STATISTICAL METHODS FOR GENETICS & GENOMICS (SMGG)
RESEARCH SEMINAR AND JOURNAL CLUB 2024-2025

(Co-ordinated with CANSSI STAGE International Speaker Seminar)

TIME and PLACE:
Friday (In Person/OnLine)
Location: Room 5-102, 60 Murray St.
(near the elevator, Prosserman Centre for Population Health Research)

September 27 9:30 am – Organizational Meeting re topics & themes for the Seminar/Journal Club this academic year

October 4 12 noon – CANSSI STAGE International Speaker Seminar

Speaker: Genevieve Wojcik, Johns Hopkins Bloomberg School of Public Health
Assistant Professor of Epidemiology

Global Diversity, Local Contexts: An Epidemiological Lens to Modeling Ancestry and Environment for Genetic Risk

https://canssiontario.utoronto.ca/event/stage-isss-genevieve-wojcik/ ** live-stream seminar **

Abstract: While there is a growing body of literature demonstrating the potential utility of polygenic scores in both precision medicine and public health genetics, there remains a lack of representation across genetic diversity that in turn limits their downstream benefits. This paucity of research in diverse groups permeates and compounds at each stage of genetic research from recruitment to discovery to translation to implementation. In this talk, I will highlight necessary considerations in addressing this gap of knowledge, specifically drawing upon examples examining genetic risk within Hispanic/Latino participants of the Population Architecture using Genomics and Epidemiology (PAGE) Study. Heterogeneity in both genetics and environment underscores the need to carefully model these interconnected factors for better models of human health with the ultimate goal of health equity.

Background Reading:

Wojcik, G.L., Graff, M., Nishimura, K.K. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* 570, 514–518 (2019). https://doi.org/10.1038/s41586-019-1310-4 https://www-nature-com.myaccess.library.utoronto.ca/articles/s41586-019-1310-4

Caliebe A et al. Including diverse and admixed populations in genetic epidemiology research, *Genetic Epidemiology* 46(7), 347-371 (October 2022). https://doi.org/10.1002/gepi.22492 https://onlinelibrary.wiley.com/doi/10.1002/gepi.22492

Raven-Adams MC, Hernandez-Boussard T, Joly Y, Knoppers BM, Chandrasekharan S, Thorogood A, Kumuthini J, Ho CWL, Gonzlez A, Nelson SC, Bombard Y, Thaldar D, Liu H, Costa A, Muralidharan V, Henriques S, Nasir J, Lumaka A, Kaiser B, Jamuar SS, Lewis ACF. Defining and pursing diversity in human genetics studies. Nat Genet. 2024 Sep 9. doi: 10.1038/s41588-024-01903-7. Online ahead of print. PMID: 39251787 https://www-nature-com.myaccess.library.utoronto.ca/articles/s41588-024-01903-7

October 11 9:30 am – Trainee Discussion: Intro to Human Genetics and GWAS

Resources:

Intro to Advanced Human Genetics (slides ADP)

GWAS: How molecular technology, biology, epidemiology, & statistical genetics came together (slides SBB)

"How to do a GWAS"
Sugolov A, Emmenegger E, Paterson AD, Sun L. Statistical Learning of Large-Scale Genetic Data: How to Run a Genome-Wide Association Study of Gene-Expression Data Using the 1000 Genomes Project Data. Stat Biosci. 2024;16(1):250-264. doi: 10.1007/s12561-023-09375-9. Epub 2023 Jul 1. PMID: 38495080; PMCID: PMC10940486.
https://pubmed.ncbi.nlm.nih.gov/38495080/

October 18 *No Seminar* – Reading Week

October 25 10 am – Seminar/Journal Club - Michael Wainberg, LTRI

Topic: **Mining the UK Biobank for novel coding-variant associations with brain diseases**

Abstract: The UK Biobank is one of the world's best-studied biomedical cohorts, featured in nearly 10,000 PubMed publications. So it seems surprising there would be much left for researchers to find – especially in the realm of genetic association studies, the raison d'être of the Biobank. In this talk, I will showcase two projects using the UK Biobank to find novel coding-variant associations with brain diseases. The first involves exome-wide association studies of 40 electronic health record-defined brain diseases and 12 quantitative brain-related traits, using carefully curated phenotype definitions and an innovative multiple testing correction procedure based on Galwey's method to boost power. The second involves using the probability of loss-of-function intolerance (pLI) score, which quantifies a gene's intolerance to mutation, to find non-recurrent copy number variants strongly associated with psychiatric disorders. I will also discuss implications of the second study on GWAS causal gene prioritization.

Background Reading: Backman, Li, Marcketta. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021). https://doi.org/10.1038/s41586-021-04103-z (*https://www.nature.com/articles/s41586-021-04103-z*)

Wainberg M, Merico D, Huguet G, *et al*. Deletion of Loss-of-Function–Intolerant Genes and Risk of 5 Psychiatric Disorders. *JAMA Psychiatry*. 2022;79(1):78–81. doi:10.1001/jamapsychiatry.2021.3211*(https://jamanetwork.com/journals/jamapsychiatry/article-abstract/2786543).*

November 1 9:30 am – Trainee Discussion: Intro to Human Genetics and GWAS (continued)

November 8 *No Seminar* **IGES November 3-5, ASHG November 5-9 **

November 15 12 noon – CANSSI STAGE International Speaker Seminar

Speaker: Elizabeth Atkinson, Baylor College of Medicine
Assistant Professor, Molecular and Human Genetics

Empowering Gene Discovery and Accelerating Clinical Translation for Diverse Admixed Populations

https://canssiontario.utoronto.ca/event/isss-elizabeth-atkinson/

Abstract: Genetic studies offer a promising basis for understanding pathophysiology and identifying new molecular targets for medicines. However, due to the lack of methodological approaches that can account for their genomic complexity, diverse admixed populations are systematically excluded from genetic epidemiology studies. Admixed populations, including many individuals who self-identify as African Americans and Hispanic/Latinos, make up more than a third of the US populace, yet face health disparities in part due to being so sorely underrepresented in research. To reap full and equitable benefits from existing and ongoing efforts to recruit more representative cohorts, there is a pressing unmet need for the development of tools permitting the study of complex traits in admixed peoples. Here, we present our novel statistical frameworks that facilitate the inclusion of admixed individuals in association studies by incorporating fine-scale consideration of genetic ancestry. We find that incorporating local ancestry into GWAS via our method, Tractor, confers several benefits beyond enabling the well-calibrated inclusion of admixed samples, including boosting power to discover ancestry-specific loci, improving the resolution of association signals, generating reliable ancestry-specific effect size estimates and p values, and identifying novel ancestry-specific associations missed by standard GWAS procedures. In conclusion, our efforts fill a gap in existing resources and will improve our understanding of complex diseases across diverse understudied populations.

November 22 9:30 am – Seminar – Highlights from IGES/ASHG
Presenters: Jianhui Gao, Eric Sanders, Henry Lu, Andrew Paterson

December 13 9:30 am – Trainee Discussion (All welcome)

Reading: Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. Verma et al., Science 385, eadj1182 (2024) 19 July 2024

https://www-science-org.myaccess.library.utoronto.ca/doi/pdf/10.1126/science.adj1182
https://www.science.org/doi/10.1126/science.adj1182

********** **2025** **********

* First week of classes January 5 -10

January 10 12 noon – CANSSI STAGE International Speaker Seminar

Speaker: Josée Dupuis, McGill University
Professor and Chair, Epidemiology, Biostatistics & Occupational Health
https://canssiontario.utoronto.ca/event/isss-josee-dupuis/

Exploiting Family History Information to Detect Rare Variant Associations

Abstract: The growing availability of sequencing data has enabled the investigation of the role of rare variants in disease etiology. However, detecting associations with rare variants or groups of rare variants requires large sample sizes for adequate power, especially for late-onset diseases, when the number of cases in cohorts of younger participants may be low. Family history (FH) contains information on the disease status of relatives, adding valuable information about the probands' health problems and risk of diseases. Incorporating data from FH is a cost-effective way to improve statistical evidence in genetic studies and overcome limitations in study designs with insufficient cases. We proposed a family history aggregation unit-based test (FHAT) and optimal FHAT (FHAT-O) to exploit available FH for rare variant association analysis. We also proposed a robust version of FHAT and FHAT-O for unbalanced case-control designs. By applying FHAT and FHAT-O to the analysis of all-cause dementia and hypertension using the exome sequencing data from the UK Biobank, we show that our methods can improve significance for known regions.

January 24 10 am – Journal Club – Michela Marcellino, Public Health

Readings: Bai, H., Zhang, X., Bush, W.S. (2023). Pharmacogenomic and Statistical Analysis. In: Fridley, B., Wang, X. (eds) Statistical Genomics. Methods in Molecular Biology, vol 2629. https://doi-org.myaccess.library.utoronto.ca/10.1007/978-1-0716-2986-4_14

Ingelman-Sundberg, Pirmohamed (2024) Precision medicine in cardiovascular therapeutics: Evaluating the role of pharmacogenetic analysis prior to drug treatment. https://doi.org/10.1111/joim.13772

January 31 *No Seminar*

February 7 12 noon – CANSSI STAGE International Speaker Seminar

Speaker: Jonathan Marchini, Regeneron Genetics Center
Head, Statistical Genomics and Machine Learning

Statistical Methods for Large Scale Genetic Association Studies

https://canssiontario.utoronto.ca/event/isss-jonathan-marchini/

Abstract: The study of rare genetic variation, which can be important in the development of complex diseases, has been increasingly carried out thanks to advances in sequencing technologies. The inherent challenges posed by the rarity of these variants and the need for large sample sizes have required the development of gene-based tests.

These tests, offering enhanced statistical power over single variant tests, aggregate information across multiple variants and can integrate external functional annotations to improve power of rare variant analysis. We will describe several existing and novel features in the association tool REGENIE and related programs which have been developed at the Regeneron Genetic Center (RGC) to carry out analyses of over 2 million exome-sequenced and genotyped individuals across a diverse set of cohorts with many thousands of phenotypes. We highlight the power of meta-analysis in genetic studies; this involves combining information across studies without requiring access to individual level data. It can be performed using summary statistics from a genome-wide scan of individual variants, or from gene-based tests that aggregate variants within a gene. The former approach is effective at identifying common variants with modest effects, while the latter boosts power for detecting rare variant associations. In this vein, we showcase REMETA, a tool designed for the efficient meta-analysis of gene-based tests in rare variant studies suitable for biobank-scale data sets. REMETA amalgamates results from multiple studies, enhancing the statistical power and reliability of the findings. We demonstrate the usefulness of these approaches for rare variant association testing through large-scale data applications.

February 21 *No seminar* – Winter Break (Reading) Week

February 28 10 am – Journal Club – Beiwen Wu, Epidemiology
Topic: Introduction to Mendelian Randomization

Reading: Davies, Holmes, Davey Smith (2018). Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018 Jul 12;362:k601. doi: 10.1136/bmj.k601. PMID: 30002074; PMCID: PMC6041728. https://pubmed.ncbi.nlm.nih.gov/30002074/

Burgess, Thompson (2021). Mendelian Randomization: Methods for Causal Inference Using Genetic Variants. Second Edition. Chapman & Hall/CRC Interdisciplinary Statistics. https://www-taylorfrancis-com.myaccess.library.utoronto.ca/reader/download/f5baee73-45e3-4d6a-a130-bed67eb32b75/book/pdf?context=ubx

March 7 12 noon – CANSSI STAGE International Speaker Seminar

Speaker: George Davey Smith, University of Bristol
Professor of Clinical Epidemiology, Bristol Medical School
https://canssiontario.utoronto.ca/event/isss-george-davey-smith/

Mendelian Randomization – what it was, what it is, and what it should become

Abstract: Mendelian Randomisation (MR) uses the special properties of germline genetic variation to strengthen causal inference regarding how modifiable exposures influence disease outcomes. Few papers were published in the first decade of its 20 year history, but since then there has been an exponential increase, which shows no sign of slowing down. The basic principles of MR will be restated – including the fundamental one of gene-environment equivalence – as will approaches to causal effect identification and estimation.

Using these it will be shown why the vast majority of current papers are, at best, uninformative, and at worst, nonsense. I will group the threats to MR under three headings: (1) Noodles: papers that are prima facie nonsense, generated from easily available two sample MR data and sub–ChatGPT text; (2) No Nulls: influential early MR findings suggested that some drug targets – such as HDL cholesterol level or C-reactive protein – were unlikely to be important. Null MR studies are now increasingly rarely seen; (3) Numb Skulls: the conceptual simplicity of the MR approach is being obscured by increasingly complex methods, the assumptions of which will be opaque to most readers and most authors of papers implementing them. The first of what should be a series of retractions of papers based on one such method has occurred. The common source of these threats will be discussed, approaches to mitigating them advanced, and some promising future directions for MR will be introduced.

March 14 10 am – Journal Club – Andrew Paterson, SickKids

Reading: Minikel *et al*. Refining the impact of genetic evidence on clinical success. *Nature* **629**, 624–629 (2024). https://www.nature.com/articles/s41586-024-07316-0

Background Reading: King, E. Stopped clinical trials give evidence for the value of genetics. *Nat Genet* **56**, 1766–1767 (2024)
https://www-nature-com.myaccess.library.utoronto.ca/articles/s41588-024-01834-3

March 28 *No seminar*

April 4 9:30 am – Journal Club – Shelley Bull, LTRI

Reading: Gaynor, Joseph, Bai *et al.* Yield of genetic association signals from genomes, exomes and imputation in the UK Biobank. *Nat Genet* 56, 2345–2351, 2024.
https://doi.org/10.1038/s41588-024-01930-4
Background: Genotype Imputation from Large Reference Panels, *Annual Review of Genomics and Human Genetics, 19, 2018.* https://doi.org/10.1146/annurev-genom-083117-021602

April 18 *No seminar (Good Friday holiday)*

April 25 10 am – Journal Club – Andrew Paterson, SickKids

Readings: Sun, B.B., Chiou, J., Traylor, M. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023). https://doi.org/10.1038/s41586-023-06592-6 https://www.nature.com/articles/s41586-023-06592-6
Eldjarn, G.H., Ferkingstad, E., Lund, S.H. *et al.* Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348–358 (2023). https://doi.org/10.1038/s41586-023-06563-x
https://www.nature.com/articles/s41586-023-06563-x

May 2 12 noon – CANSSI STAGE International Speaker Seminar
https://canssiontario.utoronto.ca/event/isss-marjanka-schmidt/

Speaker: Marjanka Schmidt, Netherlands Cancer Institute & Leiden University
Professor of Genetic Epidemiology, Head of Research Division Molecular Pathology, Netherlands Cancer Institute

Breast cancer genetics for all: Determinants of breast cancer subtypes and outcome

Abstract: Breast cancer is a heterogeneous disease influenced by a complex interplay of germline genetic and environmental factors. This presentation will address the role of germline variants, particularly in well-known breast cancer-associated genes like BRCA1, and polygenic scores (PGS) in the incidence of breast cancer subtypes, survival outcomes, and the risk of contralateral breast cancer. Large Genome-Wide-Association-Studies have been proven extremely successful in identifying coding and non-coding (common) variants for breast cancer risk, while obtaining valid results for other breast cancer related end-points has been more challenging. Moreover, risk factors for contralateral breast cancer may be different from those predicting first breast cancer. This presentation will underscore important findings and challenges in breast cancer outcome studies and the crucial need for understanding the diverse roles of genetic variants, including PGS. Increased understanding and integration of genetic information into clinical practice hold the potential to enhance early detection, inform personalised treatment approaches, and ultimately improve breast cancer outcomes.

May 9 10 am – Journal Club – Shelley Bull, *LTRI*

Reading: *Simulation panel of the STRATOS initiative (STRengthening Analytical Thinking for Observational Studies)*, Phases of methodological research in biostatistics—Building the evidence base for new methods (2023). *Biometrical Journal* 66(1) https://onlinelibrary.wiley.com/doi/10.1002/bimj.202200222

Background: *Williams et al*. Transparent reporting items for simulation studies evaluating statistical methods: Foundations for reproducibility and reliability (2024). *Methods in Ecology and Evolution* 15(1): 1926-1939 https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14415

Morris et al, Using simulation studies to evaluate statistical methods (2019). *Statistics in Medicine,* 38(11):2074-2102 https://pmc.ncbi.nlm.nih.gov/articles/PMC6492164/ https://onlinelibrary.wiley.com/doi/10.1002/sim.8086

May 23 *No seminar*

June 6 12 noon – CANSSI STAGE International Speaker Seminar

Free Hybrid (In-person/Online) Event | Registration Required

https://canssiontario.utoronto.ca/event/isss-kathleen-merikangas/

Speaker: Kathleen Merikangas, National Institutes of Mental Health (US)
Senior Investigator and Chief, Genetic Epidemiology Research Branch

Reuniting Families with Genetic Discovery

Abstract: With the advent of Genome-Wide Association Studies (GWAS), genetic research has increasingly adopted large-scale case-control study designs to identify genetic risk factors for complex diseases and traits. However, such approaches— especially those relying on Electronic Health Records (EHRs)—have notable limitations, including phenotypic heterogeneity, non-equivalent control groups, and a lack of detailed clinical and environmental exposure data. These challenges are particularly relevant for complex diseases, which often exhibit both phenotypic and etiologic heterogeneity.

In this talk, Dr. Merikangas will demonstrate how traditional tools of genetic epidemiology can be applied to address these limitations, using data from the NIMH Family Study of Affective Spectrum Disorders. The lecture will focus on three key objectives: 1. Applying the family study approach to investigate core features, sex differences, comorbidity patterns, and the influence of shared environmental factors in complex diseases through familial aggregation and co-aggregation analysis. 2. Integrating GWAS-based methods—including Polygenic Risk Scores (PRS) and Mendelian Randomization (MR)— with family-based studies to enhance our understanding of complex disease etiology. 3. Expanding the conceptualization of phenotypes from a static, cross-sectional model to a dynamic, life-span perspective.

June 13 *No seminar*

June 20 12 noon – Guest Seminar: Olga Vishnyakova, Simon Fraser University

Title: Epigenetic signature of heterogeneity in aging: findings from the Canadian Longitudinal Study on Aging

Abstract: Human aging does not follow a single trajectory. Epigenetic changes offer insight into the heterogeneity in aging by reflecting the combined influence of genetic, environmental, and lifestyle factors on the timing and progression of age-related changes beyond what chronological

age alone can explain. We investigated the role of DNA methylation in aging heterogeneity by performing epigenome-wide differential methylation and variance association analysis. We identified differentially methylated and differentially variable regions associated with health decline. Using both types of probes, which capture mean-level and variability changes, we constructed a composite epigenetic biomarker that outperformed models based on differential methylation alone, showing significant associations with mortality and the onset of chronic obstructive pulmonary disease.

Background Reading:

Review paper: C. G. Bell, R. Lowe, P. D. Adams, A. A. Baccarelli, S. Beck, J. T. Bell, B. C. Christensen, V. N. Gladyshev, B. T. Heijmans, S. Horvath, T. Ideker, J. J. Issa, K. T. Kelsey, R. E. Marioni, W. Reik, C. L. Relton, L. C. Schalkwyk, A. E. Teschendorff, W. Wagner, K. Zhang, and V. K. Rakyan. **DNA methylation aging clocks: challenges and recommendations.** Genome Biology, 20(1):249, 2019. doi: 10.1186/s13059-019-1824-y. https://pubmed.ncbi.nlm.nih.gov/31767039/

Research papers:

Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, Briem E, Zhang K, Irizarry RA, Feinberg AP. **Increased methylation variation in epigenetic domains across cancer types.**Nat Genet. 2011 Jun 26;43(8):768-75. doi: 10.1038/ng.865. https://pubmed.ncbi.nlm.nih.gov/21706001/ ;
PMCID: PMC3145050.

Lu A.T., A. Quach, J. G. Wilson, A. P. Reiner, A. Aviv, K. Raj, et al. **DNA methylation GrimAge strongly predicts lifespan and healthspan**. Aging (Albany NY), 11:303–327, 2019. https://pubmed.ncbi.nlm.nih.gov/30669119/