STATISTICAL METHODS FOR GENETICS & GENOMICS
- RESEARCH SEMINAR AND JOURNAL CLUB
2018-2019

TIME and PLACE:

Fall term 10am – 12noon Friday (The Fields Institute Room 210)
Winter term 10am - 12noon Friday (The Fields Institute Room 309)

Co-Organizers:

Shelley Bull
Professor, DLSPH
Senior Scientist,
Lunenfeld-Tanenbaum Research Institute
60 Murray Street, Box 18
Room 5-226
Email: bull@lunenfeld.ca
Phone: 416-586-8245

Andrew Paterson
Professor, DLSPH
Senior Scientist,
Hospital for Sick Children Research
PGCRL 686 Bay Street,
Room 12.9710,
Email: andrew.paterson@utoronto.ca
Phone: 416-813-6994

SEMINAR SCHEDULE

September 14 10 am – Organizational Meeting

September 21 10 am – Journal Club – Shelley Bull, LTRI

Reading: Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations, Khera et al (2018) *Nature Genetics* 50:1219-1224
https://www-nature-com.myaccess.library.utoronto.ca/articles/s41588-018-0183-z

Background Reading (UK Biobank):
Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12, e1001779 (2015).
https://journals-plos-org.myaccess.library.utoronto.ca/plosmedicine/article?id=10.1371/journal.pmed.1001779
Bycroft, C. et al. Genome-wide genetic data on ~500,000 UK Biobank participants.
Preprint at bioRxiv,
https://www.biorxiv.org/content/early/2017/07/20/166298

October 5 12 noon – STAGE International Speaker Seminar
The Fields Institute, Room 230
Speaker: Augustine Kong, Big Data Institute, University of Oxford
Talk title: *The Nature of Nurture – Effects of Parental Genotypes*
Abstract: https://stage.utoronto.ca/wp-content/uploads/2018/09/Augustine-Kong-seminar.pdf

* IGES/ASHG October 13-20, San Diego *

October 26 10 am – Reports from IGES/ASHG Meetings
(O Espin-Garcia, S Kim, Y-C Lin, R Romanescu, L Zhang, Y Zhao)

November 9 10 am – Research Seminar - Dave Soave, OICR
Topic: Population Health Cohorts and Two-Phase Studies

Abstract: Large prospective cohorts like the Canadian Partnership for Tomorrow Project (CPTP) follow participants longitudinally and capture incident cases of disease. In two-phase studies, researchers select a subset of the complete cohort based on observed outcomes and covariates and measure additional, possibly expensive, variables. In the CPTP, blood samples are collected and stored when participants enrol. During follow-up a small fraction of the cohort will present with rare diseases, and researchers often select a matched case-control sample and obtain biological information from their blood samples, seeking to discover biomarkers that predict disease. This presentation will review methods for analyzing rare events based on two-phase studies and discuss areas needing development, including study design, methods for assessing genomic and other risk factors, and predictive modelling. An illustration will be based on predicting leukaemia risk in healthy individuals.

Background Reading:

Case-cohort analysis: Samuelsen, S. O., ÅNestad, H., Skrondal, A. (2007). Stratified Case-Cohort Analysis of General Cohort Sampling Designs. *Scandinavian J of Statistics* 34, 103-119
https://onlinelibrary-wiley-com.myaccess.library.utoronto.ca/doi/10.1111/j.1467-9469.2006.00552.x

Predicting leukaemia: Abelson, S., Collord, G., Ng, S. W. K., et al. (2018). Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature* 559, 400-404
https://www-nature-com.myaccess.library.utoronto.ca/articles/s41586-018-0317-6

November 16 10 am – Seminar – Andrew Paterson, SickKids
Topic: Canadian Longitudinal Study of Ageing: Cohort overview, data availability & access procedures, some descriptives

Reading: **The Canadian Longitudinal Study on Aging:** Genome-wide Genetic Data on 9,900 Participants, Forgetta, Darmond-Zwaig, Belisle, Li, Balion, Roshandel, Wolfson, Lettre, Pare, Paterson, Verschoor, Lathrop, Raina, Richards, Ragoussis, 31 July 2018. https://www.clsa-elcv.ca/doc/2748

Abstract: Genetic and environmental factors contribute to maintaining health and in the development of disease and disability as people age. The Canadian Longitudinal Study of Aging (CLSA) is a national long-term study that will follow approximately 50,000 men and women, and presents a unique opportunity to study genetic and environmental contributions to human health and disease by providing information on the changing biological, medical, psychological, social, lifestyle and economic aspects of participants' lives. This document describes the availability and quality assessment of genetic data for 9,900 CLSA participants, comprising genome-wide directly genotyped data for 794,409 markers and whole-genome imputed data for ~39 million genetic variants. Quality assessment includes both marker- and sample-based tests, as well as analysis of sex-chromosome abnormalities, population structure, and familial relatedness. Results from exemplar genome-wide association studies for height and lipid traits are also provided. An additional 20,000 individuals are planned for a future data release of genome-wide genotypes. Qualified researchers from any country can access this genomic and phenotypic data release via the CLSA Data Access Portal.

Background: Kirkland et al (2015). Mining a Unique Canadian Resource: The Canadian Longitudinal Study on Aging. Canadian Journal on Aging / La Revue canadienne du vieillissement, 34, pp 366-377 doi:10.1017/S071498081500029X http://journals.cambridge.org/abstract_S071498081500029X

Raina et al (2009) The Canadian Longitudinal Study on Aging (CLSA). Canadian Journal on Aging / La Revue canadienne du vieillissement / Volume 28 / Special Issue 03 / September 2009, pp 221 229 DOI: 10.1017/S0714980809990055. http://journals.cambridge.org/abstract_S0714980809990055

link to publications: www.clsa-elcv.ca/stay-informed/publications
and protocols: www.clsa-elcv.ca/researchers

********** 2019 **********

February 1 10 am – Research Seminar – Lei Sun, Statistical Sciences, Toronto
Title: Back to the future: 'simple' regression models for complex genetic association studies

Abstract: Linear regression remains an important framework in the era of big and complex data. In this talk I present some recent examples where we resort to the classical simple linear regression model and its celebrated extensions in novel settings. The Eureka moment came while reading Wu and Guan's (2015) comments on our generalized Kruskal-Wallis test (Elif Acar and Sun 2013, *Biometrics*). Wu and Guan presented an alternative "rank linear regression model and derived the proposed GKW statistic as a score test statistic". Indeed, the regression framework eases the derivation and facilitates further extensions. More recently, we turned our attention to extending Levene's variance test to data with genotype uncertainty and related individuals; this test is useful for GxE interaction studies but data on E is not available. While a direct modification of the original test statistic is challenging, I will demonstrate that establishing a two-stage regression framework for the original Levene's test makes the ensuing method development quite straightforward, eventually leading to a generalized joint location-scale test (David Soave and Sun 2017, *Biometrics*). Finally, I will discuss some on-going projects, including how to robustify the allelic association test against Hardy-Weinberg disequilibrium and generalize it for quantitative traits, how to develop a flexible association test for the complex X-chromosome, and how to unify, in an analytical sense, methods developed for rare variants with the polygenic risk score analyses, among others. In each case, the crux of the work is reformulating the problem as a regression!
Recommended Reading: Any regression textbooks or your old lecture notes

March 1 10 am – Research – Fan Wang, Statistical Sciences
Topic: Co-localization Methods

Reading:
Gong, Wang, Xiao et al (2019) Genetic association & transcriptome integration identify contributing genes and tissues at CF modifier loci, PLoS Genetics
https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008007
Giambartolomei et al (2014) Bayesian test for colocalisation between paris of genetic association studies using summary statistics, PLoS Genetics.
https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004383
Hormozdiari et al (2016) Colocalization of GWAS and eQTL signals detects target genes, AJHG.
https://www.sciencedirect.com/science/article/pii/S0002929716304396

March 8 10 am – Guest Speaker – Huijue Jia, BGI Research
Topic: Nature vs. nurture in the human commensal microbiome

Bio: Dr. Huijue Jia graduated with a B. S. in Biological Sciences from Fudan University in 2005, and then a Ph.D. in Biochemistry from Case Western Reserve University in 2011. Following a brief postdoctoral training from the University of North Carolina Chapel Hill (HHMI), Dr. Jia became an editor for the then new journal of Nature Communications, responsible for manuscripts in DNA or RNA-related fields. Dr. Jia joined BGI-Shenzhen in 2013 and has since been working to establish high-quality reference datasets for the healthy human microbiome, as well as investigating microbial derangements in complex diseases and developing new probiotics.

Abstract: First published by BGI in 2012, Metagenome-wide association studies (MWAS) have received even more controversy than Genome-wide association studies (GWAS), as the gut microbiome is believed to be highly dynamic. Here through a 'M-GWAS' of whole metagenome and whole genome, we identify SNPs, genes and CNVs in the human genome that together explain a considerable fraction of the variances in gut microbial composition and functional potential. Some of the gut microbial markers for diseases such as colorectal cancer and metabolic diseases showed associations with the human genome. Less heavily studied than the fecal microbiome, the vagino-cervical microbiome is shown here to relate to both historical and recent events. The vagino-cervical microbiome is also reflected in plasma

metabolites, self-reported mental status and facial skin measurements, while its association with the fecal microbiome is generally weak.
Host: Qiang Sun, Statistical Sciences

Background Reading:
Nature (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes, doi: 10.1038/nature1145
https://www.ncbi.nlm.nih.gov/pubmed/23023125
https://www-nature-com.myaccess.library.utoronto.ca/articles/nature11450

March 15 10 am – Research – Marie Pigeyre, McMaster
Topic: Identification of novel biomarkers for type 2 diabetes through Mendelian Randomization analysis of the blood proteome

Reading:
Primary: Nature (2018) Genomic atlas of the human plasma proteome
https://www.nature.com/articles/s41586-018-0175-2
https://www-nature-com.myaccess.library.utoronto.ca/articles/s41586-018-0175-2
Secondary: Nature (2018) Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease
https://www.nature.com/articles/s41467-018-05512-x

March 22 12 noon – STAGE International Speaker Seminar
The Fields Institute, Room 230

Speaker: Stijn Vansteelandt, Ghent University
Title: *Inferring causal pathways from data: Challenges & some Solutions*
Abstract: https://stage.utoronto.ca/wp-content/uploads/2019/03/Dr.-Stijn-Vansteelandt-STAGE.pdf

March 29 10 am – Journal Club – Changchang Xu, Biostatistics
Topic: BOLT-LMM, a linear mixed model for GWAS

Reading:
Primary paper: Loh et al (2018) "Mixed-model association for biobank-scale datasets"
https://www.ncbi.nlm.nih.gov/pubmed/29892013
https://www-nature-com.myaccess.library.utoronto.ca/articles/s41588-018-0144-6
Background:
Loh et al (2015) "Efficient Bayesian mixed model analysis increases association power in large cohorts"
https://www.ncbi.nlm.nih.gov/pubmed/25642633
https://www-nature-com.myaccess.library.utoronto.ca/articles/ng.3190
Bulk-Sullivan et al (2015) "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies"
https://www.ncbi.nlm.nih.gov/pubmed/25642630
https://www-nature-com.myaccess.library.utoronto.ca/articles/ng.3211

April 5 10 am – Research – Robbie Davies, SickKids
Topic: Imputation of mother and fetus from sequence

Background Reading:
Liu et al (2018) Genomic analyses from non-invasive prenatal testing reveal genetic associations, patterns of viral infections, and Chinese population history, *Cell* 175(2), 347-359.
https://www.ncbi.nlm.nih.gov/pubmed/30290141
https://www.cell.com/cell/pdf/S0092-8674(18)31032-8.pdf

https://www-sciencedirect-com.myaccess.library.utoronto.ca/science/article/pii/S0092867418310328
Davies et al (2016) Rapid genotype imputation from sequence without reference panels,
Nature Genetics 48, 965–969
https://www.ncbi.nlm.nih.gov/pubmed/27376236
https://www-nature-com.myaccess.library.utoronto.ca/articles/ng.3594

April 12 10 am – Journal Club – Jingqi Hao, Biostatistics
Topic: Significance thresholds for EWAS

Reading:
Saffari et al (2018) "Estimation of a significance threshold for epigenome-wide association studies"
Genetic Epidemiology 42(1), 20-33
https://onlinelibrary.wiley.com/doi/full/10.1002/gepi.22086

April 26 10 am – StatGen Seminar – Research
Speaker: **Prof Jonathan Beauchamp,** Economics, University of Toronto
Topic: Recent Advances in Social Science Genetics

Abstract: Over the past few years, social scientists have used the GWAS methodology to study social scientific phenotypes and have also contributed to the development of new methods to study those phenotypes. I will summarize the results of recent GWAS of risk tolerance, educational attainment, and subjective wellbeing, and briefly discuss their implications for both geneticists and social scientists.

Background Readings:
Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences
https://www.nature.com/articles/s41588-018-0309-3
https://www-nature-com.myaccess.library.utoronto.ca/articles/s41588-018-0309-3
Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals
https://www.nature.com/articles/s41588-018-0147-3
https://www-nature-com.myaccess.library.utoronto.ca/articles/s41588-018-0147-3
Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses
https://www.nature.com/articles/ng.3552
https://www-nature-com.myaccess.library.utoronto.ca/articles/ng.3552

May 21 Tuesday 10:30am - STAGE International Speaker Seminar Series
Speaker: **Li Li ,** University of Virginia
Topic: Racial Disparities and Precision Prevention of Colon Neoplasia
https://stage.utoronto.ca/wp-content/uploads/2019/04/STAGE-Poster_Li.pdf
Location: Auditorium 6-604 Princess Margaret Hospital
610 University Ave Toronto, ON

May 23 Thursday 4:00pm – Aser Rothstein Lecture, SickKids
Speaker: **Nancy Cox,** Vanderbilt University
Topic: Pleiotropy: Geneticists Punishment for their Hubris
or a Path to Building Better Targets for Discovery
https://stage.utoronto.ca/wp-content/uploads/2019/04/Aser-Rothstein_Nancy-Cox_May-23_2019.pdf
Location: Daniels Hollywood Theatre at SickKids main campus
Room 1246, First Floor, Black Wing, 555 University Avenue, Toronto

May 26-29 Statistical Society of Canada Annual Meeting – Calgary
https://ssc.ca/en/meeting/annual/2019

June 7 Friday 12 noon - STAGE International Speaker Seminar Series
Speaker: **Gil McVean,** University of Oxford

https://stage.utoronto.ca/event/stage-seminar-dr-gil-mcvean
Location:
Robert B Salter Auditorium, Peter Gilgan Centre for Research and Learning
686 Bay Street, Toronto, On

June 14 10am – StatGen Seminar - Research
Speakers: **Prof David Meyre & Akram Alyass,** McMaster University
Location:
The Fields Institute for Research in Mathematical Sciences
Room 210, 222 College St, Toronto

June 16-19 Canadian Human & Statistical Genetics Meeting – Quebec
https://dal.cihr-ig-irsc.ca/genetics19