lip and/or palate, pigmentary retinopathy, obstructive liver disease, urogenital malformations, intestinal malrotation, and cardiovascular defects. It is caused by truncating variants in the MED12 gene on the X chromosome, and due to presumed male lethality, nearly all reported cases are de novo in females. While unbalanced X-chromosome inactivation (XCI) is considered a modifier of disease expression in females, the role of allele-specific DNA methylation (ASM) in MED12 remains unclear. Although MED12 is not classified as an XCI escape gene in genome-wide datasets, activation from variant allele could contribute to the phenotype. We examined whether ASM might influence expression in a mother–daughter pair carrying the same pathogenic MED12 variant.

Methods

Peripheral blood DNA was analyzed using nanopore long-read sequencing. Variant calling was performed with Longshot, phasing with WhatsHap, and allele-specific methylation assessed using modkit. Due to limited heterozygous SNPs, phasing and ASM analysis in MED12 were only possible in the mother.

Results

To infer XCI direction, we analyzed methylation at the RP2 gene, which undergoes complete X-inactivation. Results showed extremely skewed XCI in the mother, with preferential inactivation of the MED12 variant-carrying X. Despite this, haplotype-resolved methylation analysis of the MED12 promoter showed consistent allele-specific differences: the variant allele had lower methylation (~39%) than the reference allele (~61%). This suggests that the variant allele, although on the inactive X, may escape silencing through local hypomethylation. In the daughter, allele-specific methylation could not be evaluated due to insufficient SNPs.

Conclusion

This case suggests that hypomethylation of the pathogenic MED12 allele may allow expression from the inactive X, contributing to disease phenotypes. Long-read sequencing enabled the detection of ASM and may uncover epigenetic mechanisms in X-linked disorders.

Saturation genome editing of the non-coding snRNA RNU4-2 reveals distinct dominant and recessive neurodevelopmental disorders

Subsession Time: Friday, October 17 at 1:45pm – 1:50pm

Authors: Nicola Whiffin (Big Data Institute, University of Oxford; Centre for Human Genetics, University of Oxford), Joachim De Jonghe (The Francis Crick Institute, London, UK), Rocio Rius (Centre for Population Genomics, Garvan Institute of Medical Research and Murdoch Children's Research Institute, Australia), Alexander JM Blakes (School of Biological Sciences, University of Manchester), Hyung Chul Kim (Big Data Institute, University of Oxford; Centre for Human Genetics, University of Oxford), Ayanfeoluwa Adedeji (The Francis Crick Institute, London, UK), Elsa Leitão (Institute of Human Genetics, University Hospital Essen), Ruebena Dawes (Big Data Institute, University of Oxford; Centre for Human Genetics, University of Oxford), Yuyang Chen (Big Data Institute, University of Oxford; Centre for Human Genetics, University of Oxford), RNU4-2 Recessive Consortium, Siddharth Banka (School of Biological Sciences, University of Manchester), Julien Thevenon (Université Grenoble Alpes, France), Rodrigo Mendez (Cardiovascular Medicine, Stanford University), Daniel G MacArthur (Centre for Population Genomics, Garvan Institute of Medical Research and Murdoch Children's Research Institute, Australia), Cas Simons (Centre for Population Genomics, Garvan Institute of Medical Research and Murdoch Children's Research Institute, Australia), Christel Depienne (Institute of Human Genetics, University Hospital Essen), Caroline Nava (Laboratoire SeqOIA, Paris, France), and Gregory M Findlay (The Francis Crick Institute, London, UK)

Abstract: Background: De novo variants in RNU4-2 were recently shown to cause ReNU syndrome, a highly prevalent syndromic neurodevelopmental disorder (NDD) predicted to impact ~100,000 individuals globally. RNU4-2 encodes the U4 small nuclear RNA (snRNA) component of the major spliceosome. ReNU syndrome variants disrupt spliceosome function and alter 5’ splice-site selection. The precise relationship between RNU4-2 variation and clinical impact has yet to be comprehensively characterised.

Methods: We performed saturation genome editing (SGE) of RNU4-2 in HAP1 cells, experimentally assaying all possible single nucleotide variants across the 145 nucleotide transcript, and single nucleotide insertions and deletions in the 18 nucleotide ReNU syndrome critical region (CR). For each variant, we calculate a function score, which reflects the viability of cells with the specific variant.

Results: SGE function scores discriminate ReNU syndrome variants from those observed in the population (AUC = 0.95), dramatically outperforming in silico variant effect prediction tools (CADD AUC = 0.65). Using these data we redefine the ReNU syndrome CR at single-nucleotide resolution to two regions of nine and four nucleotides, resolve variant pathogenicity for variants of uncertain significance, and show that SGE function scores delineate variants by phenotypic severity. For example, 93% of individuals with strongly depleted SGE variants were non-verbal or spoke only a few words, compared to 5.6% with moderately depleted variants.

Finally, we identify three structural regions of RNU4-2 outside of the ReNU CR harbouring variants with significant SGE function scores. We show that biallelic variants in these regions, which are important regions for binding to spliceosomal proteins, cause a novel recessive NDD. We characterise a cohort of 38 recessive patients, revealing a striking MRI phenotype that is distinct from ReNU syndrome marked by microcystic lesions in the periventricular white matter and cerebellar atrophy.

Conclusion: We perform the first SGE of a human non-coding RNA to define the landscape of variant function across RNU4-2, providing critical insights for ReNU syndrome diagnosis. The resulting data have immediate clinical utility. In addition, we identify a novel recessive RNU4-2-associated NDD, with distinct features from dominant ReNU syndrome. Collectively, this work demonstrates the power of SGE to dissect newly identified disease loci.

Transcriptomic signatures in postmortem brains uncover tau-related neurotoxicity as a novel disease mechanism in X-linked dystonia-parkinsonism

Subsession Time: Friday, October 17 at 1:50pm – 1:55pm

Authors: Aloysius Domingo (Massachusetts General Hospital, Harvard Medical School, Broad Institute), Charles Jourdan Reyes (Massachusetts General Hospital, Harvard Medical School), Ellen B. Penney (Massachusetts General Hospital, Harvard Medical School), Ean Norenberg (Massachusetts General Hospital), Justin Han (Massachusetts General Hospital), Micaela G. Murcar (Massachusetts General Hospital), Nicolas A. Bravo (Massachusetts General Hospital), Hoang-Dai Tran (Massachusetts General Hospital), Rachita Yadav (Massachusetts General Hospital, Harvard Medical School, Broad Institute), Dadi Gao (Massachusetts General Hospital, Harvard Medical School, Broad Institute), Sid Reed (Massachusetts General Hospital, Broad Institute), Serkan Erdin (Massachusetts General Hospital, Broad Institute), Nandini Ramesh (Massachusetts General Hospital), Benjamin Wymann (Massachusetts General Hospital, Harvard Medical School), Aaron Held (Massachusetts General Hospital, Harvard Medical School), Ranee Zara Monsanto (Massachusetts General Hospital, Harvard Medical School), Laura Moran (Massachusetts General Hospital, Harvard Medical School), Hayden Wheeler (Massachusetts General Hospital, Harvard Medical School), Annie Ruan (Massachusetts General Hospital, Harvard Medical School), Grant Griesman (Massachusetts General Hospital, Harvard Medical School), Chao-Zong Lee (Massachusetts General

Hospital, Harvard Medical School), Gelbert Crescencio (Massachusetts General Hospital), Matthew Nolan (Massachusetts General Hospital, Harvard Medical School), John Lemanski (Massachusetts General Hospital), Kathleen O'Keefe (Massachusetts General Hospital), Bimal Jana (Massachusetts General Hospital), Christine Vaine (Massachusetts General Hospital), Cara Fernandez-Cerado (Sunshine Care Foundation), M. Salvie Velasco-Andrada (Sunshine Care Foundation), Gierold Paul A. Legarda (Sunshine Care Foundation), Niecy Grace Ganza-Bautista (Sunshine Care Foundation), Michelle Sy (Sunshine Care Foundation), Edwin L. Muñoz (University of the Philippines Manila), Mark Angelo C. Ang (University of the Philippines Manila), Cid Czarina E. Diesta (Makati Medical Center), Brian J. Wainger (Massachusetts General Hospital, Harvard Medical School, Harvard Medical School), Bradley T. Hyman (Massachusetts General Hospital, Harvard Medical School), Nutan Sharma (Massachusetts General Hospital, Harvard Medical School), Laurie J. Ozelius (Massachusetts General Hospital, Harvard Medical School), D. Cristopher Bragg (Massachusetts General Hospital, Harvard Medical School), Clotilde Lagier-Tourenne (Massachusetts General Hospital, Harvard Medical School), Michael E. Talkowski (Massachusetts General Hospital, Harvard Medical School, Broad Institute)

Abstract: The discovery of causal mechanisms remains a major fundamental challenge in rare diseases. In some instances, identifying shared molecular consequences with established disorders can surface novel mechanisms and prioritize functional targets. We applied this strategy to X-linked dystonia-parkinsonism (XDP), a progressive neurodegenerative disease that is rare worldwide but prevalent on one island in the Philippines. We previously discovered the cause of XDP as an SVA insertion into TAF1 intron 32. This mutation contains an expanded hexameric repeat that varies with age-at-onset and is accompanied by local intron retention and splicing defects in patient-derived cell lines. To interrogate disease signatures in tissue, we collected postmortem brains from 45 affected individuals and 10 matched controls and performed RNASeq and functional genomic profiling. Differential gene expression analysis identified neuronal loss, synaptic dysfunction, and neuroinflammation as prominent transcriptomic hallmarks, particularly in disease-involved brain regions such as the striatum, prefrontal cortex, and cerebellum. Further, we found upregulation and altered splicing of MAPT, widespread dysregulation of genes related to tau aggregation, and striking overlaps with known signatures of neurotoxicity in neurodegenerative diseases characterized by misfolded tau. Consistent with these findings, we observed inclusions containing hyperphosphorylated tau in the striatum and prefrontal cortex via targeted immunohistochemical staining. Oligomeric tau burden was associated with SVA hexameric repeat length, and pathologic four-repeat tau isoforms were overexpressed in astroglia. Linking pathology to the molecular defects in TAF1 and to global transcriptomic disruption, we recovered gene modules containing dysregulated genes and pathways that correlated with increased tau levels and intron retention. Further, protein aggregation was recapitulated in XDP neural cell lines, which showed elevated p62 levels upon autophagic challenge, a phenotype rescued by CRISPR editing of the disease-causing SVA. These convergent results establish a previously undescribed role for tau accumulation and neuroinflammation as core disease mechanisms in XDP, accelerating critically needed biomarker and therapeutic studies for this devastating disease. Further, we exemplify a blueprint to decipher novel driver pathologies through integrative functional genomic approaches in tissue.

Deficiency of the cationic amino acid transporter SLC7A1 causes spastic cerebral palsy & hereditary spastic paraplegia

Subsession Time: Friday, October 17 at 1:55pm – 2:00pm

Authors: Daniel G. Calame (Baylor College of Medicine), Mira Gandhi (Baylor College of Medicine), Jesse M. Levine (Baylor College of Medicine), Le Anh Nhung (National University of Singapore), Somayeh Bakhtiari (University of Arizona), Saeed Abdulla Boholega (King Faisal Specialist Hospital and Research Centre), Michael C. Kruer (University of Arizona), Davut Pehlivan (Baylor College of Medicine), Anaïs Begemann (University of Zurich), Katharina Steindl (University of Zurich), Ivan Ivanovski (University of Zurich), Anita Rauch (University of Zurich), Ravella Raker (University of Pittsburgh), Damara Ortiz (University of Pittsburgh), Namik Kaya (King Faisal Specialist Hospital and Research Centre), Hanan AlQudairy (King Faisal Specialist Hospital and Research Centre), Mohammed AlMuhaizea (King Faisal Specialist Hospital and Research Centre), Matteo Cagiada (University of Copenhagen), Shalini Jhangiani (Baylor College of Medicine), Long Nam Nguyen (National University of Singapore), Ehsan Ghayoor Karimiani (University College London), Maha Zaki (National Research Centre, Cairo), Reza Maroofian (University College London), Jennifer E. Posey (Baylor College of Medicine), Richard A. Gibbs (Baylor College of Medicine), James R. Lupski (Baylor College of Medicine)

Abstract: Amino acid transport across lipid bilayers is essential for cell metabolism and is mediated by solute carrier (SLC) proteins. The SLC gene SLC7A1 encodes the endothelial cationic amino acid transporter CAT1. SLC7A1/CAT1 is a high affinity arginine transporter recently found to regulate mitochondrial translation. The mouse ortholog Slc7a1 is an essential gene; Slc7a1 knockout (KO) causes perinatal lethality. Here, we provide evidence that SLC7A1 variants cause autosomal recessive spastic cerebral palsy (CP) & hereditary spastic paraplegia (HSP).

Through family-based analysis of genomic data, we characterize 13 individuals from 9 unrelated families with biallelic, ultra-rare, damaging SLC7A1 variants: p.(Arg588Gln), p.(Arg588Trp), p.(Val291Leu), p.(Val291Met), p.(Tyr301Cys), p.(Tyr420*), p. (Arg431Trp), and p.(Ser421Leu). The recurrent SLC7A1 variant p.(Arg588Gln) was identified in four unrelated families of diverse ancestries. All missense variants fall within conserved transmembrane domains and involve amino acids predicted structurally critical or functionally relevant. Phenotypes range from childhood-onset HSP to severe spastic-dystonic neurodevelopmental disorders. Other recurring features include epilepsy, ataxia, and tremor. Brain MRI was either normal or showed cerebral and cerebellar atrophy, ventriculomegaly, and white matter abnormalities. Plasma amino acids, lactate, ammonia, and urine organic acids were normal.

To better understand the cellular consequences of SLC7A1 dysfunction, HEK293T wild-type (WT) and SLC7A1 KO cells were investigated. Metabolomic profiling demonstrated an eight-fold reduction in intracellular cystine, glutathione and taurine metabolism perturbations, and no difference in arginine levels in KO versus WT cells. Other abnormalities included increased mitochondrial mass, decreased cystine uptake, and increased mTOR activation in KO cells. As in prior studies of Slc7a1 KO mouse embryonic fibroblasts, the rate of arginine uptake was similar between SLC7A1 KO and WT HEK293T cells. The expression of paralog SLC7A3/CAT3 was significantly increased in SLC7A1 KO HEK293T cells

which may compensate for SLC7A1 deficiency. Patient fibroblasts also had increased SLC7A3 expression relative to healthy control fibroblasts.

Our data implicate SLC7A1 deficiency as a cause of autosomal recessive spastic CP & HSP. Further studies will clarify the mechanisms by which SLC7A1 deficiency causes an array of neurological phenotypes.

Biallelic loss-of-function variants in the FUZ gene cause autosomal recessive skeletal ciliopathy

Subsession Time: Friday, October 17 at 2:00pm – 2:05pm

Authors: Yiyao Chen (International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University), Yi Wu (International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University), Guoqiang Li (International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University), Xu Han (International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University), Yan Xu (International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University), Hongjun Fei (International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University), Li Gao (International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University), Lei Qin (International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University), Renyi Hua (International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University), Jian Wang (International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University), Niu Li (International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University), Shuyuan Li (International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University)

Abstract: Background: Skeletal ciliopathies represent a distinct subgroup of ciliopathies. Pathogenic variants in genes encoding components of the Ciliogenesis and Planar Polarity Effector (CPLANE) complex lead to a wide variety of skeletal dysplasia with impaired ciliary functions. FUZ, a key component of the CPLANE complex, plays a crucial role in intraflagellar vesicular trafficking within primary cilia. Although the OMIM database currently lists FUZ as a susceptibility gene for neural tube defects (NTDs) with presumed autosomal dominant inheritance, biallelic FUZ variants have been identified in individuals with a range of recessive phenotypes, including short-rib polydactyly syndrome, craniosynostosis, and orofaciodigital syndrome. However, disease nomenclature remains inconsistent, and the clinical validity of the FUZ-disease association has not been systematically evaluated.

Methods: Trio-whole exome sequencing (trio-WES) was performed on the proband fetus and both parents, with identified variants confirmed by Sanger sequencing. RNA assays were conducted on the splicing variant carrier (the mother) and a healthy control. In addition, previously published evidence was integrated to evaluate the clinical validity of FUZ-related autosomal recessive skeletal ciliopathy using the ClinGen gene-disease validity framework.

Results: We report a prenatal case involving biallelic loss-of-function variants in the FUZ gene, c.20delG (p.Gly7fs*54) and c.233+1G>A (p.?). The proband presented with polydactyly, cardiac defects, and nasal bone hypoplasia. RNA analysis confirmed that the c.233+1G>A variant disrupted normal splicing, resulted in a transcript with premature termination codons. By integrating genetic and experimental data from this case and previously reports, we evaluated the clinical validity of FUZ as a causative gene for autosomal recessive skeletal ciliopathy as “strong” according to the ClinGen framework.

Conclusion: Our findings expand the prenatal phenotypic spectrum of FUZ-related disorders and provide robust genetic evidence supporting its role in autosomal recessive skeletal ciliopathy.

Gene knockouts across 138,000 individuals for novel rare disease gene discovery

Subsession Time: Friday, October 17 at 2:05pm – 2:10pm

Authors: Anthony EF McGuigan (Centre for Human Genetics, NIHR Oxford Biomedical Research Centre, University of Oxford, Big Data Institute), Alistair Pagnamenta (Department of Clinical and Biomedical Sciences, University of Exeter, Centre for Human Genetics, NIHR Oxford Biomedical Research Centre, University of Oxford), Carme Camps (Centre for Human Genetics, NIHR Oxford Biomedical Research Centre, University of Oxford), Siying Lin (NIHR Biomedical Research Centre, Moorfields Eye Hospital and the UCL Institute of Ophthalmology, Manchester Centre for Genomic Medicine, Saint Mary’s Hospital, Manchester University NHS Foundation Trust), Gavin Arno (NIHR Biomedical Research Centre, Moorfields Eye Hospital and the UCL Institute of Ophthalmology, Division of Research, Greenwood Genetic Center), Michel Michaelides (NIHR Biomedical Research Centre, Moorfields Eye Hospital and the UCL Institute of Ophthalmology), Andrew Webster (NIHR Biomedical Research Centre, Moorfields Eye Hospital and the UCL Institute of Ophthalmology), Louise C Wilson (Clinical Genetics, North East Thames Regional Genetics Service, Great Ormond Street Hospital for Children NHS Foundation Trust), Alexander Broomfield (Metabolic Department, Great Ormond Street Hospital), Jenny C Taylor (Centre for Human Genetics, NIHR Oxford Biomedical Research Centre, University of Oxford), Nicola Whiffin (Centre for Human Genetics, NIHR Oxford Biomedical Research Centre, University of Oxford, Big Data Institute)

Abstract: Background:

Structural variants (SVs) can disrupt gene function and contribute to pathogenesis of rare disorders. However, identifying disease-relevant variants is challenging due to the high rate of false positives generated during short-read SV calling. Here, we leverage the distinct read-depth signal associated with homozygous deletions to create a high-confidence knockout dataset and use this to identify novel diagnoses, identify new disease genes, and quantify tolerance to homozygous deletion genome-wide.

Material and Methods:

We analysed homozygous deletions across 138,136 individuals in the UK National Genomic Research Library (54,932 with rare disease). A read-depth threshold was applied to identify a high-confidence subset of variants from Manta/Canvas/DRAGEN calls.

Results:

We identified 1,309,894 high-confidence homozygous deletion SVs, averaging nine per participant. 5% of the genome was covered by rare deletions, with a higher proportion of the X chromosome tolerating deletion than autosomes (28.9% vs 3.3%). 15,259 unique SVs were identified solely in affected participants (n=18,906). Of these, 150 cause coding truncations of PanelApp genes fitting the individual's phenotype, indicating genetic diagnoses for 189 individuals (170 previously unsolved). The most common diagnosis was a founder variant in NPHP1 (20 affected probands). Of the remaining SVs, 934 impacted genes absent from OMIM and PanelApp. We identify 29 candidate novel disease genes with homozygous deletions in two or more individuals with matching phenotypes, including PDC (phosducin) and GCG (proglucagon). Candidate diagnostic non-coding SVs in or near to known disease genes were identified in 21 cases, with 12/21 of these being distal 5'-UTR/promoter knockouts.

Conclusion:

Analysing homozygous deletions facilitates identification of pathogenic variants, and aids discovery of both disrupted novel disease genes and essential regulatory elements.

Beyond the Germline: Mapping Somatic Causes of Rare Inflammatory Disease in Adults

Subsession Time: Friday, October 17 at 2:10pm – 2:15pm

Authors: Ashley Richardson (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Danielle Hoffman (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai) , Michael Murray (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Samira Asgari (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai)

Abstract: Background: Inborn errors of immunity (IEIs) are diseases caused by mutations in core immune genes. While genetic testing typically targets inherited variants, somatic mutations acquired later in life can also lead to IEI-like disorders, or "phenocopies," which often emerge in adulthood and worsen with age. Their true prevalence remains unclear, especially as aging populations may increase their burden. We aimed to estimate the prevalence of somatic IEI phenocopies and identify candidates for follow-up testing.

Methods: We analyzed whole-exome data from 31,250 BioMe participants using Mutect2. After validating our pipeline by detecting pathogenic UBA1 mutations that are known to cause VEXAS and reviewing clinical features via electronic health records (EHRs), we expanded to seven additional IEI genes (JAK1, KRAS, NLPR3, NARS, STA5B, TRL8). We assessed carrier phenotypes using EHRs and obtained IRB approval (STUDY-21-00804) to recontact individuals for confirmatory testing (ongoing).

Results: The mean age was 50.2 years, 39% were men, 25.48% European, 21.67% African, 32.62% Hispanic, and 20.23% other backgrounds. We identified 503 protein-altering (231 missense) somatic mutations across eight genes previously linked to IEI phenocopies. 60% were unique to a single

individual. Only one of these variants had been previously reported as likely pathogenic. Existing variant interpretation tools are limited for somatic mutations due to assumptions (e.g., germline selection) that don't apply. To infer pathogenicity, we evaluated clinical phenotypes and assessed whether symptoms correlated with variant allele fraction (VAF), which reflects the proportion of cells carrying the mutation. For instance, five patients harbored protein-altering somatic mutations in UBA1; two were novel. All five showed clinical features consistent with VEXAS syndrome (e.g., vasculitis), declining hematocrit, and elevated inflammatory markers (CRP, ESR). Similarly, we identified five individuals with somatic mutations in JAK1, none previously reported. Four had inflammatory disease diagnoses and elevated CRP/ESR, supporting their likely pathogenicity.

Conclusions: Somatic mutations likely cause underrecognized adult-onset inflammatory diseases. Estimating their prevalence is key to improving diagnosis and treatment. Our study identifies novel candidate mutations for IEI phenocopies and shows that large biobanks are valuable for discovering somatic variants in rare adult diseases. We are actively working to reduce false positives, a known challenge in somatic mutation calling, and are pursuing confirmatory clinical genetic testing.

Aberrant splicing prediction across human organ development

Subsession Time: Friday, October 17 at 2:15pm – 2:20pm

Authors: Nils Wagner (Technical University of Munich, Helmholtz Association - Munich School for Data Science - Munich), Aleksandr Neverov (Technical University of Munich), Alexandra C Martin-Geary (Big Data Institute - University of Oxford, Centre for Human Genetics - University of Oxford), Shubhankar Londhe (Technical University of Munich, Helmholtz Association - Munich School for Data Science - Munich), Carina Schröder (Technical University of Munich), Vicente Yepez (Technical University of Munich), Nicola Whiffin (Big Data Institute - University of Oxford, Centre for Human Genetics - University of Oxford, Broad Center for Mendelian Genomics - Broad Institute of MIT and Harvard), Julien Gagneur (Technical University of Munich, Helmholtz Association - Munich School for Data Science - Munich, Institute of Human Genetics - Technical University of Munich, Computational Health Center - Helmholtz Center Munich)

Abstract: Aberrant splicing can disrupt gene function and contribute to disease. We present AbSplice2, an improved model for predicting tissue- and development-specific aberrant splicing. AbSplice2 builds upon our previously published AbSplice framework with several key advancements: a refined ground truth derived from the aberrant splicing caller FRASER2, integration of state-of-the-art splicing predictions from Pangolin, the use of continuous splice site usage levels, and incorporation of transcriptomic dynamics across developmental stages. These innovations led to a two-fold increase in precision and recall compared to the original model and enabled variant effect prediction across both adult and embryonic contexts.

To model developmental alternative splicing (devAS), we constructed SpliceMaps from human RNA-seq data spanning embryogenesis to adulthood across seven tissues. This revealed that variants near devAS sites exhibit temporally varying effects, with AbSplice2 predicting potentially strong splicing disruption in

early development absent in adults; the brain showed the highest number of such variants—over 18,000.

To assess its clinical utility, we applied AbSplice2 to 64,736 participants from Genomics England. We identified 26 unique variants near devAS sites in known disease genes that are confidently linked with genetic disorders, for which AbSplice2 predicted stronger splicing disruption during early development than in adulthood. One example is a variant located in FGFR1 predicted by AbSplice2 to activate a weak, developmentally restricted splice-site in the forebrain, which would introduce a frameshift and trigger nonsense-mediated decay, consistent with loss-of-function in FGFR1, a gene known to be involved in Kallmann syndrome.

These findings illustrate how AbSplice2 adds developmental splicing context to variant effect predictions, revealing their stage-specific biological relevance. Genome-wide Pangolin and AbSplice2 predictions for all single-nucleotide variants are provided, alongside a web interface to score indels. Altogether, AbSplice2 provides a framework for predicting aberrant splicing across human organ development, with implications for rare disease research and understanding the role of developmental splicing in gene regulation.

Genetically imputed ancestry in a large clinical cohort undergoing carrier screening identifies new genes to target, but reveals potential disparity

Subsession Time: Friday, October 17 at 2:20pm – 2:25pm

Authors: Daniel E. Pineda-Alvarez (Labcorp, Formerly Invitae), Trevor J. Williams (Labcorp, Formerly Invitae), Carla Marquez-Luna (Labcorp, Formerly Invitae), Richard L. Martin (Labcorp, Formerly Invitae), Vivian Weinblatt (Labcorp, Formerly Invitae), Tali Ekstein (Labcorp, Formerly Invitae), Dana Neitzel (Labcorp, Formerly Invitae), Alix M.B. Lacoste (Labcorp, Formerly Invitae), Emily M. Russell (Labcorp, Formerly Invitae), Karen Ouyang (Labcorp, Formerly Invitae), Robert Daber (Labcorp, Formerly Invitae)

Abstract: Background:

A 2021 American College of Medical Genetics (ACMG) guideline for carrier screening (CS) of autosomal recessive and X-linked conditions recommended all pregnant patients be offered a CS panel of 113 genes, all of which should have a carrier frequency of ≥1/200 (tier 3 panel). However, it is well-established that carrier frequencies vary across different specific genetic ancestry groups. We evaluated gene and ancestry-specific carrier frequency (CF) and report-level statistics for a large cohort of individuals receiving CS at a commercial laboratory to ascertain genes meeting the ACMG CS “tier 3” criteria.

Methods:

We retrospectively analyzed CS results for individuals tested between June 2018 and December 2023. All individuals were tested using a comprehensive CS panel consisting of 287 to 570 genes. Genetic ancestry was estimated using random forests built with the 1000 Genomes dataset as a reference. We estimated

CF per gene and per gene stratified by ancestry group using the Hardy-Weinberg Equation. Statistical analyses included chi-square tests with permutation-based corrected p-values and linear mixed models.

Results:

Our cohort consisted of 102,447 individuals, 69.2% were female (n=70,899). The most represented imputed ancestry group (AG) was EUR (European) (n=56,305, 55.0%), followed by AMR (Admixed American) (n=23,039, 22.5%), AFR (African) (n=8,534, 8.3%), EAS (East Asian) (n=8,368, 8.2%), and SAS (South Asian) (n=6,201, 6.1%). Of the 287 genes for which all patients were sequenced, overall CF ranged from 0 to 11.5% (~1/9), with 49 genes having a CF ≥1/200. 18/49 genes (36.7%) were not in the ACMG tier 3 panel. When stratified by ancestry, CF ranged from 0 to 30.3% (~1/3), with 73 genes having a CF ≥1/200 in at least one AG. 34/73 were not in the tier 3 ACMG gene list (e.g, DYSF, ≥1/50, AFR; ETFDH, ≥1/150, EAS; EYS, ≥1/50, EUR; NPC1, ≥1/50, EUR; SLC22A5, ≥1/50, SAS; FANCA, ≥1/150, several AG). While mean CF did not differ across ancestry groups (F=0.7946, p=0.5287), CF within genes varied widely between ancestry groups. Additionally, the EUR ancestry group had a larger number of genes with CF ≥1/200 than others.

Conclusions:

CS of large cohorts from clinical laboratories provides additional insight into ancestry-specific CF in genes not included in the ACMG tier 3 carrier screening panel. These results suggest inequity in the ACMG tier 3 panel, and can help inform patients of personal risk, especially in historically underrepresented groups.

Session 79: Genetic Landscapes: Innovations in Population Genetics, Epidemiology, and Genomic Tools

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

Session Time: Friday, October 17 at 1:30pm – 2:30pm

Fine-mapping of loci under directional selection reveals functional architecture of human adaptation

Subsession Time: Friday, October 17 at 1:35pm – 1:40pm

Authors: Javier Maravall-López (Harvard University), Ali Akbari (Harvard University), Gaspard Kerner (Harvard University), Annabel Perry (Harvard University), David Reich (Harvard University), Alkes L. Price (Harvard University)

Abstract: Studies of directional selection inform our understanding of evolution and biology. Ancient DNA studies have shown great promise in detecting selection, with a recent study identifying 347 genome-wide significant selection signals (Akbari et al. 2024 bioRxiv). However, most studies have not fine-mapped causal variants under selection or linked variants under selection to disease GWAS.

Here, we fine-mapped causal variants at the 347 Akbari et al. loci using GWAS fine-mapping tools. We fine-mapped 53 variants with posterior probability (PIP)>0.95 (117 variants with PIP>0.5). Our results were concordant with known selected variants at well-studied loci and enriched for many functional annotations, e.g. 75.2x (s.e. 29.1x) for non-synonymous SNPs and 54.3x (s.e. 13.6x) for the cV2F variant-to-function score (Fabiha et al. 2024 bioRxiv).

We assessed the overlap between the fine-mapped variants under selection and disease GWAS. Our 117 PIP>0.5 variants had strong excess overlap with 1,275 PIP>0.5 variants from GWAS of 203 diseases/traits (114x, p<5e-9). In addition, we identified 381 GWAS trait-locus pairs (spanning 81 unique loci) for which selection and GWAS signals colocalized using coloc, including GWAS of 2 infectious diseases (Epstein-Barr virus presence and human betaherpesvirus 6B load); for 115 of the 381 trait-locus pairs, jointly fine-mapping selection and GWAS signals led to larger GWAS PIPs.

We clustered selected variants using Joint Pleiotropic and Epigenomic Partitioning (J-PEP; Kerner et al. 2024 medRxiv), obtaining 3 significant clusters: one related to pigmentation and epithelial tissue, one related to neutrophils and placental/embryonic tissue (reflecting this cell type's important role in the early stages of pregnancy), and one related to skeletal muscle and hypertension. Tissue-specific analyses further implicated endothelial and digestive tissue, in particular at immune genes. Cell type-specific analyses refined the digestive signal to mucin-secreting goblet cells, plausibly reflecting adaptation to bacterial pathogens. Our work deepens our functional understanding of variants under directional selection, furthering the emerging potential of ancient DNA to help elucidate disease biology.

A novel framework to characterize monogenic disease by leveraging founder populations in biobank-scale data

Subsession Time: Friday, October 17 at 1:40pm – 1:45pm

Authors: Christa Caggiano (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Ruhollah Shemirani (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Sarah Hanks (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Katherine E. Bonini (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Jaqueline O. Ogdis (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Priya M. Marathe (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Danielle Hoffman (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Sabrina Suckiel (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Gillian M. Belbin (Gencove), Abhi Biji (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Samira Asgari (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Vikas Pejaver (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Michael Murray (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai), Eimear E. Kenny (Institute for Genomic Health, Icahn School of Medicine at Mount Sinai)

Abstract: Intro: Identifying monogenic drivers of disease is crucial for improving health outcomes, yet current efforts are limited by reliance on specialty care and curated cohorts–restricting discovery to patients who reach specialists. Biobanks with genetic and electronic health record data offer a way to study monogenic disease at scale. However, low statistical power and insufficient clinical phenotyping hinder discovery and translational impact.

Methods: To overcome these barriers, we developed and validated a scalable framework that combines computational population genetics with systematic clinical evaluation to uncover monogenic disease drivers. We leverage founder populations, who are enriched for deleterious variants, as a tool. First, we detected pairwise identity-by-descent (IBD) sharing in 57K participants from the Mount Sinai Million (MSM) biobank using iLASH, and performed unsupervised clustering of IBD-sharing to find founder populations. Within each population, we performed phenome-wide association testing across 14021 traits, using rare variants (ExWAS), gene burden scores, and IBD-cliques defined via iCURL, our new method for haplotype clique detection. Causal variants were fine--mapped using deep learning pathogenicity predictors and clinical databases. To reduce false positives, we created a prioritization algorithm and integrated iterative clinician feedback.

Results: Founder populations were unexpectedly common, (~25% of MSM), suggesting the untapped power of founder populations in biobanks. Across 7 founder populations (n=157-8320), we identified 74129 ExWAS, 1032 gene-based, and 107373 IBD-based associations for 5037 phenotypes. Results were strongly enriched for congenital/genetic traits (p<10-16) and overlapped known recessive genes (p=8.9x10-5). Top prioritized associations represent hundreds of novel disease drivers, including musculoskeletal phenotypes in Puerto Ricans linked to RIF1.c.7105C>T (p=2.8x10-6), and LRP3.c.1846C>T associated with lipoprotein dysfunction in Ashkenazi Jews (p=1.0x10-5). We also identified clinically actionable findings, including misdiagnoses. For example, patients who were

homozygous for the causal Steel syndrome variant COL27A1.c.2089G>C were misclassified as Klippel-Feil syndrome, a disorder with similar symptoms.

Conclusions: This framework advances characterization of the genetic drivers of monogenic disease, especially for diverse populations, by harnessing the latent power of founder effects in biobanks

GATE-GENE: Rare Variant Survival Testing for Time-to-Event Phenotypes Identifies Genes Associated with Disease Onset and Progression Missed by Binary Trait Analyses in Large-Scale Biobanks

Subsession Time: Friday, October 17 at 1:45pm – 1:50pm

Authors: Ying Wang (Massachusetts General Hospital), Connor Dowd (Broad Institute of MIT and Harvard), Cindy Wen (Broad Institute of MIT and Harvard), Wenjian Bi (Peking University), Wenhan Lu (Broad Institute of MIT and Harvard), Satoshi Koyama (Broad Institute of MIT and Harvard), Alicia R. Martin (Massachusetts General Hospital), Wei Zhou (Massachusetts General Hospital)

Abstract: Chronic diseases and cancer progress through distinct biological stages and have a heritable basis, yet most genome-wide association studies (GWAS) rely on binary case-control models. They overlook the genetic basis of disease onset and progression over time, limiting insights into prevention and treatment. Time-to-event (TTE) phenotypes, curated from longitudinal electronic health records, can capture not only disease status but also age at onset, progression, and mortality. However, TTE GWAS face three major challenges: extreme censoring , population structure, and computational scalability. Extreme censoring–when most individuals do not experience the event during follow-up–can lead to inflated type I error, especially in rare variant analyses. Mixed-model methods like GATE enable single-variant TTE analysis but lack power for rare variants, which often have strong biological effects with therapeutic implications yet remain underexplored.

We present GATE-GENE, a novel, scalable framework for rare variant set-based TTE association testing. Built on the GATE framework with a mixed-effects frailty model, GATE-GENE models relatedness and population structure via a random effect and uses state-of-the-art optimizations enabling biobank-scale applications. It applies a saddlepoint approximation based on a modified Poisson distribution to handle heavy censoring. GATE-GENE, similar in design to SAIGE-GENE+ (developed for binary traits), aggregates rare variants within genes or regions and performs set-based tests (burden, SKAT, SKAT-O). It supports multiple variant masks (e.g., by allele frequency and functional annotations), and collapses ultra-rare variants (minor allele count ≤ 10) to further improve type I error control, power, and computational efficiency.

In simulations using UK Biobank exome data, GATE-GENE well controls type I error even for heavily censored TTE phenotypes (99% censoring), and has higher power than SAIGE-GENE+. We apply GATE-GENE to 71 disease onset and progression phenotypes in the All of Us exome data and uncover 24, 36 and 50 significant gene-TTE associations in African (AFR, N=79,500), American (AMR, N=71,610) and European (EUR, N=223,118) ancestry groups, respectively. We identify ancestry-enriched signals, e.g., PCSK9 for hyperlipidemia onset in AFR (event:censor=16,716:43,515) and APOB in EUR

(event:censor=70,373:92,544), highlighting the value of diverse populations in TTE genetic research. Notably, 21 novel associations are missed by SAIGE-GENE+, including C1QTNF12, an adipokine regulating insulin sensitivity, associated with progression from type 2 diabetes to diabetic retinopathy in AFR (P = 9.2E-7 vs. 1.0E-5).

GATE-GENE enables scalable, accurate rare-variant survival analysis across ancestries and disease stages, advancing stage-aware genetic discovery of complex diseases and equitable precision medicine.

PRS Without Population Labels: Accurate Multi-Ancestry Risk Prediction with ARGs

Subsession Time: Friday, October 17 at 1:50pm – 1:55pm

Authors: Nurdan Kuru (Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory), Shareef Khalid (Stony Brook University, Genetics), Adam Siepel (Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory)

Abstract: Polygenic Risk Scores (PRS) are widely used in precision medicine to predict disease risk based on genetic variation. Most methods currently require individuals to be grouped into broad continental ancestries prior to analysis. This limits their accuracy, especially in underrepresented populations due to differences in patterns of linkage disequilibrium and population structure.

To address this issue, we introduce a novel approach that models individual genomes using Ancestral Recombination Graphs (ARGs), which capture the true evolutionary history of genetic variants. Instead of relying on predefined population labels, we utilize local genealogies, enabling more precise modeling of individual-level genetic relationships.

Our method applies phylogenetic linear mixed models with ARG-derived covariance matrices: average genetic effects (fixed effects) are shared across ancestries, while individual-specific deviations (random effects) capture unobserved genetic variation, such as that arising from missing genotypes or imputation errors, by leveraging local genealogies encoded in the ARG. We estimate random effects using a mean-field variational procedure that transforms each local genealogy into a Laplacian-derived relatedness matrix, enabling scalable genome-wide inference. The framework also supports threading new individuals into the ARG, allowing prediction without reliance on population labels.

We evaluated our framework against PRS methods using population-level trees, genome-wide averaged ARGs, non-tree-based approaches, and multi-ancestry PRS tools such as PRS-CSx, PROSPER, and DendroPRS. In simulations, our ARG-based approach improved predictive accuracy by up to 50 percent, yielding higher R2 and lower prediction loss in sparse or incomplete data, and even when only trained on European participants. Notably, it showed the greatest gains in African and admixed populations, groups traditionally underrepresented in genetic research. Analyses of height and LDL in the All of Us cohort, as well as 1000 Genomes and Simons Genome Diversity Project data, confirmed consistent improvements over existing PRS methods.

By incorporating the rich evolutionary history encoded in ARGs, our approach provides a biologically informed alternative to traditional PRS methods. This framework has the potential to enhance genetic risk prediction across all populations, reduce health disparities, and improve the applicability of PRS in diverse ancestry groups.

Association of a multi-trait polygenic risk score with type 2 diabetes in the UK Biobank

Subsession Time: Friday, October 17 at 1:55pm – 2:00pm

Authors: Chi Zhao (University of Massachusetts Amherst), Konstantinos Hatzikotoulas (Institute of Translational Genomics), Alicia Huerta-Chagoya (Broad Institute of Harvard and MIT), Ravi Mandla (Massachusetts General Hospital, University of Pennsylvania), Nigel W Rayner (Institute of Translational Genomics), Lorraine Southam (Institute of Translational Genomics), Ken Suzuki (University of Tokyo, Osaka University Graduate School of Medicine), Henry J Taylor8,9,10, (National Institutes of Health, University of Cambridge), Xianyong Yin (Nanjing Medical University, University of Michigan), Raji Balasubramanian (University of Massachusetts Amherst), Zhengqing Ouyang (University of Massachusetts Amherst), Brian W Whitcomb (University of Massachusetts Amherst), Cassandra N Spracklen (University of Massachusetts Amherst)

Abstract: Type 2 diabetes (T2D) is a complex disease influenced by numerous genetic variants and multiple correlated phenotypes. The majority of prior polygenic risk scores (PRS) for T2D included only T2D-associated variants and captured only a fraction of the heritable component of T2D, limiting predictive performance and power. Integrating information from several genetically correlated traits may improve the prediction accuracy. To test this hypothesis in the context of T2D, we integrated GWAS summary statistics for 12 traits (T2D, HbA1c, fasting glucose, 2-hour glucose, fasting insulin, BMI, waist-hip ratio, LDL-C, HDL-C, triglycerides, and systolic and diastolic blood pressures) to construct a multi-trait PRS for T2D within 408,788 individuals from the UK Biobank. A PRS for each trait was calculated based on genome-wide significant variants after LD clumping. The multi-trait PRS was then constructed by: 1) generating the genetic correlation matrix of the 12 traits from LD score regression; 2) performing principal component analysis on the matrix and calculating weights by element-wise summation of the top eigenvectors explaining ≥80% variance; 3) calculating the multi-trait PRS as the weighted sum of each trait-specific PRS. Multivariable logistic regression models were used to generate adjusted odds ratios (OR), area under the receiver operating characteristic curves (AUC), and associated 95% confidence intervals (CI). Models adjusted for age, sex, genotyping chip, assessment center, and the first 10 genetic principal components. Among the selected traits, T2D is significantly genetically correlated with 6 traits after Bonferroni correction. Both T2D PRS and multi-trait PRS were significantly higher among individuals with T2D compared to controls (p<0.001). One standard deviation increase in T2D PRS was associated with a 77% increased risk of T2D (OR=1.77, 95% CI: 1.75-1.80, p<0.001), while multi-trait PRS was associated with a 19% increased risk (OR=1.19, 95% CI: 1.17-1.20, p<0.001). The corresponding AUCs were 0.71 for T2D PRS model and 0.66 for multi-trait PRS model. Overall, using the current method to integrate multiple GWAS traits to construct a multi-trait PRS did not yield improved predictive performance over the T2D PRS alone. Further modifications of PRS calculation or more sophisticated

modeling strategies may be needed to effectively incorporate genetic information from T2D-related phenotypes and enhance T2D risk prediction.

Genetic variants regulating immune cell-cell interactions are key drivers for autoimmune disease risk

Subsession Time: Friday, October 17 at 2:00pm – 2:05pm

Authors: Avantika R. Diwadkar (Penn State University)

Lida Wang (Penn State College of Medicine)

Havell Markus (Penn State College of Medicine)

Laura Carrel (Penn State College of Medicine)

Dajiang J. Liu (Penn State College of Medicine)

Abstract: GWAS has identified hundreds of variants associated with autoimmune diseases, most of which are non-coding and may function by modulating gene expression. Linking GWAS loci to the target genes has been challenging. Despite growing sample sizes and granularity of single-cell RNASeq data, many GWAS loci still fail to colocalize with bulk or single-cell (sc)-eQTLs. It motivates us to re-examine ways to integrate GWAS with sc-RNASeq data.

Importantly, immune cells do not function in isolation but interact with each other via antigen presentation, cytokine signaling, etc. When cells interact, the receptor gene expression will vary with the ligand. We hypothesize that genetic variants (which we define as cci-QTLs) may interact with ligand genes to influence receptor gene expression. We develop a method, GUCCI (Genetic determinants Underlying Cell-Cell Interaction) to identify cci-QTLs. GUCCI views the standard sc-eQTLs for receptor genes as weighted averages of cci-QTLs when the ligand gene expression is high and low. To improve power, we further propose to jointly model bulk-eQTLs with cci-QTL and sc-eQTLs, as bulk-eQTLs are weighted averages of sc-eQTLs from constituent cell types. We applied GUCCI to the OneK1K dataset across 14 immune cell types, using publicly available ligand-receptor pairs and the eQTLGen bulk-eQTL whole blood dataset. GUCCI identified 69k cci-QTLs across 154 cell type pairs and 168 ligand-receptor pairs, with only 11% identified previously as sc-eQTLs. 84% of cci-QTLs co-localize with at least one GWAS locus of autoimmune diseases. We successfully fine-mapped causal SNPs and cell type pairs for 86 ligand-receptor interactions, with 95% credible sets containing fewer than 5 variants per cell type pair. Intriguingly, cci-QTLs were enriched in regions of the lymph node spatial transcriptomics where they are known to interact e.g., cci-QTLs for B naïve cell interactions are enriched in B cell zones. Additionally, 32% of ligand-receptor pairs show significantly higher interactions in spots with fine-mapped cell types. The cci-QTLs are also enriched in ENCODE cCREs and have comparable Enformer functional scores as sc-eQTLs. Amongst our findings, interleukin-6 (IL-6) and receptor (IL6R) interact between CD4 naïve T cells and uniquely explain an ankylosing spondylitis locus, which is missed by sc-eQTLs. Our results demonstrate the validity of our method and its potential to bring unique mechanistic insight into risk genes. This is a first-of-its-kind analysis that can advance biological insights for autoimmune diseases.

Integrative Multi-omics Analyses to Identify Divergent Genetic Loci and Molecular Regulators of Crohn's Disease and Ulcerative Colitis with Real-world Data Applications

Subsession Time: Friday, October 17 at 2:05pm – 2:10pm

Authors: Chachrit Khunsriraksakul (Johns Hopkins University), Fan Zhang (Penn State College of Medicine), Havell Markus (Penn State College of Medicine), Siyuan Chen (Penn State College of Medicine), Lida Wang (Penn State College of Medicine), Dieyi Chen (Penn State College of Medicine), Xiaowei Zhan (University of Texas Southwestern), Joanna Melia (Johns Hopkins University), Ken Hui (Johns Hopkins University), Dajiang J. Liu (Penn State College of Medicine)

Abstract: Inflammatory bowel disease (IBD), which includes Crohn’s disease (CD) and ulcerative colitis (UC), presents diagnostic challenges in up to 15% of cases due to overlapping clinical, endoscopic, and histopathological features. Both conditions share common symptoms, including abdominal pain, diarrhea, weight loss, and extraintestinal manifestations, making diagnosis based solely on clinical presentation unreliable. Endoscopic findings, such as mucosal ulceration and inflammation, often appear similar, especially in colonic cases. Additionally, histopathological examination is frequently inconclusive because biopsies typically sample only the superficial mucosal layers, and hallmark features, such as granulomas, are often absent. These diagnostic uncertainties can hinder critical clinical decisions, such as the need for colectomy, underscoring the importance of molecular tools to differentiate IBD subtypes. To address these challenges, we conducted comprehensive integrative multi-omics analyses aimed at elucidating molecular distinctions between CD and UC to improve diagnostic accuracy and inform tailored treatments. Our first step involved identifying divergent genetic loci and molecular regulators between CD and UC. We created a novel genome-wide association study (GWAS) dataset that directly compared 12,194 CD cases and 12,366 UC cases. Subsequently, we performed downstream analyses, including transcriptome-wide association study (TWAS), proteome-wide association study (PWAS), statistical fine-mapping, and Mendelian randomization, to identify key molecular regulators. Our findings revealed 15 divergent genetic loci through CD vs. UC GWAS, highlighting well-known genes such as NOD2, HNF4A, and PTGE3/4. TWAS identified 27 additional novel loci, including NFKB1, while PWAS uncovered 32 additional loci, such as PLA2G10. Notably, multiple lines of evidence suggest that NF-κB may serve as a central regulator distinguishing CD from UC, supported by converging insights from PTGE3/4 (GWAS), NFKB1 (TWAS), and PLA2G10 (PWAS). Building upon these findings, we developed a predictive model that integrates epidemiological, genetic, and proteomic data to classify IBD subtypes more accurately. Using data from the UK Biobank, we identified 565 IBD patients (193 CD, 372 UC) and divided them into training and testing cohorts based on the timing of plasma protein measurements relative to diagnosis. Specifically, the training set included 87 CD and 171 UC cases, while the testing set included 106 CD and 201 UC cases. Our results demonstrate that combining polygenic risk scores (PRS) with levels of seven selected plasma proteins significantly enhances diagnostic accuracy. The AUC for PRS alone was 0.58, as was the AUC for the seven proteins alone; when combined, the AUC improved to 0.62. Importantly, our model outperformed a benchmark model based on 12 proteins previously identified in a large-scale proteomic study (SPARC IBD cohort; 636 CD, 470 UC), which achieved an AUC

of 0.56, increasing to 0.60 when combined with PRS. We hypothesized that the proteins identified through our multi-omics analyses reflect early molecular signals present in an otherwise healthy population, enabling more accurate prediction of new-onset CD and UC. Together, these findings underscore the potential of integrative multi-omics approaches to uncover early, disease-specific biomarkers and advance precision diagnostics.

Spatial Multi-omics at Scale: A Unified Framework for Human Biology, from Molecular Atlases to Clinical and Spaceflight Applications

Subsession Time: Friday, October 17 at 2:10pm – 2:15pm

Authors: Jiwoon Park (Weill Cornell Medicine), Roberto De Gregorio (Weill Cornell Medicine), Erika Hissong (Weill Cornell Medicine), Elif Ozcelik (Weill Cornell Medicine), Nicholas Bartelo (Weill Cornell Medicine), Felipe Segato Dezem (St. Jude Children's Research Hospital), Luke Zhang (St. Jude Children's Research Hospital), Maycon Marção (St. Jude Children's Research Hospital), Hannah Chasteen (St. Jude Children's Research Hospital), Yimin Zheng (CeMM), Ernesto Abila (CeMM), Junbum Kim (Weill Cornell Medicine), Jacqueline Proszynski (Weill Cornell Medicine), Akua A. Agyemang (Weill Cornell Medicine), Mohith Reddy Arikatla (Weill Cornell Medicine), Evelyn Metzger (Bruker Spatial Biology), Stefan Rogers (Bruker Spatial Biology), Prajan Divakar (Bruker Spatial Biology), Parambir S. Dulai (Northwestern University), Jason Reeves (Bruker Spatial Biology), Yan Liang (Bruker Spatial Biology), Liuliu Pan (Bruker Spatial Biology), Kimberly Young (Bruker Spatial Biology), Ashley Heck (Bruker Spatial Biology), Mithra Korukonda (Bruker Spatial Biology), Dan McGuire (Bruker Spatial Biology), Lidan Wu (Bruker Spatial Biology), Aster Wardhani (Bruker Spatial Biology), Joseph Beechem (Bruker Spatial Biology), George Church (Harvard Medical School), Sanjay Patel (Weill Cornell Medicine), Fabio Socciarelli (Weill Cornell Medicine), Sebastien Monette (Memorial Sloan Kettering Cancer Center), Brian Robinson (Weill Cornell Medicine), Massimo Loda (Weill Cornell Medicine), Olivier Elemento (Weill Cornell Medicine), Luciano Martelotto (University of Adelaide), Jasmine Plummer (St. Jude Children's Research Hospital), André F. Rendeiro (CeMM), Alicia Alonso (Weill Cornell Medicine), Robert E. Schwartz (Weill Cornell Medicine), Shauna Lee Houlihan (Weill Cornell Medicine), Christopher E. Mason (Weill Cornell Medicine)

Abstract: Spatial multi-omics is emerging as a powerful and scalable alternative to single-cell RNA sequencing for large-scale tissue characterization, preserving spatial context while capturing transcriptomic, proteomic, and morphological features at subcellular resolution. As the field advances toward profiling billions of cells and constructing virtual tissue models, spatial technologies provide an integrative framework capable of decoding complex microenvironments and cell states in health, disease, and extreme physiological conditions. Building on the Spatial Atlas of Human Anatomy (SAHA), we (1) establish a community-driven infrastructure for spatial imaging and multi-omics data generation and (2) demonstrate how spatial profiling resolves context-specific biological variation across diverse systems and exposures.

In cancer patients, spatial analysis revealed intrapatient heterogeneity (e.g., tumor-lymphocyte ratio variation within the tumor), region-specific features (immune microenvironment variations across mutational profiles), and rare molecular events (e.g., perineural invasion), enabling refined tumor

classification beyond histology or bulk profiling. These unique molecular features can be matched to histological features using a foundational model; for example, in inflammatory bowel disease (IBD), we mapped histology images to spatial transcriptomics data across the treatment timeline, showing spatiotemporal trajectories that distinguish responders from non-responders. Extending this framework to extreme phenotypes, spatial profiling of astronaut immune cells uncovered rare, adaptive cell states and colocalized transcriptional signatures of radiosensitivity that were not detected in matched single-cell or bulk datasets.

All datasets were processed using SAHA-standardized, open-source workflows, including curated metadata, scalable spatial data repositories, and foundation model–ready formats for joint training across omics and histological image modalities. By releasing annotated benchmarks, molecular atlases, and model training resources through a public-facing web portal and codebase, we provide an interoperable and reproducible spatial framework that enables next-generation discovery, diagnostics, and multi-cohort integration across terrestrial and spaceflight biology.

Single-Cell Multi-Omic Drug Response Profiling of 1 Million Cells Across 500 PRISM-Multiplexed Cancer Cell Lines Sequenced with SBX

Subsession Time: Friday, October 17 at 2:15pm – 2:20pm

Authors: Houlin Yu (Broad Institute of MIT and Harvard, Cambridge, MA, USA; Broad Clinical Labs, Burlington, MA, USA), Guoping Wang (Broad Institute of MIT and Harvard, Cambridge, MA, USA; Broad Clinical Labs, Burlington, MA, USA), Stephanie Yaung (Roche Sequencing Solutions, Inc, Santa Clara, CA, USA), Heejo Choi (Broad Institute of MIT and Harvard, Cambridge, MA, USA), Megan Rogers-Peckham (Roche Sequencing Solutions, Inc, Seattle, WA, USA), Michael Kartje (Roche Sequencing Solutions, Inc, Santa Clara, CA, USA), Matthew Rees (Broad Institute of MIT and Harvard, Cambridge, MA, USA), Paul Lund (10x Genomics, Pleasanton, CA, USA), Brian Haas (Broad Institute of MIT and Harvard, Cambridge, MA, USA; Broad Clinical Labs, Burlington, MA, USA), Charlotte Yang (Roche Sequencing Solutions, Inc, Seattle, WA, USA), Laura Doherty (Broad Institute of MIT and Harvard, Cambridge, MA, USA), Lacey McGee (Roche Sequencing Solutions, Inc, Seattle, WA, USA), Kendall Berg (Roche Sequencing Solutions, Inc, Seattle, WA, USA), Cynthia Cech (Roche Sequencing Solutions, Inc, Seattle, WA, USA), Salka Barrett (Roche Sequencing Solutions, Inc, Seattle, WA, USA), Anasha Arryman (Roche Sequencing Solutions, Inc, Seattle, WA, USA), Taylor Lehmann (Roche Sequencing Solutions, Inc, Seattle, WA, USA), Chen Zhao (Roche Sequencing Solutions, Inc, Santa Clara, CA, USA), Marc Prindle (Roche Sequencing Solutions, Inc, Seattle, WA, USA), Melud Nabavi (Roche Sequencing Solutions, Inc, Seattle, WA, USA), Carolyn Morrison (10x Genomics, Pleasanton, CA, USA), Peter Smibert (10x Genomics, Pleasanton, CA, USA), Kit Nazor (Proteintech Genomics, San Diego, CA, USA), Todd Golub (Broad Institute of MIT and Harvard, Cambridge, MA, USA; Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA), Victoria Popic (Broad Institute of MIT and Harvard, Cambridge, MA, USA; Broad Clinical Labs, Burlington, MA, USA), Katie Campbell (Broad Institute of MIT and Harvard, Cambridge, MA, USA), Jennifer Roth (Broad Institute of MIT and Harvard, Cambridge, MA, USA), Dean Proctor (Broad Institute of MIT and Harvard, Cambridge, MA, USA), Niall Lennon (Broad Institute of MIT and Harvard, Cambridge, MA, USA; Broad Clinical Labs, Burlington, MA, USA), Mark Kokoris (Roche

Sequencing Solutions, Inc, Seattle, WA, USA), Aziz M. Al'Khafaji (Broad Institute of MIT and Harvard, Cambridge, MA, USA; Broad Clinical Labs, Burlington, MA, USA)

Abstract: Inter-individual variability in drug response is a key challenge in precision oncology. Despite recent breakthroughs in targeting RAS-mutant cancers, clinical responses remain heterogeneous. To systematically dissect the molecular determinants of drug sensitivity and resistance, we developed a high-throughput single-cell multi-omic profiling framework using a PRISM pool, a barcoded co-culture platform of ∼500 human cancer cell lines spanning diverse lineages and genetic backgrounds. We treated the PRISM pool with two RAS inhibitors (BI-2865 and RMC-6236), as well as a negative control (DMSO) and a positive control (Panobinostat), and collected cells at early time points (3h and 12h) to capture initial drug response dynamics. We employed a modified 10x Genomics Flex protocol enabling simultaneous capture of the whole transcriptome, proteome (~320-plex Proteintech panel), and PRISM identity via expressed DNA barcodes. Sequencing by Expansion (SBX) was leveraged to achieve 415 billion reads to enable the analysis of 1 million high-quality single cells.

Drug treatments induced diverse, cell line–specific shifts in both transcriptional and proteomic states, with some changes strongly correlating with viability at five days post-treatment. Using hdWGCNA, we identified ~10 drug-responsive co-expressed gene modules per cell line. We then aggregated ~5,000 modules into module graphs using a KNN approach, uncovering 63 shared, lineage-specific, and/or treatment-specific clusters of modules. While modules from both sensitive and resistant lines often appeared in the same clusters, differences in the magnitude of their regulation were strongly associated with drug responses. Notably, 11 clusters showed expression changes strongly correlated with drug survival (r > 0.4) and were enriched for MAPK signaling, cell cycle regulation, and metabolic pathways. These programs could be further stratified by genetic background. Several resistance-associated features were observed only at the proteomic level, highlighting the value of dual-modality profiling in capturing functional drug response mechanisms.

Our study establishes a scalable paradigm for linking genotype, transcriptome, and proteome to pharmacologic phenotype at single-cell resolution across genetically diverse human models. These data, enabled with massively high-throughput sequencing using SBX, provide a rich resource for mechanistic discovery and rational design of combination therapies targeting the RAS pathway.

Mapping the Exposome: An integrative meta-reference across Child and Parent diverse cohorts from Human Health and Exposure Resource (HHEAR)

Subsession Time: Friday, October 17 at 2:20pm – 2:25pm

Authors: Dillon T. Lloyd (Harvard Medical School), Chirag J. Patel (Harvard Medical School)

Abstract: The exposome contains a multitude of interconnected factors that range from physical and chemical factors measured in human tissue, such as metals, plastics or organic chemicals. There is a critical gap in ascertaining the critical ranges of biomarkers of exposures as an integrated whole, akin to clinical norms used in laboratory medicine. This is especially important because while limited exposure may be best, most everyone has some level of exposure, so understanding a standard range of exposure

becomes paramount. Additionally, exposomic studies tend to not be diverse in location, participants and exposure type. As a result, it is difficult to contextualize individual exposure levels and risk, assess reproducibility, and/or compare populations from different socio demographic backgrounds. This means that we lack an understanding of what populations are at risk for higher levels of exposures, and thus it is difficult to understand who will bear the brunt of the health risks associated with environmental exposures. We built a reference cohort that combines 33 cohorts from the Human Health Exposure Analysis Resource. In this cohort, 217 targeted exposomic measures across categories such as phthalates, volatile organic compounds, trace elements, pesticides, and flame retardants that varied by cohort and 39 social exposome/demographic measures such as education, race and age in both children and adults were measured. We associated the physical exposome with the social exposome, built dynamic reference intervals for measuring the exposome after adjusting for the social exposome, built a burden metric to understand an individual's exposure profile and predicted the targeted exposome measures with untargeted exposure measurements. We found associations between all 39 social exposures with at least one targeted exposure feature, found population and demographic based differences in the dynamic reference intervals, and associations in exposure burden with age, sex, and race. Exposomics holds promise of a more precise definition of the complex phenomenon of human exposure, encapsulating large variation, across and within social groups, such as individuals with different education, income, and race. Our new cohort of exposure variation provides a guide to develop, benchmark, and contextualize biomarkers across the exposome for the first time.

Session 80: Interpreting Human Variation in Cancer Genomics and Pharmacogenomics

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

Session Time: Friday, October 17 at 1:30pm – 2:30pm

Rare Variant Burden Analysis Identifies Novel Germline Susceptibility Genes in Early-Onset Colorectal Cancer

Subsession Time: Friday, October 17 at 1:35pm – 1:40pm

Authors: Reger R. Mikaeel (Internal Medicine, Yale School of Medicine, Yale University, New Haven, CT, USA), Ruocen Song (Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA), Zhongping He (Department of Biostatistics, University of Florida, Gainesville, FL, USA), Timothy J Price 9Department of Haematology and Oncology, The Queen Elizabeth Hospital, Woodville, SA, Australia), Joanne P Young (Department of Haematology and Oncology, The Queen Elizabeth Hospital, Woodville, SA, Australia), and Xiao Fan (Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL, USA).

Abstract: Background: The incidence of early-onset colorectal cancer (EOCRC) is rising, yet the genetic factors contributing to disease risk remain largely undefined. We aimed to identify novel EOCRC risk genes through gene-based rare variant burden testing using whole-exome sequencing (WES). Methods: We analyzed 212 EOCRC cases (<55 years old) of European ancestry (174 with colorectal cancer and 38 with advanced polyps) from the South Australian Young Onset Colorectal Polyp and Cancer Study (SAYO), alongside 31,699 unaffected European ancestry controls from the Simons Foundation Powering Autism Research (SPARK) cohort. After stringent quality control for ancestry, relatedness, and sequencing artifacts, we conducted gene-based burden testing using predicted deleterious missense and loss-of-function (LoF) variants. To validate our analytic approach, we first tested gene sets previously associated with colorectal cancer (CRC). Individual gene-level associations were evaluated using binomial testing with permutation-based correction, applying a Bonferroni-adjusted threshold for significance and a predefined borderline significance threshold. Results: LoF variants in known CRC tumor suppressor genes were significantly enriched in EOCRC cases compared to controls (e.g., relative risk [RR] = 7.44 for ASCO-listed genes, P value = 4.4×10^{-6}), validating our burden testing approach. We then identified seven novel EOCRC susceptibility genes with significant or borderline-significant associations: MEIKIN, STK25, ATG3, PGBD4, DIRAS3, CD40, and RPS6KA4. MEIKIN demonstrated the strongest signal (P value = 1.0×10^{-7}) and had not previously been linked to CRC. These genes are implicated in key cancer-related pathways, including chromosomal segregation, autophagy regulation, and immune signaling. Notably, several EOCRC cases carried deleterious variants in both novel and known CRC genes, suggesting additive or synergistic genetic effects. Conclusions: This study identifies novel candidate risk genes for EOCRC through rare variant burden testing. Our findings broaden the current understanding of EOCRC

heritability and provide a basis for future functional validation and risk-based genetic screening in in young adults at risk of CRC.

Clonal Hematopoiesis of Indeterminate Potential: A Modifiable Risk Factor for Prostate and Bladder Cancer

Subsession Time: Friday, October 17 at 1:40pm – 1:45pm

Authors: Xinyu Wang (Harvard T.H. Chan School of Public Health), Linke Li (Clinical and Translational Epidemiology Unit, Massachusetts General Hospital), Md Mesbah Uddin (Broad Institute of MIT and Harvard), Abhishek Niroula (University of Gothenburg), Konrad H. Stopsack (Leibniz Institute for Prevention Research and Epidemiology - BIPS), Andrew Chan (Clinical and Translational Epidemiology Unit, Massachusetts General Hospital), Pradeep Natarajan* (Massachusetts General Hospital), Edward Giovannucci* (Harvard T.H. Chan School of Public Health), Zhi Yu* (Clinical and Translational Epidemiology Unit, Massachusetts General Hospital)

Abstract: Background: The association of clonal hematopoiesis of indeterminate potential (CHIP) with common urologic cancers remains poorly defined. This study aimed to (1) investigate the association of CHIP with prostate and bladder cancer risk; (2) prompted by recent findings that IL-6 signaling modifies TET2-associated inflammation and that metformin can modulate DNMT3A pathways, we tested these as potential gene-specific strategies to mitigate urologic cancer risk.

Methods: We conducted a prospective analysis of CHIP and its most common driver mutations (DNMT3A, TET2, and ASXL1) with incident prostate and bladder cancer in 406,799 unrelated leukemia-free participants from the UK Biobank (UKB). Significant findings were pursued in replication analyses using cross-sectional data from the All of Us (AoU) research program and the Mass General Brigham Biobank (MGBB). We tested for modification of TET2-associated risk by a common variant in IL6R p.Asp358Ala and of DNMT3A-associated risk by metformin use.

Results: Among the UKB participants (mean [SD] age: 57.1 [8.0] years; 46.2% female), 6.3% had CHIP. Over a median 13-year follow-up, CHIP was associated with an increased risk of both prostate cancer (HR: 1.12 [95% CI: 1.04, 1.21]) and bladder cancer (HR: 1.18 [1.05, 1.34]). Gene-specific analyses revealed that TET2 was most strongly associated with prostate cancer (HR: 1.21 [1.04, 1.41]), while DNMT3A showed the strongest association with bladder cancer (HR: 1.26 [1.07, 1.48]). Importantly, these gene-specific risks were modifiable. The increased prostate cancer risk from TET2 was primarily observed in non-carriers of the IL6R p.Asp358Ala loss-of-function allele (HR: 1.23 [1.09, 1.38]), while the risk was nullified in carriers (HR: 1.07 [0.97, 1.17]); this protective interaction was replicated in MGBB (Pinteraction=0.02) and showed a consistent trend in AoU (Pinteraction=0.07). Similarly, the bladder cancer risk associated with DNMT3A was apparent in metformin non-users (HR: 1.32 [1.12, 1.55]) but was attenuated among users (HR: 0.40 [0.13, 1.25]; Pinteraction=0.04).

Conclusion: Our findings establish CHIP as a novel, mutation-specific risk factor for prostate and bladder cancer. Crucially, these risks are modifiable: TET2-associated risk is blunted by IL-6 pathway disruption,

and DNMT3A-related risk by metformin use. These findings highlight promising strategies of precision prevention to mitigate urologic cancer risk in individuals with specific CHIP mutations.

Integrative analysis identifies multiple germline genetic variants and biologic pathways for chronic lymphocytic leukemia, yielding a predictive genetic risk score

Subsession Time: Friday, October 17 at 1:45pm – 1:50pm

Authors: Sonja I. Berndt (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Geffen Kleinstern (Faculty of Social Welfare and Health Sciences, School of Public Health, University of Haifa), Ricardo Cortez Cardoso Penhak (Genomic Epidemiology Branch, International Agency for Research on Cancer), Alyssa Clay-Gilmour (Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina), Simon Pahnke (Department of Medical Sciences, Uppsala University Hospital), Jojo Biel-Nielsen Dietz (Hematology, Danish Cancer Society Research Center), Josh Arias (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Hanla A. Park (Genomic Epidemiology Branch, International Agency for Research on Cancer), Charles Breeze (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Rosalie Griffin (Department of Epidemiology, University of Texas MD Anderson Cancer Center), Juan Sainz (Department of Biochemistry and Molecular Biology, Faculty of Sciences, University of Granada), Delphine Casabonne (Unit of Infections and Cancer, Cancer Epidemiology Research Programme, IDIBELL, Catalan Institute of Oncology, l'Hospitalet de Llobregat, Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP)), Ragnar Kristjansson (Hematology, Danish Cancer Society Research Center), Murat Güler (Genomic Epidemiology Group, German Cancer Research Center (DKFZ)), Olafur Davidsson (Hematology, Danish Cancer Society Research Center), Michelle A.T. Hildebrandt (Department of Lymphoma/Myeloma, University of Texas MD Anderson Cancer Center), Vijai Joseph (Cancer Biology and Genetics Program, Sloan Kettering Institute, Department of Medicine, Memorial Sloan Kettering Cancer Center), Elizabeth Brown (Department of Pathology, Heersink School of Medicine, University of Alabama), Neil Caporaso (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Mary McMaster (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Jonathan Hofmann (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Qing Lan (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Nathaniel Rothman (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health), Henrik Hjalgrim (Hematology, Danish Cancer Society Research Center), Nicola Camp (Division of Hematology and Huntsman Cancer Institute, University of Utah), Lara Sucheston-Campbell (Department of Oncology, Karmanos Cancer Institute at Wayne State University School of Medicine), James McKay (Genomic Epidemiology Branch, International Agency for Research on Cancer), Susan Slager (Department of Quantitative Health Sciences, Mayo Clinic) for the InterLymph Consortium

Abstract: Chronic lymphocytic leukemia (CLL) is the most common adult leukemia in Western countries and one of the most heritable hematological malignancies. To further understand the genetic architecture of CLL, we conducted the largest meta-analysis of genome-wide association studies to date, including 13,844 CLL cases and 1,115,090 controls, primarily from European populations. Over 27 million germline variants were imputed and evaluated for association with CLL risk, and results were combined using a fixed-effects meta-analysis. To identify independent signals, pseudo-conditional and fine-mapping analyses were conducted. A total of 137 independent significant SNPs (P<5x10-8), including 12 rare variants, across 97 loci were identified in the multi-population analysis, dramatically increasing the number of risk variants for CLL (60% novel) and demonstrating the polygenicity of the disease. Similar results were observed in the European-only analysis, with two additional loci discovered. Transcriptome-wide analyses, including in whole blood and CLL tumor cells, identified significant associations with gene expression for 26 genes, including three associations only seen with CLL tumor cells, such as reduced expression of PKNOX1, a putative tumor suppressor and transcriptional regulator. Further gene-based analyses demonstrated enrichment for antigen processing, immune system processes, and other related pathways. Using the identified loci, we constructed genetic risk scores (GRS) by summing allelic dosages weighted by the conditional per-allele log-ORs and evaluated their performance in predicting risk in two independent studies (UK Biobank and Mayo Biobank). The GRS showed strong predictive performance for CLL (C-index: 0.76, 95% CI: 0.74-0.78). Individuals in the upper decile of GRS demonstrated a 6.8-fold (95% CI: 5.41-8.58) increased risk of CLL compared to those with median values. The GRS was also associated (P=6.8x10-52) with risk of monoclonal B-cell lymphocytosis (MBL), the precursor condition to CLL, and demonstrated evidence of generalizability (P=0.009) to individuals of African ancestry, although the effect size was attenuated compared to Europeans. In conclusion, our large GWAS, coupled with integrative omics analysis, identified multiple novel risk variants and potential biological pathways underlying disease susceptibility, underscoring the complex etiology of CLL and suggesting new biological targets for therapeutic interventions and risk evaluation.

Meta-Analysis of Rare Cancers Identifies Germline Risk Variants and Functional Mechanisms in Anal Cancer and Gastrointestinal Stromal Tumor

Subsession Time: Friday, October 17 at 1:50pm – 1:55pm

Authors: Shaye Carver (DFCI), Tomin Perea-Chamblee (MSK), Kodi Taraszka (DFCI), Intae Moon (DFCI), Xinran Yu (DFCI), Jian Carrot-Zhang# (MSK), Alexander Gusev# (DFCI); (DFCI = Dana-Farber Cancer Institute, MSK = Memorial Sloan Kettering Cancer Center, #=co-last authors)

Abstract: Genome-wide association studies (GWAS) have advanced our understanding of common cancers, but rare cancers remain underexplored due to limited sample sizes and reduced power. Leveraging germline variant imputation from tumor sequencing, we conducted a two-phase GWAS to identify genetic risk factors for rare cancers. In the discovery phase, we tested common germline SNPs across 15 rare cancers in the DFCI-PROFILE cohort, followed by meta-analyses for anal cancer and gastrointestinal stromal tumor (GIST)—the two phenotypes reaching genome-wide significance—using additional data from MSK-IMPACT and FinnGen.

We identified novel germline risk loci: rs2856720-T in anal cancer (OR=1.60, $p=4.32\times10^{-17}$), near HLA-DQB1, and rs62331325-A in GIST (OR=1.96, $p=3.65\times10^{-37}$), within SLC6A18 and upstream of TERT. Both variants demonstrated replication across all three cohorts individually, with the strongest effect sizes observed in the PROFILE cohort: rs2856720 (OR = 2.07, $P = 7.44 \times 10^{-7}$) and rs62331325 (OR = 2.93, $P = 7.71 \times 10^{-15}$). Additionally, we identified a genome-wide significant association for rs4432943-T in anal cancer (OR = 1.40, $P = 3.04 \times 10^{-8}$), located intronically within MAML3, which also replicated consistently across cohorts.

The most significant GWAS hit in anal cancer was associated with Human Pappilomavirus (HPV) infection (PROFILE OR=1.47, $P=5.81\times10^{-3}$), suggesting a host-viral interaction whereby the variant increases susceptibility to HPV and facilitates HPV-driven cancer development. This variant also showed associations with other HPV-related cancers, including cervical cancer (OR=1.81, P=0.024) and head and neck carcinoma (OR=1.18, P=0.026), but not with non-HPV cancers such as colorectal cancer (OR=0.99, P=0.99). In GIST, the top germline SNP rs62331325 was associated with an increased risk of somatic KIT mutations (PROFILE OR=2.21, $P=6.50\times10^{-4}$), suggesting a germline-somatic interaction in which inherited variation may predispose tumor cells to acquire oncogenic KIT alterations during tumorigenesis. TWAS and RWAS further contextualized these variants, linking them to immune gene expression in anal cancer and TERT-associated regulation in GIST.

These findings provide the first replicated common germline associations for anal cancer and GIST, with effect sizes among the largest in cancer GWAS. Our study highlights the value of germline imputation from tumor data and meta-analysis for rare cancer discovery. It also highlights how human leukocyte antigen (HLA) variation influences host susceptibility to HPV infection, contributing to HPV-driven cancer risk, and suggests that germline variants may shape the acquisition of somatic alterations in GIST.

Saturation genome editing of BRCA1 across cell types resolves variant pathogenicity and improves cancer risk estimation

Subsession Time: Friday, October 17 at 1:55pm – 2:00pm

Authors: Gregory M. Findlay (The Francis Crick Institute), Phoebe Dace (The Francis Crick Institute), Nicole Forrester (The Francis Crick Institute), Laura Cubitt (The Francis Crick Institute), Maria Zanti (The Cyprus Institute of Neurology and Genetics), Chloé Terwagne (The Francis Crick Institute), Megan Buckley (The Francis Crick Institute), and Kyriaki Michailidou (The Cyprus Institute of Neurology and Genetics)

Abstract: Background

BRCA1 is a tumour suppressor gene with key roles in repairing DNA breaks by homologous recombination and promoting genome stability. Germline pathogenic BRCA1 variants predispose women to breast and ovarian cancer. Despite the existence of large functional data sets for variants in BRCA1, variants of uncertain significance (VUS) still account for over 50% of single-nucleotide variants (SNVs) in ClinVar. Key questions that remain include: 1) which variants outside of the RING and BRCT domains are functionally impactful, 2) how often are variant effects consistent between cell types, and 3) can data from multiple assays be leveraged to better predict cancer risk?

Methods

We performed saturation genome editing (SGE) in HAP1 cells to functionally score 3,843 variants across regions of BRCA1 not previously assayed. Furthermore, we developed a new SGE assay in human mammary epithelial cells (HMECs), allowing effects of over 2,000 variants to be compared across cell lines, drug treatments, and genetic backgrounds.

Results

In total, we scored 538 significantly depleted variants acting by diverse mechanisms in HAP1, including impacting protein-protein interactions, transcription, translation and splicing. A large portion of HAP1-depleted variants scored neutrally when tested in HMECs, including all missense variants in the coiled-coil domain, select missense variants in the RING and BRCT domains, and variants impacting specific splice sites. We identified mechanisms that explain differences between cell lines, including a role for RNF168 in facilitating homologous recombination when the BRCA1-PALB2 interaction is disrupted in HMECs.

Critically, function scores from both the HAP1 and HMEC SGE assays accurately discriminate between ClinVar-pathogenic and ClinVar-benign variants (AUCs = 0.997 and 0.977, respectively). Case-control studies suggest that variants scoring LoF in both lines confer a greater cancer risk than variants scoring LoF in HAP1 only. This finding is supported by several variants suspected to confer intermediate cancer risk that scored discordantly between lines.

Conclusions

These data will be highly valuable for aiding the classification of BRCA1 variants observed clinically. More broadly, this work illustrates how performing SGE across cell types can reveal context-specific effects that ultimately enable more accurate calibration of variant-specific cancer risk.

Identification of target genes at inherited risk loci in chronic lymphocytic leukemia

Subsession Time: Friday, October 17 at 2:00pm – 2:05pm

Authors: Huihuang Yan (Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN 55905, USA), Zhiquan Wang (Division of Hematology, Department of Medicine, Mayo Clinic, Rochester, MN 55905, USA), Alyssa Clay-Gilmour (Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208, USA), Geffen Kleinstern (School of Public Health, Faculty of Social Welfare and Health Sciences, University of Haifa, Haifa, Israel), Ricardo Cortez Cardoso Penha (Genomic Epidemiology Branch, International Agency for Research on Cancer, Lyon, France), Simon Pahnke (Department of Medical Sciences, Uppsala University, Uppsala, Sweden), Jojo Biel-Nielsen Dietz (Hematology, Danish Cancer Society Research Center, Copenhagen, Denmark), Joshua Arias (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA), Hanla A. Park (Genomic Epidemiology Branch, International Agency for Research on Cancer, Lyon, France), Rosalie Griffin (Epidemiology Department, The University of Texas

MD Anderson Cancer Center, Houston, Texas, USA), Juan Sainz (Universidad of Granada, Granada, Andalusia, Spain), Delphine Casabonne (Unit of Infections and Cancer, Cancer Epidemiology Research Programme, IDIBELL, Catalan Institute of Oncology, l'Hospitalet de Llobregat, Spain; Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), 28029 Madrid, Spain), Ragnar Kristjansson (Hematology, Danish Cancer Society Research Center, Copenhagen, Denmark), Murat Güler (Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany), Olafur Davidsson (Hematology, Danish Cancer Society Research Center, Copenhagen, Denmark), Michelle A.T. Hildebrandt (Department of Lymphoma/Myeloma, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA), Vijai Joseph (Memorial Sloan-Kettering Cancer Center, New York, NY, USA), Elizabeth Brown (Molecular & Cellular Pathology, The University of Alabama at Birmingham, Birmingham, AL, USA), Neil Caporaso (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA), Mary McMaster (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA), Jonathan Hofmann (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA), Qing Lan (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA), Nathaniel Rothman (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA), Henrik Hjalgrim (Hematology, Danish Cancer Society Research Center, Copenhagen, Denmark), Nicola Camp (Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA), Lara Sucheston-Campbell (Department of Oncology, Karmanos Cancer Institute at Wayne State University School of Medicine, Detroit, MI 48201, USA), Esteban Braggio (Department of Cancer Biology, Mayo Clinic, Phoenix, AZ 85054, USA), James McKay (Genomic Epidemiology Branch, International Agency for Research on Cancer, Lyon, France), Sonja Berndt (Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA), Susan L. Slager for the InterLymph Consortium (Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN 55905, USA; Division of Hematology, Department of Medicine, Mayo Clinic, Rochester, MN 55905, USA)

Abstract: Most disease-associated single nucleotide polymorphisms (SNPs) identified in genome-wide association studies (GWAS) are located in noncoding regions. Increasing evidence supports that these SNPs play a regulatory role by altering local chromatin state and impacting gene expression. We and others have previously identified 43 susceptibility SNPs, and with the recent largest CLL GWAS, additional 137 independent SNPs have been identified across 97 loci. To characterize the functional SNPs and potential target genes at the novel risk loci, we generated transposase-accessible chromatin with sequencing (ATAC-seq, 85 samples), cleavage under targets and tagmentation for histone marks H3K27ac (enhancer, 28 samples), H3K4me1 (enhancer, 28 samples) and H3K4me3 (promoter, 27 samples), and RNA-seq (394 samples) data in tumor cells from patients with CLL. ATAC-seq and histone marks data were used to define regulatory regions within the risk loci. To identify SNPs that impact gene expression, we performed cis expression quantitative trait loci (eQTL) analysis with index and correlated SNPs (R2 >=0.5) at 21 novel risk loci. QTLtools analysis identified 29 cis eQTLs at 10 risk loci, located within 1Mb

of transcription start sites. Of these 29 cis eQTLs, seven were located within active enhancers enriched with H3K27ac in CLL and another three in promoters enriched with H3K4me3. The same analysis identified 29 protein-coding genes (eGenes), for which the expression was affected by the identified eQTLs within 1Mb. We further analyzed publicly available chromosome interaction data from patients with CLL, lymphoblastoid cell line GM12878, and normal B cells, which revealed that, for 17 of the eGenes (58.6%), including novel target genes such as ATM at 11q22.3 and MED20 at 6p21.1, their promoters showed looping interactions with regulatory regions at six risk loci. ATM is a serine/threonine kinase with a key role in DNA damage response. Here, we identified eQTLs in enhancers that were associated with ATM expression. Chromatin interaction data supported the interactions between enhancers and ATM promoter in both CLL and GM12878. Notably, two enhancers at a risk locus on 8q24.21 targeted MYC, based on chromatin interaction data in GM12878 and CD20+ B cells. We previously reported that one of the prior risk loci tagged by rs2466029 on 8q24.21, approximately 540 kb away, also targeted MYC. MYC is overexpressed in CLL and associated with rearrangements in the immunoglobulin heavy chain locus. In summary, our integrative analyses identified additional target genes potentially contributing to the etiology of CLL.

Single-cell eQTL dataset of lung tissues from Asian never-smokers highlights the roles of alveolar cells in lung cancer etiology

Subsession Time: Friday, October 17 at 2:05pm – 2:10pm

Authors: Thong Luong (NCI, NIH), Jinhu Yin (NCI, NIH), Bolun Li (NCI, NIH), Elelta Sisay (NCI, NIH), Erping Long (Peking Union Medical College), Ju Hye Shin (Yonsei University College of Medicine), Sama Mikhail (NCI, NIH), Samuel Anyaso-Samuel (NCI, NIH), Fei Qin (NCI, NIH), Alexander Kane (NCI, NIH), Alyxandra Golden (NCI, NIH), Yoon Soo Chang (Yonsei University College of Medicine), Maria Teresa Landi (NCI, NIH), Nicholas Banovich (TGen), Nicholas Mancuso (USC Keck SOM), Nathaniel Rothman (NCI, NIH), Jinyoung Byun (UNM Comprehensive Cancer Center), Qing Lan (NCI, NIH), Christopher Amos (UNM Comprehensive Cancer Center), Kai Yu (NCI, NIH), Tongwu Zhang (NCI, NIH), Jianxin Shi (NCI, NIH), Jin Gu Lee (Yonsei University College of Medicine), Eun Young Kim (Yonsei University College of Medicine), Jiyeon Choi (NCI, NIH)

Abstract: Background: Single-cell expression quantitative trait loci (sc-eQTL) studies are instrumental in detecting context-specific genetic regulation underlying GWAS loci for common diseases such as cancers. However, sc-eQTL datasets from disease-relevant normal tissues without apparent exposures and those from non-European populations remain scarce. To enable characterization of lung cancer GWAS in diverse population, we built a lung sc-eQTL dataset of Asian never-smokers. Method: To link variant effects to transcriptomic regulation, we performed single-cell RNA sequencing using 10x Chromium Single Cell 3’ assay on tumor-distant normal lung tissues from 129 never-smoking Korean women. Flow sorting was used to enrich epithelial cells of tumor origin. Pseudo-bulked eQTL analyses were performed using TensorQTL. We performed colocalization and TWAS analyses integrating Asian lung cancer GWAS to nominate susceptibility variants and genes in cell-type specific manner. Association between pertinent variants and genes were verified via CRISPRi. Results: We identified 41 cell types in 4 categories from 360,127 cells, with an enrichment of epithelial category, ~47% of our dataset, including transitional cells.

eQTL mapping in 33 cell types revealed 2,229 eGenes and 4,372 top eQTL (top eSNP to eGene), including alveolar cell-specific eGenes unique to our data. Of the top eQTL, we found 18% are cell type specific, while 32% are category specific after effect correction by mashr. Integrating with single-cell lung ATAC-seq data, we observed that cell-type-specific eSNPs tend to be farther from transcription start sites and enriched in cell-type-specific ATAC-seq peaks compared to eSNPs common in multiple cell types. Colocalization and TWAS identified lung cancer susceptibility genes across 22 loci, validating known genes and identifying new genes. These genes were mainly detected in alveolar type I/II (AT1&2), epithelial transitional cells, and alveolar macrophages. TWAS further identified 10 novel loci previously not significant from GWAS, of which 8 were specific to one cell type. At the 6q22.1 locus, colocalization and TWAS identified ROS1 in AT2 cells, which was not apparent in bulk eQTL analyses. CRISPRi targeting a potentially functional variant, rs6937083, in lung cancer cells validated this association.

Conclusions: We created a lung sc-eQTL dataset from Asian never-smokers which enables us and others to identify context-specific susceptibility genes to further our understanding of lung cancer etiology.

Going the Extra Mile to Understand Chronic Lymphocytic Leukemia

Subsession Time: Friday, October 17 at 2:10pm – 2:15pm

Authors: Ivo Glynne Gut (Centro Nacional de Analisis Genomico), Claudia Buhigas (Centro Nacional de Analisis Genomico), Madlen Müller (Centro Nacional de Analisis Genomico), Holger Heyn (Centro Nacional de Analisis Genomico), Inaki Martin-Subero (IDIBAPS), Elias Campo (IDIBAPS), Ferran Nadau (IDIBAPS), Marta Kulis (IDIBAPS)

Abstract: Chronic Lymphocytic Leukemia (CLL) is the most common type of leukaemia in adults in Western countries. It is characterised by proliferative B lymphocytes. Genomic studies have provided a host of leads to make substantial headway to understanding CLL und with it to improved treatment. Recently we carried out an in-depth single cell investigation into CLL. First, we created a single cell atlas of a secondary lymphoid organ, the tonsils, which identified 121 distinct cell types and states providing a reference for immune cell annotation. We then carried out single cell analysis of patient progressions through the stages of CLL. Our single cell preparations were carried out using 10x Genomics, scRNA sequencing of the 3´or the 5´of the transcripts and multiome analysis for epigenetic measures, surface protein markers and T and B cell repertoires. However, the short read-length sequenced limit the ability to correlate known genomic features of patient samples. To gain more resolution we reanalysed the 10X libraries from our scRNA tonsil atlas and CLL studies with ONT long-read sequencing. We found differential usage of isoforms by many genes, for example, the commonly used surface marker CD45 is associate with different isoform usage in T and B cells, or the BLNK gene, which encodes a scaffold protein that is essential for B cell receptor signalling switches the isoform it uses as a pre germinal B cell as it transitions to mature stages.

We developed a new approach to identify single cells carrying a somatic mutation and dissecting corresponding transcript profiles. This allows us to relate the genotype of a cell to its behavior. In a case of a patient with a somatic XPO1 mutation, the mutation associate cluster exhibited a significant increase of expression of the immunoglobulin kappa joining 5 gene. In a case of a SF3B1 and ATM mutated

patient, we obtained four classes of processes that were differential between the mutation carrying and corresponding non-mutated cells. They displayed 1) complete loss of immunoglobulin kappa expression, 2) upregulation of TDRD15 and CD45, 3) a decreased ability to kill pathogens, and 4) a group of genes associated with altered adhesion, migration and inflammatory immune responses. These results demonstrate the dramatically increased level of resolution long-read single cell RNAseq analyses provide by covering the entire transcript spectrum and pointing to many novel genes associated with CLL.

Identification of CPT1A as a Blood-Based Biomarker of Mifepristone Efficacy in Alcohol Use Disorder

Subsession Time: Friday, October 17 at 2:15pm – 2:20pm

Authors: Jeesun Jung (National Institute on Alcohol Abuse and Alcoholism, NIH), Valeria Acosta (National Institute on Drug Abuse, NIH), Falk W. Lohoff (National Institute on Alcohol Abuse and Alcoholism, NIH), Barbara J. Mason (SCRIPPS Research), Leandro F. Vendruscolo (National Institute on Drug Abuse, NIH)

Abstract: Background: The glucocorticoid receptor (GR) plays a critical role in neuroadaptive stress processes implicated in the etiology of alcohol use disorder (AUD). Mifepristone, a nonselective GR antagonist, has been proposed as a potential pharmacological treatment for AUD. Studies in rodent and nonhuman primates have demonstrated that mifepristone reduced alcohol consumption in a dose-dependent manner. Similarly, a human laboratory study showed that a one-week of treatment with mifepristone reduced alcohol craving and drinking in individuals with AUD. However, the underlying biological mechanisms of mifepristone efficacy in AUD and alcohol-related behaviors have not been well studied.

Methods: In a double-blind, placebo-controlled study, individuals with AUD (n = 67) received mifepristone (600-1200 mg/day) or placebo for 7 days. Blood samples were drawn at baseline and at one week (n = 132). Bulk mRNA sequencing was conducted with blood samples. We conducted differential gene expression (DE) analysis between mifepristone and placebo groups at week 1 using DESeq2. To validate the DE results at week 1, we also conducted DE analyses at baseline between mifepristone and placebo groups, as well as DE analyses from baseline to week 1 within the mifepristone only group and within the placebo only group. Additionally, a PCR-based method was employed to validate the expression of top genes identified through the RNA sequencing analysis. Genes were considered differentially expressed if they had False Discovery Rate (FDR) p < 0.05 and an absolute log2 fold change > 0.02. Gene-set enrichment analysis (GSEA) was performed with DE genes assessed by FDR p < 0.1. A correlation study was conducted between the top significant DE genes and clinical measures over one week from baseline, including the number of standard drinks per day assessed by a timeline follow back questionnaire, the hepatic enzyme gamma-glutamyl transferase (GGT) that is sensitive to excessive alcohol drinking and cortisol level.

Results: We found three genes (CPT1A, PDK4, and TMEM135) with increased expression and four genes (CAMP, LCN2, ARG1, and CRISP3) with decreased expression in the mifepristone group compared with the placebo group at week 1 (p < 0.05). Increased expression of CPT1A was correlated with higher mifepristone concentrations (p < 0.05), reduced alcohol consumption from baseline (TLFB8-TLFB1, p

< 0.01), and elevated cortisol level from baseline (p < 0.05) after adjusting for treatment groups Notably, the RNA sequencing findings of increased CPT1A expression in the mifepristone group were validated using a PCR-based approach. The GSEA of Gene Ontology (GO) analysis identified 41 significantly enriched biological pathways (p < 0.05), including five pathways related to stress- and glucocorticoid-related signaling and lipid/immune associated processes.

Conclusion: We identified potential blood-based biomarkers of mifepristone efficacy related to alcohol-related outcomes in individuals with AUD. The findings suggest that CPT1A expression may be associated with stress-related pathways that are activated by chronic heavy alcohol consumption and are sensitive to GR antagonism.

GWAS of creatinine response to RAAS inhibition in 32,907 Finns identifies regulatory variant near LINC01267

Subsession Time: Friday, October 17 at 2:20pm – 2:25pm

Authors: Felix Vaura (Institute for Molecular Medicine Finland (FIMM), Helsinki Institute of Life Science (HiLIFE), University of Helsinki, Helsinki, Finland), FinnGen, Teemu Niiranen (Division of Medicine, Turku University Hospital, Turku, Finland; Department of Internal Medicine, University of Turku, Turku, Finland), Samuli Ripatti (Institute for Molecular Medicine Finland (FIMM), Helsinki Institute of Life Science (HiLIFE), University of Helsinki, Helsinki, Finland; Massachusetts General Hospital & Broad Institute, Cambridge, MA, USA; Faculty of Medicine, University of Helsinki, Helsinki, Finland)

Abstract: BackgroundRenin-angiotensin-aldosterone system inhibitors (RAASi) are essential medications for treating hypertension and chronic kidney disease. A modest rise in serum creatinine is common after RAASi initiation, likely reflecting changes in renal autoregulation. However, creatinine response varies between individuals, and guidelines recommend discontinuing RAASi if creatinine increases >30% after therapy initiation. We aimed to identify genetic predictors of creatinine response to RAASi initiation.

MethodsWe conducted a genome-wide association study (GWAS) in 32,907 RAASi users (mean age 64.0 years, 53.9% female) from the FinnGen Study, with laboratory measurements and nationwide electronic medication purchase register follow-up since 2014. We used serum creatinine measured within 180 days before and 120 days after RAASi initiation to compute a log-ratio response phenotype, log(creatinine_after / creatinine_before), which we then rank-normalized. We performed association testing with REGENIE, adjusting for age, sex, genotyping batch, and genetic principal components.

ResultsMedian creatinine increased from 76.5 to 78.0 µmol/L after RAASi initiation, corresponding to a median 1.5% relative increase. We identified a genome-wide signal at rs527866322 (3:14371534:C>T, GRCh38; β=0.12; $P=3.7\times10^{-9}$; effect allele frequency 3.9%), a regulatory variant near the long non-coding RNA gene LINC01267, enriched 1.5-fold in Finns compared to non-Finnish Europeans. The variant was associated with an allele dosage-dependent increase in relative creatinine levels (+1.4% per T allele, $P=1.8\times10^{-9}$), with median increases of 1.4%, 2.7%, and 4.8% for 0, 1, and 2 copies of the T allele, respectively. Phenome-wide association testing for rs527866322 revealed nominal associations with

premature separation of the placenta (β=0.42, P=1.1×10^{-4}) and reduced estimated glomerular filtration rate (β=–0.061, P=0.037) but not with baseline creatinine levels (P=0.22).

DiscussionTo our knowledge, we report the first genome-wide association of creatinine response following RAASi initiation. LINC01267 shows highest expression in the pituitary gland (1.03 transcripts per million, GTEx), a central regulator of renal hemodynamics. Replication in independent cohorts is warranted to clarify the potential role of rs527866322 in renal autoregulation.

Session 81: Linking Genetic Variation to Molecular, Cellular, and Phenotypic Function

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

Session Time: Friday, October 17 at 1:30pm – 2:30pm

Using pangenomes to characterize protein-coding and mutational constraint in high-identity duplicated genes

Subsession Time: Friday, October 17 at 1:35pm – 1:40pm

Authors: Luyao Ren (University of Washington, Howard Hughes Medical Institute), HGSVC, HPRC, Evan E, Eichler (University of Washington, Howard Hughs Medical Institute)

Abstract: Segmental duplications (SDs) are large, highly identical sequences (>90% identity and >1 kbp in length) that pose significant challenges for accurate alignment, gene annotation, and variant calling with short-read sequencing and standard computational pipelines. As a result, protein-coding genes embedded within SDs have been systematically excluded from population genetics resources (e.g., gnomAD), gene expression atlases, and disease association studies, despite their crucial roles in human evolution and disease. In this study, we employ 298 high-quality, telomere-to-telomere (T2T) human genomes from the Human Pangenome Reference Consortium (HPRC) and Human Genome Structural Variation Consortium (HGSVC) to systematically characterize 1,870 recently duplicated genes located within SDs. We use a phylogenetic-based approach to distinguish highly identical paralogs within gene families and characterize both single-nucleotide variants (SNVs) and copy number variants (CNVs) for specific copies for the first time. We show that 34.4% of recently duplicated genes are constrained for copy number and amino acid mutation. We further validate constrained genes using a database of 1.4 billion full-length cDNA (Iso-Seq) data and Oxford Nanopore Technologies (ONT) data from a panel of diverse tissues to assess methylation. Specifically, we assess whether these genes are actively expressed and show hypermethylation across gene bodies and hypomethylation across promoters. The approach identifies hundreds of previously unannotated gene models, distinguishing tissue-specific paralogs as well as those that represent human gene innovations potentially important in human brain development. In summary, our study uses the pangenome and phylogeny to establish a baseline of normal genetic variation in SD genes, distinguishing likely functional genes from pseudogenes and identifying novel protein-coding genes as candidates for future studies of disease and evolution.

A scalable framework to link rare human variants to disease phenotypes using pooled prime editing

Subsession Time: Friday, October 17 at 1:40pm – 1:45pm

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

Abstract: Large-scale sequencing projects have revealed millions of rare genetic variants in humans, yet the functional effects of most of these variants remain unknown. Here, we establish a platform to computationally prioritize and experimentally assay thousands of variants identified in whole-genome sequencing (WGS) to discover variants underlying human disease.

Our framework involves: 1) Prioritizing variants from WGS using computational predictors, 2) Conducting pooled prime editing (PE) screens to identify functionally impactful variants across each of many genes, and 3) Cross-referencing functional data with phenotypic datasets to link putatively causal variants to disease.

We initially prioritized 13,000 coding and non-coding variants across 143 disease genes from 100,000 Genomes Project participants. PE guide RNAs (pegRNAs) installing these variants plus 8,850 control variants from UK Biobank and ClinVar were assayed in PE7-expressing HAP1 cells. High editing rates (median 78.6% on day 4) allowed 81.2% of variants to be successfully scored in replicate experiments.

To interpret functional effects, we trained a supervised random forest classifier using 12 experimental features from our screen. Trained on control variants, the model achieved high performance (ROC AUC = 0.94) during cross-validation. Applying this model, we scored 11,440 variants from the Genomics England (GEL) cohort, classifying 492 variants as loss-of-function (LoF).

Among 436 ClinVar Benign and Likely Benign variants, only 2 (0.46%) were classified as LoF, compared to 22% of 138 ClinVar Pathogenic variants. Notably, we identified novel LoF variants among those of uncertain significance: 72 of 2,721 ClinVar VUS (2.6%) and 344 of 7,525 variants absent from ClinVar (4.57%) were deemed LoF, suggesting potential roles in disease.

Permutation testing revealed significant enrichment for LoF variants in genes matching participants' clinical recruitment programs. Furthermore, 126 of 410 GEL variants scoring as LoF were present in at least one carrier with a phenotype consistent with the gene’s associated disease.

This work illustrates our framework’s ability to identify loss-of-function variants observed in whole-genome sequencing at scale, suggesting potential roles in improving diagnosis and defining new genetic mechanisms of disease.

This research was made possible through access to data and findings in the National Genomic Research Library via the Genomics England Research Environment.

Comprehensive multimodal single-nucleus profiling of an iPSC reprogramming trajectory reveals variable inter-individual epigenetic differences

Subsession Time: Friday, October 17 at 1:45pm – 1:50pm

Authors: Terence Li (1,8), Justin Langerman (4), Cuining Liu (1,8), Zach von Behren (1), Min Jen Tsai (1), Yi Zhang (1), Yu Sun (4), Jason Ernst (2,4,6), Noah Zaitlen (1,6), Brunilda Balliu (2,3,5), Kathrin Plath (4), Chongyuan Luo (1)

1 Department of Human Genetics, David Geffen School of Medicine, UCLA

2 Department of Computational Medicine, David Geffen School of Medicine, UCLA

3 Department of Biostatistics, Fielding School of Public Health, UCLA

4 Department of Biological Chemistry, David Geffen School of Medicine, UCLA

5 Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA

6 Department of Neurology, UCLA

7 Department of Computer Science, UCLA

8 Bioinformatics Interdepartmental Program, UCLA

Abstract: Reprogramming of somatic cells to induced pluripotent stem cells (iPSCs) is a promising biological approach with applications in regenerative medicine and disease modeling. However, the complex molecular mechanisms underlying successful reprogramming are still unclear. Here, we present three dual-modality single-nucleus sequencing assays (10X Multiome, snmCT-seq, and snm3C-seq) applied to fibroblast-to-iPSC reprogramming across four individuals, ten timepoints, and four modalities (RNA, chromatin accessibility (ATAC), cytosine methylation, and 3D chromatin conformation). RNA-based unsupervised clustering revealed eight major cell types (35 subtypes) spanning the reprogramming trajectory. Notably, we identified substantial between-individual heterogeneity within a population of partially reprogrammed (failed) cells. Comparison of RNA and ATAC profiles shows that these failed cells maintain fibroblast-like chromatin accessibility patterns despite changes in gene expression, as evidenced by ChromHMM-based enrichment and joint RNA/ATAC gene regulatory network inference.

Strikingly, methylation-based clustering revealed that nearly all non-iPSC cell types co-cluster, and exhibit substantial individual heterogeneity. Closer inspection revealed that fibroblasts are hypomethylated and exhibit pronounced inter-individual variability that disappears upon successful reprogramming. This variability is explained by donor-specific methylation levels within partially methylated domains (PMDs). PMDs were the principal methylome features erased during reprogramming, producing hypermethylated iPSCs with reduced inter-donor variability.

Finally, we identify molecular features in other modalities that align with PMDs. 3C-based compartment analysis revealed strong concordance between B compartments and PMDs, including shared boundaries and erosion during reprogramming. Differential peak analysis revealed that many iPSC-specific ATAC peaks fall within PMDs, suggesting that PMD erosion is coupled with an increase in chromatin accessibility.

Taken together, our data provides the first quadruple-modality resource for dissecting the molecular dynamics of somatic cell reprogramming. Our high-resolution, single-cell atlas across dense timepoint

sampling captures the continuum of cellular trajectories – including both successful and failed outcomes – and offers new insights into the molecular hallmarks and bottlenecks of reprogramming.

Single-Cell Multiomic Mapping Uncovers Cell-Type-Specific Regulatory Architecture of Chronic Kidney Disease

Subsession Time: Friday, October 17 at 1:50pm – 1:55pm

Authors: Siyu Pan (University of Pennsylvania), Han Yuan (Calico Life Sciences), Xingfan Huang (Calico Life Sciences), Eunji Ha (University of Pennsylvania), Chenyu Li (University of Pennsylvania), Amin Abedini (University of Pennsylvania), Konstantin A. Kloetzer (University of Pennsylvania), Andrea Sanchez Navarro (University of Pennsylvania), Daigoro Hirohama (University of Pennsylvania), David Kelley (Calico Life Sciences), Katalin Susztak (University of Pennsylvania)

Abstract: Chronic kidney disease (CKD) affects over 800 million people worldwide, posing substantial healthcare burdens and mortality. While genome-wide association studies (GWAS) have identified more than 1,000 CKD-associated loci, over 90% of risk variants reside in noncoding regions, obscuring their functional consequences. Most existing multi-omics studies are constrained by sample mismatches across genomic modalities, with ATAC-seq and RNA-seq often derived from different donors. Previous bulk tissue analyses are unable to dissect cell-type-specific effects and directly connect regulatory variants with their target genes. Here, we performed integrated single-nucleus multiome analysis (snRNA-seq + snATAC-seq) of > 500,000 single nuclei coupled with whole-genome sequencing (WGS) on 97 healthy and CKD human kidney samples, enabling simultaneous profiling of gene expression and chromatin accessibility in the same cells at single-cell resolution while directly correlating genetic variants with their molecular phenotypes. Our analysis identified both cell-type-specific expression quantitative trait loci (eQTLs) and chromatin quantitative trait loci (caQTLs), including variants showing allele-specific effects on transcription factor binding through motif disruption. A systematic peak-to-gene linkage analysis revealed 263,008 regulatory peaks connected to 18,123 genes across cell types, with PT cells exhibiting 1.5-fold more connections than other cell types. Notably, 60% of peaks linked to a single gene, reflecting precise and cell-type-restricted regulatory architecture. By integrating these findings with CKD GWAS loci through colocalization analysis, we mapped CKD-associated GWAS variants to cell-type-specific regulatory networks, prioritizing both established and novel candidate genes. Together, our comprehensive multiomic framework not only bridges noncoding GWAS associations to cellular mechanisms in CKD by identifying functional variants that alter chromatin accessibility and their cell-type-specific target genes, but also reveals disease-relevant regulatory circuits, thereby establishing a powerful approach for translating genetic findings into mechanistic insights and potential therapeutic opportunities.

A High-Resolution Examination of Genetically and Environmentally Induced Methylation Changes in the Million Veteran Program

Subsession Time: Friday, October 17 at 1:55pm – 2:00pm

Authors: Patrick Schreiner (Booz Allen Hamilton, VA Center for Data and Computational Sciences), Kyriacos Markianos (VA Center for Data and Computational Sciences), Michael Francis (Booz Allen Hamilton, VA Center for Data and Computational Sciences), Brendan Despard (Booz Allen Hamilton, VA Center for Data and Computational Sciences), Million Veteran Program, Bryan R. Gorman (Booz Allen Hamilton, VA Center for Data and Computational Sciences), Saiju Pyarajan (VA Center for Data and Computational Sciences)

Abstract: Background. DNA methylation is a critical epigenetic programming mechanism which is influenced in an individual by both genetic inheritance and environmental exposure. DNA methylation has wide-ranging effects on cell-type definition via three-dimensional chromatin structure, chromatin accessibility, transcriptional regulation, silencing of highly repetitive DNA, and X-inactivation. It is unique relative to other NGS data given its sensitivity to modification. Aberrant DNA methylation is known to be associated with highly prevalent phenotypes including obesity, type 2 diabetes, respiratory diseases like asthma and COPD, cardiometabolic disease, cancer, and all-cause mortality. The sensitivity of DNA methylation makes it particularly useful in assessing the impacts of genetic and environmental factors on a population-scale.

Methods. Our analysis of individuals within the Million Veteran Program (MVP) represents a single-cohort examination of 42,460 individuals using 850K individual CpG sites profiled by the Infinium MethylationEPIC platform. We performed an ancestry-stratified epigenome-wide association study (EWAS) using meffil to identify methylation associations with age, smoking status, and BMI including 26,829 Europeans, 11,605 Admixed Africans, and 4,206 Hispanics. We also leverage genetic information available from these same individuals to calculate methylation quantitative trait loci (meQTLs) using tensorQTL.

Results. We analyze methylation associations with genetic variation and phenotypes including an individual's age, smoking status, and BMI. Our results confirm methylation associations with common genetic variation in the MVP that have been previously reported by GTEx and KORA. Further, we report phenotypic associations with methylation detection that agree with leading EWAS publications and identify novel phenotypic associations in the MVP. Finally, we report unexpected intra- and inter-cohort variability of methylation associations with BMI due to the sensitivity of methylation data and the complexity of the trait.

Conclusion. The comprehensive nature of our methylation profiling, alongside our uniquely large and diverse cohort including both genetic and methylation data, provides unprecedented resolution into the effects of methylation changes on a population scale. We leveraged the increased statistical power available in our unique cohort to allow for a high-resolution analysis that has identified novel methylation associations with age,smoking status, and BMI. Our results build confidence of robust methylation associations, contribute novel associations with age, smoking status, and BMI, while also serving as a cautionary tale for methylation studies reporting on complex traits.

Epigenome-wide association study of lifecourse socioeconomic position with cardiovascular and cognitive health in US Hispanics/Latinos from the Hispanic Community Health Study/Study of Latinos

Subsession Time: Friday, October 17 at 2:00pm – 2:05pm

Authors: Jee-Young Moon (Albert Einstein College of Medicine), Paola Filigrana (Albert Einstein College of Medicine), Linda C Gallo (San Diego State University), Hector M Gonzalez (UC San Diego), Myriam Fornage (University of Texas at Houston), Charles DeCarli (University of California at Davis), Krista M Perreira (University of North Carolina at Chapel Hill), Humberto Parada Jr. (SDSU), Jianwen Cai (University of North Carolina at Chapel Hill), Frank J Penedo (University of Miami), Maria M Llabre (University of Miami), Amber Pirzada (University of Illinois at Chicago), Martha Daviglus (University of Illinois at Chicago), Wassim Tarraf (Wayne State University), Freddie Marquez (University of California San Diego), Shakira F Suglia (Emory), Carmen Isasi (Albert Einstein College of Medicine)

Abstract: Background: Lifecourse socioeconomic position (SEP) has life-long influences on health, including cardiovascular health (CVH), and cognitive function. DNA methylation (DNAm) may reflect the biological embedding of SEP-related exposures. However, large-scale epigenome-wide association studies (EWAS) of SEP, particularly among US Hispanics/Latinos, are limited. We aimed to 1) identify the DNAm patterns associated with childhood and adulthood SEP, and 2) assess if these patterns are linked to gene expression, CVH, and cognitive function.

Methods: We included 1,959 individuals aged 45-74 years (mean age 57, 65% female, 91% non-US born) from the Hispanic Community Healthy Study/Study of Latinos (HCHS/SOL) at baseline (2008-2011). DNAm was assayed in baseline peripheral blood using the Illumina EPIC v1.0 BeadChip. Childhood SEP (score 0-3) was based on highest parental education (elementary, middle, high, college); adulthood SEP (score 0-7) included education, employment, household income, and homeownership. An EWAS was performed using linear regression on logit-transformed methylation values with childhood and adulthood SEP simultaneously, adjusting for age, sex, smoking, genetic ancestry (5 principal components), and blood cell composition. CpG sites with P-value <9.42x10-8 were further evaluated for associations with gene expression in blood, clinical CVH index (sum of 4 health factors in AHA's Life's Essential 8), and cognitive function (composite of verbal memory, language, processing speed, and executive function), applying Bonferroni correction.

Results: Higher childhood SEP was associated with hypomethylation at cg22762215, CLEC2L, CPT1A, PTPRB, MMP14, and hypermethylation at ADTRP. Higher adulthood SEP was linked to hypermethylation at AHRR and SOCS3. Hypermethylation at AHRR and SOCS3 was associated with lower gene expression (AHRR, SOCS3); and better cardiovascular health (SOCS3) and cognitive function (AHRR, SOCS3). Childhood SEP-related CpG sites were not consistently linked – either in direction or significance - with gene expression or health outcomes.

Conclusion: Among US Hispanic/Latino adults, hypermethylation of AHRR and SOCS3 may reflect the biological embedding of adulthood SEP, positively influencing cardiovascular and cognitive health. While

DNAm patterns linked to childhood SEP were observed, the lack of consistent directionality with health outcomes might be possibly due to a coarse proxy for early-life SEP and epigenetic plasticity.

Genetic regulation of cell-type–specific chromatin accessibility shapes immune function and disease risk

Subsession Time: Friday, October 17 at 2:05pm – 2:10pm

Authors: Angli Xue (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia), Jianan Fan (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia), Oscar Dong (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia), Hao Lawrence Huang (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia), Peter C. Allen (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia), Eleanor Spenceley (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia), Eszter Sagi-Zsigmond (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia), Blake Bowen (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia), Anna S.E. Cuomo (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia, Centre for Population Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia, Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, VIC, Australia, UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, NSW, Australia), Albert Henry (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia), Hope A. Tanudisastro (Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia, Centre for Population Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia, Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, VIC, Australia, Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia), Zhen Qiao (Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia, Genomics and Inherited Disease Program, Garvan Institute of Medical Research, Sydney, NSW, Australia), Drew Neavin (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia), Anne Senabouth (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia), Caitlin Bartie (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia), Katrina M. de Lange (Centre for Population Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia, Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, VIC, Australia), Gemma A. Figtree (Charles Perkins Centre, The University of Sydney, Sydney, NSW, Australia, Kolling Institute of Medical Research, Royal North Shore Hospital, Sydney, NSW,

Australia), Alex W. Hewitt (Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia, Department of Ophthalmology, Royal Hobart Hospital, Hobart, TAS, Australia, Centre for Eye Research Australia, University of Melbourne, East Melbourne, VIC, Australia), Daniel G. MacArthur (Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia, Centre for Population Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia, Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, VIC, Australia), Joseph E. Powell (Translational Genomics Program, Garvan Institute of Medical Research, Sydney, NSW, Australia, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia, UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, NSW, Australia)

Abstract: Understanding how genetic variation regulates gene expression at the single-cell level is critical for uncovering mechanisms of complex disease. While single-cell RNA sequencing has enabled the discovery of cell-type-specific genetic regulation, the mechanisms remain poorly understood due to limited large-scale single-cell multi-omics datasets.

Here, we introduce the TenK10K multiome initiative, which comprises chromatin accessibility profiles from 3.5 million peripheral blood mononuclear cells (PBMCs) from 1,013 donors using single-cell ATAC-seq and multiome (RNA+ATAC) with matched whole-genome sequencing. We characterise 440,996 chromatin peaks across 28 immune cell types and map 250,165 quantitative trait loci associated with chromatin accessibility (caQTLs) at FDR < 0.05, 60% of which exhibit specificity to a single cell type. We further integrate them with TenK10K scRNA-seq (5.4M PBMCs) and colocalizing caQTLs with eQTLs using coloc identifies 70,243 candidate cis- regulatory elements, 25,280 (40%) of which show evidence of causal association by Mendelian Randomization.

Integrating caQTLs with GWAS loci for 19 complex diseases and 44 blood serum traits reveals an additional 10–41% of colocalized signals not captured by using eQTL alone. These effects are most pronounced in autoimmune diseases. For inflammatory bowel disease, 39 genes show colocalized signals in CD8 effector memory T cells using only eQTLs, but it increases to 55 after incorporating caQTLs. Cell-type-specific mechanisms for these loci can be further elucidated. For instance, we identify a genetic regulatory effect via altered chromatin accessibility in CD8 effector memory T cells – but not in naïve cells – for IRGM, an interferon-inducible protein involved in autophagy.

Furthermore, we leverage a graphical neural network approach to integrate eQTLs and caQTLs and demonstrate that linking peak to gene using unpaired multiome achieves up to 80% higher accuracy compared to paired data lacking QTLs. The improved peak-to-gene links further enhance the gene regulatory network inference by identifying more regulatory edges in the network in both healthy and disease cohorts.

These findings provide an unprecedented single-cell map of chromatin accessibility and genetic variation in human circulating immune cells, establishing a foundation for dissecting cell-type-specific regulatory mechanisms and understanding how genetic variation shapes immune cell function and disease risk.

Isogenic induced pluripotent stem cells uncover distinct effects of TCF7L2 and a common diabetes GWAS variant on pancreatic beta cell differentiation and function

Subsession Time: Friday, October 17 at 2:10pm – 2:15pm

Authors: D. Leland Taylor (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Filippo Cipriani (The New York Stem Cell Foundation Research Institute),

Lori L. Bonnycastle (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Dongxiang Xue (Department of Surgery, Weill Cornell Medicine, Center for Genomic Health, Weill Cornell Medicine),

Amy J. Swift (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Stefan Semrau (The New York Stem Cell Foundation Research Institute),

Brian N. Lee (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Jeff Winchell (The New York Stem Cell Foundation Research Institute),

Henry J. Taylor (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health, British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Heart and Lung Research Institute, University of Cambridge),

Nathan Dale (The New York Stem Cell Foundation Research Institute),

Kiran Ramnarine (The New York Stem Cell Foundation Research Institute),

Jordan Guttenberg (The New York Stem Cell Foundation Research Institute),

Annie Butt (The New York Stem Cell Foundation Research Institute),

Bianca Migliori (The New York Stem Cell Foundation Research Institute),

Grayson Horn (The New York Stem Cell Foundation Research Institute),

Jordan Goldberg (The New York Stem Cell Foundation Research Institute),

Neelam Sinha (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Tingfen Yan (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Elizabeth Li (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Elizabeth Winnicki (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Erin C. Mansell (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Angela Lee (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Hannah Li (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Caleb Rathbun (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Fiona Daly (Division of Preclinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health),

NYSCF Global Stem Cell Array Team (The New York Stem Cell Foundation Research Institute),

Vukasin Jovanovic (Division of Preclinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health),

Carlos Tristan (Division of Preclinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health),

Scott Noggle (The New York Stem Cell Foundation Research Institute),

Narisu Narisu (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Dieter Egli (The New York Stem Cell Foundation Research Institute),

Leslie G. Biesecker (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Shuibing Chen (Department of Surgery, Weill Cornell Medicine, Center for Genomic Health, Weill Cornell Medicine),

Rick Monsma (The New York Stem Cell Foundation Research Institute),

Michael R. Erdos (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health),

Daniel Paull (The New York Stem Cell Foundation Research Institute),

Francis S. Collins (Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health)

Abstract: The TCF7L2 locus contains one of the first identified and strongest common variant genetic associations with type 2 diabetes (T2D). Recent fine-mapping studies have refined this signal to a single candidate causal intronic variant, rs7903146. To date, studies have examined the function of TCF7L2 across various cell types and have found multiple T2D-relevant effects, including the essential role of TCF7L2 in the differentiation of stem cells to pancreatic islet-like beta cells—a critical cell type for glucose regulation. Yet, despite progress in understanding the function of TCF7L2, the specific mechanism by which rs7903146 influences disease risk remains unclear.

In this study, we generated isogenic induced pluripotent stem cells (iPSCs) representing the full allelic series of rs7903146 (CC, CT, TT) as well as heterozygous and homozygous TCF7L2 loss of function (LoF) mutations across three parental backgrounds with multiple clones per parental line (52 lines total). Using a fully automated, robotic system, we differentiated these lines to stem cell-derived islet-like organoids (SC-islets) that contain beta cells. Throughout differentiation, we conducted extensive phenotyping, including (i) flow cytometry and immunofluorescence staining for critical, stage-specific marker genes at the definitive endoderm (S1) and pancreatic progenitor (S4) stages, (ii) cell painting at S1 and S4, (iii) multimodal single-cell sequencing of RNA and chromatin accessibility at S1, S4, and SC-islets, (iv) brightfield morphology imaging of SC-islets, and (v) insulin secretion assays of SC-islets. While data collection and analysis are ongoing, our initial results confirm a large effect of TCF7L2 LoF mutations across multiple assays, consistent with the essential role of TCF7L2 during beta cell differentiation. We also observe a strong effect of rs7903146 at later stages of differentiation across multiple assays, including SC-islet morphology and insulin secretion capacity. These results comport with previous studies that suggest rs7903146 resides within a pancreatic islet beta cell specific enhancer. To our knowledge, this study is one of the first to demonstrate a functional impact of a common, disease-associated GWAS variant in a disease-relevant cell type using isogenic iPSCs.

Comprehensive analysis of splicing snRNA genes and pseudogenes reveals new associated rare diseases genes

Subsession Time: Friday, October 17 at 2:15pm – 2:20pm

Authors: Olfa Messaoud (Center for Mendelian Genomics, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA, Harvard Medical School, Boston, Massachusetts, USA, Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts, USA, Center for Genomic Medicine, Massachusetts General Hospital, Massachusetts, USA, Biomedical Genomics and Oncogenetics Laboratory, Institut Pasteur de Tunis, Tunisia), Daniel Marten (Center for Mendelian Genomics, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA), Rahaf Tarawneh (Harvard T.H. Chan School of Public Health, Massachusetts, USA), Broad Center for Mendelian Genomics, Heidi Rehm (Center for Mendelian Genomics, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge,

Massachusetts, USA, Harvard Medical School, Boston, Massachusetts, USA, Center for Genomic Medicine, Massachusetts General Hospital, Massachusetts, USA), Christina Austin-Tse (Center for Mendelian Genomics, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA, Ambry Genetics, Aliso Viejo, CA, USA), Anne O'Donnell-Luria ((Center for Mendelian Genomics, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA, Harvard Medical School, Boston, Massachusetts, USA, Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts, USA, Center for Genomic Medicine, Massachusetts General Hospital, Massachusetts, USA)

Abstract: Splicing represents a critical molecular step in gene expression and regulation. It is considered the most complex molecular mechanism in eukaryotic cells as it involves both proteins and RNA that interact together to form ribonucleoproteins, the main actor of the splicing machinery. Recent reports have shown the involvement of small nuclear RNA (snRNA) in neurodevelopmental diseases. However, their role in the broad spectrum of rare genetic diseases remains poorly understood. In this project, we evaluated all snRNA genes that form the major and the minor spliceosome in their functional and pseudogenes forms, with a total of 1685 genes divided into 42 functional and 1643 pseudogenes in order to assess their involvement in the etiology of different forms of genetic conditions. Similar to prior efforts, our strategy is based on the hypothesis that regions depleted for variation in reference population datasets are more likely to harbor variants in patients with rare diseases. We defined depleted regions as those where the normalized observed proportion is less than -0.2 (i.e., a deviation from the median of at least 20%). Data related to allele frequency of 50,840 variants was extracted from gnomAD v4, encompassing 76,156 genomes and served to classify windows of 18 bp, each, as depleted vs non-depleted. We retained 13 depleted regions within 10 genes (previously described RNU4ATAC, RNU4-2, RNU5B-1, RNU12 and not yet disease-associated RNU5E-1, RNU1-2, RNU4-1, RNU6-9), including two originally designated as pseudogenes (RNU2-2, RNU2-63P). The size of the depleted region ranges from 18 to 42bp with a mean of 26 (±9). We assessed the presence of any rare variant in these regions of interest among affected individuals in the Broad CMG/GREGoR datasets (includes ~3,000 genomes) using the seqr analysis platform. Analysis of variant segregation was first conducted for de novo variants with a special focus on depleted regions. Later, we extended the analysis to biallelic variants within the entire gene. At least 82 candidate variants were identified (30 de novo, 16 homozygous recessive, and 36 compound heterozygous) in the 10 snRNA genes. Candidates in each of these genes are now undergoing further assessment. The recently discovered large contribution of snRNA genes in rare diseases will likely continue to expand as the phenotypes associated with other candidate snRNA genes are defined.

Addressing RNA-seq alignment bias against gene-altering SVs improves structural characterization and expression quantification

Subsession Time: Friday, October 17 at 2:20pm – 2:25pm

Authors: Bohan Ni (Computer Science, Johns Hopkins University), Alexis Battle (1. Biomedical Engineering, 2. Computer Science, Johns Hopkins University), Michael C. Schatz (Computer Science, Johns Hopkins University)

Abstract: Structural variants (SVs) often have greater functional impact than SNVs, yet transcript quantification typically relies on reference annotations, which overlooks individual gene structure variation. This conflation of altered transcript structure and abundance can obscure both SV detection and regulatory variant discovery. In our analysis of GTEx and 1000 Genomes data, over 90% of expressed genes with exonic SVs exhibit alignment biases when using the reference genome, most notably a lack of RNA-seq reads spanning SV junctions. When SV breakpoints resemble non-canonical splice sites, splice-aware aligners often soft-clip the supporting split reads, further penalizing their alignment. This absence of SV-supporting reads hampers accurate reconstruction of personal transcriptomes and reliable expression quantification.

Addressing this problem, we present STITCH, STructural variant Induced Transcript CHange, to infer personal gene models affected by SVs. First, STITCH infers SV-disrupted personal gene contig using a transcript-aware interval-adjacency graph (T-IAG). This extends the previously described IAG framework (Aganezov et al, BMC Bioinformatics, 2019) to capture SV breakpoints with exon and intron segments. In-silico splice predictions are then used to further refine the putative gene models. In parallel, we realign RNA-seq to the personal gene contigs and assemble the transcripts using StringTie. This hybrid approach allowed systematic catalogue of the variations in personal gene models.

For 56 matched long-read WGS (1000 Genomes ONT; Gustafson et al., Genome Res 2024) and short-read RNA-seq (MAGE; Taylor et al., Nature 2024) samples, STITCH identified SV-altered transcripts in 241 gene-SV pairs, 5× more than reference-only approaches, highlighting the missed transcriptomic impact of SVs. For example, a 52 bp deletion in MIB2 causes intron retention and a predicted stop-gain in three carriers, none of whom are expression outliers. STITCH also identified a truncated GGPS1 transcript (15% of full ORF) due to a rare heterozygous deletion of exons 3-4. The sample is an underexpression outlier ($z = 3.18$), and the deleted region overlaps known pathogenic variants linked to muscular dystrophy, hearing loss, and ovarian insufficiency (Foley et al., Ann Neurol 2020), suggesting it may be a risk allele. Overall, our study demonstrates that personalized annotations have the potential to improve the accuracy of transcriptomic analysis and functional SV discovery.

Session 82: Scaling Genomic Medicine: From Patient-Centered Tools to Population and Therapeutic Insights

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

Session Time: Friday, October 17 at 1:30pm – 2:30pm

The Patient-reported Genetic testing Utility InDEx (P-GUIDE): A novel measure of personal utility for pediatric clinical genetics

Subsession Time: Friday, October 17 at 1:35pm – 1:40pm

Authors: Robin Z. Hayeems (The Hospital for Sick Children, University of Toronto), Elise Poole (The Hospital for Sick Children), Stephanie Luca (The Hospital for Sick Children), Daniel Assamad (The Hospital for Sick Children, University of Toronto), Bowen Xiao (The Hospital for Sick Children), Lesleigh S. Abbott (Children's Hospital of Eastern Ontario), Linlea Armstrong (British Columbia Children's Hospital), Kym M. Boycott (Children's Hospital of Eastern Ontario), June C. Carroll (Sinai Health, University of Toronto), Lauren Chad (The Hospital for Sick Children, University of Toronto), Isabelle De Bie (McGill University Health Centre, McGill University), Avram Denburg (The Hospital for Sick Children, University of Toronto), Rebecca J. Deyell (British Columbia Children's Hospital), Alison M. Elliott (British Columbia Children's Hospital), Catherine Goudie (McGill University Health Centre), Anne-Marie Laberge (Centre Hospitalier Universitaire Sainte-Justine), Bettina Mucha-Le Ny (McGill University Health Centre, McGill University), Iskra T. Peltekova (Holland Bloorview Kids Rehabilitation Hospital), Becky Quinlan (Ontario Genetics Advisory Committee), Sarah L. Sawyer (Children's Hospital of Eastern Ontario), Maureen Smith (Canadian Organization for Rare Disorders), Anita Villani (The Hospital for Sick Children, University of Toronto), Wendy J. Ungar (The Hospital for Sick Children, University of Toronto)

Abstract: Introduction: To inform patient-centred care in genomic medicine, it is critical to understand personal utility, or how patients and families value genetic testing. To meet this need, we developed the Patient-reported Genetic testing Utility InDEx (P-GUIDE) for pediatric clinical genetics, a novel measure of personal utility.

Methods: Informed by the literature, clinical experts, and patient partners, a first draft of P-GUIDE, consisting of 55 items, was generated. Interviews with parents/caregivers of children who had genetic testing for a range of clinical indications assessed item clarity and relevance. Interviewees and clinical experts completed content validity surveys on Draft 2. Two independent samples of parents completed prospective surveys for item reduction, using exploratory factor analysis (Draft 3) and Bayesian confirmatory factor analysis (Draft 4). Analysis of performance characteristics, including construct validity and test-retest reliability, was completed on the final tool.

Results: Interviewees (n=22) found most items to be relevant but required modification. Seventeen parents and 11 clinical experts completed content validity surveys. Feedback on Draft 2 led to the modification of 29 items, removal of 11 items and addition of 3 new items. Exploratory factor analysis

with 103 parents supported a four-factor solution (KMO = .863; Bartlett's p < .001). Based on factor loadings and cross-loadings, 9 additional items were removed. Bayesian confirmatory factor analysis was conducted with a new sample of 102 parents. The final model included 19 items across three factors: (1) understanding and care implications, (2) psychosocial benefit, and (3) psychosocial concern. The model demonstrated acceptable fit: posterior predictive p-value = .05; RMSEA = .08; CFI = .91; TLI = .91. Standardized factor loadings ranged from .58 to .87, and item-level R^2 values ranged from .48 to .75. P-GUIDE scores were positively correlated to published measures of clinical utility, personal utility, knowledge, empowerment and positive emotion, demonstrating construct validity. Test–retest reliability was good to excellent overall (ICC = .86), across factors (.80–.82), and items (.71–.88).

Conclusions: P-GUIDE is a 19-item measure of personal utility for pediatric clinical genetics. P-GUIDE and emerging adaptations of this tool for a range of clinical settings represent a novel utility measurement system for genetic testing that will be available for licensed use.

Population-Scale Genomic and EHR-Driven Insights into Inherited Retinal Disease Burden and Family-based Risk in the Emirati Population

Subsession Time: Friday, October 17 at 1:40pm – 1:45pm

Authors: Budour Alkaf (M42, Emirati Genome Program), Abdelrahman Saad (M42, Emirati Genome Program), Ahmed Al Awadhi (M42, Emirati Genome Program), Albarah El-Khani (M42, Emirati Genome Program), Arif Khan (Cleveland Clinic Abu Dhabi, Emirati Genome Program), Asma Al Mannaei (Department of Health Abu Dhabi), Ayesha Al Ali (M42, Emirati Genome Program), Azza Attia (M42, Emirati Genome Program), Daniel Sanchez (M42, Emirati Genome Program), Fahed Al Marzooqi (M42, Emirati Genome Program), Gurunath Katagi (M42, Emirati Genome Program), Haiguo Wu (M42, Emirati Genome Program), Hakeem Al Mabrazi (M42, Emirati Genome Program), Hanil Sajad (M42, Emirati Genome Program), Imam Chishty (M42, Emirati Genome Program), Javier Quilez (M42, Emirati Genome Program), Joseph Mafofo (M42, Emirati Genome Program), Maaz Shaikh (M42, Emirati Genome Program), Mohamed Alameri (Department of Health Abu Dhabi), Omar Soliman (M42, Emirati Genome Program), Sahar Al Marzooqi (Department of Health Abu Dhabi), Senchao Zhang (M42, Emirati Genome Program), Shilp Purohit (M42, Emirati Genome Program), Shobith Pejathaya (M42, Emirati Genome Program), Shreyaa Chandrashekar (M42, Emirati Genome Program), Val Zvereff (M42, Emirati Genome Program), Vinay Kusuma (M42, Emirati Genome Program), Vineet Kumar (M42, Emirati Genome Program), Wadha Mohammed Abdulrahman (Department of Health Abu Dhabi), Tiago Magalhaes (M42, Emirati Genome Program)

Abstract: Inherited retinal diseases (IRDs) are a clinically and genetically heterogeneous group of disorders causing progressive vision loss. While IRDs affect ~1 in 1,000 individuals, most prevalence and penetrance estimates come from small, European-based cohorts and often overlook phenotypic overlap, gene pleiotropy, and underrepresented populations. Penetrance estimates are often inflated due to ascertainment bias in family-based studies, limiting variant interpretation, risk prediction, and clinical translation. We leveraged the Emirati Genome Program (EGP), which performs ~30x whole-genome sequencing (WGS) of all Emiratis, to estimate variant-level IRD penetrance. We analyzed WGS data from

>500,000 EGP participants linked to nationwide electronic health records via Malaffi to estimate variant-level IRD penetrance. Extended families were reconstructed across the population from genomic data, enabling scalable family-based analyses and identification of high-risk families in a consanguineous population. Vision-related ICD-10 codes (H00-H59) were extracted, excluding common conditions (≥1%). Two variant panels were used: Panel I, with 50 variants identified in Emirati IRD patients at Cleveland Clinic Abu Dhabi; and Panel II, with 16,721 pathogenic/likely pathogenic ClinVar variants in 277 RetNet genes linked to 435 MedGen vision phenotypes. We estimated penetrance (P) for each variant-ICD-10 pair, and weighted them by ICD code specificity (Pw). We identified 27,353 carriers (561 homozygous) in Panel I and 51,587 carriers (1,514 homozygous) in Panel II. Panel II showed higher overall penetrance (P=82.5%, Pw=37.6) compared to those in Panel I (P=79.9%, Pw=29.6). However, Panel I contained stronger variant-phenotype associations, notably KCNV2 (chr9:2718166) with H35 and ABCA4 (chr1:94055128) with H54 (Fisher's p<10^{-15}),>3,000 protein-coding variants in RetNet genes using CADD, PolyPhen, and SIFT scores for downstream segregation analyses. Our study provides population-scale estimates of IRD-related vision loss prevalence and variant penetrance, highlights findings in an underrepresented Arab population, and pioneers the integration of genomics, EHRs, and family analysis at national scale.</10^{-15}),>

When Recurrent Copy Number Variants Are Not the Whole Story: Diagnostic Yield from Alternative Etiologies in a Pediatric Cohort

Subsession Time: Friday, October 17 at 1:45pm – 1:50pm

Authors: Hannah R. Sandler (ASHG), Elizabeth A. Fanning (Children's Hospital of Philadelphia, ASHG), Tiffiney Hartman (University of Pennsylvania, ASHG), Laura K. Conlin (Children's Hospital of Philadelphia, Hospital of the University of Pennsylvania, ASHG)

Abstract: While recurrent copy number variants (rCNVs) contribute to several genetic disorders, the clinical relevance of many rCNVs remains uncertain, with conflicting classifications across laboratories. This discrepancy maybe due to several factors, including publication and ascertainment bias, population prevalence, incomplete penetrance, or unrecognized presence of alternative etiologies due to limited genomic sequencing. Additionally, existing classification guidelines were developed for highly penetrant Mendelian disorders, limiting their applicability to low-penetrance or risk-associated variants. Identification of rCNVs with conflicting interpretations may halt further testing, delay diagnoses, and complicate counseling. To examine the role of rCNVs with conflicting classifications, a 10-year retrospective review was conducted on individuals with rCNVs at four loci (15q11.2, 1q21.1q21.2, 16p13.11, and Xp22.31) identified by the Genomic Diagnostics Laboratory at Children's Hospital of Philadelphia. Data were collected for 384 individuals: 369 probands, 14 unaffected parents, and one parent with a history of learning disability. Most rCNVs were detected by chromosomal SNP microarray (87.6%), followed by exome/genome sequencing (10.6%) and panel testing (1.5%). The most common in this cohort were 15q11.2 duplications/deletions (47.7%). To determine whether these rCNVs were causal for the proband's reason for study, we assessed the rate of alternative diagnoses (alternative diagnostic rate) to the overall diagnostic rate of the testing methodologies. Among 369 probands, 58 (15.7%) had an additional genetic etiology, with 7.0% identified by microarray (24/343) and 26.8% by exome (22/82).

In comparison, among 15,578 individuals without these rCNVs tested during the same period, rates were 12.1% (microarray) and 18.1% (exome). When grouped by the presence of neurodevelopmental phenotypes (>2yr old), 10.3% of both groups were found to have alternative etiologies, while 22.5% of probands under age two had additional etiologies—likely prior to the emergence of neurodevelopmental symptoms. The comparable yields across groups suggest these rCNVs may not contribute meaningfully to neurodevelopmental outcomes and suggest that the rCNVs alone should not be deemed diagnostic. These findings highlight the importance of comprehensive testing, individualized strategies, and continued variant reevaluation to improve diagnoses and care for individuals with rCNVs.

Uncovering Genomic Diversity in the African Diaspora via Long-Read Nanopore Sequencing

Subsession Time: Friday, October 17 at 1:50pm – 1:55pm

Authors: Francisco M. De La Vega (Galatea Bio, Inc.), Paige E. Haffener (Galatea Bio, Inc.), Babak Moatamed (Galatea Bio, Inc.), Michael Lee (Galatea Bio, Inc.), Pavel Salazar (Galatea Bio, Inc.), Monica Campbell (Galatea Bio, Inc., University of Colorado), Deja Hadley (Galatea Bio, Inc.), Jorge Loaisiga (Galatea Bio, Inc.), Sameer Chavan (University of Colorado), Meher Boorgula (Univ. of Colorado Anschutz Med. Ctr.), Christioan Hinson (Univ. of Colorado Anschutz Med. Ctr.), Obed A. Gracia (University of Kansas), Nikki Mourtzi (Galatea Bio, Inc.), Daniel Ansong (Kwame Nkrumah Univ. of Sci. & Technology), Maria Avila-Arcos (Univ. Natl. Autónoma de México), Luis R. Caraballo (Inst. For Immunological Res.), Camila Figueiredo (Unive. Federal da Bahia), Simon Gravel (McGill University), Nadia Hansel (UC Davis), Brenna Henn (UC Davis), Edwin F. Herrera-Paz (Univ. Católica de Honduras), King Jordan (Georgia Inst. of Technology), Jennifer Knight-Madden (The Univ. of the West Indies), Alvaro Mayorga (Univ. Católica de Honduras), Sola Olopade (Univ. of Chicago), Charles Rotimi (NHGRI, NIH), Harold Watson (The Univ. of the West Indies), Margaret Taub (Johns Hopkins University), Eimear Kenny (Icahn Sch. of Med. at Mount Sinai), Jeff Wall (Galatea Bio, Inc.), Omar E. Cornejo (UCSC), Alex Ioannidis (Galatea Bio, Inc.), Nicholas Katsanis (Galatea Bio, Inc.), Carlos D. Bustamante (Galatea Bio, Inc.), Rasika Mathias (NIAID, NIH), and Kathleen Barnes (Univ. of Colorado Anschutz Med. Ctr., Oxford Nanopore Technologies, Ltd.).

Abstract: Current human genome variation resources insufficiently represent global genetic diversity, limiting genotype–phenotype studies and clinical variant interpretation, particularly for individuals of non-European or admixed ancestry. To address this bias, the Consortium on Asthma among African-Ancestry Populations in the Americas (CAAPA) previously conducted short-read whole-genome sequencing (srWGS) on 1,440 individuals from 33 populations across the Americas, the Caribbean, and Africa. Recent advances in long-read sequencing (LRS) enable discovery of structural variants (SVs) often missed by srWGS, as well as variant phasing. We sequenced 400 genomes from the CAAPA cohort using Oxford Nanopore LRS at an average depth of 30X to identify and catalog SVs undetectable by srWGS. Reads were mapped, and small variants and SVs called using Sentieon DNAscope. We identified an average of 17,883 insertions and 13,072 deletions per genome, including 155,672 SVs absent from the 5/2024 release of the 1000 Genomes Project long-read dataset. Among them, 215 hit exons of 160 OMIM and 19 dosage-sensitive genes. Notably, we observed a 972 bp deletion in exon 3 of FLG—a gene essential for skin integrity and associated with immune disorders—in 7% of subjects, and a 2,732 bp

deletion in MBL2, potentially linked to immune dysfunction, at a frequency of 1.8%. FLG exon 3 is particularly challenging to sequence using srWGS due to its repetitive structure. Pharmacogenetic star alleles were identified from long-read-phased calls using PharmCAT and, for CYP2D6, Chinook, a bespoke method. We observed novel star alleles and found that long-read phasing removes the need for a reference panel, resulting in notable differences in metabolizer status inference and highlighting inaccuracies in existing pharmacogenetic predictions for understudied populations. Reads mapped to the HLA Class I & II regions, along with unmapped reads, were realigned to the Human Pangenome v1 graph and projected to hg38 coordinates. HLA allele calls were made using SpecHLA (Class I) and SpecImmune (Class II), from long-read phased SNV and SV calls. The most frequent HLA-A alleles in our sample have been previously observed in Native American and African populations, albeit some at notably different frequencies. Finally, we show that local ancestry inference using long-read phased variants produces longer ancestry segments than unphased data, refining admixture timing estimates.

Integrating 76,000 Whole Genome Sequences from ADSP Release 6 (R6) to Accelerate Alzheimer's Disease Genetics Research

Subsession Time: Friday, October 17 at 1:55pm – 2:00pm

Authors: Wei-Hsuan Chuang (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Wan-Ping Lee (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Amanda B Kuzma (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Heather Nicaretta (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Prabhakaran Gangadharan (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Liming Qu (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Luke Carter (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Youli Ren (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Laura Cantwell (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Otto Valladares (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Naveensri Saravaran (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Yi Zhao (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Taha Iqbal (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Department of Biostatistics, University of Pennsylvania), Mike Schmidt (Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine), Pedro Mena (Department of Human

Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine), Kara L Hamilton-Nelson (Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine), Alzheimer's Disease Sequencing Project, Clifton Dalgard (Department of Anatomy, Uniformed Services University of the Health Sciences), Brian Kunkle (Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine), Will Bush (Department of Population and Quantitative Health Sciences, Case Western Reserve University; Department of Genetics and Genome Sciences, Case Western Reserve University), Eden Martin (Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine), Adam Naj (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Department of Biostatistics, University of Pennsylvania), Jonathan L. Haines (Department of Population and Quantitative Health Sciences, Case Western Reserve University; Department of Genetics and Genome Sciences, School of Medicine, Case Western Reserve University), Margaret A. Pericak-Vance (Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine), Gerard D. Schellenberg (Department of Pathology and Laboratory Medicine, University of Pennsylvania), Li-San Wang (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania), Yuk Yee Leung (Department of Pathology and Laboratory Medicine, University of Pennsylvania; Penn Neurodegeneration Genomics Center, University of Pennsylvania)

Abstract: Background: The Genome Center for Alzheimer's Disease (GCAD) leads the integration and meta-analysis of all available whole genome sequencing (WGS) data relevant to Alzheimer's disease (AD), with the aim of identifying genetic variants associated with disease risk, protection, and potential therapeutic targets. These datasets are produced through a collaborative effort between GCAD and investigators from the Alzheimer's Disease Sequencing Project (ADSP). To minimize data heterogeneity caused by differing sequencing platforms and protocols, GCAD applies standardized processing pipelines and conducts rigorous quality control (QC) procedures across all samples.

Methods: Raw sequencing data (FASTQs or BAMs) are aligned to the GRCh38/hg38 reference genome using BWA, followed by variant calling of SNVs and indels with GATK. The resulting gVCFs are then merged and jointly genotyped. GATK was used for joint genotyping in earlier releases (R1–R4); however, starting with R5, we transitioned to GLnexus for its superior scalability in large-scale genomic data processing. GLnexus will also be used to joint genotype call R6 to be released in early 2026. Variant annotation is performed using the ADSP-developed annotation pipeline and FAVOR. To improve accessibility and usability of the large joint-genotyped datasets, results are distributed in both compact VCF and GDS formats. For structural variants (SVs), sample-level calling is conducted using Smoove and Manta, followed by joint genotyping with GraphTyper2.

Results: The R6 dataset is anticipated to comprise 76,542 whole genomes from over 62 diverse cohorts. The data are deeply sequenced, with an average genome coverage exceeding 30x. In November 2024, a subset of R5 data—comprising CRAM files, GATK gVCFs, and structural variant (SV) VCFs for 58,507 samples— were made publicly available through the NIAGADS Data Sharing Service (https://dss.niagads.org/). Additional R6 samples will be released following completion of joint variant

calling. Joint-genotyped VCFs for SNVs, indels, and SVs will also be provided, following comprehensive QC and annotation.

Conclusion: The ADSP and GCAD continue to deliver high-quality SNV, indels and SV calls. The current phase involves processing more than 76,000 WGS samples, primarily generated through the ADSP Follow-Up Study, with a strong emphasis on ancestral diversity. This upcoming data release is expected to provide valuable resources for advancing genetics research on Alzheimer's disease and related dementias.

OTC-HOPE: The first in vivo, liver directed, AAV-mediated, gene insertion clinical trial in infants with Ornithine Transcarbamylase Deficiency

Subsession Time: Friday, October 17 at 2:00pm – 2:05pm

Authors: Gabriel M Cohn (iECURE, Inc., Blue Bell, PA), Julien Baruteau (Great Ormond Street Hospital for Children, London, UK), Anil Dhawan (King's College Hospital, London, UK), Anupam Chakrapani (Great Ormond Street Hospital for Children, London, UK), Stephanie Grunewald (Great Ormond Street Hospital for Children, London, UK), Molly Abbott (Great Ormond Street Hospital for Children, London, UK), Helen Ashton (Great Ormond Street Hospital for Children, London, UK), Sophie Foxall (Great Ormond Street Hospital for Children, London, UK), Ai-Ling Koh (Great Ormond Street Hospital for Children, London, UK), Christos Lazaridis (Great Ormond Street Hospital for Children, London, UK), Havea Navarro-Kennedy (Great Ormond Street Hospital for Children, London, UK), Hamza Patel (Great Ormond Street Hospital for Children, London, UK), Siyamini Sivananthan (Great Ormond Street Hospital for Children, London, UK), Eleni Tamvaki (Great Ormond Street Hospital for Children, London, UK), Katy Vecchiato (Great Ormond Street Hospital for Children, London, UK), Matthew Hall (iECURE, Inc., Blue Bell, PA), Karen Kuhn (iECURE, Inc., Blue Bell, PA), Thomas White (iECURE, Inc., Blue Bell, PA), George A. Diaz (iECURE, Inc., Blue Bell, PA)

Abstract: Background: Ornithine transcarbamylase deficiency (OTCD), an X-linked disorder, results in impaired ureagenesis and hyperammonemia. Males with severe neonatal onset OTCD often become hyperammonemic within days of birth. Despite standard of care management, recurrent hyperammonemic events (HAEs), which can be neurotoxic and hepatoxic, are common. Mortality rates are high, necessitating liver transplantation as a treatment option. ECUR-506 is an investigational, in vivo, liver-directed, gene insertion therapy being developed for the treatment of OTCD. ECUR-506 is delivered intravenously as a one-time dose, comprised of a mixture of two distinct liver-tropic, AAVrh79 vectors: one encoding a meganuclease, M2PCSK9, for editing of the PCSK9 gene, and the other, a codon-optimized OTC donor gene. Preclinical studies demonstrated that hepatocyte gene integration resulted in durable gene expression and significantly higher survival rates in a murine model of OTCD.

Methods: OTC-HOPE (NCT06255782) is a 24-week, first in human, single arm, Phase 1/2, open-label trial designed to assess the safety and efficacy of ECUR-506 in males <9 months of age with neonatal onset OTCD. Initial data from the 6-month trial and long-term follow-up are described.

Results: To date, five participants have been enrolled. Prior to undergoing gene therapy, the first participant to complete the OTC-HOPE trial experienced two HAEs by 5.5 months of age. ECUR-506 was administered at 6.5 months of age which was generally well tolerated except for Grade 3 transaminitis observed between weeks 4-8 post-infusion, which resolved with immunosuppression. Reduced glutamine levels prompted nitrogen scavenger medication discontinuation and the initiation of protein intake liberalization twelve weeks following ECUR-506 dosing. Over the ensuing 7months of post-treatment observation, the participant experienced no HAEs and mean ammonia levels were within normal limits. The participant, initially a candidate for liver transplantation, was removed from the liver transplant list, and to date has remained off nitrogen scavenger medication and dietary protein restriction.

Conclusion: These observations from the first infant to receive an in vivo, liver-directed, gene insertion, investigational product (ECUR-506) support the continued evaluation of ECUR-506 in the OTC-HOPE Study.

Imsidolimab, a novel high affinity IgG4 IL-36 Receptor Antagonist, was Effective and Well- Tolerated in Patients with Generalized Pustular Psoriasis: Results from Phase 3 trials, GEMINI-1 and GEMINI-2

Subsession Time: Friday, October 17 at 2:05pm – 2:10pm

Authors: Sandra P. Smieszek (Vanda Pharmaceuticals Inc.)

Abstract: Generalized pustular psoriasis (GPP) is a rare, severe disease characterized by debilitating flares of non-infectious pustular and erythematous skin lesions, with systemic impacts that can be life-threatening. Recurrent flares are common. The pathogenesis of GPP can be mainly attributed to excessive activity of IL-36 pathway. Mutations in the gene encoding the IL-36Ra (IL36RN) result in uncontrolled activation of the IL-36 pathway associated with the development of generalized pustular psoriasis (GPP), a rare and life-threatening disease. Imsidolimab, an investigational IgG4 antibody, binds the IL-36 receptor and antagonizes IL-36 signaling. In the Phase 2 GALLOP trial, imsidolimab showed rapid and sustained improvements in symptoms and pustular eruptions with an acceptable safety profile. Reported here are results from the Phase 3 trials, GEMINI-1 (NCT05352893) and GEMINI-2 (NCT05366855).

In GEMINI-1, a randomized, double-blind, placebo (PBO)-controlled, global trial, 45 patients with GPP received a single IV dose of imsidolimab 750mg, imsidolimab 300mg, or PBO. The primary efficacy endpoint at Wk 4 was achievement of GPP Physician Global Assessment (GPPPGA) score of clear/almost clear (0/1) collectively across all GPP disease attributes (pustulation, erythema, scaling), a stringent and comprehensive characterization of disease severity. Patients who were GPPPGA responders, partial responders, or needed rescue therapy (RT) could enroll in GEMINI-2, a follow-on trial of imsidolimab 200mg SC given every 4 wks with evaluation of safety, maintenance of response and prevention of GPP flares. Responders in GEMINI-1 were re-randomized to blinded imsidolimab or PBO, while all partial

responders received imsidolimab. Patients in the GEMINI-1 PBO group who needed RT crossed into GEMINI-2 and received imsidolimab 750mg IV followed by 200mg SC every 4 wks.

GEMINI-1: GPPPGA 0/1 was achieved in 53.3% of patients in the imsidolimab 750mg group, 53.4% in the imsidolimab 300mg group, vs 13.3% of patients on PBO. Among PBO patients that exited GEMINI-1 to receive RT in GEMINI-2, 55.6% attained GPPPGA of 0/1 at Wk 4. GEMINI-2: None of the responders re-randomized to imsidolimab maintenance dosing had a GPP flare, and all maintained a GPPPGA score of 0/1, while in the PBO group, 62.5% flared (GPPPGA ≥ 3), and 75.0% lost GPPPGA 0/1 response. None of the PBO cross-over patients had a flare, and all maintained GPPPGA 0/1. Safety: In GEMINI-1, imsidolimab 300mg and 750mg doses were well tolerated. All adverse events (AEs) in imsidolimab-treated patients were mild or moderate and balanced across imsidolimab- and PBO-treated patients. No treatment-related SAEs or severe AEs were reported. In GEMINI-2, imsidolimab 200mg SC was well tolerated. There were no SAEs leading to discontinuation, and no treatment-related SAEs. Across the trials, there was a low incidence and no elevation of infections vs PBO. There were no reported cases of Drug Reaction with Eosinophilia and Systemic Symptoms or Guillain-Barre syndrome. No infusion reactions were reported. Overall, detection of anti-drug antibodies was uncommon, and none were neutralizing.

Single IV doses of imsidolimab 300mg demonstrated clinically meaningful results and were well tolerated in patients with GPP flares. Maintenance dosing with monthly SC imsidolimab for at least 24 wks was also well tolerated, maintained GPPPGA 0/1 responses and prevented GPP flares.

Imsidolimab is a novel high-affinity humanized immunoglobulin IgG4 mAb that specifically binds the interleukin-36 receptor (IL-36R) and antagonizes interleukin-36 (IL-36) signaling with superior safety profile in comparison to existing treatment options. Targeting IL-36 signaling with imisdolimab represents a promising therapeutic option for GPP patients.

Construct Interpreted Literature Database using LLM

Subsession Time: Friday, October 17 at 2:10pm – 2:15pm

Authors: Laura Li (Breakthrough Genomics Inc.)

Abstract: Background Whole genome sequencing (WGS) and whole exome sequencing (WES) are powerful tools for diagnosing rare diseases. However, these tests generate a large number of variants that require efficient prioritization, manual review, and classification according to ACMG variant classification guidelines. ACMG variant classification involves compiling comprehensive evidence, which is especially time-consuming when it comes to literature review. Method To address this, we investigated whether large language models (LLMs) could systematically extract and summarize key findings from the literature to support automated ACMG variant classification. Specifically, we used LLMs to answer the following essential questions from variant-associated publications: How many patients with the target variant have been reported (ACMG criteria PS4 or BS2)? Are there functional studies related to this variant? What experiments were conducted, and what were the results (PS3 or BS3)? Has any segregation analysis been performed, and does the variant segregate with the disease (PP1 or BS4)? We

evaluated three LLMs—ChatGPT, DeepSeek, and Claude—using a gold-standard dataset of 5,000 unique ClinVar-cited papers for benchmarking and an additional 7,000 papers for re-alignment. These papers were selected based on the presence of variant mentions and their citation by ClinVar submitters for variant classification. Result Our benchmarking showed weighted F1 scores of 90% for Claude, 80% for ChatGPT, and 92% for DeepSeek in correctly classifying variant pathogenicity. Ultimately, we selected Claude to build our literature-based classification database, as it provided more comprehensive explanations supporting its classification decisions. Using Claude, we constructed a literature-derived database classifying pathogenicity for approximately 11–12 million variants. This resource is publicly available. Furthermore, we extended this database to perform automated ACMG classification on 180 million variants by incorporating population data, in silico prediction scores, functional evidence, and family segregation studies. Conclusion We demonstrate that LLMs can effectively automate literature review and support ACMG variant classification. This approach significantly accelerates and enhances the efficiency of clinical interpretation in next-generation sequencing (NGS).

The impact and challenges of more data: retraining of a Bayesian genotype-phenotype model on nearly one million additional patients to resolve VUS

Subsession Time: Friday, October 17 at 2:15pm – 2:20pm

Authors: Trevor J. Williams (Labcorp), Yuya Kobayashi (Labcorp), Ester Borras (Labcorp), Alexander Wahl (Labcorp), Laure Frésard (Labcorp), Toby R. Manders (Labcorp)

Abstract: Background:

Previously we introduced Clinical Variant Modeling (CVM), a machine learning (ML) pipeline trained on a very large real world dataset (3M+ patients) of paired phenotype and genotype information. CVM models covering 28 genes across 22 distinct clinical conditions have contributed to reclassification of >1200 VUS. Since their original release, these CVM models have been retrained using data from nearly 1 million additional patients.

Methods:

CVM is a two-step pipeline composed of multiple ML algorithms. First, patient-level attributes paired with molecular results are used to train a classifier estimating the probability any given patient is affected (the patient score). Second, a variant-level generative, hierarchical Bayesian inference model infers the probability that a given variant is pathogenic (PoP) based on the relevant patient scores. Both steps of the pipeline were trained or retrained using data from up to three timepoints: fall 2023, spring 2024 and winter 2024, with each model containing an initial training (IT) followed by one retraining (RT). The output of each version was vetted by a team of experts and incorporated into a classification workflow. Specifically, CVM predictions were binned into one of three categories: predicted pathogenic (PP; PoP ≥ 0.99), predicted benign (PB; PoP ≤ 0.5), or uncertain (U; 0.99 > PoP > 0.5). Changes in PoP and prediction bin among versions, as well as any resulting variant reclassification, were reviewed.

Results:

On average, the total number of patients and labeled variants per model increased by 33.6% (12.5%-87.4%) and 16.8% (5.3%-29.6%), respectively, between fall 2023 and winter 2024. Models were retrained 6.5 or 14 months post IT, resulting in initial and re-trained predictions for 61,771 variants. IT provided definitive (PP or PB) predictions for 7048 variants; RT increased this number by 57% (4022 new variants). New definitive predictions were driven mainly by: 1) new patient observations for individual variants, and 2) global model enhancements from increased data overall. Definitive predictions remained stable, with 90.1% (6350/7048) retaining the same prediction. Only 0.37% (26/7048) of variants switched from PB to PP or vice versa.

Conclusion:

We demonstrate the benefits and utility of retraining CVM models. RT resulted in large increases in available data that enhanced the global model. These changes resulted in greater numbers of classified variants and stable predictions over time.

3ASC 2.0: A unified deep learning model for comprehensive variant prioritization across WES and WGS in rare diseases

Subsession Time: Friday, October 17 at 2:20pm – 2:25pm

Authors: Kyoungyeul Lee (3billion), Ho Heon Kim (3billion), Wonseok Chung (3billion), Sukwon Kim (3billion), Seungwoo Kim (3billion), Go Hun Seo (3billion), Hane Lee (3billion)

Abstract: As deep learning continues to advance clinical genetics, particularly for rare disease diagnostics, one of the most pressing challenges is the labor-intensive interpretation of extensive variant lists—requiring concurrent evaluation of diverse variant types such as single nucleotide variants (SNVs), small insertions and deletions (INDELs), copy number variants (CNVs), short tandem repeats (STRs), inversions (INVs), breakends (BNDs), and mobile element insertion (INSs). To address this, we developed 3ASC 2.0, a variant prioritization tool trained on an integrated dataset of whole exome (WES) and whole genome sequencing (WGS) samples from 47,890 patients with suspected rare diseases. All variants were annotated and featurized following the ACMG/AMP guidelines using an internal bioinformatics pipeline, EVIDENCE.

Our model utilizes a transformer-based architecture trained with multiple instance learning (MIL) and multi-task learning (MTL) to cohesively analyze and rank diverse variant types simultaneously. This single deep learning model was ensembled with several Gradient Boosting Machine (GBM)-based models, each trained on a specific variant type and combined using different weights. The model demonstrated exceptional performance across both WES and WGS data, achieving an average Area Under the Receiver Operating Characteristic (AUROC) of 99% for discriminating causal variants. Furthermore, through probability calibration, the model categorizes variants into three confidence tiers. The 'High' confidence tier successfully identified 76% of causal variants while representing only 1% of all variants requiring review. The 'Middle' and 'Low' tiers contained 22% and 2% of causal variants in 9% and 90% of the total variants, respectively. This stratification dramatically narrows the search space for clinicians. Crucially, in

benchmarks for SNV/INDEL prioritization, 3ASC 2.0 showed superior performance, with a top-5 variant recall of 96% compared to 55% for Exomiser and 40% for LIRICAL.

This work demonstrates that a unified deep learning model can accurately prioritize a comprehensive spectrum of variants from different sequencing platforms. By offering a scalable solution proven to outperform existing tools, 3ASC 2.0 significantly reduces the manual curation burden and accelerates the diagnostic process for individuals with rare genetic disorders.