Autumn 2021, Schedule of papers

Week 1 (October 5). Everyone. Overview and motivation

Suggested paper: (Ben-Eghan et al. 2020)

This paper serves as an introduction to the topic and also provides motivation. It discusses the statistical reasons why data from minority groups are often excluded, and the ethical issues this raises when such data are discarded by default.

Week 2 (October 12). How do geneticists think about ancestry?

Suggested paper: (Mathieson and Scally 2020)

“although frequently discussed, ancestry itself is rarely defined. We argue that this reflects widespread underlying confusion about what it means in different contexts and what genetic data can really tell us.”

This paper discusses genealogical ancestry, genetic ancestry, and genetic similarity and how these concepts are addressed by researchers and considered among the wider public.

Week 3 (October 19). Methods: testing for HWE

Suggested paper: (Kwong et al. 2021)

Hardy-Weinberg equilibrium (HWE) is one of the first concepts we learn about in population genetics. Deviations from HWE can signal interesting population-level dynamics but evaluations of HWE also serve as an important bioinformatic quality-control (QC) step. However, most common tests for HWE assume homogeneous populations. This paper presents a method to test for Hardy-Weinberg equilibrium (HWE) in the presence of population structure.

See also (Meisner and Albrechtsen 2019)

Week 4 (October 26). Methods: GWAS in admixed populations

Suggested paper: (Atkinson et al. 2021)

It is well known that population structure can bias genome-wide association studies (GWAS), and many methods have been proposed to reduce this bias. Admixed individuals may be removed from GWAS due to the lack of methods and concerns over statistical bias. This paper presents Tractor, a method that uses local ancestry information to to include admixed individuals in association studies.

Week 5 (November 2). Methods: polygenic risk scores (PRS)

Polygenic risk score (PRS) "is an estimate of an individual's genetic liability to a trait or disease." These can be calculated from individual-level phenotype data and effect-size estimates taken from GWAS. However, estimating, interpreting, and applying PRS across diverse groups of individuals has proven to be difficult, with large potentials for bias.

Suggested paper: (Ruan et al. 2021)

Week 6 (November 9). Methods: reference genomes

Suggested paper: (Sherman et al. 2019)

The human reference genome has been an invaluable resource over the last ~18 years. This reference has been updated since 2003, but is still largely composed of sequences from just a few individuals. This raises the potential for reference bias, as genetic variation from around the world is not equally present in the reference genome. This paper presents 296 Mbp of reference contigs (~10% of genome length), that are present in a sample of individuals of African descent, but absent from a recent human reference genome.

See also (Ballouz, Dobin, and Gillis 2019)

Week 7 (November 16). Methods: relatedness

Suggested paper: (Freyman et al. 2021)

Inferring pairwise relatedness, either in terms of relatedness k-coefficients or familial relationships, is an important step in most genetic studies. However, recent admixture and/or population structure can confound many methods for inferring relatedness. IBD-segment based methods can be more robust to some forms of admixture and population structure than other methods that are based on allele frequencies. Here we look at a recent published IBD-based pairwise relatedness method that can be applied to biobank-sized data sets. The paper presents a method that is a modification of the Burrows-Wheeler transform, that is used to quickly find putative IBD segments between pairs of individuals.

See also: (Sticca, Belbin, and Gignoux 2021)

Week 8 (November 23). A cautionary tale

Suggested paper: (Berg et al. 2019)

Berg et al. try and fail to replicate an earlier finding of strong polygenic adaptation for height in Europe. The paper shows how the previous findings were confounded by population structure.

See also: (Sohail et al. 2019), (Refoyo-Martínez et al. 2021)

Week 9 (November 30). Study design

Suggested paper: (Wojcik et al. 2019)

This paper presents a GWAS on 49,839 “non-European” individuals from the PAGE study. They conduct a unified analysis using methods meant to account for population structure and relatedness and compare to an ancestry-stratified meta-analysis.

See also (Bien et al. 2019), (Taliun et al. 2021)

Week 10 (December 7). Ethics and health disparities of polygenic risk scores

Suggested paper: (Martin et al. 2019)

While the actual utility of Polygenic risk scores (PRS) is still mostly unproven, PRS can predict the risk of some cancers in individuals of European descent more accurately than current clinical models. However, the performance of PRS are poor in non-European populations. This paper discusses how the use of PRS may further exacerbate existing health disparities.

References

Atkinson, Elizabeth G., Adam X. Maihofer, Masahiro Kanai, Alicia R. Martin, Konrad J. Karczewski, Marcos L. Santoro, Jacob C. Ulirsch, et al. 2021. “Tractor Uses Local Ancestry to Enable the Inclusion of Admixed Individuals in GWAS and to Boost Power.” *Nature Genetics* 53 (2): 195–204.

Ballouz, Sara, Alexander Dobin, and Jesse A. Gillis. 2019. “Is It Time to Change the Reference Genome?” *Genome Biology* 20 (1): 159.

Ben-Eghan, Chief, Rosie Sun, Jose Sergio Hleap, Alex Diaz-Papkovich, Hans Markus Munter, Audrey V. Grant, Charles Dupras, and Simon Gravel. 2020. “Don’t Ignore Genetic Data from Minority Populations.” *Nature* 585 (7824): 184–86.

Berg, Jeremy J., Arbel Harpak, Nasa Sinnott-Armstrong, Anja Moltke Joergensen, Hakhamanesh Mostafavi, Yair Field, Evan August Boyle, et al. 2019. “Reduced Signal for Polygenic Adaptation of Height in UK Biobank.” *eLife* 8 (March). https://doi.org/10.7554/eLife.39725.

Bien, Stephanie A., Genevieve L. Wojcik, Chani J. Hodonsky, Christopher R. Gignoux, Iona Cheng, Tara C. Matise, Ulrike Peters, Eimear E. Kenny, and Kari E. North. 2019. “The Future of Genomic Studies Must Be Globally Representative: Perspectives from PAGE.” *Annual Review of Genomics and Human Genetics* 20 (August): 181–200.

Conomos, Matthew P., Alexander P. Reiner, Bruce S. Weir, and Timothy A. Thornton. 2016. "Model-Free Estimation of Recent Genetic Relatedness." *American Journal of Human Genetics* 98 (1): 127–48.

Freyman, William A., Kimberly F. McManus, Suyash S. Shringarpure, Ethan M. Jewett, Katarzyna Bryc, 23 and Me Research Team, and Adam Auton. 2021. "Fast and Robust Identity-by-Descent Inference with the Templated Positional Burrows-Wheeler Transform." *Molecular Biology and Evolution* 38 (5): 2131–51.

Kwong, Alan M., Thomas W. Blackwell, Jonathon LeFaive, Mariza de Andrade, John Barnard, Kathleen C. Barnes, John Blangero, et al. 2021. "Robust, Flexible, and Scalable Tests for Hardy-Weinberg Equilibrium across Diverse Ancestries." *Genetics* 218 (1). https://doi.org/10.1093/genetics/iyab044.

Martin, Alicia R., Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale, and Mark J. Daly. 2019. "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities." *Nature Genetics* 51 (4): 584–91.

Mathieson, Iain, and Aylwyn Scally. 2020. "What Is Ancestry?" *PLoS Genetics* 16 (3): e1008624.

Meisner, Jonas, and Anders Albrechtsen. 2019. "Testing for Hardy-Weinberg Equilibrium in Structured Populations Using Genotype or Low-Depth next Generation Sequencing Data." *Molecular Ecology Resources* 19 (5): 1144–52.

Ochoa, Alejandro, and John D. Storey. 2021. "Estimating FST and Kinship for Arbitrary Population Structures." *PLoS Genetics* 17 (1): e1009241.

Refoyo-Martínez, Alba, Siyang Liu, Anja Moltke Jørgensen, Xin Jin, Anders Albrechtsen, Alicia R. Martin, and Fernando Racimo. 2021. "How Robust Are Cross-Population Signatures of Polygenic Adaptation in Humans?" *bioRxiv*. https://doi.org/10.1101/2020.07.13.200030.

Ruan, Yunfeng, Yen-Chen Anne Feng, Chia-Yen Chen, Max Lam, Akira Sawa, Alicia R. Martin, Shengying Qin, Hailiang Huang, Tian Ge, and Stanley Global Asia Initiatives. 2021. "Improving Polygenic Prediction in Ancestrally Diverse Populations." *bioRxiv*. medRxiv. https://doi.org/10.1101/2020.12.27.20248738.

Sherman, Rachel M., Juliet Forman, Valentin Antonescu, Daniela Puiu, Michelle Daya, Nicholas Rafaels, Meher Preethi Boorgula, et al. 2019. "Assembly of a Pan-Genome from Deep Sequencing of 910 Humans of African Descent." *Nature Genetics* 51 (1): 30–35.

Sohail, Mashaal, Robert M. Maier, Andrea Ganna, Alex Bloemendal, Alicia R. Martin, Michael C. Turchin, Charleston Wk Chiang, et al. 2019. "Polygenic Adaptation on Height Is Overestimated due to Uncorrected Stratification in Genome-Wide Association Studies." *eLife* 8 (March). https://doi.org/10.7554/eLife.39702.

Taliun, Daniel, Daniel N. Harris, Michael D. Kessler, Jedidiah Carlson, Zachary A. Szpiech, Raul Torres, Sarah A. Gagliano Taliun, et al. 2021. “Sequencing of 53,831 Diverse Genomes from the NHLBI TOPMed Program.” *Nature* 590 (7845): 290–99.

Wojcik, Genevieve L., Mariaelisa Graff, Katherine K. Nishimura, Ran Tao, Jeffrey Haessler, Christopher R. Gignoux, Heather M. Highland, et al. 2019. “Genetic Analyses of Diverse Populations Improves Discovery for Complex Traits.” *Nature* 570 (7762): 514–18.