ASHG 2025 Plenary Abstracts

As of September 24, 2025

Instructions

1. View the Table of Contents on the next page.
    a. The Table of Contents is sorted chronologically by the presentation date/time.
2. Click on the Abstract 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 10: Featured Plenary Abstract Session I

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

Session Time: Tuesday, October 14 at 5:00pm – 6:30pm

⭐ Measuring the functional impact of 5'UTR variants in human disease

Subsession Time: Tuesday, October 14 at 5:00pm – 5:18pm

Authors: Srikar Gopinath (Yale University), Tami Gjorgjieva (Stanford University), Ethan Strayer (Yale University), Anthony Wilder Wohns (Stanford University), Kenneth Ng (Yale University), Haejeong Lee (Yale University), Monkol Lek (Yale University), Jonathan Pritchard (Stanford University), Antonio J Giraldez (Yale University)

Abstract: The 5' Untranslated Regions (UTRs) contain diverse regulatory elements that play a crucial role in determining the protein output of a gene. Variations in these elements can significantly affect protein output, with far-reaching implications for human diseases. Despite the fact that nearly 95% of disease-associated mutations occur in “non-coding regions,” including 5' and 3' UTRs, research has traditionally focused on understanding variations in the coding sequence. This is primarily due to a lack of robust methods to measure the functional effects of non-coding variations.

To bridge this gap in knowledge, we have adapted a novel massively parallel reporter assay called Nascent Peptide-Translating Ribosome Affinity Purification (NaP-TRAP) to quantify the translational consequence of over one million 5’UTR variants identified across ~ 17,000 genes from UK Biobank and gnomAD. NaP-TRAP is an immunocapture-based method that enables sensitive measurements of protein output by capturing the mRNAs associated with actively translating ribosomes. By integrating NaP-TRAP with machine learning, we have identified critical 5' UTR regulatory features and elements that modulate protein output. Specifically, our analysis reveals the functional effects of variants altering sequence motifs and novel 5'UTR structures, extending beyond the well-characterised elements such as upstream open reading frames (uORFs). Notably, we uncovered “fail-safe” mechanisms in the 5' UTR that buffer against mutations in the start codon, providing new insights into how such mutations may be tolerated in clinical contexts.

Overall, our study maps the translational impact of non-coding variants in the 5’UTR across disease-related genes and highlights candidate variants for further clinical studies. Strikingly, we find that variants with strong effects on translation in oncogenes and tumorsupressors are often catalogued as somatic variants in COSMIC, underscoring the importance of 5’UTR variants in cancer biology. We are currently using these results to develop a model to predict the effect of 5'UTR variation on protein expression to inform clinical genetics. Together, our findings emphasize the need to focus on non-coding regions in the molecular interpretation of human disease.

Comprehensive Proteomic Characterization of GLP1R Therapy: Leveraging Mendelian Randomization and Colocalization with Validation in Clinical Trials

Subsession Time: Tuesday, October 14 at 5:18pm – 5:36pm

Authors: Dmitry Shungin (Novo Nordisk A/S, Denmark), Neil Robertson (Novo Nordisk Research Centre Oxford, UK), April Elizabeth Hartley (Novo Nordisk Research Centre Oxford, UK), Joao Fadista (Novo Nordisk A/S, Denmark), Jack Bowden (Novo Nordisk Research Centre Oxford, UK, University of Exeter, UK), Yalda Jamshidi (Novo Nordisk Research Centre Oxford, UK)

Abstract: The glucagon-like peptide-1 receptor (GLP1R) plays a crucial role in metabolic regulation, and its pharmacological agonism is used in treatment of obesity and type 2 diabetes. Early identification of blood proteins causally associated with GLP1R agonism, that could potentially serve as pharmacodynamic biomarkers (BMs), is essential for enhancing therapeutic efficacy and developing novel combination therapies. We performed a comprehensive characterization of protein levels for GLP1R agonism using Mendelian Randomization (MR) & Bayesian colocalization and validated these results in blood protein data from the STEP1 & STEP2 phase 3 trials for semaglutide treatment.

To define exposures we systematically examined cis- expression quantitative trait loci for GLP1R (eQTL, +/- 500kb & $P<5x10^{-8}$) from multiple resources including 52 tissues from 14 studies and 7 meta-analysis for eQTLs; and performed MR (gIVW, GMM-PCA or Wald) against each of 2,932 blood proteins from UK Biobank (N=48,645) and 4,567 from deCODE (N=35,559) followed by colocalization (coloc). Identified proteins were validated in two phase 3 trials that measured proteomic changes from baseline to 68 weeks of treatment with semaglutide in participants with overweight or obesity, and additionally with diabetes (STEP 2, n = 645) or without (STEP 1, n = 1,311), including analyses adjusted for weight loss. Finally, we studied the causal effect of BMI on all blood proteins using MR (IVW with sensitivity analyses and Steiger filtering, LD $r^2<0.001$) and systematically compared these results with protein changes during treatment in STEP1 and STEP2.

Our analyses identified CPA1 and PNLIPRP1 proteins as two putative pharmacodynamic BMs reflective of GLP1R agonism when proxied by eQTLs ($P_{MR}<1.6x10^{-5}$ & coloc posterior probability>0.9 for both) with increased GLP1R expression leading to increased protein levels. Importantly, the direction of effect from MR was consistent with increased levels for both proteins in STEP1 and STEP2 trials after 68 weeks of semaglutide treatment ($P<3.2x10^{-42}$& $P<7.3x10^{-33}$ respectively). These two proteins were not causally modulated by BMI in MR analysis ($P_{MR}>0.05$) and were minimally affected by adjustments for weight loss in both trials, confirming that their levels are not mediated by weight loss. Further MR analysis identified 714 proteins that were causally modulated by BMI (Bonferroni $P_{MR}<2.5x10^{-6}$), and the effect size for these proteins was highly consistent both in direction and magnitude with protein changes observed due to GLP1R agonist-associated weight loss in STEP1 and STEP2 trials (P for directional enrichment=$4x10^{-19}$, Spearman Rho for effect sizes=0.83, P<0.0001), validating in trials ability of proteogenomic data to identify potential disease BMs.

Our findings have two primary implications: (1) they propose potential novel pharmacodynamic biomarkers of GLP1R agonism; and (2) our comparison of MR and colocalization results with data from two trials uniquely validates application of these approaches in molecular data for early discovery of biomarkers and drug targets.

Integrative multi-omics analysis of spaceflight-induced physiological adaptations

Subsession Time: Tuesday, October 14 at 5:36pm – 5:54pm

Authors: JangKeun Kim (Weill Cornell Medicine), Eliah Overbey (Weill Cornell Medicine, UATX) , Jeremy Wain Hirschberg (Weill Cornell Medicine), Jiwoon Park (Weill Cornell Medicine), Jacqueline Proszynski (Weill Cornell Medicine), Paul Collier (Weill Cornell Medicine), Cem Meydan (Weill Cornell Medicine), Krista A. Ryon (Weill Cornell Medicine), Christina Caragine (New York Genome Center, New York University), Neville Sanjana (New York Genome Center, New York University), Jasmine Plummer (St Jude's Children Hospital), Luciano G. Martelotto (University of Adelaide), Iwijn De Vlaminck (Cornell University), Marissa Rosenberg (SpaceX), Jaime Mateus (SpaceX), Bader Shira (King Faisal Specialist Hospital & Research Centre), Joseph Borg (University of Malta), Irina Matei (Weill Cornell Medicine), David C. Lyden (Weill Cornell Medicine), Christopher E. Mason (Weill Cornell Medicine)

Abstract: The recent expansion of commercial, private, and multinational space missions has highlighted the urgent need for precision aerospace medicine. To address this gap, we leveraged the Space Omics and Medical Atlas (SOMA) datasets from the Inspiration4, Polaris Dawn, and Ax-2 missions to define spaceflight-induced molecular signatures. We analyzed over 1,000 astronaut biospecimens using high-resolution multi-omics approaches, including proteomics (Olink Reveal and ILMN protein prep), epigenomics (whole-genome bisulfite sequencing and Oxford Nanopore long-read DNA methylation), and genomic sequencing (Ultima, Illumina, and ONT). In parallel, peripheral blood mononuclear cells underwent single-cell multi-omics profiling (DOGMA-seq, whole-blood single-cell transcriptomes, and spatial proximity assays). CRISPR knockout screens and perturbation sequencing in K562 cells were applied to identify key regulators of the cellular stress response, while complementary metabolomic, cell-free DNA/RNA, and spatial multi-omics assays provided orthogonal validation. Our integrative analysis revealed consistent spaceflight-associated alterations across missions, including modulation of stress-response and oxidative phosphorylation pathways, shifts in immune signaling, activation of DNA damage-repair mechanisms, changes in mitochondrial function, and perturbations in hematopoietic activity. Notably, we observed reproducible cytokine profile shifts, telomere length dynamics, and transcriptional rewiring in multiple astronaut cohorts. CRISPR perturbations pinpointed candidate targets for therapeutic countermeasures. These findings deliver both temporal and mechanistic insights into astronaut physiological adaptations, laying the groundwork for enhanced health monitoring, risk mitigation, and the development of targeted interventions for extended lunar and interplanetary exploration.

Identifying candidate genes and potential drug targets for central obesity through a multi-population GWAS meta-analysis in the GIANT Consortium

Subsession Time: Tuesday, October 14 at 5:54pm – 6:12pm

Authors: Emma P. Wilson (University of North Carolina at Chapel Hill), Daeeun Kim (University of North Carolina at Chapel Hill), Shreyash Gupta (Geisinger Health System), Virginia Diez-Obrero (Novo Nordisk Foundation Center for Basic Metabolic Research), Kristin L. Young (University of North Carolina at Chapel

Hill), Mariaelisa Graff (University of North Carolina at Chapel Hill), Eirini Marouli (Queen Mary University of London), Sailaja Vedantam (Broad Institute of MIT and Harvard, Boston Children's Hospital), Eric Bartell (Broad Institute of MIT and Harvard, Boston Children's Hospital, Harvard Medical School), Qianqian Liang (Geisinger Health System), Geetha Chittoor (Geisinger Health System), Victor Svenstrup (Novo Nordisk Foundation Center for Basic Metabolic Research), Alfred Pozarickij (University of Oxford),

Kari North (University of North Carolina at Chapel Hill), Ruth J.F. Loos (Novo Nordisk Foundation Center for Basic Metabolic Research), Sonja I. Berndt (National Cancer Institute), Robin G. Walters (University of Oxford), Anne E. Justice (Geisinger Health System), Karen L. Mohlke (University of North Carolina at Chapel Hill), on behalf of the Genetic Investigation of ANthropometric Traits (GIANT) Consortium

Abstract: GWAS for central obesity have identified many risk loci, but most underlying mechanisms remain to be defined. The GIANT Consortium performed GWAS meta-analyses for waist-to-hip ratio adjusted for body mass index (WHR) in 1.18 million individuals, identified conditionally distinct signals ($p<5e-9$) in 18 meta-analyses by sex and population, and identified distinct signals across meta-analyses based on linkage disequilibrium. We identified 1,009 signals, 49% of which showed significant sex effect heterogeneity (FDR 5%). These signals were enriched for expression in 38 cell types related to adipose, as expected, but also to the musculoskeletal, heart, reproductive, endocrine, and digestive systems.

To link WHR signals to genes, we used colocalization with eQTL of >1,000 samples from adipose, liver, skeletal muscle, cortex, and blood. Among the 708 multi-population sex-combined WHR signals, we colocalized 308 (44%) with eQTLs for 539 genes; 64% of these genes were identified in adipose, 18% in liver, and 27% in more than one tissue. Among the 587 female WHR signals, we colocalized 258 (44%) with eQTLs for 442 genes; 59% were identified in adipose, 22% in liver, and 24% in more than one tissue. Among the 202 male WHR signals, we colocalized 90 (45%) with eQTLs for 178 genes; 62% were identified in adipose, 21% in liver, and 31% in more than one tissue. 225 genes were only identified using sex-stratified analyses and may reflect sexually dimorphic genetic influences on WHR.

Integrating genes detected by eQTL colocalization with those detected by gene prioritization methods DEPICT and PoPS, we identified 1,339 candidate genes for WHR. Among these genes, at least 72 are approved drug targets and 47 are in advanced clinical trials; these drugs may also influence central obesity. One candidate gene, LAMB1, that encodes a laminin extracellular matrix glycoprotein, showed evidence for WHR signal colocalization with eQTLs in liver, muscle, and adipose. CRISPR interference in hepatocytes linked a variant-containing regulatory element to LAMB1 expression, and transcriptional reporter assays showed significant allelic effects across cell types. Similar integrative omics data may inform hundreds of additional molecular mechanisms for WHR.

Together, these results further our understanding of the biology of central obesity, nominating 1,339 candidate genes that may act through different tissues, explain sex differences, and contribute to WHR.

Comparing the Diagnostic Capability of Large Language Models and Clinical Geneticists

Subsession Time: Tuesday, October 14 at 6:12pm – 6:30pm

Authors: Ivy Bethea (Program in Genetic Counseling, Columbia University), Andres Morales Corado (Department of Pediatrics, Columbia University Irving Medical Center), Priyanka Ahimaz (Department of Pediatrics, Columbia University Irving Medical Center)

Abstract: As demand for genetic expertise grows, access to clinical geneticists (CGs) and genetic counselors remains limited. Non-genetics providers (NGPs) are often on the front lines of evaluating patient concerns and frequently report feeling unprepared to determine when a genetics referral is needed. Large language models (LLMs) could help fill this gap and serve as referral support tools. Assessing their diagnostic accuracy is a critical first step. Our study aimed to compare the diagnostic accuracy of three LLMs (ChatGPT-3.5, Llama, Gemini) and board-certified CGs using clinical vignettes. Forty vignettes representing a range of genetic disorders were formatted into an online quiz. A total of 312 CGs across the United States were invited to participate; 14 completed the quiz (response rate: 4.5%). Respondents were asked to provide the most likely diagnosis, a differential of 2–3 conditions, and demographic information. The total score that could be obtained with each question was 2, for a total score of 80. Responses were evaluated by three independent geneticists. The Mann-Whitney U test was used to compare scores between LLMs and CGs, and among CGs based on experience, subspecialty training, and case difficulty ($\alpha \leq 0.05$). The mean diagnostic score for LLMs was 66.6 (SD = 1.42), compared to 53.1 (SD = 7.9) for CGs ($p = 0.027$). LLMs ranked the correct diagnosis first in 71.7% of cases, compared to 57.1% for CGs ($p = 0.035$). When not ranked first, LLMs included the correct diagnosis in 13.2% of differentials, compared to 10.0% for CGs ($p = 0.407$). LLMs had fewer complete diagnostic misses (8.3%) than CGs (23.2%) ($p = 0.032$). Adjacent conditions were included in 6.6% of LLM responses vs. 9.5% of CG responses ($p = 0.242$). No significant differences were observed in performance for high-difficulty vignettes ($p = 0.1$), between CGs with >12 years of experience vs. fewer ($p = 0.6166$), or between those with and without metabolic subspecialty training ($p = 0.435$). These findings suggest that LLMs can achieve high diagnostic accuracy and may serve as valuable support tools in genetics. Their consistent inclusion of the correct diagnosis, even when not ranked first, may help broaden differential diagnoses and reduce missed referrals, especially in clinical settings with limited access to genetics professionals. Future research should explore how LLMs can be integrated into real-world workflows and how providers engage with these tools across clinical subspecialties.

Session 60: Featured Plenary Abstract Session II

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

Session Time: Thursday, October 16 at 5:00pm – 6:30pm

Supernumerary X chromosomes shape brain organoid architecture and functions in a dose-dependent fashion

Subsession Time: Thursday, October 16 at 5:00pm – 5:18pm

Authors: Veronica Astro (King Abdullah University of Science and Technology, KAUST), Angels Almenar (University of California San Diego, UCSD), Rawan Alghamdi (King Abdullah University of Science and Technology, KAUST), Kelly Yojanna Cardona-Londoño (King Abdullah University of Science and Technology, KAUST), Gabriel Herrera Lopez (King Abdullah University of Science and Technology, KAUST), Ivan Garcia Bassets (University of California San Diego, UCSD), Pierre Magistretti (King Abdullah University of Science and Technology, KAUST), Alysson Muotri (University of California San Diego, UCSD), Antonio Adamo (King Abdullah University of Science and Technology, KAUST)

Abstract: Klinefelter syndrome (KS, 47,XXY) is the most prevalent aneuploidy in males (1:400-1:600). High-grade sex chromosome aneuploidies (HGA-SCAs), such as 48,XXXY, and 49,XXXXY are rarer conditions occurring in 1:40.000-1:80.000 males. KS and HGA-SCA patients exhibit a broad spectrum of neuronal impairment, including cognitive deficits, seizures, autistic traits, and motor, speech, and language delays. While KS patients typically display a milder phenotype, HGA-SCAs are associated with profound cognitive defects. Despite the prevalence of X chromosome aneuploidies, there is a critical need for cellular models to define the transcriptional, epigenetic, and functional consequences of X chromosome overdosage during neurodevelopment. To this end, we derived cortical organoids from 47,XXY, 48,XXXY, and 49,XXXXY iPSCs. Allele-specific expression (ASE) analysis on X aneuploid organoids demonstrated a preserved epigenetic X inactivation status at different time points, from one to 12 months of differentiation in vitro. Through a multi-layered analysis integrating morphological, functional, bulk, and single-cell transcriptomics, we found that the additional X chromosomes lead to impaired neural patterning, disrupted cortical architecture, and altered electrophysiological properties of cortical organoids in a dose-dependent manner. While 47,XXY organoids are phenotypically and functionally similar to 46,XY controls, HGA-SCAs display severe functional defects and aberrant transcriptomes. Through single-cell RNA analysis, we profiled the genes that escape X inactivation in neuronal and non-neuronal cell populations and revealed a dysregulated proliferation of neural progenitor in organoids carrying supernumerary X chromosomes. Additionally, severe astrocyte differentiation defects were observed in HGA-SCAs organoids, potentially contributing to synaptic dysfunction. Moreover, high-density microelectrode arrays (MEA) analysis revealed a higher mean spike firing rate and amplitude of HGA-SCAs compared to 46,XY organoids. Finally, patch-clamp studies demonstrated significant hyperexcitability of HGA-SCA organoids and X dosage-sensitive deficits in long-term potentiation (LTP). Our work leveraged the inaugural cohort of X aneuploid cortical organoids to unravel the functional consequences of X-linked gene overdosage during neurodevelopment.

⭐ Identification of Unannotated mRNA Isoforms Driving Gene Regulation and Disease from Large-Scale RNA-seq data

Subsession Time: Thursday, October 16 at 5:18pm – 5:36pm

Authors: Carlos Fernando Buen Abad Najar (University of Chicago, Section of Genetic Medicine), Dongyue Xie (University of Chicago, Department of Statistics), Peter Carbonetto (University of Chicago, Department of Human Genetics), Ru Feng (Columbia University, Center for Statistical Genetics), Gao Wang (Columbia University, Center for Statistical Genetics), Matthew Stephens (University of Chicago, Department of Statistics), Yang I. Li (University of Chicago, Section of Genetic Medicine)

Abstract: Current RNA-seq methods for quantifying isoform expression depend heavily on pre-existing transcript annotations. Yet, thousands of unannotated isoforms—many potentially disease-causing—are actively expressed in each sample, remaining invisible to traditional approaches. To solve this limitation, we introduce Torino, a revolutionary Bayesian matrix factorization model that leverages biobank-scale RNA-seq data to directly decode transcript structures and expression levels from read coverage alone.

Torino models RNA-seq data using Poisson non-negative matrix factorization with spatial smoothness priors, enabling it to infer latent transcript structures without annotations. This approach uncovers novel isoforms shaped by alternative splicing, intron retention, and alternative polyadenylation (APA). Applied to 2,128 GTEx samples across 19 tissues, Torino accurately recovers 18,813 GENCODE isoforms spanning 15,232 protein-coding genes, while revealing extensive unannotated diversity: 23,413 cassette exon events (>10,000 novel), 83,553 intron retention events (>53,000 novel), and 8,013 APA events. Thousands of these events exhibit strong tissue specificity, suggesting functional roles.

By harnessing Torino-inferred isoform abundances as quantitative traits, we discovered a median of 2,829 isoform QTLs per tissue, demonstrating widespread genetic control over RNA processing. Colocalization with 65 GWAS traits pinpointed 815 disease-linked variants overlapping isoQTLs, implicating unannotated splicing events as hidden disease drivers. The Parkinson's-associated variant rs1045599 disrupts alternative polyadenylation in ZSWIM7, while rs4074793 triggers aberrant intron retention in ITGA1—mechanistically bridging noncoding variants to gene regulation.

Applied to 1,193 brain samples from the Alzheimer's Disease Functional Genomics Consortium, Torino uncovered a global increase in intron retention tied to AD diagnosis and Braak stage, including aberrant splicing in AD risk genes PTK2B and APBB3, implicating widespread RNA processing defects in disease progression.

Torino enables transcriptome-wide discovery of novel, functionally important isoforms and their regulation at unprecedented scale and resolution—bridging RNA processing complexity to genetic architecture and disease. By removing the constraint of annotations, it opens new frontiers in decoding the regulatory genome across tissues, species, and complex diseases.

Reimbursement of germline genetic testing for cancer is impacted by socioeconomic characteristics and insurer type

Subsession Time: Thursday, October 16 at 5:36pm – 5:54pm

Authors: Erica M. Vaccari (Labcorp (formerly Invitae Corporation)), Trevor J. Williams (Labcorp (formerly Invitae Corporation)), Edward D. Esplin (Labcorp (formerly Invitae Corporation))

Abstract: Background: Germline genetic testing (GGT) for patients with ovarian (OV), pancreatic (PANC), early-onset colorectal, endometrial, and breast cancer (CRC <50, ENDO <50, and BR ≤50) is the standard of care per clinical guidelines. We report a single national laboratory experience with GGT reimbursement for these indications and identify specific areas where gaps in reimbursement may exist.

Methods: Patients with GGT between June 2023 and July 2024 from a commercial laboratory tested with one of two multi-cancer panels were stratified by cancer type (using ICD-10s), age at testing, primary payer type "PPT" (commercial, Medicare, or Medicaid), clinician-reported race and ethnicity (R/E), and area deprivation index (ADI). Differences in reimbursement rates were assessed across these attributes. Reported p-values are from G-Tests or likelihood ratio tests.

Results: We reviewed 7,681 patients with cancer, 22.8% OV, 33.2% PANC, 13.7% CRC <50, 2.5% ENDO <50, and 27.8% BR ≤50. The cohort was 29.9% non-white R/E. Overall, GGT was not reimbursed for 22.9% across all cases. We observed differences in reimbursement by PPT, cancer type, clinician-reported R/E, and ADI.

Individuals with Medicaid PPT were less likely to have their testing reimbursed ($G = 993.0$, $p < 0.0001$). GGT was not reimbursed for 21.2%, 12.3%, and 67.1% of individuals with commercial, Medicare, and Medicaid PPT, respectively. Testing was not reimbursed for 20.6% of OV, 17.5% of PANC, 30.6% of CRC <50, 32.0% of ENDO <50, and 26.7% of BR ≤50 ($G = 108.5$, $p < 0.0001$).

Reimbursement also varied by clinician-reported R/E ($G = 112.6$, $p < 0.0001$). Individuals with white (80.0%) R/E were most likely to have their testing reimbursed while individuals with black (71.3%) or hispanic (62.6%) R/E were the least likely to have their testing reimbursed.

Individuals with higher ADI were less likely to have reimbursement ($X^2 = 70.0$, $p <0.0001$), with an OR (95% CI) = 0.9 (0.879, 0.923) for every increase of 10 points in ADI.

Conclusion: During this study period, claims for ~23% of individuals for whom clinical guidelines commonly affirm the medical necessity of GGT were not reimbursed. We observed differences in reimbursement by PPT, cancer type, clinician-reported R/E, and ADI. These findings underscore that significant gaps and complexities related to socioeconomic characteristics continue to impact the reimbursement landscape for GGT.

Pervasive interaction between HLA, KIR, and TCR for autoimmune diseases in > 1 million individuals across diverse populations

Subsession Time: Thursday, October 16 at 5:54pm – 6:12pm

Authors: F. Zhang, A.R. Diwadkar, L. Wang, L. Carrel, D Liu; Penn State Coll. of Med., Hershey, PA

Abstract: The human leukocyte antigen (HLA) genes are broadly expressed across diverse cell types and associated with virtually all autoimmune diseases. HLA class I genes interact with killer-cell immunoglobulin-like receptors (KIR), while HLA class II genes interact with T cell receptors (TCR) to influence immunity and cause autoimmune diseases. The effects of HLA alleles may differ in the presence of interacting KIR alleles or TCR repertoires. Despite their importance in autoimmune disease etiology, HLA x KIR/TCR have not been comprehensively evaluated in large datasets, which is an important gap.

To investigate these interactions, we genotyped ~300 HLA and ~100 KIR alleles with frequency > 1% using T1K in the UK Biobank and All of Us, which have close to 1 million individuals from diverse ancestries with whole genome sequences. We encode HLA and KIR alleles by their presence/absence and perform Firth corrected logistic regression to estimate the main and interaction effects for HLA and KIR alleles. Moreover, we have shown previously that the TCR repertoire forms functionally informative clusters which we call repertoire functional unit (RFUs). RFU frequencies are heritable. We also predict RFUs from DNA genotype data and examine HLA x TCR interactions in UK Biobank and All of Us as well.

Biobanks often have a limited number of disease cases and the study of HLA x KIR/TCR interactions can be underpowered. The largest GWAS datasets come from consortium studies but are available as summary statistics. To boost power, we propose a novel method HIKE (HLA interaction with KIR effects) to model the marginal HLA or KIR effects as weighted sums of conditional HLA or KIR effects given the presence or absence of interacting alleles. This new insight allows us to jointly analyze biobanks with case-control GWAS summary statistics and maximize power.

Applying HIKE to analyze All of Us, UK Biobank, and GWAS summary statistics from 14 common autoimmune diseases (with a maximal N of 2.3 million), we observe up to 95% decrease in the variance of the interaction effects equivalent to 19.34-fold increase in effective sample size. More than 90.55% of the 294,768 allele–trait combinations showing improved precision. Among 409 significant HLA–KIR–trait interactions ($P < 5\times10^{-8}$), 254 (62.1%) involved HLA alleles not previously associated with the trait in the GWAS Catalog, highlighting the discovery potential of interaction-driven models beyond marginal HLA effects. For example, we observed strong HLA–KIR interaction signals involving multiple HLA-C alleles for ankylosing spondylitis, notably HLA-C*01:02:01 and HLA-C*02:02:02, which are not reported in the GWAS Catalog. These alleles showed highly significant interactions with KIR alleles such as KIR2DL1*002 and KIR2DS4*003 (e.g., $-\log_{10}P > 17$), suggesting receptor-specific modulation of HLA-C–driven immune risk.

For the larger set of HLA models (~4.93 million), we observed up to 317 fold gain in effective sample sizes. These improvements enabled stable and interpretable inference across a wide range of traits, including previously underpowered interaction terms due to rare allele combinations or inflated standard errors. Our results demonstrate the pervasive effects of HLA x KIR/TCR interactions in autoimmune disease etiology.

Human-specific tandem repeat in CACNA1C modulates responses to neuronal stimulation

Subsession Time: Thursday, October 16 at 6:12pm – 6:30pm

Authors: Janet H.T. Song (Harvard University), Fikri Birey (Emory University), Sriram Jayabal (Stanford University), Tzu-Chiao Hung (Stanford University), Nicola A.L. Hall (University of Oxford), Catherine A. Guenther (Stanford University), Xiaoyu Chen (Stanford University), Ibrahim Alkuraya (Harvard University), Elizabeth M. Tunbridge (University of Oxford), Wilfried Haerty (University of Oxford), Jennifer L. Raymond (Stanford University), Sergiu P. Pasca (Stanford University), David M. Kingsley (Stanford University, Howard Hughes Medical Institute)

Abstract: The recent development of long-read sequencing has made it possible to catalog variable number tandem repeats (VNTRs) in the human genome. However, little is known about their functional consequences. Here, we characterized the effect of TRACT, a VNTR that is unique to humans, composed of 100-1000+ 30bp repeats, and has sequence variants linked to risk for bipolar disorder and schizophrenia. By adding or removing this VNTR in both mouse models and human organoids, we find that TRACT, which is intronic to the L-type voltage-gated calcium channel gene CACNA1C, increases the amplitude of calcium influx after stimulation and leads to widespread changes in activity-dependent transcription programs in neurons. Consistent TRACT-dependent changes are enriched for genes associated with synapse formation and plasticity, and partially recapitulate evolutionary changes in activity-dependent transcription between species. Long-read RNA sequencing revealed that these changes are likely driven by TRACT-dependent changes to CACNA1C isoform expression. Further, we find that mice containing a version of TRACT associated with risk for bipolar disorder and schizophrenia have impaired eye tracking behavior, a highly penetrant phenotype observed in both neuropsychiatric patients and their first-degree, unaffected relatives. These findings demonstrate that a single, human-specific, non-coding element can strongly affect the neuronal response to stimulation as well as disease-associated behavior, and motivate the study of VNTRs as a genetic source of phenotypic variation in both evolution and disease.

Session 83: Featured Plenary Abstract Session III

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

Session Time: Friday, October 17 at 5:00pm – 6:30pm

Haplotype-resolved chromatin differences and genome structural variation

Subsession Time: Friday, October 17 at 5:00pm – 5:18pm

Authors: Lingbin Ni (Department of Genome Sciences, University of Washington School of Medicine), Jiadong Lin (Department of Genome Sciences, University of Washington School of Medicine), Chong Li (Department of Computer and Information Sciences, College of Science and Technology, Temple University; Institute for Genomics and Evolutionary Medicine, Temple University), Xinghua Shi (Department of Computer and Information Sciences, College of Science and Technology, Temple University; Institute for Genomics and Evolutionary Medicine, Temple University), Evan E. Eichler (Department of Genome Sciences, University of Washington School of Medicine; Howard Hughes Medical Institute, University of Washington)

Abstract: Structural variants (SVs) located in distal cis-regulatory elements have the potential to disrupt 3D chromatin architecture and gene expression, leading to functional changes associated with disease and evolution (e.g., duplication near SOX9 locus; deletion near LMNB1 locus in humans; rearrangements near FOLH1 in primates). Most previous analyses map data against standard reference genomes without consideration of the SV differences or genetic diversity between datasets and the reference. Here, we specifically construct donor-specific assemblies (DSAs) and compare HiC contact maps between DSAs where maternal and paternal genomes are resolved as compared to standard reference genomes. Using 47 diverse humans from the Human Pangenome Reference Consortium (HPRC) along with matched deep Hi-C sequencing data, we observe a significant improvement (nearly 1% novel valid contacts in haplotypes) in DSAs mapping with ~15% of HiC contacts showing >10 kbp distance changes when compared to standard references. From the haplotype-resolved genomes, we discover on average 525 (paternal) and 585 (maternal) contact difference hotspots (CDHs) between haplotypes and show that these CDHs often correspond to SV differences (~45% overlap with SVs). We show that a subset of these alter topologically associating domain boundaries, suggesting potential changes of 3D genomes and gene regulation specific to haplotypes. Notably, some of these differences are specific to superpopulations, confirming the importance of using haplotype-resolved DSAs to characterize the 3D architecture of human genomes. In addition to understanding the population dynamics of the 3D architectural changes, our strategy provides a new approach to identify SVs of functional consequence with respect to disease and evolution.

Accurate representation of globally diverse human haplotypes in the second release of the human pangenome reference

Subsession Time: Friday, October 17 at 5:18pm – 5:36pm

Authors: Benedict Paten (UCSC), Julian K Lucas (UCSC), Karen Miga (UCSC), Human Pangenome Reference Consortium

Abstract: The Human Pangenome Reference Consortium (HPRC) is creating a high-quality, globally representative human reference genome that overcomes limitations of the current linear reference. We report the second HPRC data release (Release 2), a major milestone comprising phased, high-accuracy assemblies from 232 individuals, representing 470 haplotypes, and a fivefold increase from Release 1. These genomes span global ancestries, including new contributions from international partners, and incorporate multiple long-read sequencing technologies (PacBio HiFi, ONT ultralong, Hi-C, and Kinnex RNA). HPRC Release 2 provides the highest quality telomere-to-telomere assemblies to date, with substantial improvements in continuity, structural accuracy, and base-level precision. Deep learning–based polishing methods developed in collaboration with Google Research reduced base error rates. Chromosome-level scaffolding, trio-based phasing, and unified assembly pipelines enabled an unprecedented number of complete chromosomes per genome. These include centromeric and subtelomeric regions, as well as complex segmental duplications, offering a more accurate map of structural variation that is critical for both population and clinical genomics. Release 2 also introduces over 200 long-read transcriptome datasets used to refine gene annotations and full-length isoform models across ancestrally diverse backgrounds. Graph-based alignments and tools included in this release enable comparative genomics and facilitate variant discovery in genomic regions previously inaccessible to short-read or single-reference methods.

Together, these assemblies establish a foundation for a graph-based human reference that better captures the complexity of human genetic variation. By expanding global representation and enhancing the accuracy of complex genomic regions, HPRC Release 2 opens new frontiers for disease association studies, rare variant discovery, and personalized medicine. This release marks a turning point in reference genomics, offering the community a deeply annotated, structurally resolved, and ethically grounded dataset to support the next generation of human genetics research.

⭐ Local ancestry-specific genetic architecture of breast cancer risk in 40,000 women of African ancestry from the AABCG Consortium

Subsession Time: Friday, October 17 at 5:36pm – 5:54pm

Authors: Peter N Fiorica (Genetics, Genomics & Bioinformatics Program, SUNY University at Buffalo, Buffalo, NY; Department of Cancer Prevention & Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY), Lucas Mendicino (Department of Cancer Prevention & Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY), Guochong Jia (Department of Medicine, Vanderbilt University, Nashville, TN), Xiaofeng Zhu (Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH), Angela R Omilian (Department of Cancer Prevention & Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY), Elisa V Bandera (Cancer Epidemiology and Health Outcomes, Rutgers Cancer Institute of New Jersey, Brunswick, NJ), Anselm JM Hennis (Renaissance School of Medicine at Stony Brook University, Stony Brook, NY; George Alleyne Chronic

Disease Research Centre, University of the West Indies, Mona, Jamaica), Katie O'Brien (Biostatistics and Computational Biology Branch, NIEHS, Research Triangle Park, NC), Gary Zirpoli (Slone Epidemiology Center, Boston University, Boston, MA), Ebonee N Butler (Department of Epidemiology, University of North Carolina, Chapel Hill, NC), John Obafunwa (Department of Pathology and Forensic Medicine, Lagos State University Teaching Hospital, Lagos, Nigeria), Clarice R Weinberg (Biostatistics and Computational Biology Branch, NIEHS, Research Triangle Park, NC), Dezheng Huo (Department of Public Health Sciences, The University of Chicago, Chicago, IL), Bingshan Li (Department of Molecular Physiology & Biophysics, Vanderbilt Genetics Institute, Vanderbilt, University, Nashville, TN), Xingyi Guo (Department of Medicine, Vanderbilt University, Nashville, TN), Christopher Haiman (Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA), Julie R Palmer (Slone Epidemiology Center, Boston University, Boston, MA), Wei Zheng (Department of Medicine, Vanderbilt University, Nashville, TN), Christine B Ambrosone (Department of Cancer Prevention & Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY), Song Yao (Department of Cancer Prevention & Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY)

Abstract: Breast cancer mortality in women of African ancestry (AA) in the US is 40% higher than in European ancestry (EA) women, partially due to the higher risk of aggressive forms of breast cancer like estrogen receptor negative (ER-) and triple-negative (TNBC) subtypes. We hypothesize that ancestry-specific genetic loci contribute to risk for aggressive breast cancer in AA women. Based on 18,034 cases and 22,104 controls from the African Ancestry Breast Cancer Genetic (AABCG) consortium, we first analyzed subcontinental African ancestry associated with ER- and TNBC risk. We then performed local ancestry-informed genome-wide association studies (GWAS) of four breast cancer phenotypes (TNBC, ER-, ER+, and overall breast cancer) using FLARE and Tractor.

We found that West African ancestry is the dominant subcontinental ancestry in the AABCG populations, and a 10% increase in West African ancestry was associated with increased odds for TNBC (odds ratio [OR]=1.14, 1.03-1.26) relative to non-TNBC cases. In the Tractor case-control GWAS, we identified 6 loci (54 SNPs) for TNBC, 3 loci (25 SNPs) for ER- cancer, 3 loci (17 SNPs) for ER+ cancer, and 4 loci (21 SNPs) for overall breast cancer with effects specific for African local ancestry at genome-wide significance ($P<5x10-8$). Of these, 8 loci (13 SNPs) are at least 2Mb away from SNPs previously identified in AABCG or the Breast Cancer Association Consortium (BCAC) GWAS. Of 117 SNPs identified by Tractor, we found that 22 displayed substantial effect size heterogeneity (AFR:EUR β ratio >2.5 or <-1) across ancestries. Three novel TNBC SNPs (rs79385566, rs9860734, rs9813920) at 3q12 were among those 22 variants. In addition, we identified one locus (5 SNPs) for ER- cancer, 2 loci (32 SNPs) for ER+ cancer, and 3 loci (6 SNPs) for overall breast cancer with effects specific for European local ancestry at genome-wide significance. Of these, one locus (2 SNPs) at 2p16 associated with overall breast cancer and is over 2Mb away from any GWAS SNP from AABCG or BCAC. Notably, both SNPs' EUR:AFR β ratios were less than -9.0.

To conclude, in the largest AA breast cancer consortium, we find that subcontinental West African genetic components contribute to the higher risk of TNBC in AA women. Moreover, we discovered multiple new risk loci for TNBC and other breast cancer statuses with ancestry-specific effects,

highlighting the improved discovery power when leveraging local ancestry information in admixed populations.

⭐ Age- and BMI-dependent genetic architecture of blood lipids in 2.5 million individuals from globally diverse populations

Subsession Time: Friday, October 17 at 5:54pm – 6:12pm

Authors: Jacqueline S. Dron (Massachusetts General Hospital), Ida Surakka (University of Michigan), Yuxuan Wang (Boston University), Francesca Zumpano (Boston University), Jiayan Zhou (Stanford), Stavroula Kanoni (Queen Mary University of London), Seung Hoan Choi (Boston University), Margaret Sunitha Selvaraj (Massachusetts General Hospital), Satoshi Koyama (Broad Institute of MIT and Harvard), Panagiotis Deloukas (Queen Mary University of London), Michael Boehnke (University of Michigan), Yan V. Sun (Emory University), Themistocles Assimes (Stanford), Pradeep Natarajan (Massachusetts General Hospital), Gina M. Peloso (Boston University), Global Lipids Genetics Consortium

Abstract: Blood lipid levels are heritable and modifiable risk factors for cardiovascular disease (CVD). Previous work by the Global Lipids Genetics Consortium (GLGC) identified sex-specific genetic effects on lipid traits; however, the impact of other exposures, such as age and body-mass index (BMI), on the expressivity of lipid-associated genetic variants remains unclear. To address this, we performed a genome-wide association study (GWAS) and interaction study (GWIS) meta-analysis in up to 2,461,512 individuals spanning six global populations, with the largest non-European sample size to date for lipid levels: 8.0% African; 4.1% Admixed American/Hispanic; 1.9% Central and South Asian; 21.5% East Asian; 63.8% European; and 0.6% Middle Eastern. We performed GWAS for low- and high-density lipoprotein cholesterol (LDL-C and HDL-C), and triglycerides (TG), and GWIS to test for genetic interactions with age or BMI, two commonly measured variables and strong CVD risk factors. Association models included age, sex, genetic principal components, and study-specific covariates. Unique to the age-GWIS, age was included as an interaction term (1df), while for the BMI-GWIS, BMI was included as a covariate and interaction term (1df). Meta-analyses used fixed-effects inverse-variance weighting with genomic control correction on variants with MAF >0.5%. In cross-population analyses, we identified 674, 889, and 790 genome-wide significant marginal loci for LDL-C, HDL-C, and TG, respectively, representing 52.2%, 58.2%, and 64.6% increases over the previous GLGC meta-analysis. GWIS identified 130, 95, and 85 loci with age-dependent effects, and 122, 38, and 92 loci with BMI-dependent effects on LDL-C, HDL-C, and TG, respectively. Notably, 87 age- and 49 BMI-dependent genome-wide significant interaction loci were not significant in marginal GWAS analyses. Interaction loci were detected across all populations, with evidence of shared and population-specific effects. We will further annotate and fine-map these loci to resolve causal variants and identify biological pathways dependent on age or BMI. This study represents the largest and most globally representative GWAS and GWIS of lipid traits to date. Our findings expand the known genetic architecture of lipid levels and demonstrate how common risk factors like age and BMI shape genetic effects across diverse populations. These insights provide a framework for refining genetic risk prediction and advancing precision prevention of CVD.

The impact of rare deleterious mutations on human lifespan

Subsession Time: Friday, October 17 at 6:12pm – 6:30pm

Authors: Hong Gao (Illumina), Joshua G. Schraiber (Illumina), Jacob C. Ulirsch (Illumina), Shu Tadaka (Tohoku Medical Megabank Organization), Daniel M. Sanchez (M42), Shan Dong (UCSF), Jeremy McRae (Illumina), Gabrielle Lemire (Broad Institute), Rebecca Reimers (Harvard Medical School), Petko Fiziev (Illumina), Tiago R. Magalhaes (M42), Yun S. Song (UC Berkeley), Heidi L. Rehm (Broad Institute, MGH, HMS), Shamil Sunyaev (Harvard Medical School), Anne O'Donnell-Luria (Broad Institute), Stephan J. Sanders (UCSF), Kyle Kai-How Farh (Illumina)

Abstract: Natural selection has shaped the genetic history of our species, but its ongoing effects in present-day human populations remain unclear, particularly in view of recent technological and environmental changes which have markedly reduced premature mortality and doubled human life expectancy in less than 10 generations since the industrial revolution. Previous studies have identified genetic variants associated with phenotypic aspects of selection in present-day populations, including effects on reproductive success, and estimated selective constraint for protein-truncating variants. However, generalizing this work to cover all protein-coding variants has proven challenging due to the difficulty of predicting the effects of deleterious alleles, especially for missense variants, which constitute the vast majority of protein-altering variation in human cohorts.

Here, we integrate deep learning with demographic modeling to accurately estimate the heterozygous selection coefficient, PrimateAI-3D s_{het}, for nearly all ~70 million possible protein-coding single nucleotide variants in the human genome, observing the best performance when missense variants are stratified by their PrimateAI-3D pathogenicity predictions. In 454,712 individuals from the UK Biobank, we characterize the shet burden of deleterious variants per person and examine their effects on phenotypic aspects of selection. On average, a person carries 3.3 rare genetic variants with s_{het} > 2%, corresponding to a 5-month reduction in lifespan per variant. These variants act through hundreds of common diseases and often exist as intermediate-effect alleles in genes where full-penetrance variants would cause Mendelian disease.

After observing significant effects on lifespan, we explore mechanistic hypotheses for how the selective pressures on these variants arose in pre-industrial generations. We observe deleterious mutations accumulate faster than they are removed and have validated this finding in four independent large-scale cohorts: the All of Us cohort, the Vanderbilt University Medical Center cohort, the Tohoku Medical Megabank, and the Emirati Genome Program. The observed patterns were consistent across all cohorts, implying that historically, the effect of shet burden on mortality was likely the primary driver facilitating the removal of deleterious variants, whereas in contemporary populations, their effects on disease and mortality risk are insufficient to oppose the accrual of new mutations.