STATISTICAL METHODS FOR GENETICS & GENOMICS
- RESEARCH SEMINAR AND JOURNAL CLUB
2022-2023

TIME and PLACE:

Fall term 10am (sharp) – 12noon Friday (In Person/Zoom)
Room 618, Health Sciences Building
Winter term 10am - 12noon Friday (in Person/Zoom)
Room 67, University College
Spring term: 10am (sharp) – 12noon Friday (In Person/Zoom)
Room L5-102, 5th floor, 60 Murray St

Seminar: 1 hour; Small Group Discussion: 1 hour.
To be added to the e-distribution list:
please email Czarinah Pisec <cpisec@lunenfeld.ca>

SEMINAR SCHEDULE

September 16 10am – Information session for graduate students interested in registering in the CHL5228H for credit/audit (Room 618 Health Sciences Building, College St)

September 23 10 am – Organizational Meeting re topics & themes for the Seminar/Journal Club this academic year
(In person Room 618 and Zoom)

September 30 3:30pm - Distinguished Lecture in Statistical Sciences:
https://canssiontario.utoronto.ca/event/2022-dlss-xihong-lin/

Speaker: Xihong Lin, Biostatistics, Harvard T.H. Chan School of Public Health; Department of Statistics, Harvard University
Title: Ensemble Methods for Testing a Global Null Hypothesis

October 7 10am – CHL5228H Class Discussion (Room 618)
Reading: Open problems in human trait genetics, *Genome Biol* 23, 131 (2022)
https://doi.org/10.1186/s13059-022-02697-9

12 noon – CANSSI STAGE International Speaker Seminar
https://canssiontario.utoronto.ca/event/stage-isss-teri-manolio/

Speaker: Teri Manolio, National Human Genome Research Institute (US)
Director, Division of Genomic Medicine
Title: Genomic Diversity and Genomic HealthCare

October 14 No Seminar (Reading week)

October 21 10 am – Journal Club - Andrew Paterson, SickKids & Shelley Bull, LTRI
Reading: **Including diverse and admixed populations in genetic epidemiology research**, *Genetic Epidemiology*, October 2022, 46(7):347-371
https://onlinelibrary.wiley.com/doi/10.1002/gepi.22492

October 28 No Seminar **ASHG October 25-28 **

November 4 10 am – Highlights from ASHG

Presenters: Andrew Paterson, Boxi Lin, Jerry Lin, Ziang Zhang

November 11 10 am – CHL5228H Class Discussion

November 18 10 am – Research Seminar

Speaker: Qiongshi Lu, University of Wisconsin-Madison
Biostatistics & Medical Informatics

Topic: **Quantifying gene-environment interaction for human complex traits**

Abstract: The environments are often ignored or treated as nuisance parameters in human complex trait genetics research. However, in epidemiology, clinical research, and social sciences, there is a great interest in quantifying the heterogeneity of the effect of an exposure (e.g., a treatment, a major policy change, a natural experiment), and more specifically, how it interacts with genetics. However, the typical statistical methodology used in gene-environment (GxE) interaction analysis (i.e., linear models with main effects of G and E and the interaction GxE) has a number of limitations, especially in the 'omnigenic' era of complex trait genetics (we have now realized that most human traits have a large number of non-zero but weak genetic effects).

In this talk, I will introduce several recent statistical advances that reimagine the GxE analysis for 'omnigenic' human traits. First, I will introduce QUAIL, a quantile-regression-based framework to identify genetic variants associated with the variability (rather than the mean) of human traits. I will demonstrate that robust findings of variance quantitative trait loci (vQTL) can effectively prioritize candidate genetic variants in GxE analysis, and polygenic scores produced from vQTL effects (vPGS) can aggregate information across numerous genetic loci and improve both statistical power and biological interpretability of GxE studies. Next, I will discuss very recent work (and a method named PIGEON) that links two seemingly unrelated topics: GxE interaction and genetic correlation estimation. I will illustrate that current tools used for genetic correlation estimation provide an ideal alternative strategy for quantifying GxE interactions and will have a number of advantages compared to a traditional linear model with interaction effects. I will show plenty of empirical examples that involve gene-by-education-reform interaction in the UK and gene-by-sex interactions of more than 500 complex traits to showcase the performance of these approaches. Overall, new methods like QUAIL and PIGEON address critical limitations in existing methodologies and will have broad applications in future GxE studies.

Background reading:

Bulik-Sullivan, B., *et al.* (2015). An atlas of genetic correlations across human diseases and traits. *Nature genetics*, 47(11), 1236-1241.
https://www-nature-com.myaccess.library.utoronto.ca/articles/ng.3406
https://pubmed.ncbi.nlm.nih.gov/26414676/

Miao, J., *et al.* (2022). A quantile integral linear model to quantify genetic effects on phenotypic variability. *PNAS*, 119(39), e2212959119.
https://www-pnas-org.myaccess.library.utoronto.ca/doi/full/10.1073/pnas.2212959119
https://pubmed.ncbi.nlm.nih.gov/36122202/

November 25 No Seminar

December 2 12 noon – CANSSI STAGE International Speaker Seminar
https://canssiontario.utoronto.ca/event/stage-isss-andrea-ganna/

Speaker: Andrea Ganna, Harvard Medical School
FIMM-EMBL group leader at the Institute of Molecular Medicine Finland, HiLIFE, U of Helsinki.
Research associate at Massachusetts General Hospital, Harvard Medical School

Title: **Translating results from large-scale genetic association studies into public health-relevant measures**

Abstract: I will describe different efforts that our team is pursuing to translate results from genome-wide association studies into clinically and public-health-relevant applications. The talk will touch upon the use of polygenic scores for disease prediction and their value in informing patients' disease prognosis. I will discuss our recent work in mapping the impact of genetic risk factors on healthy life years (Jukarainen et al, Nature Medicine, 2022) and the use of genetics to understand the causal impact of modifiable risk factors on healthcare costs. Finally, I will show how nationwide registry data, together with genetic information, can enhance high-throughput epidemiological analyses to answer public health-relevant questions, such as COVID-19 vaccination uptake.

December 9 10 am – Research Seminar

Speaker: Nazia Pathan, McMaster University

Title: **Contribution of Rare Coding Variants to Complex Trait Heritability**

Abstract: It has been postulated that rare coding variants (RVs; MAF<0.01) contribute to the "missing" heritability of complex traits. We developed a novel framework, the Rare variant heritability (RARity) estimator, to assess RV heritability (h2RV) without assuming a particular genetic architecture. We applied RARity to 31 complex traits in the UK Biobank (N=167,348) and showed that gene-level RV aggregation suffers from 79% (95% CI: 68-93%) loss of h2RV. Using unaggregated variants, 27 traits had h2RV>5%, with height having the highest h2RV at 21.9% (95% CI: 19.0-24.8%). The total heritability, including common and rare variants, recovered pedigree-based estimates for 11 traits. RARity can estimate gene-level h2RV, enabling the assessment of gene-level characteristics and revealing 12 novel gene-phenotype relationships. Finally, we demonstrated that in silico pathogenicity prediction (variant-level) and gene-level annotations do not generally enrich for RVs that over-contribute to complex trait variance, and thus, novel methods are needed to predict RV functionality.

Background Reading: Chen *et al* (2022). Recent advances and challenges of rare variant association analysis in the biobank sequencing era. Front. Genet., 06 October 2022
DOI: 10.3389/fgene.2022.1014947

********** **2023** **********

January 13 10 am – Research Seminar - Frank Wendt, UofT Anthropology/Biostats

Title: **Large effect risk factors and intervention strategies for internalizing psychopathologies**

Abstract: Internalizing psychopathologies include diagnoses of major depressive, generalized anxiety, and post-traumatic stress disorders. These and various trans-diagnostic factors like neuroticism and suicidality link brain circuitry, behavior, and peripheral biology. The etiologies of internalizing traits also have pronounced influence from biological sex. This talk introduces my group's recent statistical genetics works related to the internalizing spectrum. Topics of this seminar include novel methodological developments to improve statistical power in GWAS and identify loci with large effect sizes, exploring sex-specific GxE across the genome and trauma-phenome, construction of genomic structural equation models of psychopathology, and using clinical and population cohorts to identify and validate psychological sources of, and non-pharmacological treatment strategies for, internalizing diagnoses.

January 20 10 am – Journal Club - Tara Henechowicz, UofT Neuroscience

Reading: **Improving polygenic prediction in ancestrally diverse populations**
Ruan et al (2022), *NatGenet* 54(5):573-580
https://pubmed.ncbi.nlm.nih.gov/35513724/

January 27 10 am – Research Seminar – Tianyuan Lu, McGill University & 5 Prime Sciences

Title: **Identifying molecular therapeutic targets in circulation using Mendelian randomization**

Abstract: Despite tremendous efforts, there are still many diseases for which treatments do not exist or are not effective, or delivery mechanisms or formulations of treatments are inadequate. New therapeutic modalities are urgently required but are difficult to identify due to the heterogeneous causes of these diseases. One category of favorable intervention targets is circulating molecules given their accessibility in circulation, relative easiness to modulate, and the possibility to provide an early read-out of therapies. While randomized controlled trials are not feasible without sufficient evidence of effectiveness, one way to study the potential causal effects of circulating molecules while reducing bias due to confounding and reverse causation in traditional observational studies is through Mendelian randomization (MR). I will describe our recent MR studies in identifying circulating proteins and metabolites that likely play a causal role in the pathogenesis of major psychiatric disorders. I will also discuss approaches to characterize mediation effects and potential non-linear effects of molecular targets in complex diseases.

February 10 10 am – Journal Club - Dingke Tang, UofT Statistical Sciences

Reading: ***Causal inference in genetic association studies***,
Bates et al (2020), *PNAS* **117**, 24117–24126.
https://pubmed.ncbi.nlm.nih.gov/32948695/ https://www.pnas.org/doi/full/10.1073/pnas.2007743117

Background on TDT: Hecker et al (2019). A comparison of popular TDT-generalizations for family-based association analysis. *Genet. Epidemiol.* 43, 300–317. doi:10.1002/gepi.22181
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6599182/

February 17 10 am – Research Seminar – Laurent Briollais & Amelia Xu, LTRI

Topic: **Bayesian analysis of rare variants using Bayes Factor: current and novel approaches**

Background Reading:
Venkataraman et al (2021*), AJHG 108(12):2354-2367.*
Bayesian model comparison for rare-variant association studies
https://pubmed.ncbi.nlm.nih.gov/34822764/
https://www.cell.com/ajhg/fulltext/S0002-9297(21)00417-1

February 24 *No seminar* – Winter Break (Reading) Week

March 3 10 am – Journal Club – Andrew Paterson, SickKids

Reading: **15 years of GWAS discovery: Realizing the promise**
Am J Hum Genet. 2023 Feb 2;110(2):179-194.
https://pubmed.ncbi.nlm.nih.gov/36634672/
https://doi-org.myaccess.library.utoronto.ca/10.1016/j.ajhg.2022.12.011

Background: 10 Years of GWAS Discovery. https://doi.org/10.1016/j.ajhg.2017.06.005
Five years of GWAS discovery. https://doi.org/10.1016/j.ajhg.2011.11.029

March 10 10 am – Journal Club - Biswajit Chowdhury, Biostatistics

Reading: **A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies**,
Nature Methods **volume 19**, pages 1599–1611 (2022)
https://pubmed.ncbi.nlm.nih.gov/36303018/ https://doi.org/10.1038/s41592-022-01640-x

March 17 10 am – Journal Club – Boxi Lin, Biostatistics

Reading: **A perspective on interaction effects in genetic association studies**, *Genetic Epidemiology* 2016 Dec; 40(8): 678-688 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5132101/

Background Reading: Update on the State of the Science for Analytical Methods for Gene-Environment Interactions, *AmJEpidemiol* 2017 Oct; 186(7): 762-770. Gauderman et al., 2017 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5859988/).

March 24 10 am – General Discussion – Andrew Paterson, SickKids

This will be a discussion/Q&A of recent journal club/seminar topics.

General Topics / Readings:

15 years of GWAS discovery: Realizing the promise. *Am J Hum Genet*. 2023 Feb 2;110(2):179-194. https://pubmed.ncbi.nlm.nih.gov/36634672/ https://doi-org.myaccess.library.utoronto.ca/10.1016/j.ajhg.2022.12.011

Open problems in human trait genetics, Brandes, N., Weissbrod, O. & Linial, M. *Genome Biol* 23, 131 (2022). https://doi.org/10.1186/s13059-022-02697-9

April 7 *No seminar* – Good Friday Holiday

April 14 12 noon – CANSSI STAGE International Speaker Seminar

https://canssiontario.utoronto.ca/event/stage-isss-manolis-kellis/

Speaker: Manolis Kellis, Computer Science, Massachusetts Institute of Technology

Topic: **From genomics to therapeutics:**
Single-cell dissection and manipulation of disease circuitry

Abstract: Disease-associated variants lie primarily in non-coding regions, increasing the urgency of understanding how generegulatory circuitry impacts human disease. To address this challenge, we generate comparative genomics, epigenomic, and transcriptional maps, spanning 823 human tissues, 1500 individuals, and 20 million single cells. We link variants to target genes, upstream regulators, cell types of action, and perturbed pathways, and predict causal genes and regions to provide unbiased views of disease mechanisms, sometimes re-shaping our understanding. We find that Alzheimer's variants act primarily through immune processes, rather than neuronal processes, and the strongest genetic association with obesity acts via energy storage/dissipation rather than appetite/ exercise decisions. We combine single-cell profiles, tissue-level variation, and genetic variation across healthy and diseased individuals to map genetic effects into epigenomic, transcriptional, and function changes at single-cell resolution, to recognize celltype-specific disease-associated somatic mutations indicative of mosaicism, and to recognize multi-tissue single-cell effects of exercise and obesity. We expand these methods to electronic health records to recognize multi-phenotype effects of genetics, environment, and disease, combining clinical notes, lab tests, and diverse data modalities despite missing data. We integrate large cohorts to factorize phenotype-genotype correlations to reveal distinct biological contributors of complex diseases and traits, to partition disease complexity, and to stratify patients for pathway-matched treatments. Lastly, we develop massively-parallel, programmable and modular technologies for manipulating these pathways by high-throughput reporter assays, genome editing, and gene targeting in human cells and mice, to propose new therapeutic hypotheses in Alzheimer's, obesity, and cancer. These results provide a roadmap for translating genetic findings into mechanistic insights and ultimately new therapeutic avenues for complex disease and cancer.

April 21 10am – Journal Club - Osvaldo Espin-Garcia, Western Univ

Reading: Zhang et al (2021) **Improved genetic prediction of complex traits from individual-level data or summary statistics** https://www.nature.com/articles/s41467-021-24485-y

Speed et al (2020) **Evaluating and improving heritability models using summary statistics** https://www.nature.com/articles/s41588-020-0600-y

April 28 10am – Seminar/Journal Club - Mohammad Akbari, Women's College Research

Reading: **Variants in ATRIP are associated with breast cancer susceptibility in the Polish population and UK Biobank,** Cybulski *et al* (2023), *AJHG*
https://www-sciencedirect-com.myaccess.library.utoronto.ca/science/article/pii/S0002929723000861
Common Variant in ALDH2 Modifies the Risk of Breast Cancer Among Carriers of the p.K3326* Variant in BRCA2, Kluznlak *et al* (2022), *JCO Precision Oncology*
https://ascopubs-org.myaccess.library.utoronto.ca/doi/10.1200/PO.21.00450

May 5 12 noon –CANSSI STAGE International Speaker Seminar
https://canssiontario.utoronto.ca/event/stage-isss-hugo-aerts/

Speaker: Hugo Aerts, Maastricht University & Harvard-MGB (Mass General Brigham)
Topic: **Artificial Intelligence for Medical Imaging**

Abstract: Technological advances in Artificial Intelligence (AI), particularly deep learning, have demonstrated remarkable progress in image-recognition tasks. Methods ranging from convolutional neural networks to variational autoencoders have found myriad applications in various medical fields, propelling it forward at a rapid pace. In this talk, Dr. Aerts will discuss recent developments from his group and collaborators performing research at the intersection of deep learning, radiology, oncology, cardiology, bioinformatics, and data science. Also, he will explore how these methods could impact multiple facets of medicine, with a general focus on applications in radiology, and demonstrate ways in which these methods are advancing the field. The presentation will conclude with a discussion on the need for open-source deep learning frameworks that are transparent and reproducible.

May 12 & 19 No Seminars

May 26 10 am – Research Seminar – Ruowang Li, Computational Biomedicine
Cedars Sinai Medical Center, California
Topic: **Statistical genomic methods for Electronic Health Record data**

Abstract: Biobank-linked electronic health record (EHR) is increasingly becoming available for genetic epidemiology and genomics research. In this talk, I will discuss several method development efforts to utilize multiple EHR data to identify cross-phenotype genetic associations and to develop genetic risk prediction models. I will present results from the UK Biobank and the Electronic Medical Records and Genomics (eMERGE) Network data to demonstrate the potential of EHR data for genomics research.

June 2 12 noon –CANSSI STAGE International Speaker Seminar
https://canssiontario.utoronto.ca/event/stage-isss-paul-pharoah/

Speaker: Paul Pharoah, Department of Computational Biomedicine
Cedars Sinai Medical Center, Los Angeles
Topic: **Inherited susceptibility to ovarian cancer: the story so far**

Abstract: Epithelial ovarian cancer has a substantial inherited genetic component. Over the past 25 years, substantial progress has been made to unravel the underlying genetic architecture of risk. Linkage studies in the 1990's in multi-case, multi-generation families were successful at identifying BRCA1 and BRCA2; rare loss-of-function allele of these genes are associated with a high risk of disease. The search for common, modest risk alleles using association study designs were initially based on candidate variant/candidate gene studies, but developments in genotyping technology together with the formation of large, international consortia ushered in an era of genome-wide association studies which have been successful in identifying multiple risk variants. In the past decade large-scale, targeted sequencing of candidate genes in case-control studies has enabled the identification loss-of-function alleles of BRIP1, PALB2, RAD51C and RAD51D that are associated

with intermediate disease risks. Future developments will be dependent on large scale exome and whole genome sequencing.

June 9 10am – Research Seminar - Brad McNeney, Statistics and Actuarial Science
Simon Fraser University

Topic: **Inference of gene-environment interaction from heterogeneous case-parent trios**

Abstract: In genetic epidemiology, log-linear models of population risk may be used to study the effect of genotypes and exposures on the relative risk of a disease. Such models may also include gene-environment interaction terms that allow the genotypes to modify the effect of the exposure, or equivalently, the exposure to modify the effect of genotypes on the relative risk. When a measured test locus is in linkage disequilibrium with an unmeasured causal locus, exposure-related genetic structure in the population can lead to spurious gene-environment interaction; that is, to apparent gene-environment interaction at the test locus in the absence of true gene-environment interaction at the causal locus. Exposure-related genetic structure occurs when the distributions of exposures and of haplotypes at the test and causal locus both differ across population strata. A case-parent trio design can protect inference of genetic main effects from confounding bias due to genetic structure in the population. Unfortunately, when the genetic structure is exposure-related, the protection against confounding bias for the genetic main effect does not extend to the gene-environment interaction term. We show that current methods to reduce the bias in estimated gene-environment interactions from case-parent trio data can only account for simple population structure involving two strata. To fill this gap, we propose to directly accommodate multiple population strata by adjusting for genetic principal components. We evaluate our approach through simulation and illustrate it on data from a study of genetic modifiers of cleft palate.

Background Reading:

Family-based Gene-by-environment Interaction Studies: Revelations and Remedies
https://journals.lww.com/epidem/Fulltext/2011/05000/Family_based_Gene_by_environment_Interaction.22.aspx
- A paper by Shi, Umbach & Weinberg that first illustrated exposure-related genetic stratification as a source of bias for inference of GxE from trios.

Adjusting for Spurious Gene-by-Environment Interaction Using Case-Parent Triads
https://doi.org/10.1515/sagmb-2013-0023
https://www-degruyter-com.myaccess.library.utoronto.ca/document/doi/10.1515/sagmb-2013-0023/html
- Shin *et al* (2012) paper. The "current method" we refer to in the abstract.

Re-analysis of a Genome-Wide Gene-By-Environment Interaction Study of Case Parent Trios, Adjusted for Population Stratification
https://www.frontiersin.org/articles/10.3389/fgene.2020.600232/full
- Ratnasekera & McNeney (2021) Focus of the presentation.

Co-Organizers:

Shelley Bull
Professor, DLSPH
Senior Scientist,
Lunenfeld-Tanenbaum Research Institute
Email: bull@lunenfeld.ca

Andrew Paterson
Professor, DLSPH
Senior Scientist,
Hospital for Sick Children Research
Email: andrew.paterson@sickkids.ca