

ABSTRACTS

The 2026 Annual Meeting of the International Genetic Epidemiology Society

9 The Apparent Causal Effect of Education on Mental Health is Driven by Shared Genetic Liability

Stephanie Sheir¹, Liza Darrous^{5,8}, Zoltan Kutalik^{5,6,7}, Margherita Malanchini^{9,10}, Neil Martin Davies^{2,3,4*}, Jean-Baptiste Pingault^{1*}

¹Department of Clinical, Educational and Health Psychology, University College London, United Kingdom; ²Division of Psychiatry, University College London, United Kingdom; ³Department of Statistical Sciences, University College London, United Kingdom; ⁴Department of Public Health and Nursing, Norwegian University of Science and Technology, Norway; ⁵University Center for Primary Care and Public Health, Switzerland; ⁶Swiss Institute of Bioinformatics, Switzerland; ⁷Department of Computational Biology, University of Lausanne, Switzerland; ⁸University of Geneva, Switzerland; ⁹School of Biological and Behavioural Sciences, Queen Mary University of London, United Kingdom; ¹⁰Social, Genetic and Developmental Psychiatry Centre, King's College, London, United Kingdom

*These authors contributed equally

Higher educational attainment is associated with lower risk for most neuropsychiatric disorders, but whether this is causal remains unclear. Genetic instruments for education can be used to investigate causality, but may act on education through different pathways. Using a range of Mendelian randomization estimators, we found widespread causal effects (largely protective) of educational attainment on 21 neuropsychiatric outcomes. We then tested the robustness of those effects using PheWAS-based clustering of Mendelian randomization instruments (PWC-MR), partitioning 269 genetic instruments for education into seven clusters, before estimating each cluster's effect on neuropsychiatric outcomes. There was substantial effect heterogeneity, driven by clusters themselves enriched for neuropsychiatric traits. For control outcomes, there were homogeneous effects across clusters, consistent with a direct effect of educational attainment. Our results suggest that the apparent effect of educational attainment on neuropsychiatric outcomes is

largely driven by shared genetic liability, which remains undetected by conventional estimators.

Keywords: Mendelian randomization, education, mental health, cognitive, causal

10 Uncovering Novel Atrial Fibrillation Genetics Through Pleiotropic Overlap With Life's Essential 8

Jingxian Wu^{1#}, Xueying Qin^{1,2#}, Shuting Xie¹, Liuyan Zheng¹, Huan Yu¹, Huairong Wang¹, Tao Wu^{1,2}, Dafang Chen^{1,2}, Yonghua Hu^{1,2}, Yiqun Wu^{1,2*}

¹Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China; ²Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China

#Contributed equally to the work

*Corresponding author: Yiqun Wu, qywu118@163.com

Atrial fibrillation (AF) is a complex polygenic disorder; its genetic architecture remains challenging to fully elucidate. In this study, we leveraged the extensive genetic overlap between AF and a spectrum of cardiometabolic and behavioral factors—collectively defined by Life's Essential 8 (LE8)—to advance our understanding of its etiology. We first estimated significant genetic correlations between AF and all LE8 components (r_g : -0.11 to 0.19) using LD score regression. We then applied conditional false discovery rate analysis and detected 970 pleiotropic loci associated with AF and at least one LE8 trait. Subsequent colocalization analysis identified 137 loci harboring shared causal variants between AF and one or more LE8 components. Through region-based annotation and functional predictors, we finally prioritized 164 candidate genes from these colocalized regions, including 40 novel genes. These candidate genes were enriched in pathways related to heart development and regulation of cardiac contraction, and were also enriched among molecular targets of vasoprotective agents. Among all LE8 components, blood pressure emerged as the central risk

factor for AF, supported by the strongest genetic correlation, highest pleiotropic enrichment, and the greatest number of shared loci with AF. Polygenic risk scores constructed from these colocated loci demonstrated robust predictive capacity not only for AF but also for a broader spectrum of cardiovascular diseases. Our findings establish a genetic pleiotropy-informed framework that enhances the discovery of novel risk loci and advances our understanding of the genetic etiology and upstream risk factors underlying AF.

Keywords: atrial fibrillation, life's Essential 8, genetic pleiotropy, genetic overlap, polygenic risk scores

11 Neurobehavioral Related Genes Exert Indirect Genetic Effects on Blood Lipid Profiles Through Behaviors: A Family-Based Study

Liuyan Zheng¹, Shuting Xie¹, Zechen Zhou^{1,2}, Jingxian Wu¹, Xueying Qin^{1,3}, Tao Wu^{1,3}, Dafang Chen^{1,3}, Yonghua Hu^{1,3}, Tao Wang⁴, Yiqun Wu^{1,3*}

¹Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing 100191, China; ²The NHC Key Laboratory of Geriatrics, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing Hospital/National Center of Gerontology of National Health Commission, Beijing, China; ³Department of Epidemiology and Biostatistics, School of Public Health, Peking University; Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China; ⁴Department of Epidemiology and Population Health, Yeshiva University Albert Einstein College of Medicine, Bronx, New York, United States of America

Correspondence: Yiqun Wu qywu118@163.com

Growing evidence reveals that indirect genetic effects (IGEs) contribute substantially to human complex traits. However, the genomic architecture and underlying mechanistic pathways of IGEs remain poorly understood. In this study, we employed a two-stage family study design to investigate IGEs and direct genetic effects (DGEs) on blood lipid profiles in the Fangshan Family-based Ischemic Stroke Study in China (FISSIC) cohort. We identified 16 IGE loci resulting in 14 candidate IGE genes, and 20 DGE loci resulting in 22 DGE genes, by an integrative functional mapping approach. The identified IGE genes were predominantly related to behavioral and neuropsychiatric phenotypes, while DGE genes were primarily involved in lipid metabolic pathways. To examine whether IGE genes are associated with lipids through behavioral factors in IGEs, we compared IGE estimates before and after adjusting for behavioral covariates, including healthy food score, drinking status, educational attainment, obesity, physical activity, sleep duration, and smoking status. Adjustment for behavioral factors attenuated the IGE associations, and mediation analysis further indicated that healthy food score partially mediated the

association between IGE-GRS and HDL-C (mediation proportion: 17.9%), whereas behavioral mediation in the DGE pathway was negligible. Our results suggest that lifestyle behaviors may substantially mediate IGE effects, indicating that they could serve as modifiable factors to mitigate IGE-related genetic susceptibility for lipid or other complex traits.

13 Investigating Age-Dependent Genetic Associations With Lung Function

Chiemela S. Odoemelam¹, Abril G. Izquierdo¹, Nick Shrine¹, Jing Chen¹, Martin D. Tobin^{1,2}, Anna L. Guyatt^{1,2}, Catherine John^{1,2}

¹Division of Public Health and Epidemiology, School of Medical Sciences, University of Leicester, Leicester, United Kingdom;

²University Hospitals of Leicester NHS Trust, Leicester, United Kingdom

Introduction: Lung function is a key indicator of respiratory health, and we can assess genetic factors affecting lung function by studying quantitative traits, such as Forced Expiratory Volume in 1 s (FEV1), Forced Vital Capacity (FVC), and the FEV1/FVC ratio, in genome-wide association studies (GWAS). These measures are used for diagnosing and grading the severity of respiratory conditions such as chronic obstructive pulmonary disease (COPD) and asthma. Age affects many biological processes, including lung function, which generally declines as people age. However, the extent to which genetic effects on lung function vary across age groups remains uncertain.

Aim(s): This study aims to improve understanding of the development of Chronic Obstructive Pulmonary Disease (COPD) by performing a genome-wide meta-regression analysis to examine how age impacts the effect of single-nucleotide polymorphisms (SNPs) on lung function traits.

Materials and Methods: GWAS summary statistics for FEV₁, FVC, and FEV₁/FVC were sourced from 45 cohorts, grouped into 10-year age bands, with adjustment for sex, height, smoking status and ancestry. Effect sizes were standardised within each cohort and age group. Meta-regression was used to meta-analyse across cohorts, incorporating modelling of changes in genetic effects with age.

Results: The analysis included checking for age effects at both previously identified lung function signals and additional variants achieving genome-wide significance in this analysis. We tested whether genetic effects differed across age groups. Significant SNP-age interactions are reported, as well as any new lung function signals from this age-stratified analysis.

Keywords: Age-dependent genetic effects-, -Lung function-, -GWAS-, -Meta-regression-, -Genetic epidemiology

15 GLEN: A Deep Learning Framework for Genetic Liability Prediction From Multimodal Data

Cue Hyunkyue Lee¹, Najmeh Abiri², Hanqing Wu³, and Iuliana Ionita-Laza^{1,3,*}

¹Columbia University, Biostatistics, New York, New York, United States of America; ²Malmö University, Computer Science, Malmö, Sweden; ³Lund University, Biostatistics, Lund, Sweden

*Corresponding Author: ii2135@cumc.columbia.edu

Electronic health records provide heterogeneous longitudinal phenotypes well suited for multi-trait risk modeling. Traditional statistical and quantitative genetic methods have been highly useful, but the use of case-control status does not fully capture genetic heterogeneity among individuals. Leveraging multimodal and cross-sectional information across traits can address this limitation, motivating flexible models under weak parametric assumptions.

We propose GLEN (Genetic Liability Estimation Net), a deep learning framework that aggregates EHR-derived phecodes and continuous clinical measurements, and incorporates imaging modalities to enhance risk prediction. GLEN estimates continuous genetic liability under the liability threshold model, linking multimodal phenotypes to liability-scale heritability. Unlike conventional DNN models optimized with a single objective, GLEN uses MGDA to jointly optimize multiple objectives reflecting liability-threshold assumptions. This multi-objective formulation links GLEN to statistical genetic liability models while guiding optimization toward Pareto-stationary solutions.

In simulations, GLEN outperforms traditional case-control GWAS and LTPI in both liability estimation accuracy and GWAS power. MGDA optimization improves performance, yielding higher correlation between simulated and predicted liabilities. We found that DNN-based liability models can capture non-target signals under some settings, which we mitigated using targeted constraints and MGDA optimization. To reduce sensitivity to hyperparameter choices, we incorporated LTPI as a surrogate objective to guide liability ranking. Application to UK Biobank psychiatric traits, including anxiety disorder and Alzheimer disease, further demonstrates improved risk prediction and increased power for genetic discovery.

16 Statistical Behavior of Ratio Traits in Genetic Studies: Impact of Distribution and Correlation

Chun-Yuan Huang^{1,2}; Chao Xing^{2,3,4}

¹Department of Statistics, Southern Methodist University; ²McDermott Center for Human Growth and Development, UT Southwestern Medical Center, Dallas, Texas, United States of America; ³Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center, Dallas, Texas, United States of America; ⁴O'Donnell School of Public Health, UT Southwestern Medical Center, Dallas, Texas, United States of America

Ratio traits such as body mass index are often analyzed in genetic association studies. However, association signals from ratio traits can be difficult to interpret from the biological

perspective, as they may be driven by either component trait or both jointly. We systematically studied how trait distribution, heritability, and genetic correlation affect statistical power in ratio-trait association studies.

Through theoretical derivations and simulation studies, we showed that association signals from ratio tests can be either numerator- or denominator- trait driven depending on their relative means and variances. Further, we demonstrated that pleiotropic effects affect statistical inference. When genetic effects on the component traits act in the same direction, association signals are weakened because of signal cancellation. In contrast, opposing genetic effects amplify association signals and increase statistical power.

Our results indicate that the power of ratio-trait genetic associations depend on the underlying genetic architecture of the component traits and their shared genetic basis. We provide practical guidelines for evaluating ratio-based analyses, including assessment of trait scale and genetic correlation. Our framework improves the interpretation of ratio-trait association studies and highlights conditions under which ratio phenotypes are informative for genetic discovery.

Keywords: association, ratio traits, statistical power

18 An End-to-End Implementation of Knockoff GWAS With Application to Primary Biliary Cholangitis Data

Richard Howey¹, Heather J. Cordell²

¹Research Software Engineering, Newcastle University, United Kingdom; ²Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, United Kingdom

Knockoff GWAS, developed by Sesia et al., is a powerful methodology for genome-wide association studies that ensures rigorous control of the false discovery rate (FDR) while properly handling the linkage disequilibrium structure of genetic data. By constructing artificial “knockoff” variants that preserve the dependency structure of the original genotypes, the method enables a direct comparison between real and synthetic features. This comparison-based approach reduces spurious findings commonly seen in traditional GWAS, improving the reliability of detected associations. Nevertheless, adoption has been limited by complex input requirements, fragmented computational steps, and the lack of a streamlined, user-friendly workflow.

Here, we present a comprehensive pipeline that implements Knockoff GWAS for a large-scale Primary Biliary Cholangitis (PBC) dataset, starting from commonly used genotype and genetic map files. The pipeline provides all necessary Bash and R components to perform format conversion, generate knockoff variables, execute statistical inference, and obtain final association results. It automates interactions with external tools, resolves compatibility issues, and standardises the entire process into a reproducible framework. A containerised version is also provided to facilitate deployment across different systems.

To further support accessibility, we include an accompanying website that delivers a detailed, stepwise tutorial spanning data preparation, quality control, knockoff generation, analysis, and result interpretation. Applying this pipeline to PBC data shows that KnockoffGWAS can be realistically implemented for large and complex genetic datasets. This work bridges the gap between theory and practice by turning a sophisticated methodological framework into a practical tool for the research community.

Keywords: KnockoffGWAS, genome-wide association studies, false discovery rate, Primary Biliary Cholangitis, statistical genetics

19 The Genetic Architecture of Premature Ovarian Insufficiency Is Mostly Shared With Common Variation Driving Age at Natural Menopause

Ana Luiza Arruda^{1*}, Sinead McGlacken-Byrne^{1*}, Marina Vabistsevits², Jack Murzynowski¹, Emily Morbey¹, James Godwin², Katherine Ruth², Felix Day¹, Ken Ong¹, Xin Wang³, Yanyu Liang³, Katelyn Bond³, 23andMe Research Team³, Anna Murray^{2*}, John Perry^{1*}

¹*Institute of Metabolic Science Epidemiology, Department of Clinical Biochemistry, University of Cambridge School of Clinical Medicine;* ²*Department of Clinical and Biomedical Sciences, University of Exeter, Exeter, United Kingdom;* ³*23andMe Research Institute, Palo Alto, California, United States of America*

*Equal contribution

Premature ovarian insufficiency (POI), defined by age at natural menopause (ANM) < 40 years is a leading cause of infertility. Its population-level genetic architecture remains poorly defined, with prior work focused on rare monogenic causes in selected pedigrees. To map the uncharacterised common-variant architecture of POI, we performed a GWAS meta-analysis of 739,350 women (17,489 with POI) using 23andMe Research Institute, UK Biobank and FinnGen data. We identify 24 independent loci associated with POI which are mostly shared with earlier ANM, demonstrating that POI represents the lower extreme of a continuous menopausal timing distribution. At the single POI-specific common-variant signal we prioritize *TOP6BL* as the likely effector gene, involved in meiotic pathway processes. We derive a polygenic score for POI, which shows consistent predictive value among 25-29, 30-34, and 35-39 years age bands for ANM. To evaluate rare genetic contributions, we analysed UK Biobank whole-exome sequencing data using burden tests (2,510 POI cases and 55,667 controls) and found no robust gene-level rare-variation specific to POI. Focusing on heterozygous variation found in POI cases only, we suggest *RRM1* and *TIA1* as novel candidate POI genes. Taken together, we reveal limited POI-specific genetic aetiology beyond determinants of menopausal timing. Our lack of POI-specific rare variant findings may suggest that truly monogenic forms of POI are early-onset (< 25 years) and driven by ultra-rare alleles or recessive mechanisms. These findings support targeted genetic testing strategies focused on early-onset or familial cases, and caution against broad clinical panels including heterozygous variants lacking robust population-level evidence.

22 CaPri: A Bayesian Mendelian Randomization Method Robust to Widespread Horizontal Pleiotropy

Ou Ren¹, Wenmin Zhang², Tianyuan Lu¹

¹*University of Wisconsin-Madison, Madison, Wisconsin, United States of America;* ²*Regeneron Genetics Center, Tarrytown, New York, United States of America*

The no horizontal pleiotropy assumption in Mendelian randomization (MR) is limiting. Several methods increase tolerance to pleiotropic instruments but rely on additional assumptions, such as InSIDE in MR-Egger or the requirement that a plurality of instruments are valid in median- and mode-based estimators and outlier-detection methods. When pleiotropy is widespread, these approaches can become biased.

We propose CaPri, a Bayesian method jointly modeling the causal effect of the exposure on the outcome and pleiotropic effects of genetic instruments via variational inference. CaPri remains valid under arbitrary proportions of pleiotropic instruments with an assumption strictly weaker than InSIDE.

We conducted simulations comparing CaPri with 10 MR methods. CaPri maintained appropriate false positive control and robustness to prior choice when instrument strength was sufficient. When pleiotropy was absent or rare, its performance matched IVW or median-based approaches. When all instruments were pleiotropic but satisfied InSIDE, its performance matched MR-Egger. Importantly, when > 50% of instruments were pleiotropic without InSIDE, CaPri substantially outperformed all competing methods in power and accuracy.

In real data analyses using tanning ability as a negative control outcome, CaPri confirmed the absence of causal effects for all 30 selected exposures, whereas several methods produced spurious associations. In positive controls, CaPri recovered expected causal directions and produced stable estimates, while mode-based methods were underpowered.

CaPri may have wide utility as GWAS for polygenic exposures continue to identify more genetic instruments, where pleiotropy is likely to be pervasive. It may also motivate more flexible MR methods with maintained robustness and statistical power.

Keywords: Mendelian randomization, Bayesian inference, Statistical genetics, Methodology

23 Local Ancestry Effects on the Relationship Between APOE Variation and Lipids and Inflammatory Markers

Elizabeth E. Blue¹, Nicola H. Chapman¹, Seth Temple², Transomics for Precision Medicine (TOPMed), Nancy Cox³, Emily Drzymalla⁴, Yun Li⁴, Laura Raffield⁴, Alex Reiner¹, Diane Xue⁵

¹*University of Washington, Washington, United States of America;* ²*University of Michigan, Ann Arbor, Michigan, United States of America;* ³*Vanderbilt University, Nashville, Tennessee, United States of America;* ⁴*University of North Carolina, Chapel Hill,*

Two missense variants in *APOE* define the $\epsilon 2$ and $\epsilon 4$ alleles with exceptional pleiotropic associations with human disease: they are common, have strong effect sizes, and clear functional consequences. However, the magnitude of their estimated effects on both peripheral lipid and inflammatory traits varies across populations. We sought to determine whether these population-specific differences can be explained by genetic ancestry at the *APOE* locus or genome-wide. The discovery sample includes 39,519 TOPMed participants. Harmonized C-reactive protein, high density lipoprotein, low density lipoprotein (LDL) and triglycerides (TG) were natural log-transformed. Genomes were phased using Beagle, and local ancestry was inferred using FLARE, and global ancestry proportions defined by the genome average. A replication sample of 100,075 All of Us (AoU) participants mimicked the discovery sample and analysis, except for global ancestry estimation with Rye and exclusion of pairs with kinship ≥ 3 rd degree relatives. Regression was performed in GENESIS, adjusting for medication use, age, sex, age * sex, global ancestry, $\epsilon 2$ dose, $\epsilon 4$ dose, and the genetic relatedness matrix. While both European- and African-like $\epsilon 2$ alleles significantly decrease lnLDL, the African-like $\epsilon 2$ effect is larger (TOPMed: 16% vs. 13%, AoU: 17% vs. 12%). Similarly, both African- and European-like $\epsilon 2$ alleles significantly increase lnTG while the European-like allele has the larger effect (7% vs. 3% in both). Four variants drove these replicated signals, including another *APOE* missense variant and established regulatory variants; 3 were pleiotropic. While global ancestry was associated with each trait, the effect was inconsistent across traits and samples.

Keywords: Genetic epidemiology, Haplotype, Population structure, Precision medicine, Risk assessment

24 The Genetic Architecture of Blood-based Biomarkers of the Central Nervous system – Insights From a Diverse Population-Based Cohort

Anna S Lorenz^{1,2}, Kimberlyn A Ellis², Hung-Hsin Chen³, Lauren E Petty², Hannah Polikowsky², Wanying Zhu², Rashedeh Roshani², Ting-Chen Wang^{1,2}, Elizabeth Frankel², Assaf Gottlieb⁴, Paul E Schulz⁵, Fadi I Musfee⁶, Kristin L Young⁶, Cassianne Robinson-Cohen⁷, Derek B Archer^{1,2}, Timothy J Hohman^{1,2}, Heather M Highland⁶, Mariaelisa Graff⁸, Joseph B McCormick⁶, Susan P Fisher-Hoch⁶, Kari E North^{6*}, Jennifer E Below^{2*}

¹Vanderbilt Memory & Alzheimer's Center, Nashville, Tennessee, United States of America; ²Vanderbilt Genetics Institute, Nashville, Tennessee, United States of America; ³Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; ⁴Bioinformatics and Systems Medicine, University of Texas Health Science Center at Houston, School of Biomedical Informatics, Houston, Texas, United States of America; ⁵Neurocognitive Disorders Center, The McGovern Medical School of UT Health, Houston, Texas, United States of America; ⁶Department of Epidemiology, University of Texas Health Science Center at Houston, School of Public Health Brownsville Campus,

Brownsville, Texas, United States of America; ⁷Department of Nephrology, Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America; ⁸Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

*Senior author

Background: Blood-based biomarkers (BBMs) of the central nervous system are increasingly used to capture neurodegenerative, inflammatory, and synaptic processes, yet the genetic architecture underlying BBM variability remains unclear, particularly in admixed populations.

Methods: We analyzed 130 brain-relevant BBMs spanning amyloid, tau, synaptic, metabolic, and inflammatory biology (NULISaseq) measured in 936 older adults (age=68.16 $\pm$ 8.41, 67.9% female) from the Border Health Research Cohort, a genetically admixed South Texas population. To identify genetic regulators of BBM abundance, we tested associations of 172 known Alzheimer's disease (AD)-risk variants with BBMs and performed conventional pQTL using GCTA and local ancestry-aware pQTL mapping using Tractor-mix. To assess shared genetic etiology, we estimated genetic covariance and performed colocalization analyses between BBMs and neurodegenerative diseases and neuropathologies.

Results: AD-risk variants showed trans associations with BBMs, including *MS4A6A* variants with *TREM2* abundance, *CELFI* variants with *S100A12* and *REST* abundance, and *APP* variants with *SAA1* abundance ($p_{FDR} < 0.05$), indicating trans-regulatory links. pQTL mapping identified 3534 cis- and 1265 trans-regulatory loci ($p < 3.85 \times 10^{-10}$), with an additional 397 NIA-specific and 427 EUR-specific pQTLs across BBMs ($p_{FDR} < 0.05$). Genetic covariance and colocalization analyses revealed widespread shared genetic architecture between multiple BBMs and AD, amyotrophic lateral sclerosis, frontotemporal dementia, multiple sclerosis, and neuropathologic traits, including A β plaques, CERAD score, hippocampal sclerosis, Braak stage, atherosclerosis, and arteriolosclerosis.

Conclusions: These findings define genetic and ancestry-aware regulatory influences on BBMs and support their use for understanding neurodegenerative disease biology in diverse populations.

Keywords: blood-based biomarkers, Alzheimer's disease, dementia, genetics, proteomics

27 Prioritizing Variants in Familial Cancer Exome Sequencing Using Transmission Patterns

Fabiha Binte Farooq¹, Angela Brooks-Wilson^{2,3}, Jinko Graham¹

¹Department of Statistics and Actuarial Science, Simon Fraser University, British Columbia, Canada; ²Department of Biomedical Physiology and Kinesiology, Simon Fraser University, British Columbia, Canada; ³British Columbia Cancer Agency, British Columbia, Canada

Exome sequencing studies in families with strong histories of disease routinely identify large numbers of genetic variants. Prioritization of candidate variants, however, often relies on heuristic filtering strategies that make limited use of inheritance information within families. We present a model-based approach for prioritizing rare variants using parent-to-child transmission patterns in disease-enriched families. The approach allows variants to contribute different levels of evidence for disease involvement based on biologically relevant features such as predicted functional impact, age at diagnosis, and disease subtype. Variants showing excess transmission to affected individuals are prioritized as candidates for disease susceptibility while borrowing information across variants to improve stability and interpretation. We demonstrate the method using simulated exome-sequencing data motivated by the BC Lymphoid Cancer Families Study. More broadly, this work illustrates how transmission information and disease context can be integrated to improve prioritization of rare disease-associated variants in family sequencing studies of complex disease.

Keywords: Rare variants; Exome sequencing; Familial cancer; Transmission patterns; Variant prioritization

28 Phenome-Derived Polygenic Scores and Social Determinants Jointly Shape Context-Dependent Disease Risk

Ying Wang^{1,2,3}, Buu Truong^{3,4,5}, Wenhan Lu^{1,2,3}, Chaimaa Fadil⁶, Yixuan He^{1,7,8}, Weijun Luo⁹, Satoshi Koyama^{3,4,10}, Peter Kraft¹¹, Pradeep Natarajan^{3,4,5,12}, Alicia R. Martin^{1,2,3}

¹Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, United States of America; ²Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America; ³Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America; ⁴Heart and Vascular Institute, Massachusetts General Brigham, Boston, Massachusetts, United States of America; ⁵Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America; ⁶Division of Medical Sciences, Harvard University, Boston, Massachusetts, United States of America; ⁷Human Genetics Center, The University of Texas Health Science Center at Houston School of Public Health, Houston, Texas, United States of America; ⁸Department of Epidemiology, The University of Texas Health Science Center at Houston School of Public Health, Houston, Texas, United States of America; ⁹College of Computing, Georgia Institute of Technology, Atlanta, Georgia, United States of America; ¹⁰Laboratory for Cardiovascular Genomics and Informatics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; ¹¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, Maryland, United States of America; ¹²Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, United States of America

Polygenic scores (PGS) are typically derived from GWAS of single observed phenotypes, yet many complex diseases arise

from shared genetic liability distributed across correlated clinical dimensions. Meanwhile, the predictive utility of PGS depends on the social context in which genetic liability is expressed. Whether broader phenome-derived representations improve genetic prediction, and how social determinants of health (SDoH) modify the realized meaning of genetic risk, remains unclear.

We constructed PGS for 35 orthogonal latent phenomic factors derived from 2,772 traits in 361,114 UK Biobank (UKB) participants and evaluated their phenomic specificity, cross-dataset portability, and predictive performance relative to disease-specific PGS across the UKB holdout set, Mass General Brigham Biobank, and All of Us (AoU). Factor-based PGS showed widespread, biologically coherent phenome-wide associations that were reproducible across biobanks and ancestries. Their predictive utility, however, was strongly disease dependent. For asthma, a respiratory factor PGS outperformed a matched disease-specific PGS and showed improved cross-ancestry portability, retaining 41.5% of European-ancestry predictive accuracy in African-ancestry individuals, compared with 22.9% for an asthma PGS derived from the largest available multi-ancestry GWAS. By contrast, disease-specific PGS remained superior for coronary artery disease and type 2 diabetes. In AoU, SDoH contributed substantial and largely independent predictive information beyond the disease-optimal genetic model and altered calibration such that the same genetic percentile corresponded to different predicted prevalences across social strata.

Together, these findings show that polygenic prediction depends on both how genetic liability is represented and the context in which it is interpreted and realized.

Keywords: polygenic scores, phenome-wide factor analysis, cross-ancestry prediction, social determinants of health, genetic risk calibration

29 Disentangling Phenotype Scale, Amplification, and Effect Heterogeneity in Polygenic Gene-by-Environment Architecture

Zhengdong Liu^{1*}, Aakarsh Anand¹, Xinyue Qie², Manuela Costantino³, Moonseong Jeong¹, Noah Zaitlen^{4,5,6}, Andy Dahl³, Sriram Sankararaman^{1,5,6}

¹Department of Computer Science, University of California, Los Angeles, Los Angeles, California, United States of America; ²Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, California, United States of America; ³Section of Genetic Medicine, University of Chicago, Chicago, Illinois, United States of America; ⁴Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, United States of America; ⁵Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, United States of America; ⁶Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, United States of America

Polygenic gene-by-environment interaction (G×E) signals are increasingly detected in biobank data, but it remains unclear whether they reflect phenotype-scale artifacts or two biologically distinct mechanisms: effect heterogeneity across environmental strata or amplification, in which genetic effects are correlated but systematically larger in one stratum. We introduce REGENA, a variance-components method that learns phenotype transformations to reduce scale-induced G×E while estimating cross-environment genetic correlation and environment-specific genetic variances.

In simulations, REGENA is calibrated, powerful, and recovers transformations that induce apparent G×E. Applied to 43 traits, seven exposures, and 291,273 unrelated White British UK Biobank participants, REGENA identified 20 of 41 original-scale G×E signals that remained significant after Box-Cox transformation, indicating scale-independent G×E. Amplification and effect heterogeneity contributed to 17 and 14 signals, respectively. Mechanisms varied across exposures: HbA1c × statin use was driven by amplification, FEV1/FVC × smoking by effect heterogeneity, and body mass index by distinct mechanisms across sex, smoking, age, Townsend deprivation, and alcohol.

Stratified analyses showed that low-frequency, low-linkage disequilibrium variants drove effect heterogeneity, whereas common, high-linkage disequilibrium variants drove variance differences. Functional annotation localized waist-to-hip ratio × sex amplification to subcutaneous adipose-expressed genes, with higher genetic variance in females than males. G×E polygenic scores improved prediction by 1.08-fold on average over additive scores, reaching 2.34-fold in males and 1.62-fold in females for waist-to-hip ratio × sex. REGENA reveals distinct polygenic G×E mechanisms with consequences for biological interpretation and stratified prediction.

30 Large-Scale Genome-Wide Association Meta-Analysis of Coronary Heart Disease in Hispanic/Latino Adults

Geetha Chittoor^{1*}, Yao Tu², Zhe Wang³, Andrea Horimoto⁴, Elizabeth Frankel⁵, Paulina Baca⁶, Jonathan Shortt⁷, Anne E. Justice¹, Mariaelisa Graff⁸, Lindsay Fernandez-Rhodes²

¹Department of Population Health Sciences, Geisinger, Danville, Pennsylvania, United States of America; ²Department of Biobehavioral Health, Pennsylvania State University, University Park, Pennsylvania, United States of America; ³Department of Epidemiology, University of Alabama at Birmingham, Birmingham, Alabama, United States of America; ⁴Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America; ⁵Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America; ⁶Universidad Nacional Autónoma de México, Mexico; ⁷University of Colorado Anschutz Medical Campus, Aurora, Colorado, United States of America; ⁸Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

Coronary heart disease (CHD) is a leading cause of mortality globally. Genetic determinants of CHD have been poorly

characterized in Hispanic/Latino (HL) populations, which are highly admixed and represent the largest U.S. ethnic minority. Thus, we assembled the largest consortium to date of HL adults (~370k, 58.4% female, 10.2% cases) across six EHR-linked biobanks and nine population-based cohorts. CHD cases were defined by ≥ 2 occurrences of ICD-9/10 conditions, revascularization or cardiothoracic surgeries in EHRs, or cohort-specific indicators, while controls had none of these. GWAS was conducted separately by study using logistic regression adjusted for sex, age, and principal components. Meta-analysis identified 15 significant loci ($p < 5E-08$), including a novel locus (rs62149753, previously BMI-associated intergenic region) and sex-specific effects at a known locus (rs113722226, $p_{\text{diff}} = 3E-03$). Sex-stratified meta-analysis revealed a novel male-specific variant (rs6096952, $p_{\text{diff}} = 8E-03$). Several loci within ±500 kb of the identified regions have previously been associated with obesity measures, suggesting that CHD-associated genes may influence disease risk through adiposity-related pathways. Conditional analysis on nearby known variants identified one novel independent signal in *CFAP161* (rs7172781, c.636+556 T > C). Phenome-wide association of HL CHD-risk loci in the full MyCode study (~110k unrelated individuals) revealed 70 significant ($p < 0.05/1,722 = 3E-05$) PheCodes associated with variants near *CELSR2*, *LDLR*, *LPA*, *CDKN2B*, and *FES*, belonging to relevant PheCode groups, i.e., hyperlipidemia, coronary atherosclerosis, and ischemic heart disease. Our study uncovers ancestry-specific aspects of CHD's genetic architecture within HL populations, demonstrating that sex-specific genetic influences and adiposity-mediated pathways are crucial for understanding disease risk.

Keywords: CHD, GWAS, PheWAS, sex-specific, EHR-linked biobanks, Hispanic/Latino

31 Reconstructing the Haplotype Transmission Dynamics of Myotonic Dystrophy Type 1 in a Founder Population

Claudia Moreau^{1,2*}, Gilles-Philippe Morin^{1,2}, Jean Mathieu^{3,4}, Élise Duchesne^{3,5,6,7}, Cynthia Gagnon^{3,4,8}, Simon Girard^{1,2,9,10}

¹Département des sciences fondamentales, Université du Québec à Chicoutimi, Saguenay, Québec, Canada; ²Centre Intersectoriel en Santé Durable (CISD), Université du Québec à Chicoutimi, Saguenay, Québec, Canada; ³Groupe de Recherche Interdisciplinaire sur les Maladies Neuromusculaires (GRIMN), CIUSSS du Saguenay-Lac-Saint-Jean, Saguenay, Québec, Canada; ⁴Faculté de Médecine et des Sciences de la santé, Université de Sherbrooke, Saguenay, Québec, Canada; ⁵School of Rehabilitation Sciences, Faculty of Medicine, Université Laval, Québec, Québec, Canada; ⁶Center for Interdisciplinary Research in Rehabilitation and Social Integration (Cirris), Capitale-Nationale Integrated University Health and Social Services Center, Québec, Québec, Canada; ⁷CHU de Québec - Université Laval Research Center, Québec, Québec, Canada; ⁸Centre de Recherche et d'innovation du CIUSSS du Saguenay-Lac-St-Jean, Saguenay, Québec, Canada; ⁹Projet BALSAC, Université du Québec à Chicoutimi, Saguenay, Québec, Canada; ¹⁰Centre de recherche CERVO, Université Laval, Québec, Canada

Myotonic dystrophy type 1 (DM1), caused by a CTG repeat expansion in *DMPK*, is highly prevalent in the Saguenay–Lac-Saint-Jean (SLSJ) founder population (158 per 100,000). Direct molecular diagnosis remains labor-intensive, limiting population-scale screening. Founder populations with extensive genealogical records provide a unique framework to integrate haplotype structure, phylogenetic inference, and transmission modeling to characterize disease dynamics. We analyzed genotype data from 226 DM1 patients. Disease-associated haplotypes were reconstructed across 2 Mb and 17 Mb regions flanking the repeat expansion locus using GeneHap R package. To characterize haplotype evolution, we reconstructed phylogenetic relationships using maximum likelihood with IQ-TREE, retaining topologies with $AU \geq 0.05$. Phylogenies were aligned with reconstructed genealogies of DM1 patients from the BALSAC database using tree-genealogy alignment software COALIGN. To quantify the expected incidence at birth, we implemented gene-dropping simulations ($n = 1,000$ replicates) over entire genealogy, propagating alleles under Mendelian inheritance, and we estimated incidence at birth for the 1931–1945 period across SLSJ subdivisions. A shared disease-associated haplotype was detected in all genotyped patients, consistent with a strong founder effect and enabling haplotype-based diagnostic. Phylogenetic-genealogical alignment supported a common ancestral origin of major haplotype groups. Gene-dropping simulations yielded an estimated incidence at birth of 197 per 100,000, exceeding observed prevalence, and revealed spatial heterogeneity across SLSJ subdivisions, highlighting the need for targeted population screening. These results support the utility of haplotype-based models for both epidemiological surveillance and potential population screening strategies.

32 Epigenetic Aging is Accelerated in Patients With X-linked Dystonia-Parkinsonism

Jonathan B. Krause^{1,2}, Silke Szymczak¹, Björn-Hergen Laabs^{1,3}

¹Institute of Medical Biometry and Statistics, University of Luebeck, Lübeck, Germany; ²Institute of Neurogenetics, University of Luebeck and University Hospital Schleswig-Holstein, Lübeck, Germany; ³Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany

X-linked dystonia-parkinsonism (XDP) is a neuro-degenerative movement disorder endemic to the Philippines, characterized by a combination of dystonia and parkinsonism. It is caused by an SVA (SINE-VNTR-Alu) insertion in the TAF1 region on the X chromosome and primarily affects men. Age at onset (AAO) shows substantial variability, with current prediction models explaining only about 67% of this variance.

DNA methylation at CpG sites changes over time and reflects biological aging processes. Epigenetic clocks leverage age-dependent methylation patterns to estimate biological age and assess disease-associated aging effects.

Methylation data from 912 Filipinos (499 XDP patients, 298 healthy controls, and 76 non-manifesting carriers, (NMCs)) were generated using the Infinium MethylationEPIC v2.0 array.

Epigenetic age was estimated using the Hannum clock based on 71 CpG sites. Linear regression and random forest models were applied to model the relationship between epigenetic and chronological age. Age acceleration was defined as the residual from these models and compared across groups using Welch's t-test.

The correlation between predicted and chronological age was 0.93 (95% CI: [0.92–0.94]). Significant differences in age acceleration were observed between patients and controls (mean difference: -0.73, CI: [-1.34; -0.13], $p = 0.017$) and between controls and NMCs (mean difference: -2.26, 95% CI: [-4.34; -0.17], $p = 0.035$), while no significant difference was found between patients and NMCs. Results were consistent across modeling approaches. These findings indicate accelerated epigenetic aging in both patients and NMCs compared to controls, suggesting a role for epigenetic mechanisms in XDP pathophysiology and their potential utility as biomarkers for disease risk and progression.

33 Genetic Influences on Dimensional Psychopathology and Its Correlation With Bipolar Disorder Liability in the Amish Mennonite Bipolar Genetics Study

Heejeong Sung¹, Diana Frank¹, Ifeoma Akwue¹, Ley N. Lacbawan¹, Layla Kassem¹, Emily Besancon¹, Thomas G. Schulze^{2,3,4}, Francis J. McMahon¹

¹Genetic Basis of Mood and Anxiety Disorder, Human Genetics Branch, National Institute of Mental Health, NIH, Bethesda, Maryland, United States of America; ²Institute of Psychiatric Phenomics and Genomics, University Hospital, LMU Munich, Munich, Germany; ³Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University, Syracuse, New York, United States of America; ⁴Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America

Bipolar disorder (BD) is highly heritable, yet genetic liability may extend beyond categorical diagnosis to broader dimensional psychopathology. We examined the heritability of dimensional psychopathology measured by the Symptom Checklist 90-Revised (SCL90-R) and its genetic correlation with BD in families ascertained through probands with BD from the Amish Mennonite Bipolar Genetics Study (AMBiGen). The analytic sample included 449 individuals with completed SCL90-R assessments and SNP genotype data. Diagnoses were established via the Diagnostic Interview for Genetic Studies, medical records, and family informants. The narrow BD phenotype (NRW) comprised bipolar I, bipolar II, and schizoaffective disorder-manic type; the broad phenotype (BRD) additionally included schizophrenia, schizoaffective disorder depressive type, major depressive disorder, and BD-NOS. Heritability (h^2) and bivariate genetic (ρ_G) and phenotypic (ρ_P) correlations between BD phenotypes and SCL90-R indices (General Symptom Index [GSI], Positive Symptom Total [PST], Positive Symptom Distress Index [PSDI]) were estimated using variance components models in SOLAR-Eclipse. Significant heritability was observed for both BD phenotypes (NRW:

$h^2=0.43$, $p=0.011$; BRD: $h^2=0.31$, $p=0.003$) and all SCL90-R global indices (PST: $h^2=0.27$, $p=0.001$; PSDI: $h^2=0.20$, $p=0.022$; GSI: $h^2=0.23$, $p=0.002$). Strong genetic correlations were found between NRW and PST ($\rho_G=0.94$, $p=0.014$) and GSI ($\rho_G=0.97$, $p=0.013$), with smaller but significant correlations for BRD. Genetic correlations consistently exceeded phenotypic correlations across all indices. These findings support a dimensional model of BD, wherein risk is expressed not only through diagnosable illness but also through broader psychopathology trait. Dimensional measures such as the SCL90-R may better capture shared psychiatric genetic liability than categorical diagnoses.

34 Genetic Confounding in Lipid-Cognition Associations: Evidence From the Canadian Longitudinal Study on Aging

Juan Pablo Tabja Bortesi^{1*}, Ellen E. Freeman^{1,2}, Mohan Rakesh¹, Christopher A. Gravel, PhD^{1,3,4}, Marie-Hélène Roy-Gagnon¹

¹School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada; ²Department of Ophthalmology, University of Ottawa, Ottawa, Ontario, Canada; ³Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, Canada; ⁴Data Literacy Research Institute, University of Ottawa, Ottawa, Ontario, Canada

Background: Evidence for associations between blood lipids and cognition is mixed. Both traits are partially heritable, but the extent to which genetic confounding contributes to these associations has not been well explored.

Methods: We applied the PolyGENic Genetic confounding Inference (PENGUIN) method to assess the extent of genetic confounding, which combines genome-wide association study (GWAS) summary statistics as well as individual-level phenotype data to estimate associations without genetic confounding. GWAS summary statistics for lipids were obtained from the Global Lipids Genetics Consortium. GWASs of cognitive test scores were conducted using baseline data from the Canadian Longitudinal Study on Aging, which were also used to assess phenotypic variation and trait relationships.

Results: Adjusting for genetic confounding resulted in a median absolute relative 28% change in effect estimates across lipid-cognition associations. Some conclusions about statistical significance changed; e.g., the association between triglycerides and Stroop test scores was no longer significant after adjustment ($\beta=0.018$, 95% CI [0.005, 0.031] vs. $\beta=0.013$, 95% CI [-0.005, 0.030]). The extent of genetic confounding varied, with instances of both attenuation and amplification of the effect. The largest change in effects was observed for high-density lipoprotein cholesterol and a test of executive function ($\beta=0.048$, 95% CI [0.035, 0.062] without adjustment vs. $\beta=0.036$, 95% CI [0.018, 0.054] with adjustment).

Conclusions: Associations between lipids and cognitive function in mid- and late life appear to be partially attributable to

shared genetic factors. An approach like PENGUIN can help mitigate the influence of genetic confounding.

Keywords: pleiotropy, cholesterol, cognitive function, residual bias

35 Partitioned Blood Pressure Polygenic Risk Reveals Differential Genetic Effects of Tissue-specific Enhancers

Xiaofeng Zhu¹, Yihe Yang¹, Dongwon Lee², Aravinda Chakravarti³

¹Case Western Reserve University School of Medicine, Cleveland, Ohio, United States of America; ²Harvard Medical School, Boston, Massachusetts, United States of America; ³NYU Grossman School of Medicine, New York, New York, United States of America

Polygenic risk scores (PRS) of a phenotype, summarizing weighted genetic effects across the genome, are important tools for predicting complex phenotypes, yet they are envisioned as a single aggregate predictor. For many phenotypes, such as blood pressure traits and their effects on cardiovascular and related outcomes, however, the genetic effects arise from different tissues of which some are tissue-specific and others tissue-agnostic. Here, we conduct a reverse analytical framework that dissects the PRS by the functional context of their constituent associated variants. We do so by partitioning different blood pressure (BP) trait PRSs additively by variant locations within tissue-specific as well as tissue-agnostic cis-regulatory elements (CREs). Besides distinct enrichments of SNP heritability, the resulting partitioned scores reveal heterogeneous effects across tissue-specific CREs despite the variant weights having been adjusted in the PRS construction. This reverse PRS framework provides a new lens through which to interpret the contribution of common genetic variants, offering mechanistic insights into disease biology and the context-dependent genetic architecture of complex traits.

36 Context-Specific Polygenic Risk Scores Provide Limited Improvement Over Traditional Approaches

Alyssa Green^{1*}, Ningyuan Wang², Yixin Zhang², Afshin Jamshidi³, James B. Meigs^{4,5,6}, Ching-Ti Liu², Josée Dupuis^{1,3}

¹Quantitative Life Sciences, McGill University, Montréal, Québec, Canada ²Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America; ³Department of Epidemiology, Biostatistics and Occupational Health, School of Population and Global Health, McGill University, Montréal, Québec, Canada; ⁴Department of Medicine, Harvard Medical School, Boston, Massachusetts, United States of America; ⁵Programs in Metabolism and Medical & Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America; ⁶Division of General Internal Medicine, Massachusetts General Hospital, Boston, Massachusetts, United States of America

Type 2 diabetes (T2D) and waist-to-hip ratio adjusted for BMI (WHRadjBMI) exhibit sex differences in disease risk, prevalence, and clinical outcomes. Polygenic risk scores (PRS) are commonly constructed using either sex-combined or sex-stratified approaches, assuming shared or distinct genetic effects across sexes. Context-specific PRS offer an alternative by incorporating sex-stratified effects when there is evidence of heterogeneity across sexes.

We compared four context-specific PRS with sex-combined and sex-stratified benchmarks for both traits. T2D PRS were derived from DIAGRAM summary statistics and evaluated in the Framingham Heart Study (FHS; $n = 8,379$) using summaryAUC, and in UK Biobank (UKB; $n = 396,783$) using AUC and Tjur R^2 . WHRadjBMI PRS were constructed from GIANT ($n = 133,723$) and evaluated in UKB using R^2 and ΔR^2 .

For T2D, context-specific PRS outperformed sex-stratified PRS across both cohorts with improvements of up to 0.003 in males and 0.009 in females in summaryAUC in FHS, up to 0.008 in males and 0.010 in females in AUC in UKB, and up to 0.003 in Tjur R^2 . Context-specific T2D PRS showed modest gains over sex-combined PRS, however, these differences were not statistically significant. For WHRadjBMI, context-specific PRS showed negligible improvement from both benchmarks, with a maximum R^2 increase of 0.004.

Context-specific PRS provided limited gains over sex-stratified and sex-combined PRS for T2D and WHRadjBMI. While incorporating sex-specific genetic effects can slightly improve PRS performance, sex-combined PRS remain a strong and generalizable benchmark when context-specific gains are small and statistically insignificant.

Keywords: Context-specific PRS, Type 2 Diabetes, Waist Hip Ratio

37 Assessing the Causal Role of Maternal BMI in the Risk of Dysglycemia in Offspring: A Mendelian Randomization Study

Isabela Morita-De Oliveira¹, Wiame Belbellaj², Melody Zuo³, Isabel Gamache¹, Marie-Julie Favé¹, Marie-France Hivert^{4,5}, Despoina Manousaki^{2,6}

¹Department of Biology, Concordia University, Montréal, Québec, Canada; ²Research Center of the Sainte-Justine University Hospital, Université de Montréal, Montréal, Québec, Canada; ³Faculty of Medicine, Université de Montréal, Montréal, Québec, Canada; ⁴Department of Population Medicine, Harvard Pilgrim Health Care Institute, Harvard Medical School, Boston, Massachusetts, United States of America; ⁵Department of Medicine, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, Québec, Canada; ⁶Departments of Pediatrics, and of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, Québec, Canada.

Background: Maternal adiposity has been associated with offspring obesity, a strong predictor of later metabolic dysfunction, but these associations may be confounded by shared genetic and environmental factors.

Objective: To investigate the causal effect of maternal BMI on offspring dysglycemia, an early stage of type 2 diabetes.

Methods: We conducted longitudinal and mother–child MR analyses in ALSPAC ($N = 8,932$ mother–child pairs), using maternal postpartum BMI and repeated measures of fasting glucose, insulin, HOMA-IR, and dysglycemia status at ages 7–24 years. Longitudinal associations were assessed using linear mixed models and generalized estimating equations. The MR instrument was a maternal BMI PRS derived from a female BMI GWAS. We applied one approach using a PRS derived from unphased maternal genotype data adjusted for offspring BMI PRS; and one using a PRS derived from phased maternal data (non-transmitted maternal alleles).

Results: In longitudinal analyses, maternal BMI was positively associated with all offspring metabolic outcomes across follow-up, with effects persisting after adjustment for lifestyle factors. In MR using unphased PRS, maternal BMI showed a potential causal effect on offspring fasting glucose ($\beta = 0.204$, 95% CI 0.02–0.39), fasting insulin ($\beta = 0.219$, 95% CI 0.02–0.42), and HOMA-IR ($\beta = 0.229$, 95% CI 0.04–0.42) at age 15 y. Higher maternal BMI also associated with dysglycemia at age 15 y (OR = 2.48, 95% CI 1.12–6.29), even after adjustment for pubertal timing and offspring BMI PRS. The direction of effects was similar using phased PRS.

Conclusion: There is a possible causal effect of higher maternal BMI on offspring dysglycemia risk.

38 Robust Mendelian Randomization Estimation Using Weighted Quantile Regression

Julien St-Pierre^{1,2*}, Archer Y. Yang³, Mireille E. Schnitzer^{1,4}, Marc-André Legault^{1,2}

¹Faculté de Pharmacie, Université de Montréal, Montréal, Québec, Canada; ²Centre de Recherche Azrieli du CHU Sainte-Justine, Montréal, Québec, Canada; ³Department of Mathematics and Statistics, McGill University, Montréal, Québec, Canada; ⁴Département de médecine sociale et préventive, Université de Montréal, Montréal, Québec, Canada

In Mendelian randomization (MR) studies, genetic variants are used as instrumental variables (IVs) to investigate causal relationships between exposures and outcomes based on observational data. However, numerous genetic studies have shown the pervasive pleiotropy of genetic variants, meaning that many, if not most, variants are associated with multiple traits, potentially violating the core assumptions of IV estimation. Uncorrelated pleiotropy occurs when genetic variants have a direct effect on the outcome that is not mediated by the exposure, while correlated pleiotropy occurs when genetic variants affect the exposure and outcome via shared heritable confounders. In this work, we propose a novel MR method, called MR-Quantile, based on weighted quantile regression (WQR) that is robust to both correlated and uncorrelated pleiotropy. We propose a procedure for selecting the optimal quantile of the ratio estimates through a likelihood-based formulation of WQR using the asymmetric Laplace distribution. Monte Carlo simulations

demonstrate the empirical performance of the proposed method, especially in settings with many invalid IVs with weak pleiotropic effects. Finally, we apply our method to study the causal effect of resting heart rate on atrial fibrillation. Genetic variants associated with heart rate were identified in a genome-wide association study of 425,748 individuals from the VA Million Veteran Program, and used as instruments in a two-sample MR analysis with summary statistics from a genetic meta-analysis of 228,926 AF cases across eight studies.

39 Development of a New US Large-Scale Online Study on the Genetics of Breast Cancer Risk

Steven Buyske^{1*}, Jonathan Marchini², Goncalo Abecasis², Jonathan Ross², the Regeneron Genetics Center, Tara Matise³

¹Statistics Department, Rutgers University, Piscataway, New Jersey, United States of America; ²Regeneron Genetics Center, Tarrytown, New York, United States of America; ³Genetics Department, Rutgers University, Piscataway, New Jersey, United States of America

The Rutgers University Genetics Coordinating Center (RUGCC), in partnership with Regeneron Genetics Center (RGC), is conducting a national, large-scale online study on the genetics of breast cancer risk. We use a direct-to-participant web app for consents, health surveys, and saliva kit ordering. Participants can optionally share electronic health records. Exome sequencing is conducted at RGC.

Participants receive estimated genetic ancestry, but no clinical findings results are returned. Participants are engaged through monthly newsletters. Recruitment includes partnerships with support groups, and we promote via ClinicalTrials.org, ResearchMatch.org, and the NJ RWJBarnabas Health MyChart app.

To date, 17,781 participants have completed enrollment, with a 79% kit return rate. Participants are 90% female and 39% cases; 28% of controls have 1st-degree family history of BrCa. We have had an average monthly accrual rate of 9%, varying from 2% to 40%. One-third of participants have been referred by our recruiting partners and one-third by other participants. Based on a haplotype matching model and a reference panel from six continental populations and an Ashkenazi Jewish population, participants matched to European (86.3%), Ashkenazi (8.1%), African (2.2%), East/Southeast Asian (1.8%), Central/South Asian (1.1%), Indigenous American (0.6%), and Middle Eastern (0.02%); greater diversity is needed. We found substantial enrichment of pLoF carrier status compared to that in UK Biobank at BRCA1 (1.6% vs 0.1%) and BRCA2 (3.4% vs 0.2%) genes, likely due to our recruitment approach. This enrichment provides an opportunity to discover protective variants in such carriers. We have recently launched similar studies on kidney diseases and on pulmonary hypertension.

40 HapEyre: Enlarging Population-Scale Genotype-Pheno-type Simulations

Victor Vera Frazão¹, Sebastian Sendel^{2,3}, Amke Caliebe^{2,3}, Michael Nothnagel¹

¹University of Cologne, Cologne Center for Genomics, Cologne, Germany; ²HMU Health and Medical University Erfurt, Department of Human Medicine, Potsdam, Germany; ³Kiel University and University Hospital Schleswig-Holstein, Institute of Medical Statistics, Kiel, Germany

Genotype and phenotype simulations are essential tools for the performance assessment and improvement of analytical methods in genetic studies, such as methods for constructing polygenic scores. Increasing variant numbers and sample sizes, as required by the consideration of whole genomes, rare variants, low-prevalence phenotypes and more, render the use of established simulation software infeasible due to a lack of scalability both in terms of memory usage and computing time. The software HAPNEST is a promising candidate for performing such large-scale simulations. It showed, however, severe limitations when upscaling variant and/or sample numbers above the published margins. These include excessive file sizes that hinder downstream processing, and problems with phenotype assignment and data validation. To enable population-scale simulations, we developed HapEyre as an extension of HAPNEST to secure its potential. HapEyre improves upon HAPNEST in several respects: it supports data partitioning instead of handling the entire dataset, incorporates appropriate validation of simulated data, enables phenotype simulation based on the modified partitioned data structure, and provides an automated pipeline for data preparation and variant filtering. As a showcase, we increased the variant size from 6.8 million to over 22 million variants to simulate one million samples based on reference data from the 1000 Genomes Project. We also simulated 11 million samples on data provided by HAPNEST, including 1.3 million variants, using HapEyre. We conclude that simulations can be upscaled to large amounts of variants and samples, thereby shifting the role of limiting factors to appropriate hardware for process parallelization and to storage.

41 Effects of Polygenic Risk Scores and Physical Activity on Blood Pressure Trajectories in Childhood, Adolescence, and Young Adulthood

Tian Xie^{1,2†}, Shili Luo^{1†*}, Chris H. L. Thio¹, Marloes Cardol¹, Lifelines Cohort Study, International Consortium of Blood Pressure, Eva Corpeleijn¹, Gerton Lunter^{1‡}, Harold Snieder^{1‡}

¹Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ²Guangzhou National Laboratory, Guangzhou International Bio Island, Guangzhou, Guangdong Province, China

[†]These authors contributed equally to this work.

[‡]These authors jointly supervised this work.

Background: Understanding effects of genetics and physical activity on blood pressure (BP) trajectories at an early age may help to improve our understanding of the pathogenesis of hypertension.

Methods: BP was assessed twice in prepuberty (GECKO Drenthe cohort; n = 1,063), puberty (Lifelines cohort; n = 2,494), and

postpuberty (Lifelines cohort; $n = 2,549$). We calculated polygenic risk scores (PRSs) for SBP and DBP based on the largest BP GWAS and examined the associations of PRSs with BP trajectories in the three age epochs.

Results: One SD of PRSs was positively associated with 1.70, 1.65 and 2.92 mmHg SBP level and 1.68, 0.85, 1.18 mmHg DBP level at ages 5, 8 and 15 years, respectively. There was evidence of associations between PRSs and rates of change in SBP during puberty and DBP during postpuberty. During puberty, individuals in the highest quartile of SBP-PRS showed a 0.32 (95% CI: 0.02, 0.63) mmHg greater increase in SBP per year compared with those in the lowest quartile, while in post-puberty, those in the highest quartile of the DBP-PRS exhibited a 0.24 (95% CI: 0.02, 0.45) mmHg/year greater increase in DBP. Children who were physically active had on average 1.90 (95% CI: -3.77, -0.03) mmHg lower initial SBP than inactive children in puberty.

Conclusions: Our study suggests that genetic influences on BP are evident as early as age 5 and extend to dynamic changes in SBP during puberty and DBP during postpuberty. Higher PA may reduce SBP levels in puberty independent of genetic susceptibility.

42 Investigating the Genetic Basis of HCC in the Context of MASLD

Rebecca Darlay^{1*}, Misti V. McCain^{2,3}, Tom Ewen^{2,3}, James Wordsworth^{2,3}, David Barnes^{2,3}, David Jamieson², Ann K. Daly², Quentin M. Anstee^{2,3}, Heather J. Cordell^{1,3}, Helen L. Reeves^{2,3}, LITMUS Investigators and HUNTER Clinical Coinvestigators Group

¹Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, United Kingdom; ²Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, United Kingdom; ³NIHR Newcastle Biomedical Research Centre Research Institute, Newcastle upon Tyne, England

Between 4% and 27% of Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD) cases with cirrhosis go on to develop Hepatocellular Carcinoma (HCC), which has the 2nd highest cancer-related mortality globally. To investigate genetic factors underlying this progression, a genome-wide association study was performed comparing 816 HCC cases (360 with MASLD) against 2409 MASLD controls and, separately, 17781 population controls. We utilised a two-stage imputation approach to address differences in genotyping arrays, whereby each cohort was first imputed separately and only the highest quality overlapping SNPs were taken forward for another round of imputation. Mixed linear modelling analysis revealed an association in rs3924920 in the first intron of *SLCO21A* ($P = 1.15 \times 10^{-8}$), whereby the A allele appears protective against developing HCC in individuals that also have MASLD. *SLCO21A* encodes the prostaglandin transporter protein (PGT), responsible for transporting prostaglandins – lipid compounds regulating inflammation and cellular growth – making it a strong candidate for future functional studies.

43 Proteomic Profiling of Disease State and Treatment Response in Primary Biliary Cholangitis

Alun Jones¹, Victoria Mulcahy^{2,3}, George F. Mells¹, Heather J. Cordell¹

¹Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, United Kingdom; ²Norfolk and Norwich University Hospital, Norwich, United Kingdom; ³Norwich Medical School, University of East Anglia, Norwich, United Kingdom

Primary biliary cholangitis (PBC) is a chronic, female-predominant autoimmune liver disease for which there is currently no immunotherapy treatment. First-line treatment is ursodeoxycholic acid (UDCA); however, not all patients respond to treatment. Studies have identified genetic associations and transcriptional changes linking PBC to genes and pathways such as JAK-STAT, IL-27 and IL-12 signalling. How these genes and pathways are reflected at the protein level is less understood. We used the 7 K SomaScan assay to examine blood serum protein levels in 172 individuals across four cohorts with the aim of identifying how associated genes, markers, and networks are reflected at the protein level. The cohorts included treatment-naïve PBC patients, patients with complete biochemical response to UDCA, patients with inadequate response to UDCA, and healthy controls. We found that protein levels measured through the ~7 K SOMAmers showed limited ability to discriminate between the cohorts, although the treatment-responsive PBC cohort clustered most closely with the controls while treatment-naïve PBC patients clustered with non-responders, suggestive of an overall correlation between protein signature and disease state. Cohort classifications for controls and UDCA non-responders were more accurate than for the other two cohorts, with proteins such as pIgR and sCD163 identified as informative predictors for these groups. This corroborates previous findings linking these proteins to disease severity and UDCA response in single-cell and serum studies. This study supports the presence of protein-level signals associated with PBC severity and UDCA response; however, further research is needed to better understand the underlying mechanisms and identify potential therapeutic targets.

44 Genomic Viewer: Visual Integration of Multi-Omics Data for Functional Validation of Genome-Wide Association Study Risk Loci

Sara Lago^{1*}, Luisa Foco¹, Andrea Minio¹, Maeregu W. Arisido², Cristian Pattaro¹, Peter Paul Pramstaller^{1,3}, Christian X. Weichenberger¹

¹Institute for Biomedicine, Eurac Research, Bolzano, Italy; ²Health Data Science Center, Human Technopole, Milan, Italy; ³Department of Neurology, General Central Hospital, Bolzano, Italy

Genome-wide association studies (GWAS) are powerful methods to identify genetic variants associated with complex traits and diseases. However, interpreting the biological relevance of GWAS risk loci remains challenging because their effect depends on the cell type and epigenetic context. Integrating GWAS findings with epigenomic datasets can help fill this gap. However, the heterogeneity and complexity of these data make their interpretation difficult, especially for users with limited programming expertise.

To facilitate integrative and locus-specific exploration of genomic data, we developed Genomic Viewer (GV), a desktop application designed to emphasize reproducibility, data privacy and accessibility. GV enables visualization of genomic, epigenomic, and annotation datasets within their genomic context, simplifying interpretation of candidate loci and without requiring programming expertise. The application also supports extraction of raw data from user-defined genomic regions and provides basic analyses of genome-wide features. Data privacy is guaranteed by dealing with local files only without sending data to external services.

We applied GV to investigate seven genome-wide significant loci associated with circulating renin–angiotensin–aldosterone system (RAAS) biomarkers identified from >35 million TOPMed-r2 imputed SNPs in 2,105 participants of the Cooperative Health Research in South Tyrol (CHRIS) study. Given the central role of the kidney in RAAS regulation, we integrated public kidney epigenomic datasets to characterize chromatin states and 3D interactions. Analyses highlighted intergenic loci near ERAP1/2 and LNPEP overlapping active chromatin and shared topologically associated domains, suggesting potential long-range regulatory interactions.

Overall, GV provides a user-friendly platform for functional interpretation of heterogeneous genomic datasets.

45 The Landscape and Phenotypic Impact of Mobile Element Variations in the UK Biobank

Dandan Tan^{1,2}, Jesse Islam¹, J Brent Richards^{1,3-6}, Clément Goubert⁷, Guillaume Bourque⁴, Nicholas F. Parrish⁸, *Guillaume Butler-Laporte^{1,2,4,9,10}

¹Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Québec, Canada; ²Quantitative Life Sciences Program, McGill University, Montréal, Québec, Canada; ³Prime Sciences Inc, Montréal, Québec, Canada; ⁴Department of Human Genetics, McGill University, Montréal, Québec, Canada; ⁵Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada; ⁶Department of Twin Research, King's College London, London, United Kingdom; ⁷R. Ken Coit College of Pharmacy, University of Arizona, Tucson, United States of America; ⁸Provisional Faculty in the Department of Surgery, Duke University School of Medicine, Durham, North Carolina, United States of America; ⁹Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; ¹⁰Division of Infectious Diseases, McGill University Health Centre, Montréal, Québec, Canada

Mobile elements (MEs) constitute around 50% of the human genome and drive structural variation. While many MEs are fixed, polymorphic mobile element variants (MEVs) contribute significantly to structural diversity. However, their evolutionary constraints and contributions to phenotypic variation remain under-characterized at a population scale.

This study presents a comprehensive landscape of MEVs across 485,836 whole-genome sequences from the UK Biobank using the Mobile Element Genotyping Analysis Environment (MEGANe), yielding 372,916 unique MEVs.

Genomic characterization revealed clear signatures of negative selection within functional regions. Because MEVs span up to kilobases in length, they capture loss-of-function variation highly effectively. MEV constraint scores correlated tightly with established metrics. MEV-constrained genes (mean LOEUF = 0.76) exhibited strong associations with neurodevelopmental disorders, whereas MEV-tolerant genes (mean LOEUF = 1.03) were linked to olfactory pathways.

To systematically map genotype-phenotype relationships, we conducted a Phenome-wide Association Study (PheWAS) against 2,922 plasma proteins, 302 binary traits, and 350 quantitative traits, identifying 4,829 variant-level and 2,258 gene-level significant associations. Notably, these analyses uncovered critical regulatory MEVs linked to oncogenesis. For instance, an *Alu* insertion at a *MAP3K1* enhancer is associated with increased breast cancer risk, whereas an *Alu* insertion at a transcription factor binding site near *PTHLH* is associated with reduced breast cancer risk. Furthermore, four MEVs located within cis-regulatory elements were significantly associated with skin cancer.

Together, these findings provide the highest-resolution map of human MEVs to date, demonstrating their enmeshment with evolutionary constraint and substantial impact on complex trait diversity.

46 Polygenic Risk for Lung Cancer is Amplified by Current Smoking: A Comparison of Two PGS in the PLCO Cohort

Jayati Sharma^{1*}, Batel Blechter¹, Shengchao A. Li¹, Shilpa Gaddam¹, Jia Liu¹, Mitchell Machiela¹, Rena R. Jones¹, Phillip Kraft¹

¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, United States of America

Lung cancer remains among the top 5 diagnosed cancers in US adults. Polygenic scores (PGS) may refine individual risk stratification, but gene-environment interactions (GxE) with smoking and ambient outdoor fine particulate matter (PM_{2.5}) exposure remain incompletely characterized.

Among 46,910 PLCO Cancer Screening Trial participants (925 incident cases), we evaluated two PGS catalog scores: PGS₁ (109 variants) and PGS₂ (655,479 variants) (Pearson $r = 0.34$). Scores were applied to TOPMed-imputed genotypes, residualized on the first 10 ancestry PCs, and z-standardized. We modeled lung cancer associations adjusted for age sex, smoking status, comorbidities, race and ethnicity, and center. We tested GxE for smoking status and residential PM_{2.5} (2000 single-year average).

Per 1-SD increase, the OR for lung cancer was 1.18 (95%CI 1.11-1.26) for PGS₁ and 1.09 (95% CI:1.02-1.17) for PGS₂. Top- versus bottom-decile ORs were 1.91 [1.40-2.62] for PGS₁ and 1.15 [0.86-1.55] for PGS₂. AUC was 0.793 for non-genetic factors, 0.794 with PGS₁ added, and 0.793 with PGS₂ added. PGS₁ showed a steeper per-SD gradient in current

smokers (1.35 [1.21-1.50]) than former (1.11 [1.02-1.21]) or never smokers (1.16 [0.95-1.42]), suggesting amplified genetic risk in active smokers (LR $p = 0.03$). Adding PGS1 to absolute 5-year risk estimation reclassified 33 former smokers above the screening threshold, expanding eligibility to a group otherwise missed by current criteria.

PGS₁ slightly outperformed PGS₂ in PLCO, with its risk-stratifying signal concentrated in current smokers. PGS-augmented absolute risk estimation reclassified high-risk former smokers above the screening threshold, suggesting PGS could expand and equitably target lung cancer screening beyond current USPSTF eligibility.

Keywords: lung cancer, smoking, gene-environment interactions, polygenic risk scores

47 Optimizing the Use of Longitudinal Data in Electronic Health Records for Genome-Wide Association Studies

Ghislain Rocheleau^{1,2*}, Ron Do^{1,2}

¹Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America; ²Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America

Longitudinal clinical measures currently available in electronic health records (EHR) are underexploited in genome-wide association studies. Proper modelling of longitudinal traits might increase the statistical power to detect novel loci by properly taking into account intra-subject variability and measurement error. In EHR data with varying number of measurements and time points, the optimal approach remains unclear due to the lack of analytical power estimates. We explored the linear mixed model approach as a potential candidate to reduce both complexity modelling and computational time. We first developed closed-form formulas of statistical power to detect a common variant effect in the ideal scenario of regularly measured time-points. Non-centrality parameters (NCP) of the Wald tests are derived and compared to the linear regression model using a single measurement. In most scenarios, the NCP of the variant effect test was primarily driven by the ratio of the variance of random intercepts over the variance of model residuals. If this ratio is large enough, a linear mixed model will increase the power of detecting a variant effect. If the ratio is too small, taking the mean of longitudinal measurements as the tested phenotype in a linear regression model might increase power by reducing the measurement error. We confirmed these observations by running genome-wide association studies in blood pressure (systolic, diastolic) and lipids (triglyceride, LDL, HDL, total cholesterol) measurements in ~50,000 BioMe participants. These results provide a framework for researchers to determine the optimal modelling approach for longitudinal traits in EHR-based genetic association studies.

Keywords: longitudinal data, electronic health records, genome-wide association study, mixed model, statistical power

48 TorchGWAS2: Cost-Effective Phenome- and Genome-Wide Association Testing in Related Samples

Mengyu Zhang^{1†}, Ziqian Xie^{2†}, Samaneh Salehi Nasab^{3†}, Nannan Wang³, Xingzhong Zhao², Taryn Alkis³, Bing Yu^{3*}, Degui Zhi^{2*}, Han Chen^{4*}

¹Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, United States of America; ²Department of Bioinformatics and Systems Medicine, McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, Texas United States of America; ³Department of Epidemiology, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, United States of America; ⁴Rory Meyers College of Nursing, New York University, New York, New York, United States of America

[†]These authors contributed equally

*Corresponding Author

Scanning thousands of phenotypes is challenging in large-scale genome-wide association studies (GWAS) on imaging and omics data, which often include related samples. Existing linear mixed model (LMM)-based tools that effectively control for sample relatedness still face a heavy computational burden when running on thousands of phenotypes. We present TorchGWAS2, a cost-effective and scalable solution that overcomes the computational bottleneck by using a deterministic variance correction algorithm for LMMs. Simulation studies showed that TorchGWAS2 had greater power than score-test-based approaches while maintaining well-controlled type I error rates in both GWAS and omics quantitative trait locus (QTL) analysis. In a large-scale GWAS with retinal image-derived endophenotypes from 64,703 participants in the UK Biobank, TorchGWAS2 finished 128 GWAS within 2.3 min by jointly analyzing retinal images from both eyes, and identified three novel genes *GABRB1*, *FAM234B*, and *MTUS2* previously missed in a GWAS running separately on left and right eyes. It was 39 times faster than REGENIE and 48 times faster than fastGWA. In a QTL analysis on 1,023 circulating metabolites from 16,352 participants in Trans-Omics for Precision Medicine (TOPMed) studies, TorchGWAS2 completed the full scan within 20 h using a single GPU on the cloud-computing platform BioData Catalyst and cost \$68, compared to 3,075 h and \$8,664 in a previous QTL analysis using GMMAT to account for extensive sample relatedness. TorchGWAS2 is applicable to both unrelated and related individuals, cross-sectional and longitudinal studies, with and without missing data. It offers a cost-effective solution for research teams with limited computational resources.

Keywords: linear mixed models, variance correction, genome-wide association study, phenome-wide, Graphics Processing Unit (GPU)-accelerated

49 Is It How Tall You Are as a Child or How Tall You End Up That Influences Socio-Economic Outcomes? A Life-Course Mendelian Randomization Analysis

Sergio Ordonez Beltran¹, Eleanor Sanderson¹

¹Bristol Medical School, University of Bristol, Bristol, United Kingdom

Objectives: Prior studies have estimated either the effect of adult or childhood height on socioeconomic outcomes, but these exposures are highly correlated — genetically and phenotypically — making it difficult to disentangle their independent contributions. Using a life-course Mendelian Randomization (MR) framework, we ask whether childhood height causally affects educational attainment (EA), cognitive performance, income, and neighborhood deprivation independent of adult height.

Methods: We apply univariable MR (UVMR) to estimate total effects and multivariable MR (MVMR) to estimate direct effects of genetically predicted height at age 10 and in adulthood. Robustness is assessed through complementary MR estimators, MR-Clust, MR-Corge, a PheWAS-based clustering analysis, and reverse-direction UVMR.

Results: In UVMR, both exposures were positively associated with all outcomes. In MVMR, adult height showed robust direct effects ($\beta \approx 0.26$ – 0.32 for EA, cognitive performance, and income; $\beta = -0.171$ for neighbourhood deprivation). The childhood height conditional estimate was not robust under any sensitivity specification. Reverse-direction UVMR found no evidence that deprivation causally influences height.

Discussion: Childhood height appears to affect socioeconomic outcomes largely indirectly through adult stature, with no robust evidence of an independent early-life effect. Limitations include retrospective self-report of childhood height, coarse exposure categorisation, and substantial heterogeneity in instrument-specific estimates.

50 Adiposity-Partitioned Genetics of T2D Disentangling Adiposity-Related and Non-Adiposity-Related Pathways for Type 2 Diabetes

Chen-Yang Su, PhD^{1*}, Tianyuan Lu, PhD^{1,2,3,4,5*}

*Presenting author

¹Department of Population Health Sciences, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, United States of America; ²Department of Biostatistics and Medical Informatics, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, United States of America; ³Center for Precision Medicine, University of Wisconsin-Madison, Madison, Wisconsin, United States of America; ⁴Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, Wisconsin, United States of America; ⁵Center for Demography of Health and Aging, University of Wisconsin-Madison, Madison, Wisconsin, United States of America

Background: Type 2 diabetes (T2D) shares substantial genetic architecture with body mass index (BMI), which may obscure therapeutic targets acting through pathways not primarily mediated by adiposity. We applied GWAS-by-subtraction and proteome-wide Mendelian randomization (MR) to identify circulating proteins associated with BMI-subtracted T2D liability.

Methods: We used European ancestry GWAS summary statistics for T2D ($n = 74,124$ cases, 824,006 controls) and BMI ($n = 681,275$) to partition T2D genetic liability into BMI-related and BMI-subtracted components using genomic structural equation modeling. We then performed proteome-wide MR using *cis*-pQTL from four proteomic resources. Robust protein associations were defined by study-wide Bonferroni significance and *cis*-pQTL colocalization with BMI-subtracted T2D using PWCoCo or SharePro. We further evaluated tissue-of-action using *cis*-eQTL colocalization across GTEx, pancreatic islets, subcutaneous adipose tissue, and whole blood, and assessed therapeutic actionability using the druggable genome and Open Targets Platform.

Results: GWAS-by-subtraction attenuated the genetic correlation between BMI and T2D from 0.54 to 0.35, while the BMI-related component remained perfectly correlated with BMI. Proteome-wide MR prioritized 29 proteins for BMI-subtracted T2D, including GCKR, HHIP, CTRB2, NOTCH2, RSPO3, NCAN, and PAM. Thirteen proteins showed eQTL colocalization in diabetes-relevant tissues, implicating hepatic, pancreatic-islet, adipose, and blood regulatory mechanisms. Sixteen proteins were classified within druggable-genome tiers, and existing compounds were identified for five targets.

Conclusion: Integrating GWAS-by-subtraction with proteome-wide MR identifies candidate protein targets for T2D pathways not fully explained by adiposity, supporting complementary therapeutic strategies beyond weight-centered mechanisms.

Keywords: Type 2 diabetes, GWAS-by-subtraction, Mendelian randomization, plasma proteomics, drug target prioritization

51 Dissecting Glycemic and Non-Glycemic Genetic Components of HbA1c Improves Diabetes Risk Stratification in Prediabetes

Yanzhen Cheng¹, Shihui Peng¹, Ou Ren¹, Tianyuan Lu¹

¹University of Wisconsin-Madison, Madison, Wisconsin, United States of America

Hemoglobin A1c (HbA1c) is widely used to monitor glycemia and diagnose diabetes, but it reflects both glycemic and non-glycemic pathways, which may complicate interpretation. We hypothesized that among individuals with borderline HbA1c values (prediabetes), those with lower genetic predisposition to non-glycemic HbA1c elevation would have worse underlying glycemic dysfunction and greater risk of future diabetes.

We performed a GWAS meta-analysis for HbA1c in approximately 1 million individuals of European ancestry from MAGIC, Million Veteran Program, NIH All of Us, and FinnGen. HbA1c-associated

variants were contrasted against GWAS of fasting glucose, fasting insulin, 2-h glucose, type 1 diabetes, and type 2 diabetes to separate glycemic and non-glycemic pathways. Variants associated with HbA1c but not glycemic traits were used to construct a non-glycemic polygenic risk score (ngPRS). Functional enrichment analyses and validation in UK Biobank were performed.

We identified 316 genome-wide significant HbA1c loci, including 83 designated as non-glycemic loci. Enrichment analyses demonstrated that non-glycemic HbA1c loci were strongly enriched for erythrocyte-related rather than glycemic pathways. In UK Biobank, ngPRS and conventional HbA1c PRS demonstrated distinct biomarker association patterns, consistent with different underlying biology. Among individuals with prediabetes defined by HbA1c, lower ngPRS was associated with increased risk of incident type 2 diabetes and any incident diabetes (both HR = 1.14, 95% CI: 1.11–1.17). Importantly, these associations remained consistent after adjustment for measured HbA1c, indicating that ngPRS provides information beyond the observed HbA1c level.

These findings demonstrate that genetically informed partitioning of HbA1c can improve risk stratification and may refine diabetes screening and prevention strategies.

52 Power Comparison of X-chromosome Association Methods - Application in Asthma

Mozart Nerva Deneus¹, Zhonglin Li¹, Nathalie Gaudreault¹, Alexandre Bureau^{2,3}, Sébastien Thériault^{1,4}, Yohan Bossé^{1,5}, Aida Eslami^{1,3}

¹Institut Universitaire de Cardiologie et de Pneumologie de Québec, Université Laval, Québec, Québec, Canada; ²Centre de Recherche Université Laval Robert Giffard, Québec, Canada; ³Department of Social and Preventive Medicine, Université Laval, Québec, Québec, Canada; ⁴Department of Molecular Biology, Medical Biochemistry and Pathology, Université Laval, Québec, Québec, Canada; ⁵Department of Molecular Medicine, Université Laval, Québec, Québec, Canada

Context: Asthma is a common respiratory disease influenced by both genetic and environmental factors. Genome-wide association studies (GWAS) have identified approximately 200 genetic loci associated with asthma. However, most published asthma GWAS have focused on autosomal variants and have routinely excluded the X chromosome. Analysis of the X chromosome is challenging because of differences in allele dosage between males and females, male hemizyosity, X-chromosome inactivation (XCI) in females, and the presence of different XCI patterns, including random, skewed, and escape from inactivation. Several statistical methods have been proposed for X-chromosome association analysis under different XCI assumptions.

Aim: To compare the statistical power of existing X-chromosome association methods under different XCI scenarios.

Methodology: We evaluated four X-chromosome association methods (Clayton, PLINK, XCMAX4, and Xlink) using simulated datasets. These methods were then applied to the Québec

City Case-Control Asthma Cohort (QCCAC), which included 1,618 French-Canadian participants comprising 1,089 asthma cases (660 women and 429 men) and 529 healthy controls (330 women and 199 men), and 14,129 genotyped X-chromosome SNPs. Association analyses were adjusted for age, sex, and the first 10 ancestry-based principal components.

Results: Power comparisons based on simulated data showed that Xlink was the most powerful method across all XCI scenarios. In the QCCAC dataset, the most strongly associated SNP was rs12396431 ($P = 1.42 \times 10^{-4}$, and $P = 6.58 \times 10^{-5}$ when adjusted only for sex), which was identified under a random XCI model. The minor allele frequency for this SNP was 0.261 in asthma cases and 0.199 in controls.

Conclusion: Including the X chromosome in GWAS may enhance understanding of the biological mechanisms underlying complex diseases.

53 Polygenic Mapping and CRISPR-Based Imaging of Latent Mechanisms Underlying Type 2 Diabetes Reveal Distinct Disease Risk and Cellular Programs

Satoshi Yoshiji^{1,2}, Rachel Diao³, Patrick Smadbeck², Trang Nguyen², Oliver Ruebenacker², Masashi Hasebe¹, Amit Majithia³, Jason Flannick^{2,4,5}

¹McGill University, Montreal, Québec, Canada; ²Broad Institute, Cambridge, Massachusetts, United States of America; ³University of California San Diego, San Diego, California, United States of America; ⁴Harvard Medical School, Boston, Massachusetts, United States of America; ⁵Boston Children's Hospital, Boston, Massachusetts, United States of America

Background: Precision medicine requires stratifying individuals according to the underlying mechanisms of disease and aligning therapeutics to those mechanisms. Yet for complex traits such as type 2 diabetes (T2D), these mechanisms are latent, and genetic discovery is often limited to clustering known loci into small sets, and supervised methods to generate mechanism-specific GWAS (mGWAS) from such clusters remain limited.

Materials and Methods: We developed ORIOLE, a supervised Bayesian multi-trait statistical method that generates mGWAS. ORIOLE learns multi-trait association signatures from pre-specified “seed” variants derived from T2D loci clustering and projects genome-wide variants onto learned signatures. We generated mGWAS for three major T2D mechanisms across 2.3 million variants. We constructed mechanism-specific polygenic scores (mPS) with PRS-CS and tested phenome-wide associations in All of Us ($n = 280,521$; 1,147 outcomes). We performed CRISPR perturbation of lipodystrophy-prioritized genes in SGBS preadipocytes with high-content imaging (3,752 features).

Results: The three mGWAS identified 176 (β -cell), 92 (obesity), and 111 (lipodystrophy) loci, including 46 missed by the largest T2D GWAS, with distinct tissue- and cell-type enrichment. Individuals in the top decile of each mPS were largely missed by an overall T2D genetic risk score (15.5–20.2% overlap), yet showed substantial mechanism-specific cardiovascular risk.

With CRISPR perturbation, putative lipodystrophy genes produced marked adipose phenotypes; *CITED2*, *CMIP*, and *FST* reduced lipid droplet area more than *PPARG* (a known T2D target) and showed coordinated actin remodeling.

Conclusion: ORIOLE provides a generalizable framework to map latent mechanisms at GWAS resolution and connect mechanism-anchored loci to risk stratification, therapeutic targets, and cellular programs.

Keywords: Type 2 diabetes; GWAS; Multi-trait Bayesian method; Polygenic risk score; CRISPR high-content imaging

54 Identifying Shared and Distinct Biology Across Cancers Using Pathway-Based Polygenic Risk Scores

Jiachen Zhang^{1*}, Austin H. Suger², Shuyan Cheng¹, Boya Guo³, Harriett Fuller³, Louisa B. Goss³, Tabitha A. Harrison¹, Sara Lindström², Burcu F. Darst³

¹*Institute of Public Health Genetics, University of Washington, Seattle, Washington, United States of America;* ²*Department of Epidemiology, University of Washington, Seattle, Washington, United States of America;* ³*Public Health Sciences, Fred Hutchinson Cancer Center, Seattle, Washington, United States of America*

Polygenic risk scores (PRS) aggregate genome-wide genetic effects into a single liability measure, potentially obscuring heterogeneous mechanisms underlying cancer risk. We derived pathway-based PRS (pPRS) from 21 cancer genome-wide association studies (GWAS) using PRSet across 50 Hallmark and 658 Kyoto Encyclopedia of Genes and Genomes (KEGG) MEDICUS pathways, then tested associations with 17 cancer phenotypes in the All of Us Research Program (v8; median 2,022 cases; range 948–15,494). Logistic models adjusted for age, sex where applicable, and 10 principal components. In stage one, pPRS were tested against their corresponding cancers with false discovery rate (FDR) correction (FDR < 0.05); in stage two, significant pPRS were tested across remaining cancers with Bonferroni correction ($p < 0.05$).

Stage one identified 986 significant pPRS associations across 14 cancers, recovering established biology including breast cancer FGFR-to-RAS/ERK signaling (OR = 1.10, 95% CI = 1.08–1.13, $P = 3 \times 10^{-23}$) and prostate cancer androgen response (OR = 1.18, 95% CI = 1.15–1.21, $P = 2 \times 10^{-36}$), which was absent from cross-cancer results. Stage two identified 274 significant cross-cancer associations. Telomere Elongation was the most recurrent shared pathway (51 associations across 15 cancers); the kidney cancer-derived score was associated with melanoma, leukemia, and thyroid cancer. Melanoma and non-melanoma skin cancer showed extensive bidirectional pathway sharing, including Fanconi anemia and microtubule-related pathways.

These findings show that pPRS can decompose cancer genetic architecture into interpretable shared and cancer-specific

mechanisms, highlighting telomere maintenance and Fanconi anemia DNA repair as pleiotropic cross-cancer mechanisms.

Keywords: Pathway-based PRS, cross-cancer pleiotropy, All of Us, cancer susceptibility, telomere maintenance

55 A Methodological Framework for a pan-European Genomic Reference Resource: Data Harmonization, Federated Analyses, and Early Results from the Genome of Europe Project

Anthony F. Herzig¹, Hugo Martiniano^{2,3}, Sumudu R. Samarasinghe¹, Helen Ray-Jones⁵, Jeroen van Rooij⁵, Emmanuelle Génin^{1,4}, Astrid Vicente^{2,3}, André G. Uitterlinden⁵, the Genome of Europe consortium⁶

¹*Inserm, Univ Brest, EFS, UMR 1078, GGB, Brest, France;* ²*Instituto Nacional de Saúde Doutor Ricardo Jorge, Av Padre Cruz, Lisbon Portugal;* ³*Biosystems and Integrative Sciences Institute (BioISI), Faculdade de Ciências da Universidade de Lisboa, Lisboa Portugal;* ⁴*CHRU Brest, Brest, France;* ⁵*Laboratory for Population Genomics, Erasmus MC, Rotterdam, The Netherlands;* ⁶*Genome of Europe Consortium, Digital Europe Program, European Union*

Within the 1 million genomes initiative, the digital Europe funded project Genome of Europe (GoE) is currently collecting whole-genome sequencing (WGS) data from >100,000 European citizens across 51 contributing partners from 27 countries. By 2028, this will form a new resource of reference genomes, including a more complete representation of minority groups than in previous European panels.

GoE operates in a federated environment, which raises technological and methodological challenges. We first tested statistical methods for federated principal components analysis to describe the global structure across the contributing cohorts. In parallel, we developed bioinformatic pipelines to harmonize variant-calling and aggregate allele frequencies across the genome, including specifically calling haplotype-frequencies in pharmacogenes with known clinical consequences.

We demonstrate the importance of variant selection and data harmonization for federated principal components analyses, and on cataloguing and describing allele-frequency variations across the continent. We were able to observe significant heterogeneity in allele-frequencies in important pharmacogenes in an initial pilot study on ~15,000 individuals. In conjunction with the European Genomic Data Infrastructure, we have developed an interface for allele-frequency queries, adapted for wide-scale WGS data. We then extend to descriptions of polygenic risk distributions for established genetic scores for cancers; to determine how they can covary at a fine resolution with ancestry gradients across cohorts. We also propose and test workflows for phasing and anonymisation, as to create imputation reference panels suitable for collaboration with the GDPR-compliant Helmholtz imputation server. Finally, we detail a roadmap for integrating and jointly analysing the envisaged ~30,000 long-read samples that will be generated in the initiative.

56 Familial aggregation, Familiarity and Genome wide associations for Dizziness in the Lifelines Cohort

Bandana Mainali^{1*}, Martje Bos², Ilja M. Nolte³, Harold Snieder³, Sonja J. Pyott^{1,4}

¹Department of Otorhinolaryngology, University Medical Center Groningen, The Netherlands; ²Department of Psychiatry, University Medical Center Groningen, The Netherlands; ³Department of Epidemiology, University Medical Center Groningen, Groningen, The Netherlands; ⁴Department of Neuroscience, University of Oldenburg, Germany

Background: Dizziness is a common and heterogeneous symptom. Twin studies suggest heritability for chronic dizziness (27%–47%). However, studies on familial aggregation in the general population remain scarce, and genome-wide association study (GWAS) findings are limited. We evaluated familial aggregation, familiarity, and GWAS for multiple dizziness phenotypes in a large multi-generational cohort.

Methods: We analyzed self-reported dizziness from up to 141,751 adults in the Lifelines Cohort. Dizziness severity was measured on a five-level scale at baseline and recoded into four trait definitions for familial analyses using different thresholds. Dizziness with falls was defined using baseline and longitudinal binary responses. Familial aggregation was assessed using recurrence risk ratios in first- and second-degree relatives (FDRs and SDRs, respectively), and familiarity was estimated as the proportion of trait variance (liability scale) explained by familial factors. GWASs were performed with Regenie adjusting for age, sex, and principal components.

Results: For FDRs (SDRs) the recurrence risk ratios ranged from 1.23 (1.15) for any kind of dizziness to 1.96 (1.60) for the most severe dizziness categories. For dizziness with falls they were similar for FDRs and SDRs (1.52 and 1.55, respectively). Familiarity estimates were up to 24.4% for dizziness severity and 17.2% for dizziness with falls. GWASs identified significant hits on chromosome 4 for dizziness severity and on chromosome 6 for dizziness with falls.

Conclusion: Both dizziness and dizziness with falls are for a moderate part explained by genetic factors. GWAS identified a few genomic loci, which require post-GWAS follow up.

Keywords: Dizziness, Familiarity, Familial aggregation, GWAS, Prospective cohort

57 Personalised Polygenic Score Through Genetic Ancestry Continuum

*Chuan Fu Yap¹, Andrew P Morris¹

¹Arthritis UK Centre for Genetics and Genomics, The University of Manchester, Manchester, United Kingdom

A polygenic score (PGS) is a sum of alleles at SNPs, weighted by their estimated effect sizes, used to predict an individual's

genetic liability to a disease or trait. Current approaches typically utilise effect size estimates derived from population-based summary statistics. As a result, these models often represent average genetic effects in the population and may not adequately capture individual-level genetic heterogeneity. This can limit predictive performance and transferability across genetically diverse individuals. To address this limitation, we propose a method for PGS that accounts for ancestry through the “genetic ancestry” continuum. The method derives ancestry-adjusted personalised PGS based on the genome-wide genetic background of each individual by allowing allelic effect heterogeneity in the weights assigned to variants included in the score. This is achieved using a continuous multi-dimensional representation of genetic diversity by projecting discovery and target data onto genetic principal components (PCs) derived from the Human Genome Diversity Project. The resulting PCs are used to train SNP×PC interaction terms (e.g., SNP×PC1–3) on SNPs from the discovery genome-wide association study (GWAS). The ancestry-adjusted weight for the i th individual in the target data at the j th variant is given by, where α_j denotes the main allelic effect of the variant j , β_{kj} denotes the interaction effect between variant j and the k th axis of genetic variation, and z_{ik} is the position of individual i along the k th genetic axis, estimated from the discovery GWAS. The utility of this approach is demonstrated through application to type 2 diabetes.

58 Cross-tissue Ancestry-Specific TWAS Refines Type 2 Diabetes GWAS loci

Inti Pagnuco^{1*}, Stephen Eyre¹, Magnus Rattray², Andrew P. Morris¹

¹Arthritis UK Centre for Genetics and Genomics, Centre for Musculoskeletal Research, Division of Musculoskeletal and Dermatological Sciences, The University of Manchester, Manchester, United Kingdom; ²Division of Informatics, Imaging and Data Sciences, The University of Manchester, Manchester, United Kingdom

Type 2 diabetes (T2D) is a metabolic disorder characterized by hyperglycemia and insulin resistance. Although genome-wide association studies (GWAS) have identified numerous T2D risk loci, the causal genes and tissues mediating these associations remain unclear.

To address this, we performed tissue-specific, ancestry-aware transcriptome-wide association studies (TWAS) across six T2D-relevant tissues: subcutaneous adipose, visceral adipose, hypothalamus, liver, skeletal muscle and pancreas. We conducted ancestry-specific UTMOST-TWAS in European (EUR) ancestry data using EUR-GWAS summary statistics (242,283 cases; 1,569,734 controls) and gene expression prediction models derived from GTEx EUR individuals ($n = 689$).

In EUR analyses, we identified 684–750 significant gene-T2D associations per tissue ($P < 1.91 \times 10^{-6}$), implicating established and novel candidate genes. Of these, *JAZF1* showed consistent

association signals across all six tissues, whereas *TCF7L2* exhibited signals restricted to a subset of T2D-relevant tissues, with heterogeneous effects across tissues. Conditional analyses refined these findings to 289–322 independent signals per tissue.

Cross-ancestry replication of EUR-derived TWAS signals was evaluated in African American (AFA) ancestry datasets. We conducted AFA-TWAS using AFA-GWAS data (50,251 cases; 103,909 controls) in combination with prediction models trained in GTEx individuals of AFA ancestry ($n = 111$). Significant enrichment of EUR-derived TWAS signals was observed in AFA-TWAS across subcutaneous and visceral adipose, skeletal muscle, and pancreas, while replication was weaker in liver, likely due to smaller sample size.

Overall, tissue-specific and ancestry-aware TWAS improves identification of causal T2D genes, with cross-ancestry replication supporting signal robustness and highlighting the need for larger, more diverse ancestry-specific datasets.

59 An Assessment of the Contribution of Structural Variation to Known Pulmonary Fibrosis Loci

Power H J^{1,2}, Joof E^{1,2}, Massen G M³, Quint J K³, Jenkins R G⁴, Fawcett K A^{1,2}, Allen R J^{1,2}, Wain L V^{1,2}

¹*Division of Public Health and Epidemiology, School of Medical Sciences, University of Leicester, Leicester, United Kingdom;* ²*University Hospitals of Leicester NHS Trust, Leicester, United Kingdom;* ³*School of Public Health, Imperial College London, London, United Kingdom;* ⁴*National Heart & Lung Institute, Imperial College London, London, United Kingdom*

Pulmonary fibrosis (PF) is the scarring of lung tissue, leading to lung function impairment. 62 single-nucleotide polymorphism (SNP) associations with PF have been reported, but the effects of structural variants (SVs: genetic variants ≥ 50 base pairs) are largely unexplored. We aimed to assess whether previously-associated SNPs may be tagging SVs by investigating linkage disequilibrium (LD).

As SVs were called at the individual participant level in UK Biobank, SVs were merged across participants using bedtools intersect and a map of SVs from the All of Us research program. Variants were merged if their reciprocal overlap exceeded 50%. Using genotype matrices, LD was calculated between the 62 SNPs and all SVs. Pairs were considered to be in LD if $r^2 \geq 0.8$. Conditional analyses were performed for these pairs: the SNP was tested for association with PF with and without incorporating the SV as a covariate. Unrelated individuals of European ancestry were included in the study. PF cases were selected using International Classification of Diseases, Tenth Revision (ICD-10) codes, with seven controls selected per case, matched on age and sex (1,755 cases and 12,285 controls). Association testing was performed using PLINK2, with the first ten ancestry-based principal components as covariates.

Two PF-associated SNPs, rs242561 and rs112572874, were found to be in LD with two intronic deletions on known SV hotspot *17q21.31* ($r^2 \geq 0.959$). Conditional analyses showed attenuation of both associations upon SV incorporation, suggesting the variants may represent a shared signal, though further research is needed to disentangle causality.

Keywords: Linkage disequilibrium, Structural variation, Large-scale biobanks, Respiratory system, Genetic epidemiology

60 SOLAMAME: Software for Large-scale Multiallelic Variant and Meta Analyses

Yuriko Katsumata^{1,2*}, Khine Zin Aung^{1,2}, Inori Tsuchiya^{2,3}, Xian Wu^{1,2}, Qi Qiao³, Erin L. Abner^{2,4}, Peter T. Nelson^{2,5}, Shubhabrata Mukherjee⁶, David W. Fardo^{2,3}

¹*Division of Biomedical Informatics, University of Kentucky, Lexington, Kentucky, United States of America;* ²*Sanders-Brown Center on Aging, University of Kentucky, Lexington Kentucky, United States of America;* ³*Department of Biostatistics, University of Kentucky, Lexington, Kentucky, United States of America;* ⁴*Department of Epidemiology and Environmental Health, University of Kentucky, Lexington, Kentucky, United States of America;* ⁵*Department of Pathology, University of Kentucky, Lexington, Kentucky, United States of America;* ⁶*Department of Medicine, University of Washington, Seattle, Washington, United States of America*

Multiallelic variants represent a substantial component of human genetic diversity, yet they are frequently excluded from genetic association studies because most existing analytical tools are designed for biallelic loci. We developed Software for LArge-scale MultiAllelic variant and MEta analyses (SOLAMAME), a Python package for multiallelic variant association testing and meta-analysis that preserves the allelic structure across multiple studies while enabling both global and allele-specific inference. We evaluated the SOLAMAME performance using simulated phenotypes and multiallelic genotype data from the 1000 Genomes Project Phase 3. In addition, we applied SOLAMAME to whole-genome sequencing data from the Alzheimer's Disease Sequencing Project and Alzheimer's disease-related neuro-pathological endophenotypes from the National Alzheimer's Coordinating Center dataset. Simulation studies examined empirical type I error rates and statistical power for both continuous and binary outcomes under mega-analysis and meta-analysis frameworks. Power was evaluated across 10,000 simulations through three scenarios: (1) a primary alternative allele as the causal allele, (2) a secondary alternative allele as the causal allele, and (3) multiple causal alternative alleles with effects in opposite directions. Real-data analyses tested associations between multiallelic variants and both a continuous Alzheimer's disease related latent trait and dichotomized Braak neurofibrillary tangle stage (V–VI vs. 0–IV). SOLAMAME demonstrated well-controlled type I error rates and strong statistical power across simulation scenarios. In real-data applications, the framework identified multiallelic variants associated

with Alzheimer's disease-related phenotypes, highlighting its robustness, scalability, and utility for comprehensive analysis of multiallelic variation in large-scale genomic studies.

61 Fine-Scale Regional Ancestry Classification: A Genealogically Validated Benchmark With Applications to Rare Variant Enrichment

Gilles-Philippe Morin¹⁻³, Claudia Moreau^{1,2}, Amadou Barry²⁻⁴, Simon L. Girard^{1-3,5,6*}

¹Département des sciences fondamentales, Université du Québec à Chicoutimi, Saguenay, Québec, Canada; ²Centre Intersectoriel en Santé Durable (CISD), Université du Québec à Chicoutimi (UQAC), Saguenay, Québec, Canada; ³Unité mixte de recherche INRS-UQAC en santé durable, Saguenay, QC, Canada; ⁴Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique (INRS), Laval, Québec, Canada; ⁵Projet BALSAC, Université du Québec à Chicoutimi (UQAC), Saguenay, Québec, Canada; ⁶Centre de recherche CERVO, Université Laval, Québec, Québec, Canada

*Corresponding author: simon2_girard@uqac.ca

Population biobanks often lack precise information on the deep geographic origins of their participants, relying instead on recent self-reported data that may not reflect historical population structure. In Québec's CARTaGENE (CaG) cohort, this is particularly relevant for isolating individuals whose ancestry traces to distinct regions such as Saguenay-Lac-Saint-Jean. Existing studies have addressed this through clustering and classification approaches, but these methods are often based on coarse continental background rather than fine-scale population structure, applied inconsistently across literature, and rarely validated beyond a single population.

CaG is uniquely positioned to overcome this gap. A substantial subset of its genotyped participants have been linked to BALSAC, a population register encompassing over three centuries of Québec demographic history reconstructed from civil records, allowing ancestral regional origins to be established with genealogical precision, a demographic ground truth unavailable in most biobanks.

We leverage this resource to classify all of CaG in a two-part study. First, we systematically benchmark published clustering and semi-supervised classification methods, evaluating concordance with genealogically-verified regional origins in individuals with ancestry concentrated to a single region. Second, we apply the best-performing classifier to rare variant data to identify variants likely enriched in specific regions, and validate these findings against published variant descriptions.

Beyond its direct applications to carrier frequency estimation and rare disease genetics, this framework establishes a genealogically-grounded standard for regional ancestry classification, that, once validated, can be exported to any population-scale biobank without genealogical data.

Keywords: fine-scale ancestry classification, population structure, rare variant enrichment, genealogical validation, population biobank

62 Gene-Based and Set-Based Tests of Primary Ciliary Dyskinesia Gene Variants and Asthma Susceptibility in UK Biobank

Kishore Palanisamy¹, Ian P. Hall¹, Matthew J. Saward², Ian Sayers¹, Harriet Holme³, Katherine A. Fawcett^{2*}

¹Centre for Respiratory Research, NIHR Biomedical Research Centre, School of Medicine, Biodiscovery Institute, University of Nottingham, Nottingham, United Kingdom; ²Division of Public Health and Epidemiology, School of Medical Sciences, University of Leicester, Leicester, United Kingdom; ³PCD Research, United Kingdom

Background: Primary ciliary dyskinesia (PCD) is a rare autosomal recessive condition caused by mutations in genes that encode motile cilia, leading to impaired mucociliary clearance. Although asthma is more common in individuals with PCD, the role of heterozygous PCD gene variants in asthma risk is not well understood.

Objective: To investigate whether heterozygosity for known or likely pathogenic variants in PCD genes is associated with susceptibility to childhood-onset asthma (COA) or adult-onset asthma (AOA).

Methods: Exome sequencing from UK Biobank was analysed to test associations with COA (9,971 cases, 349,163 controls) and AOA (27,640 cases, 339,616 controls) separately. Gene-based tests were performed for 46 autosomal PCD genes using rare (minor allele frequency ≤ 0.01) loss of function and missense variants annotated by UK Biobank to maximise power. Where sufficient numbers of known or likely pathogenic variants were available, additional gene-based and set-based analyses were performed.

Results: Large numbers of rare variants were identified across the 46 PCD genes; however, no gene showed a statistically significant association with COA or AOA in gene-based analyses of loss of function and missense variants. Gene-based and set-based analyses restricted to known or likely pathogenic mutations revealed weak associations of *DNAH5* and *DNAH11* with COA (OR = 1.61, $p = 0.012$; OR = 1.72, $p = 0.015$ respectively).

Conclusion: This large population-based study provides weak evidence that heterozygosity for pathogenic PCD variants in *DNAH5* and *DNAH11* increases susceptibility to COA. There

was no evidence for high-effect variants across other PCD genes. Modest effects of very rare pathogenic variants cannot be excluded and warrant further study.

63 Leveraging Runs of Homozygosity to Identify Rare Susceptibility Loci for Cardiometabolic Traits in Middle Eastern individuals

Abdullah Shaar¹, Areeba Irfan², Asma Saeed², Sidonie Foulon³, Hervé Perdrey³, Anne-Louise Leutenegger³, Mohamad Saad¹

¹*Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar;* ²*College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar;* ³*Université Paris-Saclay, UVSQ, Inserm, CESP, Paris, France*

Background: Long autozygous segments known as runs of homozygosity (ROH) result from high rates of consanguinity/inbreeding. ROH segments might harbor rare variants with recessive effects that impact complex traits. Our goal is to use ROH segments to identify rare variants with recessive effects in the Middle East populations.

Methods: We analyzed 24,699 individuals with WGS from the Qatar Precision Health Institute. ROH was inferred from SNVs using *Fantasio2* (Hidden Markov Model-based approach). Association analysis was performed on inbred individuals between ROH and cardiometabolic traits (e.g., T2D, Obesity, HbA1C, Glucose, LDL, ECG PR interval, BMI, and DBP). Regions with $P < 5 \times 10^{-6}$ were considered for fine mapping where additive and recessive association testing and functional annotation were performed on rare variants.

Results: 48% of individuals were inbred. Two ROH genomic regions were prioritized for binary traits: chr2:240.7–241.0 Mb for T2D, which includes *CAPN10* (implicated in non-insulin-dependent diabetes), and chr6:67.27–67.42 Mb for Obesity. For quantitative traits, a few regions reached $P < 5 \times 10^{-6}$: chr14:53.72–54.67 Mb for HbA1C and Glucose (*BMP4* and *CDKN3*); chr6:0.23–0.25 Mb for DBP (*DUSP22* and *IRF4*); and chr6:55.71–55.75 Mb for PR interval. The fine mapping using single marker association prioritized significant additive rare variants within *BMP5* for PR, *CROCC2* and *KIF1A* for T2D, *SAMD4A* and *CGRRF1* for HbA1C and Glucose. A few enriched variants in our cohort showed recessive effects such as *EXOC5* for HbA1C, *EXOC2* for T2D, and *BMP5* for PR interval.

Conclusion: Homozygosity mapping in a highly consanguineous Qatari cohort showed potential to uncover novel susceptibility variants for cardiometabolic traits.

65 Large-scale Multi-omics Analyses of Cardiometabolic Traits in the Middle East

Areeba Irfan¹, Asma Saeed¹, Abdullah Shaar², Khalid Kunji², Faaiz Joad², Parvathy Premkumar², Amna Mohammad Salim², Tamara Saidelartah¹, Ehsan Ullah², Georges Nemer¹, Mohamad Elsharif², Mohamad Saad²

¹*College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar;* ²*Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar*

Background: Genome-wide association studies (GWAS) have improved understanding of cardiometabolic trait genetics, but biological interpretation remains challenging. To address this, we developed a multi-omics framework combining genomics, metabolomics, proteomics, and cardiometabolic traits in an underrepresented population.

Methods: Our study focuses on a Middle East cohort of 24,000 whole-genome sequences from the Qatar Precision Health Institute. We conducted GWAS for 24 cardiometabolic traits, 783 metabolites, and 1305 proteins. We performed colocalization analyses of trait genome-wide significant (GWS) loci using *coloc*. SNV-based heritability and genetic correlation for cardiometabolic traits, metabolites, and proteins were performed. A *PheWeb* platform for summary statistics visualization across traits and omics was developed.

Results: We identified 279 lead SNVs across cardio-metabolic traits (20 novel SNVs in 10 traits). 33 SNVs were GWS pQTLs and 9 were GWS mQTLs. Colocalization analyses showed high-confidence colocalization ($PP_4 > 0.99$) for 14 trait-metabolite and 12 trait-protein. High-confidence trait-metabolite colocalizations were found at the *GCKR* triglycerides (TG) locus. Trait-protein colocalizations were observed for four genes: *CELSR2* for LDL, *FXYP6*, *APOC1/APOE*, and *GCKR* for TG. Estimates of SNV-based heritability for cardiometabolic traits showed highest heritability for HDL ($h^2 = 0.39$). Binary traits such as HTN and T2D showed 13% and 22% heritability. Proteins showed higher heritability estimates ($h^2 = 0.04$ – 0.904) than metabolites ($h^2 = 0.05$ – 0.765).

Conclusion: This multi-omics framework links genetic variation to molecular mechanisms, improving biological interpretation and target prioritization for cardiometabolic traits in Qatari population. Ongoing analyses include Mendelian Randomisation and Transcriptomic-wide association studies (TWAS).

Keywords: Genome-wide association studies, Multi-omics, Cardiometabolic Traits, Middle East

66 Integrative Multi-Omic and Epidemiologic Analyses Reveal Convergent Immune and Transcriptomic Dysregulation in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome

R. Moslehi^{1,2}, A. Kumar¹, and A. Dzutsev²

¹*College of Integrated Health Sciences, University at Albany, State University of New York, Albany, New York, United States of America;* ²*National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America*

Background: Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a disabling multisystem disorder with no established etiology or approved treatment. More than 24

million people are affected globally and the COVID-19 pandemic may substantially increase the disease burden. We conducted a molecular epidemiologic study to identify risk factors and biologic mechanisms for ME/CFS.

Methods: Our clinic-based case-control study included 60 carefully characterized ME/CFS patients and 61 matched healthy controls. We evaluated: (1) prevalence of autoimmune disease (AID) and cancer among first-degree relatives; (2) epidemiologic risk factors; (3) serum levels of 48 cytokines; and (4) whole-blood RNA-seq gene expression profiles. Conventional statistical and machine learning approaches were used to identify associative and predictive signatures.

Results: First-degree relatives of ME/CFS cases were more likely than those of controls to have AID [relative risk (RR) = 3.52, $p = 0.0014$] and early-onset cancer (< 60 years) (RR = 2.24, $p = 0.034$). Epidemiologic analyses identified several significant risk factors, including history of allergies requiring medication [odds ratio (OR) = 6.00, $p < 0.0001$] and contaminant exposure (OR = 4.35, $p = 0.0002$). Cytokine profiling identified key predictors including IL-27, IP-10, RANTES, and Fractalkine. Whole-blood RNA-seq analysis identified 115 differentially expressed genes, enriched in pathways related to immune regulation, mitochondrial function, metabolism, and cellular stress responses. FGSEA further highlighted coordinated alterations in related pathways.

Conclusions: This multidimensional analysis identified previously unreported risk factors, familial associations with AID and cancer, and convergent immune and transcriptomic alterations that provide insight into ME/CFS etiology and highlight potential diagnostic biomarkers and therapeutic targets.

67 Novel regulatory X-chromosome CpG loci Associated With Recurrent Venous Thromboembolism Are not Explained by Variable Escape From X-chromosome Inactivation

Ohanna C. L. Bezerra^{1*}, Qicheng Zhao², Kathleen Oros Klein³, Marc Rodger⁴, Michael J. Kovacs⁵, France Gagnon^{1,6}, Celia M. T. Greenwood^{2,3,7}

¹Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; ²Department of Epidemiology, Biostatistics & Occupational Health, McGill University, Montreal, Canada; ³Lady Davis Institute, Sir Mortimer B. Davis Jewish General Hospital, CIUSSS du Centre-Ouest-de-l'Île-de-Montréal, Montreal, Canada; ⁴Department of Medicine, McGill University, Montreal, Canada; ⁵Department of Medicine, Western University, London, Canada; ⁶University of Toronto Mississauga, Toronto, Canada; ⁷Gerald Bronfman Department of Oncology, McGill University, Montreal, Canada

Sex differences in recurrent venous thromboembolism (rVTE) remain poorly understood. Although first VTE incidence is similar between sexes, males have ~2.5-fold higher risk of

recurrence after anticoagulant discontinuation. Given the enrichment of immune- and coagulation-related genes on the X chromosome, sex-dependent X-linked regulatory mechanisms may contribute to this disparity. However, the X chromosome is frequently excluded from epigenome-wide association studies (EWAS) due to challenges related to X-chromosome inactivation (XCI) and variable escape from XCI.

We performed a sex-stratified X-EWAS using Illumina EPIC methylation data from 417 REVERSE I participants, applying survival models adjusted for age, methylation principal components, and cell-type composition. We additionally applied a novel framework developed by our group to detect methylation patterns consistent with locus-specific variable escape from XCI (LVE-X). RNA-seq data from rVTE cohorts (GSE19151) were used for functional comparison.

We identified hypomethylated CpGs associated with time to rVTE at several X-linked loci, including *GAB3*, *UBA1*, *GRIA3*, and *BCOR*. The strongest female-specific association mapped to an active transcription factor binding region overlapping a predicted CTCF binding site at *GAB3*, a regulator of cytokine signaling and macrophage differentiation implicated in inflammatory coagulation pathways. *GAB3* expression was stable across rVTE groups in females, whereas males showed higher expression in recurrent cases. *UBA1*, an established XCI escapee gene associated with VEXAS-related thrombosis, also emerged as a candidate locus. Although these loci showed methylation patterns suggestive of variable XCI escape, estimates did not differ between recurrent and non-recurrent VTE groups, supporting alternative sex-specific regulatory mechanisms underlying rVTE risk.

68 Genetic Evidence Supports a Protective Role of Immunoglobulin M in Alzheimer's Disease

Shihui Peng^{1*}, Guillaume Butler-Laporte^{2,3,4}, Sterling C. Johnson^{5,6,7}, Corinne D. Engelman¹, Tianyuan Lu^{1,8,9,10,11}

¹Department of Population Health Sciences, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, United States of America; ²Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada; ³Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Québec, Canada; ⁴Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; ⁵Division of Geriatrics and Gerontology, Department of Medicine, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, United States of America; ⁶Wisconsin Alzheimer's Disease Research Center, Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, United States of America; ⁷Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, United States of America; ⁸Department of Biostatistics and Medical Informatics, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, United States of America; ⁹Center for Precision Medicine, University of Wisconsin-

Madison, Madison, Wisconsin, United States of America;¹⁰Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, Wisconsin, United States of America; ¹¹Center for Demography of Health and Aging, University of Wisconsin-Madison, Madison, Wisconsin, United States of America

Immune dysfunction has been implicated in Alzheimer's disease (AD), motivating therapeutic interest in intravenous immunoglobulin (Ig) treatments. However, clinical trials have been largely null, potentially reflecting IgG-dominant formulations that overlook other Ig classes with distinct biological roles. We integrated complementary approaches across multiple cohorts to evaluate genetically predicted IgG, IgA, and IgM levels in relation to AD as well as cerebrospinal fluid (CSF) and plasma biomarkers.

Two-sample Mendelian randomization (MR) were performed using largest-to-date GWAS of circulating Ig, AD, and CSF AD biomarkers. Polygenic risk scores (PRS) for Ig classes were optimized in All of Us Research Program and tested for associations with AD in 355,241 unrelated European ancestry UK Biobank participants using Cox regression. Associations with plasma AD biomarkers were evaluated in 1,413 participants from the Wisconsin Registry for Alzheimer's Prevention (WRAP) and Wisconsin Alzheimer's Disease Research Center (ADRC).

MR analyses demonstrated that higher genetically predicted IgM levels were associated with lower AD risk (IVW OR = 0.81 per SD increase, $P = 2.6 \times 10^{-8}$) and favorable CSF biomarker profiles ($\beta_{A\beta_{42}} = 0.12$, $\beta_{p\tau_{181}} = -0.07$, $\beta_{total-\tau} = -0.08$; all $p < 0.05$). Results were highly consistent using alternative MR methods. In UK Biobank, IgM PRS was associated with lower AD hazards (HR = 0.92, $P = 2.0 \times 10^{-5}$) and consistent after adjusting for potential socioeconomic confounders. In WRAP and Wisconsin ADRC, higher IgM was associated with favorable plasma biomarkers ($\beta_{A\beta_{42}/40} = 0.08$, $P = 5.0 \times 10^{-3}$; $\beta_{p\tau_{217}} = -0.10$, $P = 2.3 \times 10^{-4}$). Contrarily, IgG and IgA showed no consistent associations.

These findings provide converging evidence supporting a protective role of IgM in AD and suggest IgM-related pathways as potential targets for prevention and early intervention.

69 AfriGWAtlas: An Atlas of Genome-wide Associations in African-Ancestry Populations

Thais S. Tavares¹, Victoria E. Brown¹, Thiago P. Leal¹, Ana Laura Grossi de Oliveira², Adebawale A. Adeyemo³, Charles N. Rotimi³, Daniel Shriner³, Rick A. Kittles⁴, Karlijn A.C. Meeks⁵, and Mateus H. Gouveia^{1,3*}

¹Department of Public Health Education, Morehouse School of Medicine, Atlanta, Georgia, United States of America; ²Infectious Diseases and Tropical Medicine Program, Universidade Federal de Minas Gerais, MG, Brazil; ³Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, Maryland, United States of America; ⁴Department of

Community Health and Preventive Medicine, Morehouse School of Medicine, Atlanta, Georgia, United States of America; ⁵Departments of Medicine and Epidemiology & Public Health, University of Maryland School of Medicine, Baltimore, Maryland, United States of America

African-ancestry populations (AFR) harbor the greatest human genetic diversity and capture much variation present in non-African populations (non-AFR). Yet genome-wide association studies (GWAS) remain dominated (~88%) by European-ancestry populations (EUR), with less than 4% of participants being AFR. Existing AFR genome-wide associations have not been systematically compiled, limiting understanding of their trait distribution, allele frequency (AF) differences across populations, and potential utility for discovery (meta-analysis and replication). We conducted a systematic review of GWAS, GWAS meta-analyses, and admixture mapping studies reporting AFR associations since 2004 (PROSPERO: CRD420251158914), resulting in AfriGWAtlas, a curated atlas of genome-wide significant associations identified in AFR. Notably, two-thirds of included GWAS (247; 64%) were not indexed in the GWAS Catalog and required manual curation. Compared with EUR studies from the GWAS Catalog, AFR studies were enriched for substance use (3.6-fold), infectious disease (2.7-fold), and cardiovascular traits (1.7-fold), whereas EUR studies were enriched for cancer (2.7-fold) and immune-inflammatory traits (2-fold). We identified 1,879 independent associations/loci across traits, predominantly endocrine/metabolic (748; 39.8%) and cardiovascular (300; 16.0%). Among these loci, 1,162 (61.8%) showed substantial AF differences ($\geq 10\%$) between AFR and non-AFR, 918 (48.9%) showed $\geq 5\%$ AF variation across AFR sub-continental groups, and 123 (6.5%) were AFR-specific. The strongest subcontinental AF differentiation was observed at *APOL1* rs73885319 (43.1% in Nigerians, 22.2% in Gambians, and 4.5% in Kenyans). AfriGWAtlas provides the first systematically curated, scalable, and updatable resource of genome-wide significant associations in AFR.

70 Characterizing Sex Differences in the Effect of APOE on Lipid Profiles and Cognitive Trajectories

Sara N. L. Gunderson^{1,2*}, Freida Blostein³, Sarah A. Gagliano Taliun^{4,5}, Jessica K. Dennis^{1,2}

¹Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; ²British Columbia Children's Hospital Research Institute, Vancouver, British Columbia, Canada; ³Windreich Department of Artificial Intelligence and Human Health, Icahn School of Medicine, Mount Sinai, New York City, New York, United States of America; ⁴Department of Medicine, Université de Montréal, Montreal, Québec, Canada; ⁵Department of Neuroscience, Université de Montréal, Montreal, Québec, Canada

The $\epsilon 4$ allele of the apolipoprotein E gene (*APOE*) is the strongest genetic risk factor for late-onset Alzheimer's disease (AD), the most common form of dementia. Female carriers are especially at

risk, experiencing earlier symptom onset and steeper cognitive decline compared to male carriers. However, the biological mechanisms driving these sex differences remain poorly understood. Elevated non-high-density lipoprotein (non-HDL) cholesterol is one of the top modifiable risk factors for AD, and a potential contributor to these sex differences. *APOE* regulates lipid transport and metabolism in the periphery and the brain. *APOE* $\epsilon 4$ is associated with higher non-HDL compared to $\epsilon 3$ and $\epsilon 2$. Lipids also vary by sex across the life course. We hypothesize that non-HDL partially mediates the association between *APOE* $\epsilon 4$ and cognitive decline in a sex-specific manner. The Canadian Longitudinal Study on Aging (CLSA) is a national prospective cohort study of ~26,000 genotyped adults aged 45–85 at baseline, who are followed every 3 years until 2033 or death. We used linear and linear mixed-effects regression to study how associations between *APOE*, non-HDL, and cognitive trajectories differed between sexes and across the life course. In females, we observed stronger protective effects of *APOE* $\epsilon 2$ on non-HDL (p value < 0.05) and more damaging effects of *APOE* $\epsilon 4$ on cognitive trajectories (p value < 0.05) than in males. We are replicating these findings across large-scale cohort studies and electronic health record-linked biobanks across North America, totaling > 1 million individuals. This work will highlight sex-differential windows for therapeutic intervention.

71 Menopause-Associated Biological Age Acceleration and Its Contribution to Increased Blood Pressure

Jennifer Zevounou^{1,2}, Guillaume Lettre^{1,2}

¹*Institut de Cardiologie de Montréal, Montréal, Canada;* ²*Université de Montréal, Montréal, Canada*

Biological age reflects physiological rather than chronological aging. Age acceleration (AA), defined as the difference between biological and chronological age, is associated with cardiovascular disease (CVD) risk. However, its role in menopause-related increases in CVD risk remains unclear despite postmenopausal females exhibiting both increased AA and higher CVD risk.

Conventional and organ specific AA were estimated using a plasma protein-based aging clock in participants from the UK Biobank Pharma Proteomic Project ($N = 54,219$). Focusing on female, we evaluated the impact of menopause on AA, systolic and diastolic blood pressure (SBP, DBP), adjusting for relevant cofounders. We also re-examined previously reported BP-associated loci identified in female-specific GWAS to test whether the menopause status or AA modified these associations.

We observed increased AA after menopause in both the conventional ($\Delta_{\text{POST-PRE}} = 0.92$ year, p -value < 0.001) and some organ specific models (eg. Kidney AA $\Delta_{\text{POST-PRE}} = 0.32$ year, p -value < 0.001). Stepwise regression identified chronological age, body mass index, testosterone, kidney AA, and conventional AA as the best BP descriptors (all $p < 0.001$ in

multivariate models). Although menopause status was associated with BP in univariate models, this association was non-significant after adjusting for chronological age, suggesting menopause effects are difficult to distinguish from overall aging. Among 391 BP-associated loci from female-specific GWAS, five loci showed significant interaction with menopause (Bonferroni p -value < 0.05). However, these loci also interacted with chronological age but not with conventional or kidney AA.

In conclusion, biological age acceleration increases after menopause and is associated with BP, although menopause-related BP changes appear substantially driven by chronological aging.

Keywords: menopause, blood pressure, biological age

72 Interaction Index: A Novel Unified Measure for Interaction

Zeluan Li¹, Zewei Xiong², Thuan Quoc Thach¹, Evangelos Vassos³ and Pak Sham^{1*}

¹*Department of Psychiatry, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong;* ²*Department of Mathematics, Faculty of Science, Hong Kong Baptist University, Hong Kong;* ³*Institute of Psychiatry, Psychology and Neurology, King's College London, United Kingdom*

*Corresponding author email: pcsham@hku.hk

The assessment of statistical interaction between risk factors for disease is highly dependent on the choice of scale, which may lead to conflicting and potentially misleading conclusions. Here, we introduce the Interaction Index, derived by rescaling the relative excess risk due to interaction (RERI). This index assesses interaction on a unified continuous scale and classifies different interaction modes (additive and multiplicative risk effects) for two binary risk factors. We evaluated the theoretical behaviour of this index under different etiological scenarios in a liability-threshold framework, and conducted simulation studies to assess the statistical performance of the estimator of this index. The theoretical behaviour of this index was influenced by the residual liability distribution and disease threshold, and how causal pathways interact when the two risk factors belong to different pathways. The estimator of this index exhibited upward bias, especially in unbalanced case-control designs. We implemented both the jackknife and delta methods to reduce estimation bias, and the jackknife method with Efron's correction to obtain the sampling variance of the estimate. We recommended minimum values for terms in the denominator of the index to ensure to ensure its reliable application. The proposed Interaction Index provides a rigorous and interpretable measure to assess interaction in complex disease epidemiology. We have developed an R package to facilitate the use of the index.

Keywords: statistical interaction, liability-threshold model, case-control study, bias correction, variance estimation

74 Retrospective Modeling of Rare Variants: Impact of Including Unaffected Individuals and Functional Prioritization in a Founder effect Family Cohort

Ariane T. Ange^{1,2*}, Simon Girard^{2,3}, AnneMarie Madore², Catherine Laprise², Alexandre Bureau^{1,3}

¹Département de médecine sociale et préventive, Université Laval, Québec, Canada; ²Centre intersectoriel en santé durable (CISD), Université du Québec à Chicoutimi, Saguenay, Québec, Canada; ³Centre de recherche CERVO, Québec, Canada

Common complex diseases such as asthma arise from genetic and environmental interactions, yet much heritability remains unexplained. Rare variants may account for this gap. This project develops family-based models integrating affected and unaffected relatives to test rare variant associations.

The impact of including unaffected individuals in a retrospective family-based model was assessed through simulation studies comparing an affected-only model with a combined model. Performance, measured using the score statistic, its variance, and relative efficiency, showed that incorporating unaffected relatives increases information and improves precision. Simulations conducted on small families show that adding unaffected individuals improves estimation precision for odds ratios close to 10 when prevalence ranges from 5% to 20%, with marked gains in the power of the score and Wald tests.

In continuity with this work, two methods based on the affected plus unaffected model are being developed: a variant-by-variant analysis and a regional analysis. Before statistical testing, variants most likely to be deleterious were selected. Variants present in the autosomal sequence of 670 Saguenay-LacSaintJean participants with minor allele frequency < 1% in CARTaGENE were retained (8,115,826 SNV and short indels). After annotation and filtering, 6,964,154 variants remained, including 48,455 with CADD ≥ 20 and 126,553 with 15 < CADD < 20, enriched in loss of function and missense categories.

These results demonstrate the value of including unaffected individuals in family based retrospective analyses and provide a prioritized set of rare variants for upcoming studies in the Saguenay-Lac-Saint-Jean founder cohort. Preliminary findings from individual rare variant analyses will also be presented.

75 Multivariate Analysis of Frailty Improves Genomic Discovery and Provides New Insights Into How Frailty Influences Brain Aging

Andrea KM Drilon¹, Ngoc Khanh Huyen Nguyen¹, Jonny P Flint², Andrew Rutenberg³, Kenneth Rockwood^{4,5}, Andrew D Grotzinger^{6,7}, Isabelle F Foote^{1,8*}

¹Centre for Preventive Neurology, Wolfson Institute of Population Health, Queen Mary University of London, London, United

Kingdom; ²Faculty of Health and Life Sciences, University of Exeter, Exeter, United Kingdom; ³Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada; ⁴Nova Scotia Health, Halifax, Nova Scotia, Canada; ⁵Division of Geriatric Medicine, Dalhousie University, Halifax, Nova Scotia, Canada; ⁶Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, Colorado, United States of America; ⁷Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, Colorado, United States of America; ⁸Academic Centre for Healthy Ageing, Whipps Cross Hospital, Barts Health NHS Trust, London, United Kingdom

Frailty is a multifaceted clinical state associated with accelerated aging and adverse health outcomes. It is routinely measured using single aggregate measures such as the Frailty Index, which is a proportional score based on the number of 'deficits' that an individual has. Whilst such measures are clinically useful, the aggregation of highly heterogeneous deficit profiles in genetic studies obscure important insights into the underlying biology of frailty. To address this, we performed the first multivariate genome-wide association study of the shared genetic architecture between 30 frailty deficits using genomic structural equation modelling. This identified 408 genomic risk loci for frailty, vastly improving genomic discovery compared to previous univariate studies. Our latent model identified risk loci for general frailty pathways that represent shared genetics between all 30 measured deficits, as well as risk loci for shared genetic signal between 6 distinct subgroups of deficits, including frailty pathways linked to limited social support, unhealthy lifestyle, multimorbidity, metabolic problems, poorer cognition and disability. We demonstrate the added value of these six factors in refining our knowledge of frailty by using transcriptomic, functional and pathway analyses to explore how different components of frailty impact brain health and lead to an increased risk of neurodegenerative diseases. Using Mendelian randomization, we examine how frailty subdomains differentially affect distinct features of brain aging. Taken together, we show that nuanced multivariate models of frailty are key to understanding its causes and how frailty leads to worse health, which provides promising new avenues for the prevention of age-related diseases.

Keywords: Frailty, multivariate GWAS, accelerated aging, neurodegeneration, genomic structural equation modelling

76 Phenome-wide Atlas of HLA-regulated T-cell Receptor Variation Underlying Immune-mediated Traits and Diseases

Aaron L Meyers^{1,2}, Khalid Mahmood^{2,3,4}, Daniel D Buchanan^{2,4,5}, Mark A Jenkins^{1,2}, James G Dowty^{1,2}

¹Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Victoria, Australia; ²Centre for Genomic Cancer Medicine, Victorian Comprehensive Cancer Centre, Victoria, Australia; ³Melbourne Bioinformatics, University of Melbourne, Victoria, Australia; ⁴Colorectal Oncogenomics Group, Department of Clinical

Background: HLA loci influence more traits than any other genomic region, yet are often excluded from follow-up analyses due to complex diversity and unique immunobiology. HLA alleles are thought to shape T-cell receptor (TCR) features that modulate antigen recognition; indeed, TCR quantitative trait loci ('tcrQTL') mapping has found strong associations with autoimmune HLA alleles. Here, we present ImmunoXcan, a method to identify TCR features mediating trait associations by integrating tcrQTL mapping with HLA association statistics.

Method: Using HLA genotypes, ImmunoXcan fits multi-variate penalised models to predict TCR features, including TRBV gene usage and CDR3 composition, which determines antigen recognition. ImmunoXcan then extends transcriptome-wide association study (TWAS) and Bayesian fine-mapping to model TCR-mediated and direct HLA effects on a trait, adjusting for linkage disequilibrium and horizontal pleiotropy.

Results: Using blood TCR sequencing (N ~ 1,500), we trained ~3,000 TCR feature prediction models (mean 62% heritability captured). Across diverse simulated trait architectures, ImmunoXcan achieved calibrated false discovery and competitive power relative to leading TWAS methods. In positive-control analyses, ~80% of identified credible sets for Epstein-Barr virus (EBV)-specific antibody responses in UK Biobank contained TCRs validated to recognise cognate EBV antigens. Further, phenome-wide analysis in FinnGen yielded at least one credible set for 314 traits (mean 4 TCR features per set), with enrichment in autoimmune diseases. Identified TCR-trait pairs reflected canonical aetiology (e.g., EBV-specific TCRs and lymphoma) and therapeutically-tractable signals (e.g., TRBV9 usage and autoimmune diseases).

Conclusion: ImmunoXcan is a powerful framework for resolving TCR drivers of complex traits, with implications for immunotherapy development.

Topic areas: 1st topic area: Causal inference (incl. Mendelian randomization); 2nd topic area: Immunogenetics and infectious diseases

Keywords: HLA, T-cell receptor (TCR), Transcriptome-wide association study (TWAS), Molecular quantitative trait loci (mQTL), Phenome-wide association study (PheWAS)

77 Towards Robust Epigenetic Age Estimation From Incomplete DNA-methylation Data

Christelle N. Kemda^{1*}, André Scherag¹, Lutz Leistritz¹

¹Institute of Medical Statistics, Computer and Data Sciences, Jena University Hospital, Jena, Germany

Introduction: An accurate estimation of DNA-methylation based biological age is highly sensitive to missing data and the choice of imputation strategy [1]. We evaluated four imputation approaches for epigenetic age prediction using Steve Horvath's epigenetic clock in 3,402 individuals profiled with the Illumina HumanMethylation450K BeadChip array [2, 3]. The assessed methods included multiple-sample imputation approaches methyLImpute and impute. knn, as well as a single-sample imputation method OSMI with an its advanced variant, which prioritizes imputation within CpG islands (advancedOSMI) [4]. Due to computational resource constraints, methyLImpute was limited to 353 Horvath's clock CpGs. For comparability, impute. knn was also restricted to these CpGs (impute. knn353), and additionally applied to the entire 450K dataset (impute. knn450K).

Result: Accuracy of age estimation was assessed by comparing estimated DNA methylation age with chronological age using median absolute error (MedAE in years). Impute. knn450K achieved the lowest MedAE (4.6 years), followed by impute. knn353 (5.0 years). OSMI and advancedOSMI showed similar performance (≈ 5.4 years), whereas methyLImpute performed worst (6.1 years). The Friedman test indicated significant differences in MedAE across methods ($p < 0.001$). Post-hoc Wilcoxon tests revealed that methyLImpute performed significantly worse than all other approaches ($p < 0.001$), while impute. knn450K outperformed all other methods ($p < 0.001$). There was no evidence that OSMI, advancedOSMI, and impute. knn353 differed in MedAE.

Although OSMI did not achieve the highest accuracy method in terms of MedAE, its significance may lie in its applicability to personalized medicine, where reliable imputation at individual-sample level is particularly valuable [5].

78 Pleiotropic Effects of Asthma Genetic Risk Scores Partitioned Using Cell Specific Epigenomic Profiles

Matthew J. Sward^{1*}, Lianne Trap^{2,3}, Ralph Stadhouders², Liam G. Heaney⁴, Ian Sayers⁵, Katherine A. Fawcett¹

¹Division of Public Health and Epidemiology, School of Medical Sciences, University of Leicester, Leicester, United Kingdom;

²Department of Pulmonary Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands;

³Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands;

⁴Wellcome-Wolfson for Experimental Medicine, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, United Kingdom; ⁵Centre for Respiratory Research, National Institute for Health Research Nottingham Biomedical Research Centre, School of Medicine, Biodiscovery Institute, University of Nottingham, Nottingham, United Kingdom

Asthma is a heterogeneous condition varying by age-at-onset, severity, and response to treatments. Key drivers of heterogeneity are thought to be underlying inflammatory pathways

(including type 2, non-type 2 and mixed/dynamic endotypes), in which different cell types dominate. We investigated whether known genetic risk factors for asthma, partitioned by their epigenomic profiles in asthma-relevant cell types, are associated with distinct clinical traits and comorbidities.

Based on overlap with predicted active or poised gene regulatory elements obtained from epigenomics analysis of 19 asthma-relevant cell types, five partitions of asthma risk variants were previously reported by Stikker and colleagues. Using Deep-PheWAS, we tested partition-specific genetic risk scores (GRS) for association with 1,899 traits in UK Biobank European participants with asthma ($n = 57,113$).

T helper cell, myeloid and ILC2-mast cell partitions showed strongest association with eosinophil count, a marker of type 2 inflammation ($P = 4.74 \times 10^{-94}$, 1.9×10^{-45} and 2.25×10^{-38} respectively). However, the broader lymphocyte partition showed strong association with increased neutrophils, immune cells that are linked to poor treatment response ($P = 1.89 \times 10^{-6}$) but no association with eosinophils. The epithelial cell partition showed association with hypertension ($P = 2.18 \times 10^{-5}$) and digestive traits, such as gastritis ($P = 9.11 \times 10^{-6}$). All partitions had distinct patterns of association with non-respiratory traits and comorbidities.

GRS partitioned by epigenomic profiles across asthma-relevant cell types show distinct patterns of association with key immunological parameters and comorbidities. Further exploration of the mechanisms of underlying variants and implicated genes may provide biological insights into asthma endotypes and inform precision medicine.

79 Feature Stability in Multi-omics Integration Models: A Comparative Study of MOFA and DIABLO

Ibrahim Moussa¹, François Ged², Vincent Calvez², Ozvan Bocher¹

¹Univ Brest, Inserm, EFS, UMR 1078, GGB, F-29200 Brest, France; ²Univ Brest, CNRS, UMR 6205, LMBA

Molecular or “omics” profiles, now characterized for hundreds of individuals, provide opportunities to decipher the biological mechanisms underlying complex human traits. Beyond the separate analysis of individual omics layers, such as transcriptomics or metabolomics, integrating multiple omics datasets is needed to understand biological interactions and obtain a comprehensive view of how they shape disease risk. To address this challenge, numerous multi-omics integration methods have emerged, many of which rely on dimensionality reduction techniques to identify latent molecular structures and features associated with a phenotype. However, evaluating the performance and stability of these models for out-of-sample prediction remains challenging, partly due to the lack of realistic synthetic datasets.

In this study, we focus on two widely used multi-omics integration frameworks, MOFA and DIABLO, and investigate their assumptions, differences, and stability. Preliminary analyses on real datasets suggest that, for omics layers whose variance can be explained by a limited number of latent factors, the selected features remain relatively stable across subsamples of the original data. In contrast, for high-dimensional and highly variable omics layers such as transcriptomics, feature stability decreases substantially with reduced similarity between training and testing subsets. Similar patterns are observed for DIABLO, which still maintains better stability likely due to its supervised implementation. Synthetic datasets are currently being developed to further characterize the parameters influencing model stability. Overall, this work provides insights into the behaviour of current multi-omics integration approaches and lays the foundation for ongoing efforts aimed at improving their robustness through the incorporation of prior genetic information.

80 Longitudinal GWAS Reveals the Genetic Architecture of Kidney Function Across the Life Course in the Population and Disease

I.M. Heid¹, M. Gorski¹, S. Mozaffari², H. de Hessel¹, B. Fröhlich^{1,3}, M. Behr³, A. Justice⁴, A. Khan⁵, K. Stark¹, A. Köttgen⁶, C. Pattaro⁷, A. Teumer^{8,9}, T. Winkler¹, K. Estrada²

¹Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany; ²Maze Therapeutics, South San Francisco, California, United States of America; ³Faculty of Informatics and Data Science, University of Regensburg, Regensburg, Germany; ⁴Department of Population Health Sciences, Geisinger, Danville, Pennsylvania, United States of America; ⁵Department of Medicine, Division of Nephrology, Vagelos College of Physicians and Surgeons, Columbia University, New York, United States of America; ⁶Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany; ⁷Eurac Research, Institute for Biomedicine (affiliated with the University of Lübeck), Bolzano, Italy; ⁸Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany; ⁹DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany

Genome-wide association studies using longitudinal data (longGWAS) can help answer important questions about age-related disease and disease progression. One example is chronic kidney disease (CKD), characterized by reduced kidney function. While kidney function declines naturally with age, accelerated kidney function decline implies high risk of kidney failure necessitating dialysis or kidney transplantation. Only few genetic loci for kidney function decline have been identified so far.

We conducted longGWAS in 38 cohort studies with longitudinal follow (1,295,542 individuals, 10,402,884 assessments of serum creatinine based estimated glomerular filtration rate, eGFR; age 18 to 105 years; follow-up time 4 to 68 years) using linear mixed models. Meta-analyses identified 58 independent

genetic variants with significant age-/time-interaction association with eGFR, which reflect genetics of eGFR decline (44 novel signals). We observed different types of life course of genetic effects, accelerating or attenuating eGFR-decline or protective in young and adverse in old, with differential biological and clinical implications. Our results also underscore that genetics of eGFR decline in the general population overlaps with genetics of eGFR decline in persons with CKD or diabetes, rendering identified genes as compelling drug targets to help decelerate kidney function decline in these high-risk groups.

Overall, our results reveal genetics of kidney aging and kidney function decline in high-risk groups. We also highlight the implications for downstream analyses like age-informed fine-mapping and polygenic score analyses as part of a framework for GWAS in longitudinal data.

81 New Proteomic and Rare Variant Analyses to Map Genetic Variants Associated With Lung Function to Causal Genes and Pathways

Nick Shrine^{1,2}, Alex Williams^{1,2}, Jing Chen^{1,2}, Kayesha Coley^{1,2}, Samuel T. Moss^{1,2}, Catherine John^{1,2}, Anna L. Guyatt^{1,2}, Louise V. Wain^{1,2}, Ian P. Hall³, Ian Sayers⁶, Emma Rawlins^{4,5} and Martin D. Tobin^{1,2}

¹Department of Population Health Sciences, University of Leicester, Leicester, United Kingdom; ²Division of Respiratory Medicine and NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, United Kingdom; ³National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, University of Leicester, Leicester, United Kingdom; ⁴Wellcome Trust/CRUK Gurdon Institute, University of Cambridge, Cambridge, United Kingdom; ⁵Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom; ⁶Centre for Respiratory Research, National Institute for Health Research Nottingham Biomedical Research Centre, School of Medicine, Biodiscovery Institute, University of Nottingham, Nottingham, United Kingdom

Genome-wide association studies (GWAS) have identified 1,020 genetic variants associated with lung function, implicating 559 genes via variant-to-gene mapping. Translating variant associations to causal genes and biological mechanisms remains challenging. Emerging large-scale proteomic data and rare variant analyses from whole-exome sequencing (WES) provide complementary approaches to link GWAS signals to downstream molecular effects and coding variants that directly perturb gene function.

We updated the variant-to-gene mapping with new evidence from (i) GWAS of plasma protein levels for 2,923 proteins in 46,595 European ancestry UK Biobank participants, using an updated colocalisation framework to assess shared genetic signals between protein abundance and lung function; (ii) reanalysis of expression data using the same colocalisation framework; (iii) rare variant associations (allele frequency < 1%), incorporating meta-analysis of single-variant associations across 13 cohorts (N = 433,869) and gene-based tests in European UK Biobank WES data (N = 356,860). These results

were integrated into an extended variant-to-gene mapping framework also incorporating five previous independent lines of evidence.

The number of genes supported by at least two lines of variant-to-gene evidence increased from 559 to 600. Rare-variant analyses identified 13 novel genes, as well as novel exonic variants in nine previously implicated genes. Pathway analysis highlighted 260 biological pathways at FDR < 5%, and we generated pathway-specific phenome-wide association studies to understand the effect of intervening in these pathways.

By integrating proteomic and rare variant analyses with GWAS, we refined the mapping of lung function loci to causal genes and pathways which could represent potential therapeutic targets for lung health and COPD.

82 Genotype-Phenotype Associations in Parkinson's Disease: A Systematic Review With Identification of Biological Pathways

Jean-François Nepveu^{1*}, Jacopo Cristini², Celia M. T. Greenwood^{1,3,4}, Christina Wolfson^{1,5,6}

¹Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Québec, Canada; ²School of Physical and Occupational Therapy, McGill University, Montreal, Québec, Canada; ³Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Québec, Canada; ⁴Quantitative Life Sciences Program, McGill University, Montreal, Québec, Canada; ⁵Neuroepidemiology Research Unit, Research Institute of the McGill University Health Centre, Montreal, Québec, Canada; ⁶Department of Medicine, McGill University, Montreal, Québec, Canada

Parkinson's disease (PD) is a multifactorial neuro-degenerative disorder with substantial clinical heterogeneity. Genome-wide association studies and candidate gene approaches have established genetic contributions to disease risk, and genetic variation may influence clinical variability. This systematic review synthesizes evidence on genotype-phenotype associations in PD, evaluates consistency of reported findings, and identifies biological pathways underlying clinical heterogeneity.

Following PRISMA guidelines, we searched MEDLINE and Embase for human genetic association studies examining genotype-phenotype relationships in PD and conducted functional enrichment analyses to identify overrepresented biological processes across implicated genes.

Of 5,734 publications identified, 33 met inclusion criteria. Most studies focused on monogenic PD genes, particularly *GBA* and *LRRK2*. *GBA* variants were consistently associated with earlier onset, hyposmia, autonomic dysfunction, and a greater non-motor burden; *LRRK2* variants with later onset, preserved olfaction, and milder non-motor features. Enrichment analyses revealed pathway-level patterns: genes influencing age at onset mapped to neuronal maintenance and degeneration, those linked to olfactory phenotypes to autonomic and gliotic processes, genes related to non-motor symptoms to dopaminergic

and catecholamine signaling, those associated with motor signs to synaptic transmission and oxidative stress, and genes linked to disease severity to cellular stress and protein stability pathways.

Distinct genetic variants are associated with specific clinical manifestations, and enrichment analyses can generate hypotheses regarding molecular pathways contributing to clinical heterogeneity. Integrative genomic and longitudinal studies are needed to improve patient stratification and inform personalized therapeutic strategies.

83 Indirect-Direct Genetic Association Utilizing Partitioned Family-Based SNP Effects in MoBa Reveals Parental Genetic Nurture Traits Putatively Influencing Offspring Birth Weight While Controlling for Assortative Mating and Population Stratification

Annabel Perry^{1,2*}, Laura Hegemann³, Sara Stinson⁴, Ilse Krätschmer⁵, Michael C. Honigberg^{2,6,7}, Ole A. Andreassen⁴, David Reich^{1,2,6,8}, Nick Patterson², Richard Border⁹, Alkes L. Price^{2,10}

¹Department of Human Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America; ²Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, United States of America; ³PsychGen Centre for Genetic Epidemiology and Mental Health, Norwegian Institute of Public Health, Oslo, Norway; ⁴Centre for Precision Psychiatry, Division of Mental Health and Addiction, Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Oslo, Norway; ⁵Institute of Science and Technology Austria, Klosterneuburg, Austria; ⁶Harvard Medical School, Boston, Massachusetts, United States of America; ⁷Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, United States of America; ⁸Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, United States of America; ⁹Ray and Stephanie Lane Computational Biology Department, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America; ¹⁰Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America

Parental genetic nurture occurs when parents' heritable traits impact offspring traits through environmental pathways. We developed indirect-direct genetic association (idGA), a framework leveraging genetics to identify candidate parent traits impacting offspring traits. idGA takes as input per-SNP direct genetic effects on hypothesized parental traits and the association of each SNP on offspring phenotype by way of genetic nurture; inputs were estimated using family-based, parental sex-aware methods. idGA regresses the genetic nurture effects on the direct genetic effects. idGA distinguishes itself from other genetic nurture inference methods by controlling for assortative mating (AM) and identifying the sex of the parent mediating genetic nurture.

Simulations found idGA well calibrated, well powered, and robust to AM and population stratification. We applied idGA to

25,163 parent-offspring trios from the Norwegian Mother, Father, and Child Cohort Study (MoBa), with offspring birth weight from the Medical Birth Registry of Norway. We evaluated 27 parental traits with plausible impact on fetal growth, including anthropometric, cardiovascular, immune, and socioeconomic traits.

The strongest signals were maternal body mass index (BMI; $Z = 3.25$; $BH_{jp} = 0.02$) and paternal cigarettes per day ($Z = -2.96$; $BH_{jp} = 0.04$). Maternal, but not paternal, BMI was associated with offspring birth weight, consistent with prior findings of maternal size influencing fetal growth independently of direct effects and supporting idGA's control for AM. In contrast, paternal, but not maternal, cigarettes per day were associated with lower birth weight, highlighting partner smoking as a potential environmental pathway affecting intrauterine growth.

Keywords Statistical genetics, Gene environment interaction, Teratogens, Reproductive genetics, Women's health

84 Cerebrovascular Age to Map the Trajectory of Brain and Vascular Diseases

Cienna Gin-Naccarato^{1,2*}, Thomas M. Zheng^{2,3}, Anna Zaidi^{2,3}, Arthur Richard^{2,3}, Zhenqian Wang^{1,2}, Sirui Zhou^{1,2,3}

¹Victor Phillip Dahdaleh Institute of Genomic Medicine, McGill University, Montréal, Canada; ²Department of Human Genetics, McGill University, Montréal, Canada; ³Quantitative Life Science Program, McGill University, Montréal, Canada

Cognitive decline is a hallmark of aging, but chronological age alone does not capture the biological and environmental factors underlying healthy and unhealthy aging. Biological clocks estimate biological age from physiological features. The brain age gap (BAG), defined as the difference between chronological and biological age, is a promising marker for disease risk. Neuroimaging-derived BAGs have been linked to cognitive performance and risk of neurodegenerative disease; however, cerebrovascular aging BAGs remain poorly characterized. We therefore developed a cerebrovascular aging clock to determine whether cerebrovascular aging improves prediction of age-related disease risk.

Cerebrovascular imaging data from the UK Biobank ($N = 71,970$) was used to train an aging clock using a LASSO regression model. The training cohort included 3,000 age- and sex-balanced individuals without stroke or neurodevelopmental disorders. LASSO selected 62 of 104 imaging-derived features from T1-, T2-, subcortical volume, susceptibility-weighted, and diffusion MRI modalities. BAGs were used in phenome-wide association studies across 106 stroke and neurological disease codes ($N = 68,970$), including 11,023 prevalent cases and 3,613 incident cases. Restricting analyses to diseases with at least 30 incident cases yielded 23 diseases positively associated with BAG, including nontraumatic intracerebral hemorrhage ($OR = 1.050$; $p = 2.99 \times 10^{-2}$), ischemic stroke ($OR = 1.088$; $p = 5.64 \times 10^{-24}$) and sequelae of cerebrovascular disease ($OR = 1.116$; $p = 2.85 \times 10^{-10}$). Proteomic analysis identified 509 significant proteins associated with BAG.

Our results show that cerebrovascular BAGs are associated with increased risk of stroke and stroke-related cognitive impairments, supporting BAG as an important contributor to neurological disease risk, and earlier detection of individuals at risk of cognitive decline.

85 Proteome-Wide Genetic Analyses Identify ASGR1 as a Potential Mediator of Smoking on Ischaemic Heart Disease

Alexander C. Tinworth¹, Pang Yao¹, Andri Iona¹, Alfred Pozarickij¹, Adam Von Ende¹, Iona Millwood¹, Robin G. Walters¹, Robert Clarke¹, Fiona Bragg^{1,2}, Zhengming Chen¹

¹Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom; ²Health Data Research United Kingdom Oxford, University of Oxford, United Kingdom

Tobacco smoking increases risk of multiple diseases, but the biological mechanisms involved remain incompletely understood. Recent developments in affinity-based assay proteomics have enabled the quantification of thousands of circulating plasma proteins in large prospective cohort studies. However, observational associations are vulnerable to reverse causation and residual confounding. Mendelian randomization (MR) uses genetic variants to mitigate these biases and provides a framework for strengthening causal inference. Using publicly available data from large genome-wide association studies, we applied MR to identify circulating proteins mediating the effects of smoking on disease. In European ancestry populations, genetically-predicted smoking exposures were robustly associated with 93 (total: 2922) proteins, including higher levels of asialoglycoprotein receptor 1 (ASGR1). Higher circulating ASGR1 levels were associated with ischaemic heart disease (IHD) risk, an association supported using Bayesian colocalization (PP.H4 = 0.91), and was found to mediate 6.35% (95%CI 2.61-10.09%, $p = 8.77 \times 10^{-4}$) of smoking's effects on IHD in two-step MR analyses. Multi-variable MR analyses adjusting for body mass index (BMI) and type 2 diabetes (T2D), strengthened the association between smoking and ASGR1, whilst the univariable association between BMI and ASGR1 levels attenuated to the null. In network analyses, genetically-predicted ASGR1 levels were inversely associated with >100 circulating proteins, including growth arrest-specific 6 (GAS6). These associations were driven by rs55714927, the *cis*-protein quantitative trait loci [pQTL] for ASGR1, acting as a shared trans-pQTL. These findings delineate the smoking-induced proteomic profile and identify ASGR1 as a key molecular intermediary affecting a broad network of circulating proteins and linking smoking to increased risk of IHD.

86 Cardiovascular and Pharmacogenomic Responses to Beta-Blockers Under Extreme Temperatures

Élyse Faucher^{1,2,3*}, Yanis Sid⁴, Marc-Olivier Pilon^{1,2,3}, Marie-Christyne Cyr^{1,2}, Louis-Philippe Lemieux Perreault^{1,2}, Amina Barhdaji^{1,2}, Jean-Claude Tardif^{1,4}, Daniel Gagnon^{1,5}, Simon de Denus^{1,2,3}, Marie-Pierre Dubé^{1,2,4},

¹Montreal Heart Institute, Montreal, Canada; ²Beaulieu-Saucier Pharmacogenomics Centre, Montreal, Canada; ³Faculty of Pharmacy, Université de Montréal, Montreal, Canada; ⁴Department of Medicine, Université de Montréal, Montreal, Canada; ⁵EPIC Center, Montreal, Canada

Climate change and intensification of extreme heat events pose growing risks to the health of older adults. Many aged ≥ 65 years receive beta-blockers for chronic cardiovascular diseases, and these medications may alter physiological responses to heat, including heart rate regulation and peripheral hemodynamics. This study evaluates whether extreme heat modifies cardiovascular and pharmacogenomic responses among beta-blocker users.

The study included 13,958 participants from the Montreal Heart Institute Biobank with a median follow-up of 12 years and a mean of 4 measurements per patient. This resulted in 61,928 observations of blood pressure and heart rate, of which 49% were recorded during beta-blocker use, and 21% within 48 h following outdoor temperatures $\geq 27.5^\circ\text{C}$. Linear mixed-effects models with patient-level random intercept were used and adjusted for age, sex, comorbidities, deprivation indices and care location to assess the heat $\times$ beta-blocker interaction. Outcomes included blood pressure and heart rate. Analyses of ECG parameters, electrolyte markers, and genetic analyses of CYP2D6 involved in metoprolol metabolism are ongoing.

Results showed that during extreme temperatures, systolic blood pressure decreased by 1.7 mmHg among beta-blocker users ($p < 0.001$) and by 3.0 mmHg among non-users ($p < 0.001$) with significant heat $\times$ beta-blocker interaction ($P_{\text{int}} = 0.002$). Diastolic blood pressure increased by 0.6 mmHg among beta-blocker users ($p = 0.002$), whereas no significant variation was observed among non-users ($P_{\text{int}} = 0.02$). Heart rate increased by 0.3 beats per minute among beta-blocker users ($p = 0.03$), while no significant variation among non-users or interaction was observed.

These findings suggest that patients using beta-blockers have different hemodynamic responses during hot days compared to non-users.

87 Links Between Prediagnostic Exposures, Germline Genetics, and Somatic Mutational Profiles in Triple-Negative Breast Cancer

Deborah A. Tadesse¹, Clara Bodelon², Cheng Peng^{3,4}, Kristen D. Brantley^{4,5}, Margaux Delporte⁶, Yujing J. Heng^{4,7}, Lauren R. Teras², Rulla M. Tamimi⁶, Peter Kraft^{1*}

¹Trans-Divisional Research Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, United States of America; ²Department of Population Science, American Cancer Society, Atlanta, Georgia, United States of America; ³Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States of America; ⁴Harvard Medical School, Boston, Massachusetts, United States of America; ⁵Dana-Farber Cancer Institute, Boston, Massachusetts, United States of America; ⁶Department of Population Health Sciences, Weill Cornell Medicine, New York, New York, United States of America;

Triple negative breast cancer (TNBC) is an aggressive subtype of breast cancer with poor prognosis and limited treatment options. Reproductive, behavioral, and genetic risk factors for TNBC differ from other types of breast cancer, and TNBC exhibits distinct evolutionary pathways reflected in heterogeneous mutational profiles. To understand the relationships between prediagnostic exposures, germline genetics, and mutational profiles in TNBC, the PRedagnostic Exposures, Mutations, Immune SignaturEs-Triple Negative (PREMISE-TN) project performed whole exome sequencing of matched formalin-fixed paraffin embedded tumor tissue and germline DNA samples from 322 TNBC patients from four prospective cohort studies, the Nurses' Health Studies (NHS, NHS-II) and Cancer Prevention Studies (CPS-II, CPS3). Genes in several cancer pathways were frequently mutated, including DNA damage response and apoptosis (*TP53*), homologous recombination repair (*BRCA1*, *RBI*), PI3K/AKT/mTOR signaling (*PIK3CA*, *PTEN*), Notch signaling (*NOTCH1*) and extracellular matrix integrity (*USH2A*, *DMD*, *HMCN1*). Non-negative matrix factorization estimated sample-specific activity for five single base substitution (SBS) signatures, including clock-like signatures (SBS1, SBS5), APOBEC activity (SBS2, SBS13), and homologous repair deficiency (HRD; SBS3). After adjusting for age at diagnosis, tumor stage was positively associated with clock-like signatures; menopausal hormone therapy was positively associated with the HRD signature and negatively associated with APOBEC signatures. Ongoing work is assessing associations between tumor somatic mutations and inherited germline variation, including rare pathogenic variants in cancer predisposition genes (e.g. *BRCA1/2*) and common-variant polygenic scores. These analyses help elucidate the interplay between inherited genetics, exposures, and tumor evolution in TNBC.

88 Leveraging a Multivariate Genomic Framework to Identify the Shared Genetic Architecture Between Osteoarthritis and Alzheimer's Disease

Abhishek Choudhary^{1*}, Jared Nakahara¹, Paul C. Okoro¹, Michelle S. Yau^{1,2}

¹*Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, Massachusetts, United States of America;* ²*Department of Medicine, Harvard Medical School, Boston, Massachusetts, United States of America*

Osteoarthritis (OA) and Alzheimer's disease (AD) are both age-associated diseases that present significant global healthcare challenges. Although emerging epidemiological evidence suggests possible links to cardiometabolic syndrome, the shared genetic etiology between OA and AD remains poorly characterized. We used the Genomic SEM approach to generate a multivariate genomic framework to identify the latent genetic architecture across OA, AD, and cardiometabolic phenotypes, including BMI, HDL, and triglycerides. We then conducted a multivariate GWAS to identify variants associated with this latent factor. We removed highly heterogeneous variants across traits, then conducted conditional analysis to identify

independent loci. We integrated eQTLs across all tissues in GTEx to perform pathway enrichment analyses of GO and KEGG networks. The structural equation model revealed highly significant factor loadings across all five traits. The latent factor was characterized by strong positive loadings for BMI ($\lambda = 0.85$, $P = 3.2 \times 10^{-117}$), triglycerides ($\lambda = 0.45$, $P = 5.1 \times 10^{-75}$), and OA ($\lambda = 0.39$, $P = 1.1 \times 10^{-103}$), paired with inverse loadings for HDL ($\lambda = -0.51$, $P = 9.99 \times 10^{-115}$) and AD ($\lambda = -0.17$, $P = 4.8 \times 10^{-8}$). Our multivariate GWAS identified 477 independent genome-wide significant variants, many of which were novel and not previously associated with any one trait. Downstream functional mapping identified high-confidence risk genes that were significantly enriched for biological networks including central nervous system neuron differentiation, regulation of synapse assembly/organization, and social behavior. Further work will consider alternative two factor models that partition the genomic space into distinct *metabolic* versus *aging/inflammation* components. This will ultimately serve as a roadmap for discovering novel therapeutic targets for multiple age-related traits.

89 Genetic Contributions to Early Life Growth Progression

Fasil Tekola-Ayele¹, Guisong Wang², Kunal Kathuria¹

¹*Epidemiology Branch, Division of Population Health Research, Division of Intramural Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, United States of America;* ²*The Prospective Group (TPG), contractor for Division of Population Health Research, Division of Intramural Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, United States of America*

Early life growth is foundational to lifespan health and is associated with future disease risk. While some studies have examined the genetic contributions to mean differences in growth measures at a single time-point, genetic factors associated with longitudinal change in growth trajectories are sparse. Identifying genetic factors that regulate early life growth changes, including genes with developmental period-specific effects, can offer a window of opportunity for early diagnosis and development of targeted therapeutics. Here we performed a multi-ancestral GWAS of longitudinal changes in early growth and development. First, we evaluated different models to determine fit on the mean and intra-individual variability of the early growth measures assessed at multiple time points using models with knots with random effects and correlation structure. The optimal models were selected based on AIC and fit into a genome-wide model with adjustment for fetal sex, gestational age and genotype principal components. Genome-wide significance was determined if $p < 5 \times 10^{-8}$. Polygenic scores were computed based on previously known GWAS associations with single-time growth measures. The association of the polygenic scores with the growth measures' mean shift and within-individual variability were tested to evaluate transferability of genetic signals of cross-sectional growth and longitudinal change in growth. We found genome-wide significantly associated loci at genes important for linear growth, growth

hormone binding, and central nervous system regulation of body weight. The findings suggest that the genetics of growth progression can provide insights that may not be obtained from growth measured at a single snap point.

Keywords: early life growth, childhood, longitudinal (repeated) measures, GWAS

91 Multi-Ancestry Fine-Mapping in Two Million Participants Improves Breast-Cancer Credible Set Resolution

Peter Kraft^{1*} (on behalf of the Confluence Project), The Confluence Project - National Cancer Institute¹, African-ancestry Breast Cancer Genetic Consortium², Breast Cancer Association Consortium³, Consortium of Investigators of Modifiers of BRCA1/2³, Latin America Genomics of Breast Cancer Consortium⁴, Male Breast Cancer Genetics Consortium⁵

* Presenting author

¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, United States of America; ²Department of Medicine, Division of Epidemiology, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America; ³Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; ⁴University of California Davis Comprehensive Cancer Center, University of California, Davis, Davis, California, United States of America; ⁵The Johnston Cancer Research Centre, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, United Kingdom

Background: Genome-wide association studies (GWAS) have identified hundreds of breast cancer susceptibility loci, but causal variants at most loci remain unresolved. The Confluence Project, the largest multi-ancestry breast cancer GWAS to date, includes >420,000 breast cancer cases and 1.6 million controls, with greater ancestral diversity than prior GWAS, including African (AFR, $n = 24,000$), East Asian (EAS, $n = 49,000$), and Admixed American (AMR, $n = 28,500$) cases. We report preliminary multi-ancestry (MA) fine-mapping results refining candidate causal variants (CCVs) at 545 overall breast cancer risk loci.

Methods: We prioritized candidate MA loci showing nominal association evidence ($P < 1 \times 10^{-5}$) within ± 1 Mb of the sentinel variant in at least two ancestries. We applied four multi-ancestry fine-mapping (MAFM) methods (MESuSiE, MultiSuSiE, SuSHiE, and SuSiEx) and compared their credible sets (CS) with the union of ancestry-specific CS and EUR-only fine-mapping.

Results: Among 545 evaluated loci, 313 met the MA criterion; 118 (37.7%) were novel Confluence loci. MAFM reduced CS size by an average of 52.6% relative to the ancestry-specific CS union and by 46.2% relative to EUR-only fine-mapping. Among novel loci, CS size was reduced by an average of 36.9% and 49.2% versus these baselines, respectively. For loci with a nominal AFR signal, reduction averaged 55.9% across methods and 33.0% among novel loci.

Conclusion: MAFM can substantially reduce CS size and improve resolution of breast cancer risk variants. Ongoing analyses are applying these methods to breast cancer subtypes, including individual-level data using multinomial regression, and incorporating functional annotations, enabling a richer catalog of breast cancer CCVs and regulatory mechanisms.

92 Can I Borrow Your Instruments? Robust and Efficient Mendelian Randomization by Sharing Across Outcome Traits

Anagh Chattopadhyay¹ and Nilanjan Chatterjee^{1,2*}

¹Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America; ²Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, Maryland, United States of America

Background: Mendelian randomization (MR) uses genetic variants as instruments to estimate causal effects of exposures on disease, but horizontal pleiotropy biases even robust estimators, which lose power as invalid instruments accumulate—limiting reliable causal inference.

Methods: We show that pleiotropy across outcome traits, usually a nuisance, can be exploited: related outcomes share confounders and pathways, and therefore valid/invalid instruments, for a given exposure. We introduce co-heterogeneity, an exposure-anchored summary statistic—with consistency and asymptotic theory as instruments grow—that ranks auxiliary outcomes by this pathway-based similarity. Co-heterogeneity anchors the Instrument Borrowing (IB) framework, which borrows strength from an auxiliary outcome. We instantiate IB in two methods: IB-Mode (extending MR-Mode via a joint bivariate mode) and IB-PRESSO (extending MR-PRESSO via joint outlier detection).

Results: In simulations, IB methods improved power while preserving type-I error, with gains growing in the degree of shared invalid instruments. Applied to Million Veteran Program GWAS—selecting auxiliary outcomes from 40 cardiometabolic traits—across 48 established exposure-outcome pairs, IB-Mode was the only method of ten to correctly classify both HDL→CAD and HDL→stroke as non-causal, matching randomized-trial evidence where even robust methods failed. For confirmed risk factors, IB-Mode delivered efficiency gains up to ~200% (e.g., BMI). Re-examining the contested vitamin D hypothesis, IB-Mode found no causal effect on CAD or type 2 diabetes, consistent with null trials, whereas many methods declared significance.

Conclusion: Instrument Borrowing makes MR estimation more robust and efficient, while co-heterogeneity maps the shared pathways linking an exposure to related outcomes, with potential for systematic drug-repurposing prioritization.

Keywords: Mendelian randomization, horizontal pleiotropy, co-heterogeneity, instrument borrowing, causal inference

93 TL-TWAS: Transfer Learning for TWAS Using Large-Scale Transcriptomic Resources in Underpowered Target Context

Dongsoo Lee^{1*}, Zhaotong Lin³, Chang Su^{1,2}

¹Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, United States of America; ²Department of Human Genetics, Emory University, Atlanta, Georgia, United States of America; ³Department of Statistics, Florida State University, Tallahassee, Florida, United States of America

Background: Accumulating transcriptomic resources across technologies and populations enable powerful transcriptome-wide association studies (TWAS). However, existing methods poorly transfer regulatory information across heterogeneous contexts, especially when small-sample targets could benefit from larger sources. For example, cell-type-specific TWAS (CTS-TWAS) requires single-cell resolution yet could leverage larger bulk expression quantitative trait locus (eQTL) data. Similarly, TWAS in underrepresented populations could benefit from large European-ancestry eQTL studies.

Method: We present TL-TWAS, a transfer-learning framework that improves TWAS in small-sample targets by transferring information from larger sources. It incorporates source eQTL information via SNP-level penalty factors, penalizing SNPs with stronger source eQTL evidence less. This directs variable selection toward SNPs with consistent source regulatory signals. A subsequent debiasing step uses only target data to correct source-target mismatch, allowing TL-TWAS to borrow source strength and preserve target-specific signals. TL-TWAS avoids relying on a matched linkage disequilibrium reference and applies to bulk-to-single-cell, cross-tissue, and cross-ancestry TWAS.

Results: We applied TL-TWAS to CTS-TWAS using ROSMAP single-cell data ($n = 424$) as the target and MetaBrain cortex eQTL ($n = 2,683$) as the source. Across five cell types, TL-TWAS increased predictable genes by 27-34% over single-cell-only models. In Stage 2, it identified 24 additional Alzheimer's disease (AD)-associated genes, with the largest gains in astrocytes (45%) and microglia (15%). In microglia, TL-TWAS prioritized AD-relevant genes *FCER1G* and *RIN3* and showed enrichment for AD risk loci ($OR = 2.77$) with immune signatures.

Conclusion: TL-TWAS leverages large-scale source resources to improve TWAS in small-sample targets.

94 Investigating Allele Count Cutoffs for Safe Sharing of Sequencing-Based Genome-Wide Association Study (GWAS) Summary Statistics

Hongyu Xiao^{1,2*}, Daniel Taliun^{1,2}

¹Department of Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montreal, Québec, Canada; ²Victor Phillip Dahdaleh Institute of Genomic Medicine at McGill University, Montreal, Québec, Canada

Data sharing is central to genome-wide association studies (GWAS), enabling replication, variant interpretation, and aggregation of larger sample sizes for novel findings. As GWAS increasingly rely on high-depth genome sequencing, which allows assessment of rare variants, sharing summary-level results including rare variants raises concerns about potential participant re-identification. Overly restrictive policies may limit opportunities to study rare variants, whereas more permissive approaches may risk participant privacy. Although this challenge is widely recognized, there is currently no clear, data-driven guidance on allele count (AC) thresholds for safe sharing, particularly in sequencing-based GWAS.

We present the GWAS Sharing Assessment Framework (GWAS-SAFE), a machine-learning approach that quantifies re-identification risk across AC thresholds. GWAS-SAFE integrates features such as variant sharing patterns and genetic ancestry to predict relatedness using both observed and simulated relationships to estimate the probability of re-identifying participants or their relatives under different sharing scenarios. We evaluated GWAS-SAFE using high-depth whole-genome sequencing data from the 1000 Genomes Project ($N = 3,202$). At singleton resolution ($AC = 1$), an individual or their relatives up to third degree could be identified in summary-level statistics with accuracy and recall approaching 1.0. This risk decreased with increasing AC; the model's performance in re-identifying an individual or related samples approached random-chance levels only when AC exceeded 50.

These results suggest that widely used AC thresholds may not adequately protect participant privacy in sequencing-based GWAS. Instead, GWAS-SAFE provides a practical, data-driven tool to guide sharing decisions tailored to specific datasets, rather than relying on fixed, potentially insufficient thresholds.

Keywords: Ethical, legal, and social implications; statistical genetics; genome-wide association study (GWAS); exome/genome sequencing; computational tools

95 Identification of Novel Genetic Associations With COVID-19 Severity Through Analysis of Mobile Element Variants

Jasmin J. Lee^{1,2*}, Shohei Kojima³, Bessie Wang⁴, Erika Garg⁵, Rohan Abraham⁶, Steven J.M. Jones^{1,7,8}, Maria Tokuyama⁴, Jessica Dennis^{1,2}

¹Department of Medical Genetics, The University of British Columbia, Vancouver, British Columbia, Canada; ²BC Children's Hospital Research Institute, Vancouver, British Columbia, Canada; ³Human Biology Microbiome Quantum Research Center, Keio University, Tokyo, Japan; ⁴Department of Microbiology and Immunology, The University of British Columbia, Vancouver, British Columbia, Canada; ⁵The Hospital for Sick Children, Toronto, Ontario, Canada; ⁶BC Cancer Research Institute, Vancouver, British Columbia, Canada; ⁷Genome Sciences Centre, Vancouver, British Columbia, Canada; ⁸Provincial Health Services Authority, Vancouver, British Columbia, Canada

Mobile genetic elements (ME) are heritable DNA sequences that insert into new genomic locations over evolutionary time, generating structural variation across human genomes. Composing ~44% of the human genome, their contribution to health and disease remains poorly understood. Recent advances in whole genome sequencing and bioinformatics have enabled the identification of ME-derived variants (MEV) at the population level, fueling association studies with complex traits in large biobanks.

We leveraged whole genome sequencing and SARS-CoV-2 outcome data from the Canadian HostSeq biobank (n = 8,735) to characterize MEV distribution across ancestry groups and test associations with COVID-19 severity within a genome-wide association study framework. We detected 80,494 MEVs, of which 6,909 had allele frequency (AF) > 1%. Given differences in MEV presence and frequency across ancestries, we performed pooled and ancestry-stratified analyses. Four MEVs passed the Bonferroni-corrected significance threshold in the pooled analysis, including an L1HS insertion on chromosome 2 previously identified in the 1000 Genomes Project and BioBank Japan. Loss of this insertion was associated with increased risk of severe COVID-19 (AF_{case}: 0.120, AF_{control}: 0.053, OR = 1.49, $P = 2.80 \times 10^{-7}$), with the strongest association in European ancestry samples (71.3% of samples; OR = 1.92; $P = 1.06 \times 10^{-8}$). The L1HS MEV was not in high linkage disequilibrium with any nearby SNPs, suggesting an independent signal.

Our findings highlight the value of incorporating ME-derived structural variants in disease association studies, with implications for biomarker discovery and understanding of COVID-19 outcomes.

96 Genetic Architecture of Pathway-Specific Metabolomic Aging Clocks in the Canadian Longitudinal Study on Aging

Thomas M. Zheng^{1,2*}, Alejandro Mejía-García³, Cienna Narratto³, Anna Zaidi¹, Arthur Richard¹, Dandan Tan^{1,2}, Jesse Islam², Kevin Y.H. Liang², Sirui Zhou^{1,3}

¹Quantitative Life Sciences, McGill University, Montreal, Québec, Canada;

²Lady Davis Institute, Montreal, Québec, Canada;

³Human Genetics, McGill University, Montreal, Québec, Canada

Background: Aging is associated with a loss of function and susceptibility to many diseases. However, chronological age is an imperfect metric for determining risk. Metabolomic clocks enable us to investigate how our metabolism impacts aging. We developed and tested metabolomic aging clocks and found putative causal genetic variants in a Canadian population-based biobank (CLSA).

Methods: We developed eight pathway-specific age clocks using the CLSA, leveraging 735 metabolites in 9,881 individuals. We trained LASSO models in a CLSA subset (n = 1,000) to predict chronological age. We tested these predicted metabolic ages against cross-sectional and longitudinal diseases, and age-related phenotypes such as clinical frailty (n = 8,881). We performed European ancestry-specific GWAS, followed by fine-mapping, pQTL analysis, and constructed an aging PRS.

Results: We found that our clocks outperformed chronological age in predicting phenotypes such as clinical frailty and that metabolomic age was significantly associated with increased risk of heart disease (HR = 1.084, P value) and type 2 diabetes (HR = 1.045, P value = 3.4). Additionally, we identified 9 genome-wide significant loci corresponding to both missense and intronic SNVs. We found three missense variants located in *TYMP*, *AGXT2*, and *ACADS*, which reduce the function/expression of their coded proteins.

Conclusion: Metabolomic aging clocks can be used in disease monitoring and prediction. We have also shown that these clocks better capture disease phenotypes than chronological age alone. Finally, we have identified specific variants that allow us to find causal links between our biological age and potential drug targets.

97 Shared Genetic Architecture and Novel Pleiotropic Variants Across Solid Cancers

Stacey J. Winham¹, Xunxuan Chen, Gamaliel T. Taengwa¹, Brandon J. Coombes¹

¹Department of Quantitative Health Sciences, Mayo Clinic, Rochester, Minnesota, United States of America; ²Department of Cancer Biology, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with individual cancer risk, and polygenic risk scores (PRS) have been derived from these cancer-specific variants. While prior pan-cancer studies have shown shared genetic architectures across cancers, latent factors underlying cancer risk have been unexplored. Identifying these shared factors may improve cancer risk prediction by leveraging shared variants across cancer types. We conducted an integrative pan-cancer GWAS of seven solid tumors (breast, ovarian, endometrial, pancreatic, lung, colorectal, melanoma) to identify shared latent genetic factors and novel pleiotropic variants.

Using large-scale GWAS summary statistics from 400,000 cases and 1.5 million controls of European ancestry, we used Genomic Structural Equation Modeling to construct three latent factors: Common Cancer Factor, Female-Specific Cancer Factor and Non-Sex Specific Cancer Factor. Although the Common Cancer Factor explained some of the shared risk, a correlated two-factor model distinguishing female-specific from other cancers showed stronger fit. Factor-specific GWAS identified 233 genome-wide significant loci ($p < 5 \times 10^{-8}$) with 24 distinct loci showing shared susceptibility across cancers, with results driven by the largest GWAS. Pathway enrichment analysis highlighted common oncological pathways, including Keratinization, TGF- β and KLF-related cancers. We also identified eight novel pleiotropic variants not previously associated with cancer risk, many in intronic and intergenic regions.

Our findings reveal a shared genetic architecture across solid tumors and uncover novel pleiotropic variants with potential roles in oncogenesis which need to be further investigated. This

study provides insight into the genetic basis of cancer susceptibility, supporting future development of multi-cancer PRS.

98 Age-Informed Finemapping Improves Prioritization of Likely Causal Variants for Kidney Function Decline

Benedikt Fröhlich^{1,2*}, Mathias Gorski¹, Thomas W. Winkler¹, Hannah C. de Hesselde¹, Iris M. Heid¹, Merle Behr²

¹Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany; ²Faculty for Informatics and Data Science, University of Regensburg, Regensburg, Germany

Finemapping prioritizes variants in genome-wide association studies (GWAS) by selecting credible sets of likely causal variants in the associated locus. Existing Bayesian finemapping approaches assign a posterior inclusion probability of causality (PIP) to each variant within a given locus and allow for identifying multiple causal variants per region. However, they do not account for interactions with age. When opposing effects across the lifespan attenuate or mask effects captured by marginal association effects, conventional finemapping may limit informative prioritization of likely causal variants.

We developed an interaction-aware finemapping approach that incorporates age-dependent genetic effects. For each locus, we identified the most informative age for the primary index variant (“best age”) and derived age-specific association Z scores for all variants in the region. We then performed finemapping using SuSiE based on these best-age Z scores.

We applied our age-informed finemapping to 49 associated loci from a recent longitudinal GWAS meta-analysis on kidney function decline (more than one million individuals, more than 10 million estimated glomerular filtration rate measurements), focusing on European-ancestry individuals.

Our approach identified 721 variants in 95% credible sets (38 with PIP of 50% or greater) compared with 972 variants from conventional finemapping (i.e., using marginal effect estimates, 26 with PIP of 50% or greater), indicating improved resolution. Top-prioritized variants from best-age finemapping showed stronger enrichment for interactions with age (interaction P value $< 1 \times 10^{-4}$ in 75% of credible sets, compared with 57.5%). These findings suggest that interaction-aware finemapping improves prioritization of likely causal variants for kidney function decline.

99 Multi-Omic Integration Reveals Rare Variant Effects on Blood DNA Methylation and Gene Expression in Alzheimer's Disease Loci

Jeewoen Shin^{1*}, Han Chen², Josée Dupuis³, Chloé Sarnowski¹

¹Department of Epidemiology, Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, United States of America; ²Rory Meyers College of Nursing, New York University, New York, New York,

United States of America; ³Department of Epidemiology, Biostatistics and Occupational Health, School of Population and Global Health, McGill University, Montreal, Québec, Canada

GWAS have identified many loci associated with Alzheimer's disease (AD), but identifying causal variants and estimating effect sizes for low allele frequency variants (minor allele frequency, MAF $< 5\%$) remains a challenge. Omics data have the potential to help elucidate the biological mechanisms underlying rare variant associations. We applied JEM [PMID: 33038041] to jointly evaluate the effects of multiple genetic variants ($0.005 \leq \text{MAF} \leq 0.02$) on blood DNA methylation (N = 2,489) and gene expression (N = 5,082) data from the Framingham Heart Study.

We first applied MiST [PMID: 23483651] to test the joint effects of rare variants on gene expression (eQTLs) and DNA methylation (meQTLs) within established AD loci [PMID: 35379992]. We then applied JEM to the significant eQTLs and meQTLs to estimate individual rare variant effect sizes.

We identified significant joint effects of rare variants for 20 transcripts and 28 CpG sites (Bonferroni-adjusted P-value $< 1.0 \times 10^{-4}$ and $< 9.1 \times 10^{-5}$, respectively). JEM identified 10 rare variants acting as both eQTLs and meQTLs (suggestive association: $p < 0.01$): one variant in *C10orf57* (*TMEM254*), six in *ANK3*, two in *C6orf10* (*TSBP1*), and one in *TREML4*. Notably, rs185539955 (MAF = 0.007) had significant negative association with *C10orf57* expression ($P = 9.5 \times 10^{-9}$) and positive association with cg06079468 methylation ($P = 1.2 \times 10^{-6}$) after Bonferroni correction. eQTL and meQTL effect estimates had opposite signs for variants in *C10orf57*, *C6orf10*, and 2/6 *ANK3* variants, but the same direction for the variant in *TREML4*. By applying JEM to blood multi-omics, we identified rare variants acting as both eQTLs and meQTLs in AD loci. Further studies are needed to investigate their relation with AD risk.

Keywords: Rare variant association, Multi-omics, eQTL, meQTL, Alzheimer's disease

100 Rare Population-Specific *HOXB13* Variants and Prostate Cancer Risk in The All of Us Research Program

Shuyan Cheng^{1,2*}, Rohini S. Janivara^{3,4}, Louisa B. Goss², Austin Hammermeister Suger¹, Boya Guo², Sara Lindström^{1,2}, Joseph Lachance⁴, Burcu F. Darst^{1,2}

¹School of Public Health, University of Washington, Seattle, Washington, United States of America; ²Division of Public Health Sciences, Fred Hutchinson Cancer Center, Seattle, Washington, United States of America; ³School of Medicine, Yale University, New Haven, Connecticut, United States of America; ⁴School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, United States of America

HOXB13 is a major prostate cancer (PCa) risk gene harboring rare population-specific variants, including missense variant G84E (minor allele frequency (MAF) $\sim 0.24\%$) in Scandinavian populations and stop-lost variant X285K (MAF $\sim 0.22\%$) in Western African populations. To better characterize *HOXB13*

PCa risk variants, we performed validation and haplotype analyses, scanned for additional risk variants, and explored trait associations using multi-ancestry *All of Us* (AoU) Research Program.

We conducted ancestry-stratified single-variant association tests across *HOXB13* in AoU v8 using REGENIE. Ancestral origin of X285K was investigated via ADMIXTURE, and identity-by-descent (IBD)-based haplotype analyses using hap-ibd with Eagle and YRI genetic maps. Associations between G84E and X285K with 1,660 traits were evaluated via phenome-wide association study (PheWAS).

G84E was associated with higher PCa risk in European ancestry individuals (OR = 4.19, 95% CI = 3.01–5.83, $P = 1.81 \times 10^{-17}$; 5,757 cases/54,751 controls), and X285K with higher PCa risk in African ancestry individuals (OR = 2.63, 95% CI = 1.25–5.55, $p = 0.011$; 1,212 cases/20,193 controls). X285K MAF was highest (0.41%) in individuals with >65% Western African ancestry. IBD analyses revealed longer spanning than non-spanning segments (Eagle: 0.808 vs. 0.645 cM; YRI: 0.712 vs. 0.637 cM). Four additional rare deleterious *HOXB13* variants were nominally associated with PCa risk (splice donor, stop-lost, two frameshifts; minor allele count $\leq$ two each). PheWAS identified nominal risk-increasing associations for both variants with several other cancers.

We validated PCa risk associations of G84E and X285K and traced X285K to a Western African founder haplotype. Additional rare *HOXB13* variants may contribute to PCa risk, and larger studies are needed to evaluate roles in other cancers.

101 A Pangenomic Approach Reveals New Structural Variants Associated With Gene Expression in the CARTaGENE Cohort

Rose Laflamme^{1,2*}, Cristian Groza², Guillaume Lettre^{1,2}

¹Department of Medicine, Université de Montréal and Montreal Heart Institute, Montréal, Québec, Canada; ²Montreal Heart Institute, Montréal, Québec, Canada

Structural variants (SVs) are the largest source of genetic diversity. However, they are not well captured by linear genome references (eg. hg38). Efforts have been made to use pangenome reference because they include multiple alleles at variable sites. Its use increases the number of discovered variants, especially structural variants. We used pangenomic approaches to study their phenotypic impact in the CARTaGENE cohort (CaG). CaG is a population-based cohort of 43,000 individuals aged between 40 and 69 at recruitment from the province of Québec (Canada).

We used the Human Pangenome Reference Consortium pangenome and PanGenie to genotype 645,680 SVs in 2,173 CaG participants with short-read whole-genome sequencing data. In 505 participants, we identified 3,081 cis-eQTLs. To observe the gain of using pangenomic approaches, we conditioned these cis-eQTL on variants identified by methods based on hg38. We found 186 eQTLs that remained significant, implicating 171 SVs and 101 genes. 53/171 SVs intersect an ENCODE candidate

cis-regulatory element (binomial p-value 8.9×10^{-7} , OR = 2.81) and 13/53 are linked to the same gene through enhancer-to-gene predictions.

To explore whether the 171 SVs are associated with human phenotypes, we extracted the small variants from the pangenome that are in strong linkage disequilibrium with the SVs and queried them in the Open Targets GWAS credible set database. We found 36/171 SVs with at least one variant proxy in a credible set, with associated traits such as cholesterol levels and blood pressure. Our work illustrates the value of using pangenomes to discover new genetic variants associated with complex human traits.

Keywords: Pangenomic, Structural Variants, eQTLs, Complex Traits, CARTaGENE Cohort

103 Multi-Ancestry Genome-Wide Association Study Identifies Novel Genetic Contributors to Dyslipidemia Risk in Childhood Cancer Survivors

Michael J. Betti^{1*}, Kendrick Li¹, Kateryna Petrykey¹, Stephanie B. Dixon^{1,2}, M. Monica Gramatges³, Yutaka Yasui¹, Kirsten K. Ness¹, Melissa M. Hudson^{1,2}, Gregory T. Armstrong¹, Yadav Sapkota¹

¹Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, Tennessee, United States of America; ²Department of Oncology, St. Jude Children's Research Hospital, Memphis, Tennessee, United States of America; ³Department of Pediatrics, Division of Hematology-Oncology, Baylor College of Medicine, Houston, Texas, United States of America

Introduction: Childhood cancer survivors are seven times more likely to die from cardiovascular disease compared with the general population, and dyslipidemia is a common long-term risk factor among these individuals. Although treatment exposures are known to influence dyslipidemia risk, the contribution of genetic factors remains understudied.

Methods: In the St. Jude Lifetime Cohort Study ($n = 4,375$), we performed multi-ancestry genome-wide association studies (GWAS) of four baseline lipid traits (total cholesterol, HDL cholesterol, LDL cholesterol, and triglycerides). Linear models were adjusted for age at cancer diagnosis and measurement, sex, treatment exposures, and the first 10 principal components. Associations reaching suggestive significance ($P < 1 \times 10^{-6}$) were evaluated for association ($p < 0.05$ and concordant direction of effect) with dyslipidemia (defined using Common Terminology for Adverse Events as Grade ≥ 2 hypercholesterolemia or triglyceridemia) using Cox regression and subsequently tested for replication in the Childhood Cancer Survivor Study.

Results: We identified four total cholesterol-associated loci, nine with HDL cholesterol, 11 with LDL cholesterol, and 10 with triglycerides. Among these, 23 unique loci were associated with dyslipidemia, 14 of which replicated. The strongest overall association was at 19q13.32 (rs7412, total/LDL cholesterol, $\beta = -0.58$, $P = 9.28 \times 10^{-57}$) within *APOE*. Nine replicated loci

had no previously reported association with dyslipidemia in the general population. The strongest novel association was at 6p24.2 (rs2273884, triglycerides, $\beta = -0.11$, $P = 2.08 \times 10^{-7}$), overlapping *SMIM3*.

Conclusions: In addition to established treatment exposures and genetic variants previously identified in the general population, there are novel genetic factors influencing long-term risk for dyslipidemia among childhood cancer survivors.

104 Preliminary Rare Variant Analysis Identifies a Novel Gene Associated With Survival in Idiopathic Pulmonary Fibrosis

Dominic Sayers^{1,2}, Iain Stewart³, Toby M. Maher^{3,4,5}, Phil Molyneux^{3,5}, Gisli Jenkins³, Louise Wain^{1,2}, Kath Fawcett^{1,2} and Richard Allen^{1,2}

¹Division of Public Health and Epidemiology, University of Leicester, Leicester, United Kingdom; ²University Hospitals of Leicester NHS Trust, Leicester, United Kingdom; ³National Heart and Lung Institute, Imperial College London, London, United Kingdom; ⁴Division of Pulmonary and Critical Care Medicine, University of Southern California, Los Angeles, California, United States of America; ⁵Royal Brompton and Harefield Hospitals, London, United Kingdom

Idiopathic pulmonary fibrosis (IPF) is a progressive lung disease characterised by tissue thickening and scarring, leading to breathlessness and fatigue. Median survival following diagnosis is 3–5 years. Current antifibrotic treatments, including nintedanib and pirfenidone, can slow disease progression but are not curative, while lung transplantation is limited by significant risks. Although common genetic variants have been associated with IPF survival, the contribution of rare genetic variants remains poorly understood. This study aims to investigate how rare genetic variants influence IPF survival.

Rare variant survival analyses were conducted using the time-to-event framework implemented in Regenie v4.0, incorporating genetic and longitudinal health data from the PROFILE cohort. Whole Genome sequencing data underwent stringent quality control, variant annotation and gene-based aggregation. Single-variant survival analyses were conducted using Cox proportional hazards models, adjusting for age, sex, and principal components. Gene-based burden tests focussed on predicted loss of function variants and deleterious missense variants across four allele frequency thresholds. Statistical significance was determined using Bonferroni correction.

The PROFILE study included 449 individuals with whole genome sequencing. In the single-variant analysis, one variant in chromosome 10 between LINC02944 and FAM204A showed a significant protective association with survival. Gene-based analyses identified AGMO as significant at the 0.1% allele frequency threshold.

These preliminary findings suggest that whole-genome sequencing analyses can identify novel genetic loci associated with survival in IPF, highlighting the potential contribution of rare variants to disease outcomes. Replication analyses are currently underway to validate these findings.

105 The Developmental GTEx Resource Enables the Discovery of Disease-Associated Genes

Sofia Salazar-Magaña¹, Temidayo Adeluwa², Sarah Sumner¹, Winona Oliveros-Diez³, Rebecca Keener⁴ & Hae Kyung Im¹ on behalf of the dGTEx consortium

¹Department of Medicine, Section of Genetic Medicine, The University of Chicago, Chicago, Illinois, United States of America; ²Genetics, Genomics, and Systems Biology, The University of Chicago, Chicago, Illinois, United States of America; ³Department of Gene Technology, SciLifeLab, KTH Royal Institute of Technology, Solna, Sweden; ⁴Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, United States of America

The Developmental Genotype-Tissue Expression (dGTEx) project collected pediatric human tissues to study transcriptome regulation in childhood. However, small sample sizes preclude traditional transcriptome-wide association studies (TWAS), limiting disease-gene discovery. We circumvented this by developing SNP-based predictors of gene expression for 72 contexts, spanning 24 tissues and four developmental windows (ages 0–2, 2–8, 8–12, and 12–18).

Using the scPrediXcan approach, we trained deep-learning predictors of dGTEx expression, achieving high correlations with observed expression across the genome (median $r = 0.85$). To enable population-scale analyses, we approximated these as SNP-based linear predictors. Deep-learning and linear models remained highly correlated (median $r = 0.89$).

We performed TWAS of 136 traits (99 adult, 37 pediatric) across seven tissues and four developmental windows, benchmarking dGTEx scPrediXcan models against elastic net and scPrediXcan models trained on adult GTEx.

Both dGTEx and GTEx scPrediXcan discovered $\geq 20\%$ more genes than GTEx elastic net, with comparable precision, both exceeding elastic net. Interestingly, dGTEx scPrediXcan outperformed adult-trained GTEx models across all developmental windows, (median gains: 8.98% for adult traits, 16.01% for pediatric traits, 9.82% overall). Traits with larger gains ($\geq Q50$) tended to be more pediatric (OR = 1.84, $p = 2.87 \times 10^{-3}$), and genetic predictability decreased with age, suggesting early-life expression improves TWAS performance.

In summary, deep-learning prediction enables effecting model training models from limited samples. Developmental predictors improve TWAS power over adult-trained models for both pediatric and adult-onset disease, likely because gene expression is more genetically predictable early in life, making pediatric transcriptomes broadly informative.

106 S4-Oligo: Optimizing Polygenic Scores for Clinical Deployment

John Baierl¹, Jonathan P. Tyrer², Paul D. P. Pharoah¹

¹Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, United States of America; ²Department

Polygenic scores (PGSs) have progressed toward clinical utility in recent years, demonstrating odds ratios that are comparable with both lifestyle-related and monogenic risk factors for several complex diseases. However, PGS implementation relies on developing analytically and clinically valid assays. Methods for developing polygenic models (PGMs) that incorporate tens of thousands to millions of variants require either genome-wide chip genotyping combined with imputation or whole genome sequencing, limiting their portability. Additionally, variants may be in regions where the local genomic sequence makes it difficult to assay, and multi-allelic or strand ambiguous variants can make bioinformatics more complicated. Because of these challenges, attempts to develop a reliable assay for published PGMs can produce multiple variations from the same underlying model. This hinders clinical deployment and incorporation into multi-factorial risk prediction models. In this analysis, we present an adjustment to the previously published select and shrink with summary statistics (S4) PGM method designed to optimize performance for realistic clinical use by restricting to a pre-specified number of available SNPs that are well-captured by genotyping arrays. We compare the performance of this oligogenic S4 method (S4-Oligo) to currently published PGSs as well as the full S4 model for prediction of breast, colorectal, ovarian, and prostate cancers in All of Us, Biobank Japan, and UK Biobank. We find that S4-Oligo matches or exceeds the accuracy of published PGMs at comparable model sizes while retaining between 88% and 99% of the performance of the original S4 model when limited to just 1,000 variants.

107 Polygenic Scores for Targeted Atrial Fibrillation Screening: Prime Time for the Clinic?

Cristian Riccio^{1*}, Marcus Vollmer^{2,3*}, Daniel Engler^{4,5}, Alexander Teumer^{3,6}, Stefan Gross^{3,7}, Raphael Twerenbold^{4,5}, Stefan Blankenberg^{4,5}, Andreas Ziegler^{1,4,5,8}, Tanja Zeller^{9,10**}, Renate Schnabel^{4,5**}

¹Cardio-CARE, Medizincampus Davos, Davos, Switzerland;

²Institute of Bioinformatics, University Medicine Greifswald, Greifswald, Germany; ³German Center for Cardiovascular Research (DZHK), partner site North, Greifswald, Germany;

⁴German Center for Cardiovascular Research (DZHK), partner site North, Hamburg, Germany; ⁵Department of Cardiology, University Heart & Vascular Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; ⁶Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany; ⁷Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany; ⁸Discipline of Statistics, School of Agriculture and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa; ⁹German Center for Cardiovascular Research (DZHK), partner site North, Lübeck, Germany; ¹⁰Institute for Cardiogenetics, University of Lübeck, Lübeck

Background: Atrial fibrillation (AF) often goes undetected until complications arise. Polygenic scores (PGS) may enhance AF screening beyond established clinical risk algorithms such

as the CHA₂DS₂-VA score. Whether PGS can predict screen-detected AF in people aged 45 years or older is not known.

Methods: All 60 AF polygenic scores available in the PGS Catalog were computed in participants of the German population-based Hamburg City Health Study (HCHS) and the Study of Health in Pomerania (SHIP) and evaluated for their ability to predict screen-detected AF (HCHS: n = 33/7,050, 0.5%; SHIP: n = 63/11,134, 0.6%).

Results: HCHS participants were approximately 49% women with a median age of 62 years. SHIP participants were approximately 53% women with a median age of 60 years. In both cohorts, CHA₂DS₂-VA had higher point estimates than the best-performing PGS alone (PGS005313): AUROCs were 0.64 [95% CI: 0.53–0.75] versus 0.58 [0.53–0.64] in HCHS, and 0.76 [0.64–0.89] versus 0.74 [0.68–0.80] in SHIP. A combined PGS + CHA₂DS₂-VA model achieved higher AUROCs of 0.73 [0.64–0.82] and 0.84 [0.76–0.93], respectively. Results in participants aged 65 years or older were consistent with the overall pattern.

Conclusions: The PGS alone did not outperform CHA₂DS₂-VA, but combining the two yielded higher AUROC point estimates in both cohorts. The wide confidence intervals reflect the small number of screen-detected AF events, and further research is needed to establish whether these improvements translate into clinical practice.

108 Cell-Type-Informed DNA Methylation Clocks Reveal Immune Aging Signatures Across Immune-Mediated Diseases

Shuangyu Zhao^{1*}, Xiaoyu Liang¹

¹Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, Michigan, United States of America

Bulk blood DNA methylation (DNAm) profiles reflect signals from mixed leukocyte populations. This cell mixture can obscure cell-subtype-related methylation aging patterns in conventional blood-based analyses. Cell-type-informed DNAm clocks may help identify age-related functional shifts in particular subtypes, potentially predicting immune-mediated diseases involving those subtypes. We developed cell-type-informed DNAm clocks from peripheral blood DNAm profiles (N = 2,711) to estimate biological age for major immune cell populations including neutrophils (Neu), naïve/memory CD8 + T cells (CD8Tnv/mm), naïve/memory B cells (Bnv/mm), T regulatory cells (Treg), natural killer cells (NK) and naïve/memory CD4 + T cells (CD4Tnv/mm). Candidate CpGs were identified from age-associated cell-type-informed methylation sites and prioritized using repeated subsampling-based feature selection. Cell-type-informed clocks were developed and tuned to estimate biological age, then assessed in independent cohorts. Clocks were further applied to immune-mediated diseases cohorts including rheumatoid arthritis (RA, N = 689), inflammatory bowel disease (IBD, N = 384) and asthma (N = 641) to evaluate associations between disease status and epigenetic age acceleration (EAA). RA was significantly associated with EAA from Neu-clock ($P = 2.27\text{E-}02$, $P = 6.78\text{E-}03$) and CD8Tnv/mm-clock ($P = 3.86\text{E-}02$, $P = 9.91\text{E-}03$), all as protective factors. IBD was significantly

associated with EAA from Bnv/mm-clock ($P = 4.34\text{E-}03$), Treg-clock ($P = 8.06\text{E-}03$) and NK-clock ($P = 9.64\text{E-}03$), all as risk factors. No significant association was found between asthma and EAA from those clocks. These findings suggest that cell-type-informed blood DNAm clocks may capture disease-associated immune aging patterns that differ across immune-related disease.

109 Pleiotropic Effects of HLA Alleles on Immune, Thyroid, and Lipid Traits in a Population-based Canadian Cohort

Peyton McClelland^{1*}, Frida Lona Durazo², Sedat Demiriz¹, Hongyu Xiao¹, Claude Bh  rer¹, Sarah Gagliano Taliun^{2,3}, Daniel Taliun¹

¹Department of Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montreal, Qu  bec, Canada; ²Research Centre, Montreal Heart Institute, Montreal, Qu  bec, Canada; ³Department of Medicine, Department of Neurosciences, Faculty of Medicine, Universit   de Montr  al

Autoimmune thyroid diseases are complex disorders that occur more often in females and are linked to dysregulated lipid levels. Genetic association studies have linked HLA alleles with thyroid disease, but rarely report sex-biased effects.

We investigated sex-effects and sex-by-gene interactions (SxG) of HLA alleles on blood cell, thyroid hormone, and lipid levels, in the Canadian CARTaGENE study (N = 27,036 European; 54.3% female) and UK Biobank (UKB; N = 409,199 white British; 54.1% female). In CARTaGENE, HLA-B*08:01 showed associations with absolute leukocyte ($P = 4.1 \times 10^{-8}$, $\beta[\text{SE}] = -0.11[0.02]$), monocyte ($P = 6.0 \times 10^{-10}$, $\beta = -0.12[0.02]$), and lymphocyte count ($P = 2.2 \times 10^{-5}$, $\beta = -0.09[0.02]$), and thyroxine ($P = 4.9 \times 10^{-5}$, $\beta = 0.06[0.02]$), alongside male-specific effect on HDL cholesterol ($P_{\text{SxG}} = 6.9 \times 10^{-3}$, $P_{\text{male}} = 1.4 \times 10^{-4}$, $\beta = -0.08[0.02]$; $P_{\text{female}} = 0.9$, $\beta_{\text{female}} = -0.002[0.02]$). HLA-C*16:01 and HLA-DPA1*02:02 showed SxG effects on LDL cholesterol ($P_{\text{SxG}} = 1.3 \times 10^{-4}$, $\beta = -0.2[0.05]$; $P_{\text{male}} = 7.9 \times 10^{-4}$, $\beta_{\text{male}} = 0.13[0.04]$; $P_{\text{female}} = 0.1$, $\beta_{\text{female}} = -0.06[0.04]$) and relative lymphocyte counts ($P_{\text{SxG}} = 1.3 \times 10^{-4}$, $\beta = -0.24[0.06]$; $P_{\text{male}} = 3.1 \times 10^{-3}$, $\beta_{\text{male}} = 0.13[0.04]$; $P_{\text{female}} = 3.3 \times 10^{-3}$, $\beta_{\text{female}} = -0.13[0.05]$), respectively. In UKB, HLA-B*08:01 showed consistent sex-combined associations with blood cell counts and HDL cholesterol, but the reported sex-differential effects could not be replicated in the UKB.

We report pleiotropic associations with HLA-B*08:01 and immune, lipid, and thyroid-related traits across two biobanks. However, sex-differential effects were not consistent across studies, highlighting the difficulties in identifying sex-differential effects in the HLA region, and suggesting the need for further investigation.

110 Deep Learning for Gene-Environment Interaction Analysis of Complex Traits

Jessica George¹, Juan Pablo Lewinger¹, W. James Gauderman¹

¹Division of Biostatistics and Health Data Science, Department of Population and Public Health Sciences, University of Southern California, Los Angeles, United States of America

Many complex human diseases and traits are shaped by interactions between genetic factors and environmental exposures (G  E). Traditional regression approaches require pre-specified interactions, limiting their ability to capture nonlinear effects in high-dimensional data. We propose DLG  E, a feed-forward neural network that uses only genetic and environmental inputs to directly learn complex G,E  Y relationships. In simulation studies spanning low-, medium-, and high-complexity settings, with 100 replicated datasets of 10,000 individuals per scenario, we benchmarked DLG  E against linear and penalized (LASSO, elastic-net) regression, random forest (RF), gradient boosting (LightGBM), and a tabular prior-data fitted network (TabPFN). All models were fit using additive genetic encodings and environmental features only, with no interaction terms provided. A ‘‘gold standard’’ linear regression model reflecting the full simulation design, including additive, dominant, and recessive genetic effects, environmental main effects, and G  E interactions, was used as a reference upper bound. For interpretability, we applied SHapley Additive exPlanations (SHAP) to quantify feature contributions. Across scenarios, deep learning approaches (DLG  E and TabPFN) achieved the strongest performance. In the high-complexity setting, DLG  E closely matched the gold standard ($R^2 = 0.80$ vs. 0.82), outperforming RF ($R^2 = 0.74$) and linear and penalized regression models ($R^2 = 0.56$). SHAP analyses indicated that flexible ML methods correctly prioritized genetic, environmental, and G  E effects. These results highlight the advantages of deep learning for modeling complex G  E relationships while retaining interpretability. We applied DLG  E to predict cancer-related biomarkers, substantially improving performance over standard regression and identifying potentially biologically meaningful G  E interactions.

111 Low-Dimensional Linkage Disequilibrium Matrices as a Scalable Framework for Multi-Ancestry Fine-Mapping

Nadia Alipour^{1*}, Cristian Pattaro¹

¹Institute for Biomedicine, Eurac Research, Via Volta, 21 – 39100 Bolzano, Italy

Fine-mapping of genome-wide association study (GWAS) results is sensitive to mismatches between external reference panel linkage disequilibrium (LD) and in-sample LD. Using external reference-panel LD in place of in-sample LD can inflate spurious credible sets, while violations of positive semi-definiteness (PSD) of the LD matrix invalidate the underlying likelihood. The high dimensionality of LD further impedes sharing across large GWAS meta-analyses, particularly in multi-ancestry settings. Low-dimensional LD (ld-LD) approximations may represent a useful compromise to distribute in-sample LD information across studies, enhancing the use of in-sample LD without affecting computational resources.

The method review of existing ld-LD approximations suggest that they face a structural trade-off: entrywise thresholding and hard banding compress LD but routinely break PSD, while block-diagonal approximations preserve PSD only via coarse partitions that discard long-range correlation. Our review

suggests that a more efficient representation is possible that captures both local LD decay and long-range structure, guarantees PSD, and yields compact, shareable LD objects. We evaluate its performance in multi-ancestry fine-mapping along two axes: (i) computational and storage cost and (ii) accuracy and mismatch robustness, against dense, block-diagonal, and shrinkage baselines. Benchmarks on 1000 Genomes data assess coverage, resolution, calibration, runtime, and storage.

Our review and application show that methods are available that guarantee PSD while preserving long-range LD structure within a low-dimensional representation. They may offer a promising solution to fine-mapping of multi-ancestry meta-analyses of whole-genome sequencing association studies.

Keywords: Fine-mapping, Linkage disequilibrium, Positive semi-definiteness, multi ancestry, genome-wide association studies

112 Multi-Ancestry Genome-Wide Association Study of Longitudinal BMI Variability and Trajectory

Authors: Chang Liu¹, Greg Linchangco¹, Qin Hui^{1,2}, Jin J Zhou^{3,4}, Yuk-Lam Ho⁵, Daniel C Posner⁵, J Michael Gaziano^{5,6}, Kelly Cho^{5,7}, Peter W F Wilson^{2,8}, Yan V Sun^{1,2,8}

¹Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, Georgia, United States of America; ²Atlanta VA Healthcare System, Decatur, Georgia, United States of America; ³Department of Medicine and Biostatistics, University of California, Los Angeles, California, United States of America; ⁴Veterans Affairs Phoenix Healthcare System, Phoenix, Arizona, United States of America; ⁵Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System, Boston, Massachusetts, United States of America; ⁶Division of Aging, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America; ⁷Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America; ⁸Division of Cardiology, Department of Medicine, Emory University School of Medicine, Atlanta, Georgia, United States of America

Body mass index (BMI) is often treated as a static measure of obesity, but longitudinal BMI changes may provide additional insights. We conducted a multi-ancestry GWAS of two longitudinal BMI traits: BMI coefficient of variation (CV), defined as within-person BMI standard deviation divided by mean BMI, and BMI slope, defined as the annual rate of change in BMI per person estimated by a mixed effects model. The analysis included 543,633 Veterans enrolled in the Million Veteran Program with ≥ 3 BMI measurements (median 29) spanning ≥ 2 years (median 10.4 years). Ancestry-specific GWAS were performed in European (N = 373,471), African (N = 111,031), Admixed American (N = 53,217), and East Asian (N = 5,914) ancestral groups, adjusted for age, sex, and top 10 genetic principal components, followed by random-effects meta-analysis.

We identified 24 genome-wide significant loci for BMI CV and 21 for BMI slope, which were concordant across ancestries (mean heterogeneity $I^2 = 0.17$ and 0.2). The *APOE4* locus and lead variant rs429358 had the strongest association with BMI CV ($P = 2.85 \times 10^{-33}$) and BMI slope ($P = 9.16 \times 10^{-108}$). Additional loci included established adiposity and metabolic genes, including *FTO*, *TCF7L2*, *GCKR*, *PPARG*, and *IGF2BP2*. Baseline BMI had moderate genetic correlation with BMI CV (0.483, SE 0.019) but not BMI slope (-0.036 , SE 0.038), suggesting distinct genetic architecture for long-term BMI change.

Collectively, these findings show that EHR-derived longitudinal BMI variability and trajectory have robust, partly shared but distinct genetic architectures across ancestries, highlighting longitudinal BMI traits as complementary phenotypes for understanding obesity-related disease risk beyond single BMI measurement.

Keywords: Longitudinal study, BMI variability, BMI trajectory.

113 Investigating the Genetic Architecture of Juvenile Myoclonic Epilepsy Through Genome-wide Association Meta-analysis

Brendan A. Newton^{1,2}, Eric J. Sanders^{2,3}, Delnaz Roshandel², Fan Lin², the BIOJUME Consortium, Deb K. Pal^{4,5,6}, Lisa J. Strug^{2,3,7,8}

¹Institute of Medical Science, Temerty Faculty of Medicine, The University of Toronto, Toronto, Canada; ²Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, Canada; ³Division of Biostatistics, Dalla Lana School of Public Health, The University of Toronto, Toronto, Canada; ⁴Department of Basic & Clinical Neurosciences, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom; ⁵MRC Centre for Neurodevelopmental Disorders, King's College London, London, United Kingdom; ⁶King's College Hospital, London, United Kingdom; ⁷Departments of Statistical Sciences and Computer Science, The University of Toronto, Toronto, Canada; ⁸The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada

Juvenile myoclonic epilepsy (JME) is an adolescent-onset subtype of genetic generalized epilepsy (GGE) with a presumed complex genetic basis. JME-specific genetic architecture remains incompletely characterized: previous GWAS aggregated heterogeneous GGE subtypes and had modest JME case counts, with three reported loci at 8q23.1, 4p12, and 16p11.2 not independently replicated. European JME cases from the BIOJUME Consortium were imputed to TOPMed r3 and matched 1:5 in PC space to European controls from Spit for Science (Ontario pediatric cohort). Association was tested using logistic regression score tests in REGENIE, adjusting for sex and the first ten PCs. Summary statistics were meta-analyzed with the most recent International League Against Epilepsy (ILAE) JME GWAS using an inverse variance-weighted fixed-effect model ($n = 2,434$ cases; total $n = 48,380$). Meta-analysis identified six genome-wide significant loci, three novel to JME: rs264976 (intergenic, $= -0.0920$ [0.0169], $p = 4.983 \times 10^{-8}$); rs2383435 at ESYT2 ($= 0.1082$ [0.0198], $p = 4.687 \times 10^{-8}$); and rs133565 near

FAM19A5 ($= -0.1085$ [0.0180], $p = 1.604 \times 10^{-9}$). Three ILAE-reported loci reached genome-wide significance: rs3019359 at TMEM74 ($= -0.0952$ [0.0171], $p = 2.624 \times 10^{-8}$); rs11191193 at KCNIP2 ($= -0.1087$ [0.0178], $p = 1.092 \times 10^{-9}$); and rs1046276 near STX1B ($= 0.1117$ [0.0177], $p = 2.520 \times 10^{-10}$). As the largest JME genetic study to date, this work identifies three novel risk loci not significant in either cohort alone. Two ILAE JME loci (8q23.1, 16p11.2) remain genome-wide significant while 4p12 attenuates to suggestive significance. Relevant genes implicate membrane tethering (ESYT2), autophagy (TMEM74), ion channel regulation (KCNIP2), synaptic integrity, and neuroinflammation (FAM19A5). These findings consolidate and expand the genetic architecture of JME.

115 A Multi-Organ Vascular Aging Clock to Map Trajectories of Systemic Disease

Anna Zaidi^{1,3*}, Thomas M. Zheng^{1,3}, Cienna Gin-Naccarato^{2,3}, Arthur Richard^{1,3}, Zhenqian Wang^{1,2}, Sirui Zhou^{1,2,3}

¹Victor Phillip Dahdaleh Institute of Genomic Medicine, McGill University, Montréal, Canada; ²Department of Human Genetics, McGill University, Montréal, Canada; ³Quantitative Life Science Program, McGill University, Montréal, Canada

Although aging is one of the strongest determinants of chronic disease risk, individuals of the same chronological age can have different biological age, which also varies between organ or systems. Vascular age reflects the health of cardiovascular system, it can be used as a risk indicator of heart disease and stroke, and usually estimated via specific clinical tests in combination of risk factors. Our project aims to leverage multimodal imaging data to develop a comprehensive vascular aging clock integrating brain, cardiac MRI and carotid ultrasound and examine its predictability to disease risks, molecular correlates, and longitudinal trajectories.

We selected 37 multimodal IDPs from UK Biobank spanning carotid ultrasound (N = 8), cardiac MRI (N = 6), and brain MRI (N = 23) measures in 64,349 individuals to develop a vascular clock using a LASSO regression framework. The model demonstrated strong predictive performance in cross-validation ($r = 0.843$, MAE = 4.22 years) and independent holdout validation ($r = 0.691$, MAE = 4.66 years).

Phenome-wide association analyses, including Cox proportional hazards analyses, revealed significant associations between elevated vascular age and multiple cardiometabolic, vascular, and neurological diseases, including hypertension (HR = 1.047, $p = 7.89 \times 10^{-27}$), diabetes mellitus (HR = 1.067, $p = 7.17 \times 10^{-14}$), and Alzheimer's disease (HR = 1.168, $p = 1.66 \times 10^{-10}$). Longitudinal analyses demonstrated stability of vascular age across imaging visits ($r = 0.736$). Proteomic analyses found inflammatory and extracellular matrix pathways, including GDF15 and HAVCR1, that are associated with increasing of vascular age. These results suggest that imaging-derived vascular age captures biologically meaningful variation linked to systemic aging and future disease risk and can be used to explore potential therapeutic targets for vascular health.

116 Reduced Cancer Incidence Associated With Metformin Use in DNMT3A-mutant Clonal Hematopoiesis: An Entropy Balancing Approach

Sam Pedro Galilee Ayivi^{1,2,3*}, Lambert Busque^{4,5}, Marie-Christyne Cyr^{1,2}, Maurel Dylane Tchuisse IIMbiagoup^{1,2,3}, Louis-Philippe Lemieux Perreault^{1,2}, Johanna Sandoval^{1,2}, Amina Barhdadi^{1,2}, Simon de Denus^{1,2,6}, Jean-Claude Tardif, MD^{1,4}, Marie-Pierre Dubé, PhD^{1,2,4}

¹MontrealHeartInstitute, Montreal, Canada; ²Université de Montréal Beaulieu-Saucier Pharmacogenomics Centre, Montreal, Canada; ³Department of Biochemistry and Molecular Medicine, Université de Montréal, Montreal, Canada; ⁴Department of Medicine, Université de Montréal, Montreal, Canada; ⁵Hôpital Maisonneuve Rosemont, Montreal, Canada; ⁶Faculty of Pharmacy, Université de Montréal, Montreal, Canada

Background: Clonal hematopoiesis of indeterminate potential (CHIP) is an age-related condition characterized by the expansion of hematopoietic stem and progenitor cells harboring somatic mutations that confer a competitive advantage. CHIP is associated with increased risks of age-related diseases, including hematological malignancies and solid tumors. CHIP mutations, particularly DNMT3A-mutant CHIP, are prevalent among individuals with type 2 diabetes (T2D) and may create a synergistic environment promoting increased cancer susceptibility. However, the role of DNMT3A-mutant CHIP in modulating the response to T2D treatments remains poorly characterized.

Methods: Our analysis included 18,394 cancer-free T2D participants from the UK Biobank cohort, of whom 11,392 were metformin users. To control for confounding and enable robust comparisons of treatment effects, we applied entropy balancing, a covariate reweighting approach that assigns weights to participants to achieve balance in baseline covariates between metformin users and non-users. Covariate balance was assessed using standardized mean differences and compared to other covariate balancing approaches including propensity scores. The association between metformin use and cancer incidence according to CHIP status was estimated using weighted Cox models.

Results: We observed a significant 48% reduction in cancer incidence among large DNMT3A clone carriers (variant allele frequency $\geq 10\%$) who were using metformin compared to non-users (hazard ratio = 0.52; 95% CI: 0.32–0.82; $p < 0.01$; P for interaction = 0.04). No association was observed in the overall T2D population or among non-DNMT3A CHIP carriers.

Conclusion: These findings suggest that metformin may modify cancer risk in DNMT3A-mutant CHIP carriers.

117 Genetic Architecture of Height in Longitudinally Phenotyped Canadians From the Canadian Longitudinal Study on Aging (CLSA)

Abhinav Kansal¹, Myriam Brossard², Shelley B Bull^{1,2}

¹Dalla Lana School of Public Health, University of Toronto; ²Lunenfeld-Tanenbaum Research Institute, Sinai Health

Background: Genome-wide association studies (GWAS) predominantly examine phenotypes cross-sectionally, leaving the genetic architecture of within-person change underexplored. Two complementary longitudinal approaches exist to test for genetic association: regression of a derived phenotype summarising subject-level repeated measures, and linear mixed models (LMM) of repeated measures with fixed and random age effects.

Objective: Compare cross-sectional and longitudinal GWAS approaches to characterize aging-related genetic architecture in a Canadian aging cohort.

Methods: We analysed European-ancestry participants from the CLSA with height measured at three visits spanning 6 years. Two analytical datasets were constructed: DS1 (baseline only, $N = 22,914$) and DS2 (all three visits, $N = 13,749$). In DS1, single-variant GWAS of imputed variants was performed using PLINK2 to fit cross-sectional regression models including covariates for sex, linear and quadratic age, age-by-sex interaction, batch, and ten ancestry principal components. In DS2, an LMM with random intercepts and slopes was fit using lme4 to test for trends in height across visits. Subject-specific intercepts and slopes extracted from this LMM were then used as derived phenotypes in separate PLINK2 GWAS.

Results: Cross-sectional GWAS of baseline height and GWAS of LMM-derived subject-specific mean height identified concordant genome-wide significant loci (339 and 327 respectively), with consistent effect sizes, confirming the genetic stability of adult height ($ICC = 0.952$). GWAS of subject-specific rates of height change with age identified no genome-wide significant loci ($\lambda = 0.998$).

Conclusions: Cross-sectional GWAS captures the dominant genetic signal for height, consistent with the high ICC. Subject-specific intercepts and slopes derived from the LMM provide GWAS-feasible longitudinal phenotypes.

Keywords: GWAS, longitudinal, aging, height, repeated measures

119 Subtype-Specific Polygenic Scores Reveal Genetic Heterogeneity Across Dystonia Phenotypes

Jonathan Krause^{1,2}, Sebastian Sendel^{3,4*}, Björn-Hergen Laabs^{2,5}, Katja Lohmann¹, Amke Caliebe^{3,4}

¹Institute of Neurogenetics, University of Lübeck, University Medical Center Schleswig-Holstein, Campus Lübeck, Lübeck, Germany; ²Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany; ³Epidemiology, Medical Biometry and Medical Informatics, Health and Medical University Erfurt, Erfurt, Germany

⁴Institute of Medical Statistics, Kiel University and University-Hospital Schleswig-Holstein, Kiel, Germany; ⁵University Medical Center Göttingen, Department of Medical Statistics, Göttingen, Germany

Polygenic scores (PGS) are increasingly used to investigate the genetic architecture of complex diseases. Nevertheless, their utility in dystonia remains unclear due to the clinical and etiological heterogeneity of the disorder. In this study, we generated and evaluated multiple dystonia PGS in 1398 individuals with dystonia and 1333 controls of European ancestry. Two genome-wide dystonia PGS were developed using LDpred2 based on two independent dystonia genome-wide association studies (GWAS1 (Laabs et. al 2024) and GWAS2 (Sun et al. 2021): *DYT-PGS1* and *DYT-PGS2*. Both scores showed limited discriminatory power between dystonia cases and controls (*DYT-PGS1*: $AUC = 0.522$; *DYT-PGS2*: $AUC = 0.552$).

A previously published *mDYT-PGS* (Putzel et. al 2016) was derived from a musician's dystonia GWAS and trained on laryngeal dystonia patients from comparatively small datasets (GWAS: 127 cases, 984 controls; PGS development: 57 laryngeal dystonia cases, 30 controls). We applied this score to the subset of individuals of our cohort with musician's dystonia (87 cases, 1333 controls), resulting in a markedly increased AUC of 0.683. Performance varied considerably across additional dystonia subtypes using the same control cohort, including hand dystonia (101 cases, $AUC = 0.582$) and writer's dystonia (38 cases, $AUC = 0.496$).

Our findings support substantial genetic heterogeneity across dystonia subtypes and suggest that subtype-specific PGS may better capture disease risk than scores derived from mixed dystonia cohorts.

Keywords: polygenic scores, gwas, dystonia

120 Population-Scale DNA Methylation Profiling From Long-Read Sequencing Data and EWAS of Aging in the All of Us Cohort

Sarbani Thakur^{1,2*}, Sabrina Chang¹, Brandon L. Pierce^{1,2,3,4}

¹Department of Public Health Sciences, University of Chicago, Chicago, Illinois, United States of America; ²University of Chicago Comprehensive Cancer Center, University of Chicago, Chicago, Illinois, United States of America; ³Committee on Genetics, Genomics, Systems Biology, University of Chicago, Chicago, Illinois, United States of America; ⁴Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America

Long-read whole genome sequencing (lrWGS) enables simultaneous profiling of structural variants and 5-methylcytosine (5mC) DNA methylation (DNAm) in a haplotype-resolved manner. We developed and validated a DNAm extraction pipeline for lrWGS data (v8) from 1,490 All of Us (AoU) participants on two PacBio platforms (Revio and Sequel-II) across four sequencing centers. CpG methylation tags were extracted from GRCh38-aligned BAM files using pb-CpG-tools, parallelized via Cromwell on the AoU Terra. We evaluated the mod-score, a Primrose model-based probability calibrated on the HG002 cell line, and the discretized-mod-score, an empirical read-count fraction. The latter is preferable for stability, although we identified an integer discretization artifact, whereby samples with even-numbered read depth at a CpG show beta value

clustering at 50%. After QC, all 1490 samples and 71.8% of CpGs were retained on average. We performed an epigenome-wide association study (EWAS) of chronological age in 1,466 samples across 25,701,541 autosomal CpGs, adjusting for sex, body mass index (BMI), predicted ancestry, sequencing platform, sequencing center, and ten surrogate variables. We identified 28,082 differentially methylated positions (DMPs) with a Bonferroni *P* value of less than 0.05. Of the 353 Horvath and 514 Zhang clock CpGs, 350 and 507 were detected, respectively. Notably, 14 and 94 reached Bonferroni significance, with directional consistency of 94.4% and 96.3%, respectively. We will scale this pipeline to the future AoU lrWGS release (v9), exceeding 10,000 participants. This work establishes that AoU PacBio HiFi lrWGS provides accurate DNAm estimates, positioning AoU as an important population-scale epigenomic resource across diverse ancestries.

Keywords: Long-read whole genome sequencing, DNA methylation, All of Us cohort, Epigenetic Aging, EWAS

121 Genetic Epidemiology of Rare Genetic Diseases in the Punjabi Population

Iqra Tabassum^{1,2}, Muhammad Shafique¹, Muhammad Shoaib Akhtar^{3,4}

¹Forensic DNA Typing Laboratory, National Center of Excellence in Molecular Biology, University of Punjab, Lahore 53700, Pakistan; ²Punjab Thalassemia and Other Genetic Disorders Prevention and Research Institute, Sir Ganga Ram Hospital, Lahore 54000, Pakistan; ³Department of Cell Biology and Human Anatomy, School of Medicine, University of California Davis, Davis, California, United States of America; ⁴Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwa, Kashiwa 277-8562, Chiba, Japan

Rare genetic diseases (RGDs) impact millions globally, yet data from South Asian populations are significantly lacking in worldwide databases. This study provides a genomic and epidemiological overview of RGDs in the Punjabi population based on data collected from 167 families registered in the Punjab Thalassemia and Other Genetic Disorders Prevention and Research Institute (PTGDPRI) registry between 2021 and 2023. The registry identified 72 distinct pediatric-onset RGDs, with a high burden of disease, with an average of 0.81 affected children per family. Consanguinity was present in 89% of families, with 77% involving first-cousin marriages, and was significantly associated with RGD incidence. The most common diseases were Niemann-Pick disease (NP), progressive familial intrahepatic cholestasis (PFIC), and mucopolysaccharidosis (MPS). Molecular analysis revealed 131 pathogenic variants, including copy number variations (CNVs) and single nucleotide polymorphisms (SNVs), classified as pathogenic, likely pathogenic, or variants of unknown significance across 110 genes, including 24 South-Asian enriched variants. We also identified 19 variants with shared ancestry between South Asian and non-Finnish European populations. Two other variants shared ancestry between the South Asian and Middle Eastern populations. Four other variants were shared among the South Asian, European, and Middle Eastern populations. RGDs carrier families were highly literate

and belonged to low-income groups. This low-income group limits access to basic healthcare for diagnosis and treatment. The findings reveal how genetic, socioeconomic, and cultural factors converge to amplify the RGD burden and highlight the need for affordable molecular diagnostics, inclusive genomic databases, and regional genomic surveillance initiatives in South Asia.

Keywords: rare genetic diseases (RGDs), Copy Number Variations (CNVs), single nucleotide variations (SNVs), South Asia, South Asian Enriched Variants, Consanguinity, Variant of Unknown Significance (VUS)

122 Nationwide Insights Into Pharmacogenomics and Medi-cation Use in France

Marc Gros-La-Faige^{1*}, the POPGEN study group², GOLD consortium³, Emmanuelle Génin^{1,4}, Anthony F. Herzig¹

¹Inserm, Univ Brest, EFS, UMR 1078, GGB, Brest, France; ²Inserm, Brest, France, c19-39.isp@inserm.fr; ³Inserm, Paris, France, contact.ptvg-gold@inserm.fr; ⁴CHRU Brest, Brest, France

* presenting author

Pharmacogenomics (PGx) is crucial in the development of personalized medicine by enabling drug prescription to be tailored to each patient's genomic profile. To assess the potential benefits of implementing PGx screening at the population level, it is important to first describe pharmacogene diversity in the target population and then link this information with drug consumption data. Until now, this has not been feasible in France due to the lack of genomic data representative of the general population.

Our study is encompassed within the POPGEN project, a wide investigation of genomic diversity across metropolitan France, involving SNP-array data from 9,598 individuals, 3,998 of whom were whole-genome sequenced. To better capture pharmacogene variability affected by structural variants (SVs), we developed and evaluated a novel PGx-specific SV imputation method for *CYP2A6*, *CYP2B6*, and *CYP2D6*. Pharmacogenomic profiles were then linked to individual electronic health records through the GOLD-GENOPHENOMET project and the linkage of the Constances cohort with national healthcare databases.

We found several pharmacogenes harboring actionable alleles with region-specific frequency differences. For a list of 50 drugs, we found that a substantial proportion of individuals had been prescribed drugs poorly suited to their PGx profile. For nine common medications including tramadol and statins, we observed significant associations between number of prescriptions and predicted metabolism-based therapeutic recommendations; potentially suggesting adverse effects that could have been avoided through PGx profiling.

Our study provides a first overview of the PGx diversity and its consequences in France and open the debates of its implementation in medical care.

Additional information: Fundings: The POPGEN Project: French Ministry of Research PFMG2025; The CONSTANCES cohort: ANR-11-INBS-0002; Inserm cross-cutting program Genomic variability 2018 GOLD

123 Real-World Impact of Comprehensive β -Thalassemia Prevention Services in Punjab

Yasmeen Ehsan¹, Iqra Tabassum^{1,2}, Amna Tariq²

¹*Punjab Thalassemia and Other Genetic Disorders Prevention and Research Institute, Fatima Jinnah Medical University/Sir Ganga Ram Hospital Lahore, PTGD, FJMU, Lahore;* ²*National Center of Excellence in Molecular Biology (CEMB), University of Punjab, Lahore, Pakistan*

Beta Thalassemia (b-Thal) is a common autosomal recessive disorder characterized by reduced or absent β -globin chain production, leading to severe anemia. In Pakistan, the carrier frequency ranges from 5% to 7%, with approximately 12.8 million carriers nationwide. This study evaluates preventive measures implemented by the Punjab Thalassemia & Other Genetic Disorders Prevention & Research Institute (PTGDPRI) from 2012 to March 2026. The program introduced voluntary premarital and extended family-based carrier screening after extensive awareness campaigns about b-Thal. Identified carrier couples received genetic counseling (GC), and prenatal diagnosis (PD) was offered using trans-abdominal chorionic villus sampling (CVS). Approximately 501,425 individuals underwent GC, including extended carrier screening (ECS) for families. Premarital screening of 151,718 individuals revealed a carrier frequency of 3.5-5% across 36 districts in Punjab. Among 14,460 families with affected children, the carrier frequency was 30-40%. PD was performed on 12,308 at-risk pregnancies, detecting 3,148 (25.58%) fetuses with b-Thal Major, 6,054 (49.19%) with b-Thal trait, and 3,106 (25.24%) normal fetuses. These results confirm that targeted carrier screening, combined with GC and PD, effectively reduces b-Thal births. The program's integration of modern diagnostics, real-time monitoring, and psychosocial support has improved service uptake and quality. This model presents a scalable, cost-effective approach for thalassemia prevention in resource-limited, high-prevalence settings, addressing the significant public health challenge posed by b-Thal in Punjab due to high carrier rates, population size, and consanguinity.

Keywords: Punjab Thalassemia and Other Genetic Disorders and Prevention Institute (PTGD), Prenatal Diagnosis (PND), Chorionic Villous Sampling (CVS), extended carrier screening (ECS), premarital screening

124 Objective Multi-criteria Optimisation Yields a Fine-scale, Validated Ancestry Reference Panel that Improves Local Ancestry Inference for Admixed and Underrepresented Populations

Taghrid M. Aloraini^{1,2}, Maiah Walters³, Indhu-Shree Rajan-Babu⁴, René L. Warren⁵, Lauren Coombe⁵, Jan M. Friedman⁴, Inanç Birol⁵.

¹Graduate Program in Bioinformatics, University of British Columbia, Vancouver, Canada; ²Center for Genomic Medicine,

Riyadh, Saudi Arabia; ³Graduate Program in Medical Genetics, University of British Columbia, Vancouver, Canada; ⁴Department of Medical Genetics, The University of British Columbia, and BC Children's Hospital Research Institute, Vancouver, British Columbia, Canada; ⁵Canada's Michael Smith Genome Sciences Centre, BC Cancer Research Institute, Vancouver, Canada

Introduction: Self-reported population labels remain the dominant framework for stratifying genomic data, yet they conflate social, geographic, and genetic identity in ways that often mismatch underlying genetic structure. The five-to-seven continental super-populations used by 1000 Genomes, HGDP, and gnomAD place genetically distinct populations (Finnish with Spanish, Yoruba with Maasai) into single groups, obscuring haplotype, linkage disequilibrium, and admixture signals. The 2023 National Academies recommendation to move toward genetic-similarity-based classification motivates a higher-resolution, data-driven alternative.

Methods: We aggregated 3,767 unrelated individuals from 146 populations across six harmonised datasets (1000 Genomes, HGDP, SGDP, KPGP, SSMP, UPR; 966,683 LD-pruned autosomal SNPs). We built an ancestry reference panel by spectral clustering of L2-normalised PC embeddings, selecting the pipeline configuration by objective multi-criteria decision-making over 1,120 candidates using two complementary weighting schemes (CRITIC and Shannon entropy) that eliminate analyst-imposed priors. We then integrated the allele-frequency reference into the HPRC pangenome and ran ntRoot local ancestry inference (LAI).

Results: Both schemes independently ranked the same 48-cluster configuration first, roughly ten times current resolution. The panel achieved 98.3% continental homogeneity, recovered ten documented isolates as pure clusters, and captured admixture gradients. Four validations confirmed it: Hopkins 0.967, mean FST 0.039, Mantel $r = 0.43$ ($p < 0.001$), leave-one-population-out ARI = 0.862. ntRoot recovered three-way admixture in Puerto Rican and Colombian samples at within-continent resolution.

Conclusion: Objective optimisation yields a fine-scale validated panel that improves local ancestry inference for admixed and underrepresented populations, with direct implications for clinical variant interpretation.

Keywords: ancestry reference panel; population genetic structure; spectral clustering; local ancestry inference; pangenome; allele frequencies; underrepresented populations; clinical variant interpretation

125 Multi-Ancestry Genome-wide Association Meta-Analysis of Wearable-derived Physical Activity Identifies Novel loci in up to 214,000 Individuals

Charilaos Zisou^{1,2*} on behalf of the ActiGen consortium

¹Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom; ²Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, United Kingdom

Introduction: Physical activity is a modifiable risk factor for several common diseases, and identifying its genetic determinants could provide insights into some of the underlying mechanisms. However, existing genetic studies have mostly relied on self-reported measures, which are prone to reporting biases and recall errors. Here, we conducted a multi-ancestry meta-analysis of genome-wide association studies (GWAS) of wearable-derived physical activity phenotypes.

Methods: We meta-analysed GWAS summary statistics from up to 214,020 adults across eight cohorts from six countries (85.1% European, 7.5% East Asian ancestry) for wearable-derived phenotypes of overall activity, step count, peak 1-min cadence, moderate-to-vigorous physical activity (MVPA), and sedentary behaviour. Fixed-effects inverse-variance weighted meta-analysis was performed within and across ancestries. SNP-based heritability and genetic correlations with other traits were estimated using LD score regression.

Results: European-ancestry and multi-ancestry meta-analyses across phenotypes identified 30 significant loci (8 for overall activity, 6 for step count, 8 for cadence, 7 for MVPA, and 7 for sedentary behaviour), of which 12 were novel. SNP-based heritability ranged from 6.7% (MVPA) to 11.2% (overall activity). MVPA was genetically correlated with a range of health-related traits, including educational attainment ($r_g=0.47$, $P=2.81\times10^{-84}$), body mass index ($r_g=-0.46$, $P=1.05\times10^{-85}$), systolic blood pressure ($r_g=-0.22$, $P=5.55\times10^{-15}$), ischaemic heart disease ($r_g=-0.39$, $P=3.09\times10^{-43}$), stroke ($r_g=-0.52$, $P=5.18\times10^{-15}$), heart failure ($r_g=-0.47$, $P=7.55\times10^{-24}$), type 2 diabetes ($r_g=-0.33$, $P=3.81\times10^{-32}$), and major depressive disorder ($r_g=-0.33$, $P=7.06\times10^{-20}$). Functional follow-up analyses are ongoing.

Conclusion: Our findings provide new insights into the biology of physical activity and highlight shared genetic architecture with educational attainment, cardiometabolic traits and mental health.

126 Copula Regression Improves the Ability to Detect Modulation of Promoter-enhancer Dependence

Bowei Xiao¹, Kathleen Oros Klein², Qihuang Zhang³, Josée Dupuis^{3,4}, Mathieu Blanchette⁵, Celia M.T. Greenwood^{2,3,6}

¹Quantitative Life Sciences, McGill University, Montreal, Québec, Canada; ²Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Québec, Canada; ³Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Québec, Canada; ⁴School of Population and Global Health, McGill University, Montreal, Québec, Canada; ⁵School of Computer Science, McGill University, Montreal, Québec, Canada; ⁶Gerald Bronfman Department of Oncology, McGill University, Montreal, Québec, Canada

While promoters and enhancers are the primary cis-regulatory elements to regulate transcription, such regulation can vary not only through marginal changes in promoter or enhancer but also through characteristics of the combination, i.e. their joint

distribution. Using extracted gene-level promoter and enhancer data from 6 different epigenetic profiles (DNA methylation, 4 histone marks, ATAC-seq) of macrophage cells before and after infection with influenza A (Aracena et al. 2024), we propose a novel analytic pipeline built on copula regression models that robustly estimate promoter and enhancer marginal patterns as well as their dependence. This pipeline tests whether sample-level covariates (e.g. age, ancestry, infection status) modify the dependence and whether such dependence is associated with RNA expression. Statistical evidence is combined using a composite hypothesis testing approach where significance is only declared if 1) promoter-enhancer dependence varies with at least one covariate and 2) the estimated dependence is associated with RNA expression.

We tested this pipeline for 15,551 genes in each of the 6 epigenetic data types. The new framework identified 123 gene-datatype pairs (using a composite hypothesis test and Bonferroni correction) where at least one covariate modified the promoter-enhancer dependence AND influenced gene transcription. In contrast, a standard linear model with interaction only found 10 such patterns. Furthermore, we have also demonstrated that our proposed model has improved out-of-sample RNA expression prediction in an independent validation set. Overall, our results demonstrate that direct modelling of covariate-modulated dependence can be a practical and interpretable method to discover co-regulation activity.

127 Accounting for the Moderating Effect of Polygenic Risk Scores in Association Studies of Rare Variants With a Trait in Families

Saa Moussa Tenguiano¹, Loïc Mangnier², Michel Maziade³, Simon Girard⁴, Alexandre Bureau¹

¹Laval University, Department of Social and Preventive Medicine: Québec City, Québec, Canada; ²University of Lille, Department of Health Engineering and Management: Lille, Hauts-de-France, France; ³Laval University, Department of Psychiatry and Neuroscience: Québec City, Québec, Canada; ⁴University of Québec at Chicoutimi, Department of Fundamental Sciences: Saguenay (Chicoutimi), Québec, Canada

Complex diseases are polygenic and multifactorial, with a heterogeneous genetic architecture. Although genome-wide association studies have identified numerous variants and enabled the development of polygenic risk scores (PRS), a substantial proportion of genetic risk remains unexplained. In the Eastern Québec Schizophrenia and Bipolar Disorder (EQSBD) study, some families exhibit low PRS despite strong familial aggregation of psychiatric disorders, suggesting a potential role for rare variants (RVs). However, their detection remains challenging and requires relevant functional annotations.

Few approaches combine RVs and PRS in family contexts, although RV effects may be modulated by PRS. To address this limitation, we propose an interaction model between RVs and PRS in families. This approach extends the retrospective likelihood to genotypes of multiple RVs within a genomic region, weighted by functional annotations and conditioned on affected

family members, and incorporates an interaction term between weighted rare variant burden and PRS. Score statistics for both main and interaction effects were derived and evaluated using realistic simulations.

Simulations based on 1000 Genomes Project data show that the Type I error rate of the interaction test is well controlled when at least 15 carrier families are included. For 2% causal variants, across three scenarios (100%, 75%, and 50% within annotated regions), the ACAT test shows higher power for the main effect than for the interaction, as expected since detecting interactions is more difficult, particularly because the main effect test is one-tailed whereas the interaction test is two-tailed. Preliminary results on EQSBD sequencing data will be presented.

Keywords: Complex diseases, Family history, Heritability, Polygenic risk score, Rare variants.

128 Linear Mixed and Threshold Linear Mixed Model Analyses of DNMT3A-associated TET2 Clonal Expansion in Clonal Hematopoiesis Under Colchicine Treatment

Maurel D. Tchuise II Mbiagou^{1,2,3,*}, Amina Barhdadi^{1,2}, Sam Pedro Galilée Ayivi^{1,2,3}, Johanna Sandoval^{1,2}, Louis-Philippe Lemieux Perreault^{1,2}, Ian Mongrain^{1,2}, Essaid Ousaid^{1,2}, Lambert Busque^{4,5}, Jean-Claude Tardif^{2,4}, Marie-Pierre Dubé^{1,2,4}

¹Université de Montréal Beaulieu-Saucier Pharmacogenomics Centre, Montreal, Québec, Canada; ²Montreal Heart Institute, Montreal, Québec, Canada; ³Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montreal, Québec, Canada; ⁴Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Québec, Canada; ⁵Maisonnette Rosemont Hospital, Montreal, Québec, Canada

Clonal hematopoiesis (CH) is an age-related expansion of somatic mutations in hematopoietic stem cells, increasing cardiovascular risk at variant allele frequencies (VAF) $\geq 2\%$. *DNMT3A* and *TET2*, the most frequently mutated CH genes, promote inflammation. In a sub-study of the COLCOT trial in patients with recent myocardial infarction, randomized to low-dose colchicine or placebo, we showed that colchicine reduced *TET2*-mutant clonal expansion over 19.5 months of follow-up in 848 patients.

Using this dataset, we assessed the effect of *DNMT3A* mutations on *TET2* clonal expansion and its modulation by colchicine. We used linear mixed models to evaluate changes in *TET2* VAF, adjusting for age and follow-up duration. Baseline *DNMT3A* VAF was evaluated as a predictor of *TET2* VAF change within each treatment arm and overall using a treatment $\times$ baseline *DNMT3A* VAF interaction term. Threshold analyses investigated potential baseline VAF thresholds for *DNMT3A* and *TET2*, associated with these effects. To account for non-standard asymptotic distributions and data clustering, parameters and confidence intervals were estimated using cluster bootstrapping.

Results showed that, in the placebo group ($n = 417$), higher baseline *DNMT3A* VAF was associated with a greater increase in *TET2* clonal expansion ($\beta = 4.52 \times 10^{-3}$; $p < 0.001$). In contrast, in the colchicine group ($n = 431$), the association between *DNMT3A* VAF and *TET2* clonal expansion was not significant. The difference between treatment arms was significant ($P_{\text{int}} < 0.001$).

These findings suggest *DNMT3A*-mutant clones may promote *TET2* clonal expansion, a relationship attenuated by colchicine. Ongoing analyses will determine if specific baseline *DNMT3A* or *TET2* VAF thresholds identify subgroups with differential susceptibility to this effect.

129 GWAS of Quantitative Traits With Repeated Measures: Alanine Aminotransferase (ALT) Levels in the Canadian Longitudinal Study on Aging (CLSA)

Authors:

Wendy Shue^{1*}, Myriam Brossard², Shelley B. Bull^{1,2}

¹Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; ²Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Canada

Recent UK Biobank longitudinal genome-wide association studies (GWAS) have identified genetic variants influencing baseline differences and dynamic trajectories of quantitative traits, including alanine aminotransferase (ALT). Motivated by these findings, we compared longitudinal analytical frameworks to evaluate genetic influences on ALT in the Canadian Longitudinal Study on Aging (CLSA).

We analyzed log-transformed ALT across three visits in 10,545 European-ancestry CLSA participants (~5.78 million SNPs). Genetic associations were assessed using linear regression (LR) for: (1) visit-specific and 3-visit average log ALT, adjusted for age, sex, and principal components; and (2) best linear unbiased predictors (BLUPs) for random intercept and age slope derived from covariate-adjusted linear mixed models (LMM). Single-variant gene-by-age interactions were evaluated for model (1). RegionScan assessed region-level associations using linkage disequilibrium-defined regions.

Single-variant GWAS using both LR models identified genome-wide significant signals ($P < 5e-8$) on chromosomes 8 and 22. LMM-derived BLUPs for random intercept and age slope produced results comparable to the 3-visit average log ALT. Suggestive evidence ($P < 1e-5$) was observed for gene-by-age interactions, indicating age-dependent genetic effects. Using a region-level significance threshold ($P < 5e-7$), RegionScan corroborated single-variant findings and identified additional associations on chromosomes 4, 10, 17, and 22.

Detecting genetic effects on longitudinal aging trajectories remains challenging due to limited power. Single-variant and region-level analyses provided complementary insights for repeated ALT measures and supported previously reported

ALT-associated loci, encouraging ongoing analyses of age-group trends and gene-by-age interactions in both single-variant and region-level GWAS.

130 Accounting for Selected Genetic Regulators of Prostate-Specific Antigen Levels Enhances Prostate Cancer Prediction

Mykhaylo M. Malakhov^{1*}, Martina Fu¹, Brenda Xiao¹, Thomas J. Hoffmann^{2,3}, Rebecca E. Graff², Daniela Fusco⁴, Zhiyu Yang⁴, Andrea Ganna⁴, John S. Witte^{1,5,6}, Linda Kachuri¹

¹Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, California, United States of America; ²Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, United States of America; ³Institute for Human Genetics, University of California San Francisco, San Francisco, California, United States of America; ⁴Institute for Molecular Medicine Finland, Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland; ⁵Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, California, United States of America; ⁶Department of Genetics, Stanford University School of Medicine, Stanford, California, United States of America

Elevation of prostate-specific antigen (PSA) can be indicative of prostate cancer, but its use for population-based screening remains controversial. The same PSA level can have different interpretations depending on genetic predisposition, potentially leading to missed cases or overdiagnosis. Here we present an approach for obtaining genetically informed PSA values and validate it in the FinnGen biobank. Unlike previous methods that remove the component of PSA captured by a genome-wide polygenic score (PGS), we explore PGS partitioning to provide a more precise correction with fewer off-target genetic effects on prostate cancer risk. We first found that although the *cis* region explains only 20% as much variation in PSA as a genome-wide PGS, PSA adjusted by the *cis*-region PGS predicted prostate cancer nearly as effectively as with genome-wide PGS adjustment (OR = 5.92, $P = 1.09\text{e-}444$, AUC = 0.840 vs. OR = 6.07, $P = 3.52\text{e-}453$, AUC = 0.837). Both models improved upon observed pre-diagnostic PSA (OR = 5.83, $P = 3.05\text{e-}440$, AUC = 0.838). Next, we partitioned the PGS into linkage disequilibrium blocks to identify *trans* loci that enhance the accuracy of genetic adjustment in an independent cohort. Applying the resulting filtered PSA score to FinnGen yielded an even stronger association with prostate cancer than genome-wide adjustment, particularly in men with PSA ≤ 4 ng/ml (OR = 6.07, $P = 5.42\text{e-}182$, AUC = 0.783 vs. OR = 5.87, $P = 3.35\text{e-}179$, AUC = 0.779). This work demonstrates the importance of carefully selecting genetic signals when recalibrating noncausal disease biomarkers.

131 MR2G: A Novel Framework for Causal Network Inference Using GWAS Summary Data

Zhaotong Lin^{1*}, Wei Pan², Haoran Xue³

¹Department of Statistics, Florida State University, Tallahassee, Florida, United States of America; ²Division of Biostatistics and Health Data Science, School of Public Health, University of

Minnesota, Minneapolis, Minnesota, United States of America; ³Department of Biostatistics, City University of Hong Kong, Hong Kong, China

Inferring a causal network among multiple traits is essential for unraveling complex biological relationships and informing interventions. Mendelian randomization (MR) has emerged as a powerful tool for causal inference, utilizing genetic variants as instrumental variables (IVs) to estimate causal effects. However, when the directions of causal relationships among traits are unknown, reconstructing the underlying causal network becomes challenging. In particular, the presence of cycles or feedback loops, which are common in biological systems, poses additional challenges for causal network inference, and remains largely under-studied with standard MR approaches and existing IV-based network inference methods.

To address these issues, we introduce MR2G, a new statistical framework that enables robust inference of causal networks, including those with cycles, directly from GWAS summary statistics. MR2G is built on a formally defined recursive causal graph model that rigorously links direct causal effects to (univariable) MR estimands. It recovers a biologically interpretable causal network from pairwise MR effect estimates, while incorporating a network-informed IV screening strategy to reduce pleiotropic bias and improve robustness.

132 Fast and Powerful Gene Discovery Using pQTL Weights: A Summary-Statistics-Based Method for Complex Trait Analysis

Samantha Gonzales^{1*}, Jane Zizhen Zhao^{2*}, Cheng Gao^{3*}, Christopher M. Wu⁴, David A. Bennett⁵, Jinying Zhao^{3†}, Xuexia Wang^{1†}

¹Department of Biostatistics, Florida International University, Miami, Florida, United States of America; ²Department of Neuroscience, University of North Carolina, Chapel Hill, North Carolina, United States of America; ³Health Informatics Institute, University of South Florida, Tampa, Florida, United States of America; ⁴Department of Physiology and Aging, University of Florida, Florida, United States of America; ⁵Rush Alzheimer's Disease Center, Rush University Medical Center

*Equal contributions

†Correspondence to Xuexia Wang - Email: xuexwang@fiu.edu

or Jinying Zhao - jinying.zhao@epi.usf.edu

Despite the success of genome-wide association studies (GWASs) in identifying variants associated with complex traits, much heritability remains unexplained. We propose the Protein-Weighted Omnibus Test (POT), a pQTL-informed gene-based association method that uses GWAS summary statistics and a Cauchy combination test to integrate burden, quadratic, and adaptive tests across multiple pQTL-derived weights. POT calculates p-values analytically, enabling scalable analysis of large genomic datasets. Simulation studies demonstrated that POT effectively controls the type I error rate and achieves

higher statistical power than the three comparison methods across most simulation settings. We applied POT to Alzheimer's disease (AD) GWAS summary statistics from Bellenguez et al. (111,326 cases and 677,663 controls) and pQTL summary statistics from Wingo et al. (2025), identifying 51 AD-associated genes at BH-adjusted p -value < 0.05 , including 24 novel genes and 27 previously reported AD genes, such as ADAM10 and PTK2B. Using ROSMAP brain proteomics data, we further evaluated associations between the identified genes and three AD neuropathological phenotypes: AD diagnosis, β -amyloid plaque burden, and PHFtau neurofibrillary tangle density with sample size as 896, 891, and 890. Twenty-eight genes, annotated to 28 proteins, were significantly associated with at least one phenotype, and 20 were associated with all three phenotypes (FDR < 0.05). SMR analysis identified four gene-protein pairs with evidence supporting potential causal roles in AD. Gene Ontology analysis revealed enrichment in signal transduction and protein-containing complex assembly pathways. These findings demonstrate that POT is a fast, powerful, and robust framework for improving genetic discovery and elucidating the genetic architecture of complex traits and diseases.

135 Multi-Ancestry GWAS of Triglyceride-to-HDL Ratio in 1.2 Million Individuals Reveals Distinct Mechanisms of Insulin Resistance and Cardiometabolic Risk

Urvashi Singh^{1*}, Pik Fang Kho², Rachel Diao³, Jiayan Zhou², Leighton Smith¹, Siqi Hu³, Sirui Zhou¹, Themistocles L. Assimes², Amit R. Majithia³, Satoshi Yoshiji^{1,4}

¹Department of Human Genetics, McGill University, Montreal, Québec, Canada; ²Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, California, United States of America; ³Division of Endocrinology, Department of Medicine, University of California San Diego School of Medicine, La Jolla, California, United States of America; ⁴Programs in Metabolism and Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America

Background: The triglyceride-to-high-density lipoprotein cholesterol ratio (TG:HDL) is a marker of insulin resistance and cardiometabolic risk, but its genetic architecture across diverse ancestries is uncharacterized.

Methods and Results: We performed a multi-ancestry genome-wide association study (GWAS) meta-analysis of TG:HDL in 1,182,592 individuals of European, African, East Asian, Admixed American, and South Asian ancestry from UK Biobank, All of Us, BioBank Japan, the Canadian Longitudinal Study on Aging, and the Million Veteran Program. We identified 1,430 independent signals across 390 loci, including 200 novel loci. Ancestry-stratified analyses showed broadly concordant effects at shared loci and identified 26 African ancestry-specific signals, including rs75405157, an intronic *HKDC1* variant with cis-eQTL evidence in African American individuals. Multi-ancestry fine-mapping improved resolution

at 448 loci, reducing the median 99% credible set size by 29% and increasing median maximum posterior inclusion probability from 0.47 to 0.62. At *BNC2*, adding non-European ancestry data narrowed the credible set from 51 to two variants. To test whether TG:HDL loci represent distinct mechanisms of insulin resistance, we clustered loci using association patterns across 33 metabolic traits and identified five mechanistic clusters, including obesity and lipodystrophy clusters. Partitioned polygenic scores showed distinct associations with 405 binary outcomes and 2,920 circulating proteins, revealing cluster-specific patterns of disease risk, including type 2 diabetes and coronary artery disease, and distinct proteomic signatures.

Conclusion: Multi-ancestry TG:HDL GWAS improves fine-mapping, identifies ancestry-specific signals, and captures mechanistic clusters with distinct clinical and proteomic profiles, supporting precise risk stratification across populations.

136 Evaluating Polygenic Score Performance Across the Menopause Transition and Demographic Characteristics in the All of Us Research Program

Adam Giacalone¹, Jack W. Staples², Samantha L. White² Meng Lin², Nikita Pozdeyev², Mary D. Sammel^{3,4}, Barbara E. Stranger^{2,5}, Nanette Santoro⁴, Audrey E. Hendricks^{1,2,3,5}

¹Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, Colorado, United States of America; ²Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, Colorado, United States of America; ³Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, Colorado, United States of America; ⁴Department of Obstetrics and Gynecology, University of Colorado School of Medicine, Aurora, Colorado, United States of America; ⁵Ludeman Family Center for Women's Health Research, University of Colorado Anschutz Medical Campus, Aurora, Colorado, United States of America

Polygenic scores (PGS) measure baseline genetic risk for a trait or condition. While PGS performance is known to vary by genetic ancestry and other contexts (e.g., age and sex), how performance varies over the menopause transition, a time of significant physiological change for women, is unknown. Here, we leverage the All of Us Research Program (AoURP) to assess 81 cardiometabolic PGS across the menopause transition and by sociodemographic strata in up to $N \approx 250,075$ women using survey or electronic health record (EHR) derived menopause information. We used three approaches to disentangle chronological age from menopause: age adjustment, subsets (< 40 , $40-60$, > 60 for pre-, peri-, post-menopausal respectively), and matching. Of the 81 PGS, 15 had incremental AUC or in the full AoURP. Of these, seven demonstrated FDR significant differential performance by menopause status with most having better performance before menopause. Four were significant in only the survey, one in

only the EHR, and two in both the survey and EHR. Age sensitivity analyses yielded consistent results, suggesting observed differences are driven by menopause status rather than age. Within demographic strata, FDR significant results were observed in 12/56 (21.4%) age quartile, 17/84 (20.2%) ancestry, 7/112 (6.3%) education, 12/28 (42.9%) ethnicity, 6/56 (10.7%) income quartile, and 11/84 (13.1%) race comparisons. These findings suggest PGS performance changes across the menopause transition and that this change varies across demographic contexts. Improved context-aware PGS models, including by menopause status, is likely to support accurate and equitable personalized healthcare for all people across the life course.

Keywords: Women's Health, Polygenic risk scores, Risk assessment, Precision Medicine

138 Estimation of Joint SNP Effects from GWAS Summary Statistics: Use of Reference Panel Genotype Frequencies

Fatemeh Yavartanoo^{1*}, Myriam Brossard², Wonil Chung¹, Shelley B. Bull^{2,3}, Andrew D. Paterson^{3,4}, Yun Joo Yoo⁵

¹Department of Statistics and Actuarial Science, Soongsil University, Seoul, South Korea; ²Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Ontario, Canada; ³Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; ⁴Program in Genetics & Genome Biology, Hospital for Sick Children Research Institute, Toronto, Ontario, Canada; ⁵Department of Mathematics Education, Seoul National University, Seoul, South Korea

Genome-wide association studies (GWAS) frequently provide marginal summary statistics, limiting joint multi-marker analyses when individual-level data are unavailable. Existing summary-statistics-based approaches reconstruct joint single nucleotide polymorphism (SNP) effects using linkage disequilibrium (LD) from a reference panel; however, these methods often approximate genotype variance under Hardy-Weinberg equilibrium (HWE), potentially introducing bias when genotype frequencies depart from HWE.

We propose a regression-based framework to reconstruct joint SNP effects and their covariance structure from GWAS summary statistics and reference panel genotype data without relying on HWE assumptions. Instead of estimating genotype variance from allele frequencies, the framework estimates the squared sum of genotypes directly from genotype frequencies in the reference panel and integrates this information with local LD structure to reconstruct joint regression coefficients.

We conducted simulation studies with 1,000 phenotype replications using BigLD-defined genomic regions from the Canadian Longitudinal Study on Aging (CLSA). Compared with a previous summary-statistics-based approach, the proposed framework demonstrated reduced bias in reconstructed joint SNP effects, with the largest improvements observed in regions exhibiting stronger departures from HWE. When genotype and

reference panels were matched, reconstructed coefficients closely matched the true regression effects.

We illustrate the framework using GWAS summary statistics for verbal numerical reasoning from the UK Biobank and European-ancestry individuals from the CLSA as the reference panel. These findings suggest that incorporating genotype-frequency information improves the accuracy of joint SNP effect estimation from GWAS summary statistics.

Keywords: GWAS summary statistics, Joint SNP effects, Hardy-Weinberg equilibrium

140 Identifying Genes Associated With Occurrence of Leg Ulcers in Sickle Cell Disease Patients

Anvita Kulshrestha^{1*}, Joan Bailey-Wilson², Allison Ashley-Koch¹, Dayne Okuhara³, Hasmin Ramirez⁴, Devan Stoeling⁵, Ashley Buscetta⁶, Cheedy Jaja⁷, Marilyn Telen⁸, Caterina Minniti⁹, Vence Bonham¹⁰

¹Duke Molecular Physiology Institute, Duke University Medical Center, Durham, North Carolina, United States of America; ²National Human Genome Research Institute, Bethesda, Maryland, United States of America; ³Booz Allen Hamilton, Lexington, Massachusetts, United States of America; ⁴Behavioral and Community Health, University of Maryland School of Public Health, College Park, Maryland, United States of America; ⁵Prevention and Community Health, The George Washington University Milken Institute School of Public Health, Washington, District of Columbia, United States of America; ⁶National Cancer Institute, Bethesda, Maryland, United States of America; ⁷College of Nursing, University of South Florida, Tampa, Florida, United States of America; ⁸Division of Hematology, Duke University Medical Center, Durham, North Carolina, United States of America; ⁹Albert Einstein College of Medicine, Bronx, New York, United States of America; ¹⁰Diaspora Human Genetics Institute, Meharry Medical College, Nashville, Tennessee, United States of America

Sickle cell disease (SCD) is a debilitating genetic disorder caused by a single point mutation in the beta-globin chain of hemoglobin. Leg ulcers, a complication of SCD, affect a significant number of patients. The pathophysiology of leg ulcers is incompletely understood, but it likely involves vaso-occlusion, impaired endothelial function, and chronic inflammation. Together, these processes result in reduced blood flow to the skin, leading to tissue damage and ulcer formation.

We conducted a gene-based rare variant (<1% minor allele frequency) burden test using EPACTS in three different SCD cohorts. Combined Multivariate and Collapsing (CMC) test was performed to analyze rare variants within a gene or region, covarying age, mean squared age, sex, and genetic principal components. Summary statistics from each of the individual cohort analyses were then meta-analyzed using Fisher's test.

The meta-analysis resulted in the identification of *TAF6* as the most significant gene, with a meta *P* value of 1.77×10^{-5} . Using a subset of genes with meta *P* values < 0.01, we conducted a

gene enrichment analysis using GOrilla. The top three gene ontology (GO) biological processes were hyaluronan catabolic process, epithelial cell proliferation, and endothelial cell proliferation, all of which are critical to angiogenesis and wound healing.

This study highlights the utility of performing a gene-burden test for successfully identifying risk genes (and rare variants) associated with our clinical outcome of interest, leg ulcers. The GO processes identified through this analysis suggest which biochemical and structural activities may be dysregulated in patients with leg ulcers and SCD.

141 Omics Mendelian Randomization Analysis for Cardio-metabolic Traits in the Middle East

Mohamed Elsharif^{1*}, Faaiz Joad¹, Abdulla Shaar¹, and Mohamad Saad¹⁺

¹*Qatar Computing Research Institute, Hamad Bin Khalifa University, Qatar*

*Presenting author

+Corresponding author

Background: Large-scale biobanks with comprehensive phenotyping and genome-wide profiling are becoming increasingly available for performing a variety of genetic analyses, including Mendelian Randomization (MR). Middle Eastern populations, despite their large size, are underrepresented in causal analysis research. Our aim is to investigate the causal relationship between metabolite and protein levels against 24 cardiometabolic traits in a Middle Eastern dataset utilizing MR.

Methods: Genetic instruments were identified using summary-level data from 783 metabolomic genome-wide association studies (mGWAS) and 1,305 proteomic GWAS (pGWAS) using 13890 participants with whole genome sequence data at the Qatar Precision Health Institute. We comprehensively analyzed 24 cardiometabolic traits, which included 21 quantitative traits and 3 binary diseases. Causal associations were evaluated using five different one-sample MR approaches using strong independent genetic instruments ($P < 5 \times 10^{-8}$, LD r^2 threshold < 0.001). Sensitivity analyses with robust one-sample MR methods were performed to account for pleiotropic effects.

Results: Our results identified several metabolites and proteins that are “causally” associated with cardiometabolic traits. Among these, we observed associations between *PSME1* protein and hydroxypalmitoyl sphingomyelin (d18:1/16:0(OH))** metabolite and low-density lipoprotein-cholesterol ($P = 2.20 \times 10^{-54}$ and $P = 3.81 \times 10^{-36}$). Additionally, IL27RA protein levels is associated with hypertension ($P = 3.71 \times 10^{-5}$). Similarly, F5 protein is associated with perturbing the ECG JT interval, which is used for evaluating patients with cardiac arrhythmias.

Conclusions: This study identified multiple plasma proteins and metabolites that may play causal roles in cardiometabolic traits for Middle Eastern populations. Future research is needed to validate our findings.

142 Multi-Omic Signatures of Hepatic Steatosis in an Under-represented Hispanic/Latino Population

Rashedeh Roshani^{1,2}, Annie Green Howard³, Jessica Sprinkles⁴, Katie Meyer⁴, Susan Fisher-Hoch⁵, Joseph B. McCormick⁵, Suzanne R. Sharpton², Jennifer E. Below^{1,2}, Kari E. North⁵, Heather M. Highland^{5*}

¹*Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America;* ²*Department of Medicine, Division of Genetic Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America;* ³*Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America;* ⁴*Department of Nutrition, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America;* ⁵*Department of Epidemiology, Border Health Research Center, School of Public Health, Brownsville Regional Campus, The University of Texas Health Science Center at Houston, Brownsville, Texas, United States of America*

Introduction: Liver diseases, including steatosis and fibrosis, affect more than 25% of United States adults and can progress to cirrhosis, hepatic carcinoma, and liver failure. Steatosis is increasingly common in individuals with excess adiposity, obesity and other metabolic conditions, such as type 2 diabetes, dyslipidemia, and hypertension. Hispanic/Latino populations have a high burden of these cardiometabolic risk factors and therefore bear a disproportionate burden of disease yet remain understudied in biomedical research.

Methods: We compared plasma proteomics (p, Olink Explore HT) and whole blood transcriptomics (t) with cross-sectional liver ultrasound (FibroScan) in the Border Health Research Cohort (BHRC) from the Texas-Mexico border. Steatosis was defined using FibroScan measures of controlled attenuation parameter (CAP) ≥ 288 dB/m ($N_p=566$, $N_t=433$) and controls CAP < 248 dB/m ($N_p=323$, $N_t=246$).

Results and discussion: In a model adjusting for age and sex, 48 transcripts and 413 proteins were significantly associated with steatosis, after FDR correction. Findings include pathways central to hepatic biology: alcohol metabolism [ADH4 (P value_p = $1.8e-27$)], amino acid metabolism [FTCD (P value_p = $6.3e-35$)], and fatty acid oxidation [CPT1A (P value_t = $1.4e-8$)]; mediators linked to steatosis such as leptin [LEP (P value_p = $1.8e-20$)], reflecting adiposity; and candidates with unestablished roles in steatosis [CISH (P value_t = $1.1e-6$)]. These markers refine our understanding of steatosis biology that can contribute to our understanding of progression to fibrosis, liver failure, or cancer.

145 Brain Single-Nucleus and Multi-Omic Integration Reveals Female-Predominant MS4A Locus for AD CSF Biomarkers

Ting-Chen Wang^{1,2*}, Jigyasha Timsina^{3,4}, Knight ADRC, EADB Consortium, EPAD-EMIF Consortia, ACE, ALFA, and MISSION-AD Study Groups, Carlos Cruchaga^{3,4,5}, Timothy J. Hohman^{1,2}, Jennifer E. Below², Logan Dumitrescu^{1,2}

¹Vanderbilt Memory and Alzheimer's Center, Vanderbilt Health, Nashville, Tennessee, United States of America; ²Vanderbilt Genetics Institute, Vanderbilt Health, Nashville, Tennessee, United States of America; ³Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, United States of America; ⁴NeuroGenomics and Informatics Center, Washington University School of Medicine, St. Louis, Missouri, United States of America; ⁵Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, Missouri, United States of America

Emerging clinical trials and research studies on Alzheimer's disease (AD) reveal prominent sex differences in disease pathology. We aim to leverage three cerebrospinal fluid (CSF) biomarkers, amyloid- β 42 (A β 42), total Tau (t-tau), and phosphorylated tau (pTau181) in 18,491 unrelated European ancestry individuals to elucidate the genetic architecture of sex-specific AD pathology.

We performed colocalization using sex-stratified eQTL summary statistics (males = 404, females = 579) for six brain cell types (astrocytes, excitatory and inhibitory neurons, microglia, oligodendrocytes, and oligodendrocyte precursor cells) from the SingleBrain atlas, alongside sex-stratified summary statistics for z-score standardized CSF protein levels from 22 harmonized cohorts (N = 6,785; 51.84% male; mean age = 68) and six summary statistics (N = 11,706; 45.27% male; mean age = 69). PP.H4 $\geq$ 0.8 defined robust colocalized signals. CSF genome-wide significant index variants were further cross-referenced in the NIAGADS FunGen-AD xQTL portal to assess additional QTL associations.

Colocalization and xQTL queries revealed tissue- and cell type-specific transcriptomic and epigenetic architectures. A female-predominant CSF t-tau index variant, rs1582763, robustly colocalized as an eQTL for MS4A4E in microglia (PP.H4 = 0.92). Non-sex-stratified xQTL analyses suggest rs1582763 extends regulatory effects across the MS4A cluster, acting as a tissue-specific methylation QTL (e.g., mQTL for MS4A6A in dorso-lateral prefrontal cortex (DLPFC) only, and for MS4A4A in DLPFC, parahippocampal gyrus, and parietal cortex).

These findings identify a robust female-predominant microglial MS4A4E regulatory signal for CSF t-tau and reveal broader brain-region-specific epigenetic architecture at the MS4A locus, highlighting sex-, tissue- and cell-type-dependent regulation in AD-related CSF biomarkers.

146 Effects of Homozygosity on Fitness-Related Traits in the Taiwan Biobank

Mohan Rakesh^{1*}, Emily Tran¹, Kelly M Burkett², Marie-Hélène Roy-Gagnon¹, Meng-Che Tsai³

¹School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada; ²Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, Canada; ³Department of Pediatrics, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

Children of related individuals are known to have reduced survival and fitness. In populations where consanguineous marriages are common, rates of birth complications, fetal death, and genetic diseases, are higher. Inbreeding effects can also be observed in biobank samples using genomic inbreeding coefficients. Runs of homozygosity (ROHs) are stretches of an individual's diploid genome where both chromosomal copies can be inferred to be identical by descent (inherited from a common ancestor), and the proportion of an individual's genome in ROHs can then be used as an inbreeding coefficient (F_{ROH}). Recent studies have shown that higher F_{ROH} is associated with deleterious effects on many fitness-related traits. In this work we use genetic data from the Taiwan Biobank (TWBv2 array: ~690,000 markers, n = 87,824 individuals) and we use the IBDLD software to identify ROHs. We find significant ($p < 0.05$) associations between F_{ROH} levels and fitness-related traits (height, bone density, LDL-C level, hemoglobin level, red blood cell count, total cholesterol). These associations are driven by individuals with a total ROH length > 50 Mb (top 1.1% of our sample and ~8x median value). We also identify certain genomic regions where the presence of ROHs is significantly ($P < 10^{-5}$) associated with change in phenotype. Some of these regions corroborate known regions from the literature and others are novel. Our results contribute to a growing body of work attempting to characterize inbreeding effects on complex traits.

147 Genetic Architecture and Disease Tracking of a Multi-Ancestry Biological Aging Clock

Alejandro Mejia-Garcia^{1,2,3,4}, Chen-Yang Su^{2,3,5}, Thomas M. Zheng^{2,3,5}, Hsuan Megan Tsao^{1,2,3}, Arthur Richard^{2,3,5}, Dylan Hamitouche⁴, Satoshi Yoshiji^{1,2,3,5,6}, Vincent Mooser^{1,2,3}, Adil Harroud^{1,3,4,7,8} and Sirui Zhou^{1,2,3,5}

¹Department of Human Genetics, McGill University, Montréal, Québec, Canada; ²Canada Excellence Research Chair in Genomic Medicine Program, McGill University, Montréal, Québec, Canada; ³Victor Phillip Dahdaleh Institute of Genomic Medicine, McGill University, Montréal, Québec, Canada; ⁴The Neuro (Montreal Neurological Institute-Hospital), Montréal, Québec, Canada; ⁵Quantitative Life Sciences, McGill University, Montréal, Québec, Canada; ⁶Programs in Metabolism and Medical & Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America; ⁷Center for Molecular Medicine, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden; ⁸Department of Neurology and Neurosurgery, McGill University, Montréal, Québec, Canada

Background: Aging is accompanied by a progressive decline in physiological function that contributes to chronic disease development. Chronological age fails to capture inter-individual variation in aging. Biological clocks estimate “biological age”, and the difference from chronological age (biological age gap, “BAG”) provides a more granular measure

of the aging processes. Current clocks are predominantly trained on Europeans using expensive omics, limiting cross-ancestry generalizability and clinical applicability. Here, we developed a practical biological age clock using routine clinical biomarkers, assessed its predictive ability for neurological disease, and investigated genetic determinants of aging across diverse populations.

Materials and Methods: Using the All of Us biobank, we trained a biological clock via Lasso regression on 22 routine blood and anthropometric biomarkers in 14,328 age- and sex-balanced individuals without neurological/psychiatric disorders. We performed a Phenome-wide association study across 434 traits (N = 152,733). We conducted a GWAS in European (N = 74,675), African (N = 22,315), and Admixed American populations (N = 19,940), fine-mapping and single-cell enrichment (1.9 M cells).

Results: Elevated BAG was significantly associated with 158 neurological phenotypes; increasing dementia (HR = 1.020) and Parkinson's risk (HR = 1.014), although reducing migraine risk (HR = 0.991). We identified ancestry-specific genetic loci for aging: HBB in African ancestry and FADS1/FADS2 in European ancestry. Single-cell enrichment revealed that BAG-associated genes in Europeans are overexpressed in double-negative (DN) T cells of older adults (65-85). Conclusion: We present a multi-ancestry biological age clock predicting neurological disease risk. Our findings uncover population-specific genetic drivers such as erythropoiesis and DN T-cell-mediated neuroinflammation

Keywords: biological aging, multi-ancestry GWAS, All of Us Biobank, heterogeneity.

148 Genomic Atlas of Cardiac and Adiposity Imaging Phenotypes

Hsuan M. Tsao^{1,2*}, Leighton Smith^{1,2}, Arthur Richard^{2,3}, Urvasi Singh^{1,2}, Masashi Hasebe^{1,2,4}, Yefeng Yang^{2,3}, Hongyu Xiao^{1,2}, Jason Flannick^{5,6}, Tianyuan Lu^{7,8}, Satoshi Yoshiji^{1,2,5}

¹Department of Human Genetics, McGill University, Montreal, Québec, Canada; ²Canada Excellence Research Chair Program in Genomic Medicine, McGill University, Montreal, Québec, Canada; ³Quantitative Life Sciences Program, McGill University, Montreal, Québec, Canada; ⁴Department of Diabetes, Endocrinology and Nutrition, Kyoto University Graduate School of Medicine, Kyoto, Japan; ⁵Programs in Medical & Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America; ⁶Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts, United States of America; ⁷Department of Population Health Sciences, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, United States of America; ⁸Department of Biostatistics and Medical Informatics, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, United States of America

Cardiac and adiposity imaging-derived phenotypes (IDPs) quantify organ structure, function, and body composition, but their shared and trait-specific genetic architecture remains

incompletely defined. We constructed a whole-genome sequencing atlas of 42 IDPs in UK Biobank, including 11 cardiac magnetic resonance imaging traits (N up to 75,562) and 31 dual-energy X-ray absorptiometry adiposity traits (N up to 66,194). Genome-wide association analyses identified 1,305 independent loci. Integrating fine-mapping, FLAMES, combined SNP-to-gene, Open Targets locus-to-gene scores, and rare-variant burden testing nominated candidate effector genes at 1,151 loci, implicating 751 unique genes. Genetic correlations were extensive within cardiac and adiposity domains but selective across domains, supporting structured pleiotropy. In 257,716 All of Us participants, IDP polygenic scores (PGS) showed distinct disease spectra. In a larger dilated cardiomyopathy (DCM) prediction analysis of 410,287 participants, including 2,665 DCM cases, adding a left ventricular (LV) imaging composite PGS to a DCM PGS improved discrimination beyond the disease PGS alone (AUC = 0.703 to 0.711; $p = 0.001$). The LV imaging composite PGS also stratified DCM risk: compared with the middle 40-60%, the top 5% had higher DCM odds (OR = 1.76; 95% CI, 1.52-2.05), whereas the bottom 5% had lower odds (OR = 0.54; 95% CI, 0.40-0.72). Clustering LV end-diastolic volume loci resolved five disease-association clusters; one enriched for cardiomyopathy and heart failure nominated *FBXO32*, a cardiomyocyte-expressed regulator of cardiac mass, as a candidate effector gene. Single-nuclei and spatial transcriptomics supported *FBXO32* prioritization. This atlas links imaging endophenotypes to effector genes and cardiometabolic disease mechanisms.

149 Identifying Density-Correlated Genes From Single-Cell-Resolution Spatial Transcriptomics With DenMark: A Bayesian Hierarchical Model

Mingchi Xu¹, Alexandra M. Schmidt¹, Qihuang Zhang^{1*}

¹Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Québec, Canada

Single-cell-resolution spatial transcriptomics (ST) profiles gene expression and preserves cell coordinates, revealing how tissue organization shapes molecular state. Local cell density varies across microenvironments and can co-vary with transcriptional programs, yet estimation of the density-expression relationship with calibrated uncertainty is lacking. Conventional spatially variable gene (SVG) methods detect spatial structure but cannot separate density-linked signal from generic heterogeneity. We introduce a unified framework that jointly models cell locations and gene expression as a marked point process. Cell positions follow an inhomogeneous point process whose intensity captures where crowding occurs. Expression at each cell is a mark whose mean shares latent spatial components with that intensity. Spatial variation is decomposed into a component shared between density and expression and a gene-specific residual, separating density-driven effects from other heterogeneity and supporting formal testing for density-correlated genes (DCGs) with calibrated uncertainty. A low-rank Hilbert-space Gaussian process (GP) approximation scales the model to millions of transcripts and hundreds of thousands of cells. In simulations, the estimator recovers density-expression coupling accurately, yields well-

calibrated intervals, and runs orders of magnitude faster than full-GP baselines. On multiplexed error-robust fluorescence in situ hybridization (MERFISH) mouse brain data, it recovers known density-correlated markers and detects signals missed by SVG pipelines. On a 10x Xenium breast cancer dataset, it identifies density-associated programs tracking immune infiltration. By making density a covariate of transcriptional state, DenMark provides a principled lens on how tissue architecture relates to gene programs in development and disease.

Keywords: spatial transcriptomics, cell density, spatially variable genes, Bayesian hierarchical model, marked point process

150 VACANT: A Tool for Annotation-Stratified Rare Variant Association Analysis and Effect Size Estimation

Shu-Hsien Cho^{1,2*}, Yu Yao², Ryan Bohlender², Chad Huff²

¹The University of Texas MD Anderson Cancer Center UTHealth Houston Graduate School of Biomedical Sciences, Houston, Texas, United States of America; ²Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America

Functional annotation is increasingly recognized for improving the power to detect disease associations in rare variant association studies, but the majority of the rare variant association approaches underutilize this information and provide no within-gene effect estimates. Also, estimation of disease risk conferred by a previously unobserved variant, for instance, a novel variant identified from clinical testing, is a critical need that is difficult to address with existing functionally informed rare variant association tests.

VACANT (Variant Annotation Clustering Association Test) leverages functional annotation by applying Gaussian mixture modeling to cluster rare variants (MAF < 0.5%) into severity tiers from one or more continuous scores, such as CADD, then fits a joint Firth regression with a penalized likelihood ratio test for gene-level significance. The same model reports each tier's variant count, score range, carrier frequency, and OR with 95% CI. Tier boundaries are cohort-independent, enabling tier-level meta-analysis, and a prediction mode assigns any new variant to a tier from its score. VACANT is implemented in R with a command-line interface for high-performance computing.

In the UK Biobank 200,000 whole-exome sequencing cohort across five breast cancer genes, VACANT achieved 17 to 38 percentage-point higher power than STAAR-O at genome-wide significance. For BRCA2, VACANT selected three severity tiers: OR ranged from 1.05 (0.95–1.17) among 477 low-score variants (score 0–5; 6.8% case carriers) to 7.39 (5.27–10.45) among 69 high-score variants (score 20–67; 1.2% case carriers). These per-tier estimates give each variant, including one from clinical testing, an interpretable risk that a gene-level association cannot provide.

Keywords: Rare variant association, functional annotation, per-tier odds ratio, biobank-scale case-control

151 Refined Phenotyping and Cross-Biobank GWAS Meta-Analysis Reveals New Genetic Insights into Multisite Chronic Pain

Amandeep Kaur^{1,2*}, Etienne Kabongo^{2,4}, Marc Parisien^{2,3}, Audrey V. Grant^{2,3,4}

¹Department of Human Genetics, McGill University, Québec, Canada; ²Alan Edwards Centre for Research on Pain, McGill University, Québec, Canada; ³Department of Dental Medicine and Oral Health Sciences, McGill University, Québec, Canada; ⁴Department of Anesthesia, McGill University, Québec, Canada

Chronic pain (CP) is the leading cause of disability globally, affecting ~1.5 billion people with associated costs exceeding \$600 billion annually. The most heritable CP trait, multisite chronic pain-binary (MCP-B) is defined as persistent pain at ≥ 2 body sites for 3+ months, capturing the systemic nature and genetic overlap of CP conditions. Despite implicating brain cell types and axonal guidance as key biological processes, prior MCP GWAS locus discovery has remained limited by single-biobank analyses.

We developed MCP-B as a harmonized phenotype across individuals of European ancestry in the UK biobank (UKB), Canadian Longitudinal Study of Aging (CLSA), CARTaGENE (CaG), and All of Us (AoU). This definition was refined using multimodal data across multiple time points, including pain medication use and comorbid conditions like depression, to improve discovery of genetic risk factors. Genetic correlation using LD score regression (LDSC) revealed $r_g = 68\%$ (P value = 6×10^{-8} , SE = 0.12) between CLSA ($h^2 = 17\%$) and UKB ($h^2 = 12\%$), exceeding previous reports of CP traits. Further, GWA meta-analysis using the inverse-variance weighted method identified 41% additional independent loci ($r^2 < 0.2$), including 6 independent significant SNP's not previously reported. The tissue-expression analysis and gene-set analysis revealed significant pituitary enrichment alongside brain tissues highlighting neuroendocrine and synaptic mechanisms in CP pathophysiology.

Our findings suggest that harmonized MCP-B phenotype across diverse biobanks substantially enhances genetic discovery, with higher potential to identify likely causal variants through fine-mapping. These results provide a scalable framework for advancing precision-based diagnostic and therapeutic strategies for CP globally.

152 BRIDGE: A New Epigenetic Clock for Mid-to-Late Life

Denitsa Vasileva^{1*}, Celia M. T. Greenwood², Denise Daley¹

¹Center for Heart Lung Innovation, Faculty of Medicine, University of British Columbia, Vancouver, Canada; ²Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Canada

Introduction: The epigenetic clock leverages DNA methylation to calculate epigenetic age (EA) and may be a biomarker for age-related conditions. Available clocks (e.g., Centenarian, Horvath, and Skin & Blood) underestimate EA (i.e.,

EA < chronological age (CA)) in older adults, causing false phenotypic associations.

Objective: To build a clock using older samples and flexible algorithms to mitigate underestimation.

Methods: Methylation was assayed by Illumina arrays in 4376 blood samples from the Canadian Longitudinal Study on Aging (CLSA), Alzheimer's Disease Neuroimaging Initiative (ADNI), and four GEO datasets. The CLSA and GEO samples were split into training (1976, range: 30-101 years) and validation sets (495, range: 31-103); ADNI was an independent test set (1905, range: 55-96). BorutaSHAP selected 600 CpGs that were used to build 11 linear and non-linear models. For each, median absolute error (AE = |EA-CA|) was calculated in the validation set. The most accurate clock was applied to ADNI and its AE and error (EA-CA) compared to the centenarian, Horvath and Skin & Blood clocks. Error < 0 indicated underestimation.

Results: Ridge regression yielded the clock with the lowest median AE (2.17 years) in the validation set. The resulting BRIDGE (“**B**oruta+**R**idge”) clock outperformed (median AE = 2.82 years) existing clocks in ADNI (Centenarian: 6.35, Horvath: 7.15, Skin & Blood: 5.29) and showed the least underestimation bias in the oldest decade (86-96 years; median error=0.22 years).

Conclusions: The BRIDGE clock has greater accuracy and lower bias at ~50-90 years and could be a more reliable biomarker at this life stage.

Keywords: Epigenetics, DNA methylation, Machine learning algorithms.

153 Can Generative AI Learn Gene Expression Like Language? A framework for Self-Supervised Representation Learning for Cancer Genomics

SeyedMohsen Hosseini¹, Divya Sharma^{1,2,3}

¹Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada; ²Department of Biostatistics, Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada; ³Department of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

Generative AI models are rapidly gaining popularity in biology, with growing applications in protein design, single-cell genomics, drug discovery, and biomedical language modeling. Yet bulk RNA sequencing, despite remaining a cornerstone of translational genomics, has received comparatively less attention. Inspired by the way large language models like ChatGPT learn meaningful structure from sequences of words, we developed a self-supervised generative AI framework that learns the “language” of gene expression from bulk RNA-seq profiles.

Our approach converts each unordered expression profile into a sample-specific gene sequence using term frequency-inverse document frequency ranking, prioritizing genes highly

expressed within a sample and selectively expressed across the cohort. A transformer model is then trained to predict masked genes from their transcriptomic context, learning inter-gene relationships without reconstructing raw expression values. This reduces dependence on expression scale, binning, pre-defined gene panels, or external biological embeddings.

We trained and evaluated the model using 9,201 TCGA Pan-Cancer samples spanning 33 cancer types, five RNA-seq representations, and 10,000 genes. On test samples, the model achieved 90.83% cancer-type classification accuracy and 0.996 Area Under the Curve. The learned representations also captured pathway-level biology, predicting 1,387 PARADIGM pathway activities with mean sample-wise and pathway-wise Pearson correlations of 0.754 and 0.762 respectively. In an independent evaluation on 500 GTEx healthy tissue samples across 27 tissue types, the model preserved tissue-level genomic organization without retraining.

These findings suggest that self-supervised generative AI can learn reusable, normalization-robust representations of bulk RNA-seq data for genomic discovery and precision medicine.

Keywords: Generative AI; Cancer genomics; Bulk RNA-sequencing; Pan-cancer classification; Transformer model

155 Genetic Similarity Among 178 Disease Phenotypes Predicts Therapeutic and Side Effects for 1,711 Drugs

Panagiotis Nikolaos Lalagkas², Rachel Melamed¹

¹Department of Biological Sciences, University of Massachusetts, Lowell, Massachusetts, United States of America

Human genetics demonstrates great potential for drug discovery, but challenges in identifying causal genes limit its clinical translation. Pleiotropy, the phenomenon where genetic variants or genes influence multiple traits, has been previously used to identify drug targets shared between phenotypically similar diseases. Here, we expand the use of pleiotropy to develop and evaluate a gene-agnostic method that predicts novel drug therapeutic and side effects across the phenome. We hypothesize that diseases with high genetic similarity to a drug's known indications can point to new drug uses. To test this, we develop five metrics to quantify the genetic similarity between pairs of 178 diseases integrating genome-wide genetic correlation, gene-level associations, and tissue-specific gene regulation. Comparing these metrics with data on indications and side effects of 1,711 common drugs, we find that more genetically similar diseases tend to share more drugs. We then use genetic similarity between phenotypically dissimilar to predict drug therapeutic effects using a Bayesian logistic regression model: our predictions with probability > 0.1 show a 2.03-fold increased likelihood of progressing from Phase I clinical trials to regulatory approval. As well, predictions with probability > 0.2 are 1.42 times more likely to correspond to true adverse effects. Notably, the indications model predicts side effects better than expected by chance, and vice-versa, implying a shared genetic basis for therapeutic and adverse drug effects. Together, our

results suggest that genetic similarity can reveal new drug-disease links, putting forward a new use of genetics that bypasses the need for disease and drug target identification.

156 Coxformer-Surv: Transformer-based Survival Modelling of Longitudinal Health Records for Calibrated Multi-disease Risk Prediction

Lap Sum Chan¹, Nilanjan Chatterjee¹

¹Johns Hopkins University, Baltimore, Maryland, United States of America

Longitudinal electronic health records contain temporally ordered diagnoses, procedures, measurements, and risk factors that may improve disease risk prediction beyond static baseline models. However, many transformer-based health models are optimized for sequence or next-event prediction rather than fixed-horizon incident disease risk. We developed Coxformer-Surv, a transformer-based survival model that maps age-ordered health histories to dynamic subject representations and combines them with disease-specific piecewise-exponential hazards on the attained-age scale. We evaluated Coxformer-Surv in held-out UK Biobank participants using the current all-token, 1-year age-bin v4 model selected by validation-based calibration-first criteria. For each disease and prediction horizon, prevalent cases at the prediction/index age were excluded, and hazard-derived finite-horizon probabilities were evaluated for incident disease prediction.

In preliminary analyses of five diseases—prostate cancer, breast cancer, colorectal cancer, coronary artery disease (CAD), and type 2 diabetes (T2D)—Coxformer-Surv showed strong discrimination across 2-, 5-, and 10-year horizons. Five-year area under the receiver operating characteristic curve (AUC) values were 0.825 for prostate cancer, 0.835 for breast cancer, 0.796 for colorectal cancer, 0.803 for CAD, and 0.823 for T2D. Corresponding 2-year AUCs were 0.844, 0.851, 0.819, 0.825, and 0.851, respectively; 10-year AUCs were 0.791, 0.793, 0.746, 0.746, and 0.764. These results support the feasibility of using transformer-based survival models for multi-disease fixed-horizon risk prediction from longitudinal health histories. Future analyses will evaluate external generalization in All of Us and assess calibration across ancestry, age, and clinical subgroups.

157 Novel Alzheimer's Disease-Associated Variants and Genetic Interactions Identified from UK Biobank Whole-Exome Sequencing Data Using IBI-DT

Jin Ren¹, Rui Yin², Jinying Zhao³, Jinling Liu^{1*}

¹Department of Epidemiology, University of Florida, Gainesville, Florida, United States of America; ²Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, Florida, United States of America; ³Health Informatics Institute, University of South Florida, Tampa, Florida, United States of America

Although genome-wide association studies (GWAS) have revealed many Alzheimer's disease (AD) loci, population-level marginal testing remains poorly suited for detecting rare, heterogeneous, and interaction-related genetic signals. We applied Individualized Bayesian Inference-Decision Tree (IBI-DT), a Bayesian ensemble tree method that prioritizes subgroup-specific variants and hierarchical interaction patterns, to whole-exome sequencing (WES) data from 8,292 unrelated White British participants in UK Biobank. Fisher exact test-based GWAS was used as a marginal-testing comparator. Replication was evaluated in an independent ADSP cohort of 1,560 unrelated individuals.

IBI-DT identified 178 empirically significant variants mapping to 173 genes, compared with 16 significant GWAS variants mapping to four genes. The gain was not simply an expansion of correlated signals: IBI-DT prioritized lower-frequency variants and less redundant linkage patterns than GWAS. The top IBI-DT results recovered established AD biology while highlighting additional genes with plausible neurodegenerative relevance. Cross-cohort evaluation supported three variants and four genes, including *KIF14* and *ZNF90*, which were replicated at the gene level only by IBI-DT. IBI-DT also detected gene-gene and gene-environment interaction patterns enriched in biologically relevant pathways, supporting its ability to move beyond one-variant-at-a-time discovery. In prediction analyses, neural-network models using IBI-DT-prioritized variants outperformed GWAS-based variant sets for AD status classification (area under the receiver operating characteristic curve, 0.67 vs. 0.63).

These results demonstrate that IBI-DT can complement GWAS by recovering known AD signals while identifying rare, less redundant, and interaction-related variants with replication and predictive support.

159 A Multi-Ancestry Optimization of Polygenic Risk Prediction for Type 1 Diabetes

Wiame Belbellaj¹, Brandon Jansonius^{1,2}, Despoina Manousaki^{1,2,3}

¹Research Center of the Sainte-Justine University Hospital, Montreal, Québec, Canada; ²Department of Biochemistry and Molecular Medicine, Université de Montréal, Montreal, Québec, Canada; ³Departments of Pediatrics, Université de Montréal, Montreal, Québec, Canada

Background: Type 1 Diabetes (T1D) is a heritable autoimmune disease for which polygenic risk scores (PRS) can identify high-risk individuals before onset, allowing preventive therapies. However, most PRSs are based on European populations, limiting their portability across diverse populations. Although efforts like the TA-PS improve trans-ancestry prediction, their HLA component remains European-derived. Here, we improve T1D risk prediction by integrating a multi-ancestry HLA score (HLA-GRS) into the TA-PS framework.

Methods: We used a case-control T1D cohort of 4,522 individuals from five ancestries. Data were split into training (60%), testing (20%), and validation (20%) sets to preserve ancestry representation and avoid overfitting. We evaluated three integration approaches: an unweighted additive model

(HLA-GRS + TA-PS), logistic regression weighting, and elastic net regularization. Predictive performance was assessed using receiver area under the receiver operating curve (AUC), and DeLong's test for pairwise model comparisons.

Results: The unweighted additive model showed an AUC of 0.872 (95% CI: 0.849–0.895), comparable to elastic net and logistic regression. All integrated approaches significantly outperformed existing PRSs, including the European-based GRS2x (AUC = 0.821, 95% CI: 0.793–0.848) and TA-PS (AUC = 0.858, 95% CI: 0.834–0.882). Compared with GRS-HLA alone (AUC = 0.861, 95% CI: 0.837–0.885), the integrated additive model yielded a modest but statistically significant improvement ($\Delta\text{AUC} = 0.0113$, $p = 0.0135$). Similarly, significant gains were observed relative to GRS2x ($\Delta\text{AUC} = 0.0516$, $p = 3.5 \times 10^{-5}$) and TA-PS ($\Delta\text{AUC} = 0.0548$, $p = 1.2 \times 10^{-5}$). Validation in independent multi-ancestry cohorts is pending.

Conclusion: Integrating multi-ancestry HLA and non-HLA PRS components improves T1D risk prediction across diverse populations.

Keywords Polygenic risk score, Type 1 Diabetes, Multi-ancestry Genetics, HLA Region, Screening

160 A Novel Statistical Method for Unmasking Sex-Specific Genomics Signatures in Complex Traits

Samaneh Mansouri^{1,2,3}, Mélissa Rochette³, Qingrun Zhang^{4,5,6}, Ting-Huei Chen³

¹Université de Sherbrooke Faculté de Médecine et des Sciences de la Santé, Québec, Canada; ²Department of Social and Preventive Medicine, Faculty of Medicine, Université Laval, Québec City, Québec, Canada; ³Department of Mathematics and Statistics, Université Laval, Québec City, Québec, Canada; ⁴Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada ⁵Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta, Canada; ⁶Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada

Correspondence: Ting-Huei Chen (ting-huei.chen@mat.ulaval.ca)

Genotype–phenotype association studies have advanced our understanding of complex traits but often overlook sex-specific genetic signals. The growing awareness of sex-specific influences on human traits and diseases necessitates tailored statistical methodologies to dissect these genetic intricacies. Rare genetic variants play a significant role in disease development, often exhibiting stronger per-allele effects than common variants. In sex-dimorphic analysis, traits are viewed as having two sex-specific subsets rather than being uniformly defined. Existing methods for gene-based analysis of rare variants across multiple traits can identify shared genetic signals but cannot reveal the specific subsets from which significant signals originate. This means that when a significant signal is detected, it remains unclear whether it arises from the male samples,

female samples, or both. To address this limitation, we propose SubsetRV, a new methodology capable of identifying genes associated with specific traits or diseases in males, females, or both. SubsetRV can also be applied to broader applications in multiple traits analysis. Simulation studies have demonstrated SubsetRV's reliability, and real data analysis on bipolar disorder and schizophrenia has revealed potential sex-specific genetic signals. SubsetRV offers a valuable tool for identifying sex-specific genetic candidates, aiding in understanding disease mechanisms. An R package for SubsetRV is available on GitHub. It can be accessed directly through this link: <https://github.com/Mansouri-S/SubsetRV>.

Keywords: rare variants, sex-specific genetic factors, SKAT statistics, subset-based analysis, SubsetRV

161 Cell-type-aware Transcriptome-wide Association Studies Using Paired Bulk Gene Expression and Digital Pathology

Nicholas B. Larson*, Shannon K. McDonnell, Yijun Tian, Liang Wang

¹Department of Quantitative Health Sciences, Mayo Clinic College of Medicine and Science, Rochester, Minnesota, United States of America; ²Department of Tumor Microenvironment and Metastasis, Moffitt Cancer Center, Tampa, Florida, United States of America

Transcriptome-wide association studies (TWAS) integrate GWAS summary statistics with genetically regulated expression (GREX) prediction models to identify trait-associated genes. However, bulk tissue expression profiles represent heterogeneous mixtures of multiple cell types, and both gene expression and genetic regulation may vary substantially across cellular contexts. Consequently, latent cellular heterogeneity may reduce the accuracy and interpretability of conventional TWAS. Recent approaches attempt to address these limitations using single-cell RNA-seq data. However, single-cell assays are sparse, require inferred cell-state classification via clustering, and remain costly to generate for large-scale applications. Advances in digital pathology enable cost-effective cell segmentation and classification from whole-slide tissue images (WSIs). Here, we combine bulk tissue gene expression profiling with paired digital pathology data to perform cell-type-specific ct-GREX model training using sparse group LASSO regression for corresponding TWAS analyses (ct-TWAS). Using $n = 471$ normal prostate tissue samples with paired WSIs, we performed nuclear segmentation to derive cellular proportions for three primary cell-type classes: epithelial, stromal, and immune-related. We fit both conventional GREX models and cell-specific ct-GREX models for the three cell types of interest. We then performed respective TWAS and epithelial-specific ct-TWAS analyses for prostate cancer risk. Comparable numbers of genes had successfully trained models by either approach, and TWAS results were broadly consistent among ~8000 overlapping genes (TWAS Z-score Pearson $R = 0.82$). However, we identified several significant results unique to both analyses, including multiple findings from ct-TWAS with direct cancer relevance. These results support usage of digital pathology as a scalable alternative to single-cell data for cell-type aware TWAS analyses.

162 Genomic Analysis Identifies Polygenic and Region-specific Contributions to ADHD-migraine Comorbidity

Yaxin Luo^{1,2}, Christina Dardani^{1,2,3,4}, Robyn E Wootton^{2,3,4,5}, Evie Stergiakouli^{1,2}

¹Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom; ²Medical Research Council Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Bristol, United Kingdom; ³Research Department, Lovisenberg Diaconal Hospital, Oslo, Norway; ⁴PsychGen Centre for Genetic Epidemiology and Mental Health, Norwegian Institute of Public Health, Oslo, Norway; ⁵School of Psychological Science, University of Bristol, Bristol, United Kingdom

Attention-deficit/hyperactivity disorder (ADHD) and migraine frequently co-occur, yet their shared genetic architecture remains unclear. Using genome-wide association summary data of ADHD and migraine, we applied a multi-layered genetic analysis framework. Conjunctional false discovery rate was used to identify shared pleiotropic variants. Local genetic correlation was performed within semi-independent regions. Mendelian randomization (MR) used eQTLs as instruments to assess overlap of genetically predicted gene expression on ADHD and migraine in relevant brain and blood tissues. Colocalization analyses were conducted to assess whether shared association signals were driven by the same underlying variants. We identified 27 pleiotropic variants shared between ADHD and migraine, 20 with concordant effect directions. Local genetic correlation analysis identified a single shared region on chromosome 11 with evidence of local heritability for both ADHD and migraine and a positive local genetic correlation. Cis-eQTL MR of druggable genes identified multiple genes with evidence for causal effects of their expression on both ADHD and migraine across brain cortex and blood. Genetically predicted *MANBA* expression showed consistent associations with both traits in brain cortex and blood. Colocalization for *MANBA* supported a shared causal variant in cortex but not in blood, suggesting tissue-specific mechanisms. Current findings provide evidence for shared genetic architecture between ADHD and migraine across variant, regional, and gene-expression levels. Among our findings, genetically predicted expression of *MANBA* in the brain cortex appeared to be a potential shared biological contributor to ADHD and migraine.

163 Using the Genotype-Phenotype Map (GPMaP) to Uncover Shared Skin Biology and Prioritize Drug Targets

Lavinia Paternoster¹, Katie Watts¹, Andrew R. Elmore^{1,2}, Aimee L. Hanson¹, Genevieve M. Leyden¹, James Johnson¹, George Davey Smith¹, Tom R Gaunt¹, Gibran Hemani¹

¹MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; ²NIHR Bristol Biomedical Research Centre, University Hospitals Bristol and Weston, NHS Foundation Trust and University of Bristol, Bristol, United Kingdom

Mapping pleiotropic genetic variation on biological processes and complex phenotypes is key to extracting translational insight from genome-wide association studies (GWAS). The

Genotype-Phenotype Map (GPMaP) maps pleiotropy via genetic colocalization of 15,997 complex traits and 2.7 million molecular measurements. We use GPMaP to identify shared biological mechanisms between skin diseases and likely safety profiles of drug targets.

GPMaP identified 97,393 colocalization groups (including 1512 skin disease loci). Pleiotropy was widespread: 81% of GWAS loci colocalized with at least one other trait (56% with a molecular trait). We identified 11 shared signals between psoriasis and eczema, conditions that rarely co-occur and are typically driven by distinct immune pathways.

One shared locus strongly colocalized with *IL7R* splicing QTLs and *IL7R* protein levels (psoriasis colocalizing posterior probability=0.86). *IL7R* acts upstream of JAK signalling, consistent with the efficacy of JAK inhibitors in both conditions. However, *IL7R* showed high pleiotropy (colocalization with 10 trait categories), suggesting it may not be an optimal therapeutic target, due to potential off-target effects. Another shared signal, with opposing effects between conditions, involved *IL13*. IL-13 signalling is inhibited by dupilumab, an approved eczema drug. Psoriasis-like inflammation has been reported as a side effect, consistent with this finding.

Other genes at shared loci also exhibited high pleiotropy, indicating they may be less suitable as direct targets, but could highlight pathways containing safer therapeutic opportunities.

In the Skin Genetics Consortium, we are generating GWAS summary statistics across all common skin conditions (total N ~ 5 million). Through investigation of shared biology and pleiotropy, we aim to support drug target prioritization in areas of unmet clinical need.