Sociogenomics – Integrating Social Science Research and Genomics

Jinyuan (Chi) Qi

Course Overview:

Since the completion of the Human Genome Project in 2003, advances in the field of genomics over the past twenty years have led to substantial reductions in the cost of genome sequencing (Wetterstrand, 2018). Recently, many social science research datasets have added biodata collection, including genetic sequence data, such as the Add Health, the Health and Retirement Study, Fragile Families and Child Wellbeing Study, etc. With emerging interests and increasingly available data to study classical social science questions with new insights from genetics, this course provides an overview of sociogenomics and intends to teach and discuss the following issues:

- How can genetics be informative for social science research?
- What are the theories, methods, and tools for integrating these two areas?
- What are the challenges and difficulties in integrating the two areas?
- How did the state-of-the-art journal articles address those challenges and difficulties?
- What would be the new directions for future research in sociogenomics?

Core Learning Objectives:

1. Students will have an intermediate-level understanding of essential concepts and theories in genetics and biology relevant to sociogenomics.
2. Students will be able to understand and analyze current research in sociogenomics from both theoretical and empirical perspectives.
3. Students will be able to propose research questions that integrate some aspects of genetics and social science research
4. Students will have basic knowledge of how to use existing methods to answer their research questions in sociogenomics.

Note:

- Russel Sage Foundation Summer Institute in Social Science Genomics 2019 Reading List has influenced this syllabus

Week 1 Topic: Overview – How genetics can inform social science research

Main goals:

Students could

- Understand the motivation and structure of the course
- Recognize, and describe the most basic concepts used in sociogenomics
- Gain an overview of how genetics can provide novel insights into social science research
- Recognize the new questions that could be answered by integrating two fields
- Acquire a basic understanding of what are the challenges and pitfalls of integrating genetic data
- Explore sociogenomic studies in different disciplines of social science

Discussion of the readings:

The first week's readings give an overview of how genetics can provide novel insights to social science research, and answer questions that cannot be answered without genetic data, and motivate learning by giving interesting topics and findings from existing research in various fields. First, fundamental concepts of human genome are introduced by the first chapter in the Introduction to Statistical Analysis (2020) to help students better understand the required readings. The first week's readings include important summary pieces discussing general challenges and opportunities in these fields. For example, Plomin (1994) summarized that complex human behaviors are caused and influenced by a large number of genes with very small effects. This genetic basis is further proved (Boyle, & Pritchard, 2017), and becomes a big concern for power issues (small effects). The optional readings provide examples of sociogenomic studies in different social science disciplines, including health, psychology, economics, sociology and politics. Some optional readings would require knowledge that will be covered in later weeks, so these readings are only included to show what kind of studies and topics that are being explored using genetic data in social science.

Questions for students:

- Try to propose three research questions that you are interested to study in sociogenomics (regardless of data availability or lack of analysis skills).

Required Readings:

Mills, M. C., Barban, N., & Tropf, F. C. (2020) Chapter 1 Introduction: Fundamental Concepts and the Human Genome. *An Introduction to Statistical Genetic Data Analysis*. The MIT Press, pp. 1-32.

Conley, D., & Fletcher, J. (2017). Chapter 1 MOLECULAR ME**.** *The Genome Factor: What the social genomics revolution reveals about ourselves, our history, and the future*. Princeton University Press, pp. 1-11.

Plomin, R., Owen, M. J., & McGuffin, P. (1994). The genetic basis of complex human behaviors. *Science*, *264*(5166), 1733-1739.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current directions in psychological science*, *9*(5), 160-164.

Freese, J. (2008). Genetics and the social science explanation of individual outcomes. *American Journal of Sociology*, *114*(S1), S1-S35.

Conley, D. (2009). The promise and challenges of incorporating genetic data into longitudinal social science surveys and research. *Biodemography and Social Biology*, *55*(2), 238-251.

Benjamin, D. J., Cesarini, D., Chabris, C. F., Glaeser, E. L., Laibson, D. I., Age, G., ... & Smith, A. V. (2011). The promises and pitfalls of genoeconomics. *Annual Review of Economics*, 2012(4:1), 627-662.

Optional Readings:

Robinson, G. E., Grozinger, C. M., & Whitfield, C. W. (2005). Sociogenomics: social life in molecular terms. *Nature Reviews Genetics*, *6*(4), 257-270.

Health:

Thompson, O. (2014). Genetic mechanisms in the intergenerational transmission of health. *Journal of Health Economics*, *35*, 132-146.

Psychology:

Tucker-Drob, E. M., Briley, D. A., & Harden, K. P. (2013). Genetic and environmental influences on cognition across development and context. *Current directions in psychological science*, *22*(5), 349-355.

Sociology:

Boardman, J. D., Domingue, B. W., & Fletcher, J. M. (2012). How social and genetic factors predict friendship networks. *Proceedings of the National Academy of Sciences*, *109*(43), 17377-17381.

Conley, D. & Domingue, B. (2016). "The Bell Curve Revisited: Testing Controversial Hypotheses with Molecular Genetic Data," *Sociological Science* 3, 520–539.

Belsky D. W., Domingue B. D., Weedow R., Arseneault L., Boardman J., Caspi A., …& Harris K. M. (2018). "Genetic analysis of social-class mobility in five longitudinal studies." *Proceedings of the National Academy of Sciences*, 115(31):E7275-E7284.

Economics:

Conley, D., & Fletcher, J. (2017). Chapter 6 The wealth of nations: something in our genes?. *The Genome Factor: What the social genomics revolution reveals about ourselves, our history, and the future*. Princeton University Press, pp. 113-135.

Fletcher, J. M. (2011). The promise and pitfalls of combining genetic and economic research. *Health economics*, *20*(8), 889-892.

Ashraf, Q., & Galor, O. (2013). The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, *103*(1), 1-46.

Guedes, J. D. A., Bestor, T. C., Carrasco, D., Flad, R., Fosse, E., Herzfeld, M., ... & Patterson, N. (2013). Is poverty in our genes? A critique of Ashraf and Galor," The'out of Africa'hypothesis, human genetic diversity, and comparative economic development," *Current Anthropology*, *54*(1), 71-79.

Politics:

Benjamin, D.J., Cesarini, D., van der Loos, M.J.H., Dawes, C.T., et al. (2012). "The genetic architecture of economic and political preferences," *PNAS,* 109 (21).

Smith, Kevin, John R. Alford, Peter K. Hatemi, Lindon J. Eaves, Carolyn Funk, and John R. Hibbing. 2012. "Biology, ideology, and epistemology: How do we know political attitudes are inherited and why should we care?" American Journal of Political Science 56:17-33.

Fowler, James H., and Christopher T. Dawes. "In defense of genopolitics." American Political Science Review 107.02 (2013): 362-374.

Week 2 Topic: Fundamental concepts in genetics

Main goals:

Students could

- Define, recognize and describe the fundamental terminology used in genetics
- Understand the significant drop in costs in genomic data and the limitation of genotyping arrays and gain a basic understanding of genotyping and sequencing technologies that produce genomic data
- Define genetic polymorphisms and the term allele, SNP, MAF, and the unique identifier
- Comprehend the organization of DNA in the nucleus of a human cell and the terms genome, gene and chromosome
- Acquire a basic knowledge of transcriptional regulations in molecular biology
- Identify the fundamentals of evolution, natural selection, fitness, types of selection, and related terminology
- Comprehend how evolution can also concur via genetic drift in the form of a bottleneck or founder effects
- Understand the assumptions, notation, and implications of the Hardy-Weinberg equilibrium
- Grasp the basics about linkage disequilibrium and haplotype blocks

Discussion of the readings:

This week's readings provide essential genetic knowledge and concepts to students who have minimal biology training to conduct interdisciplinary research. First, fundamental concepts (e.g., what are DNA, genes, SNPs, and MAF) introduced in Week 1 are reviewed in the first article. Second, a primer of human evolution, including the Hardy-Weinberg Equilibrium and linkage disequilibrium is introduced. The optional readings provide further information about human genetic variation, genotype, haplotype, and copy-number variation and sequencing techniques. Third, chapters from Molecular Biology of the Gene textbook introduced the transcriptional regulation of genes and the techniques that are used for sequencing to give students a better idea of how genes work and how genetic information is collected. This reading might be hard for students without much biological background. Students can study materials from Khan Academy to better understand the gene expression central dogma.

Questions for students:

- Use Ensembl browsers and (https://useast.ensembl.org/index.html) and SNPedia (https://www.snpedia.com/index.php/SNPedia) to explore any gene or SNP you are interested and find out the relationship between genotype and phenotype.

For example:

- rs7412 and rs429358 can raise the risk of Alzheimer's disease by more than 10x
- rs6152 can influence baldness
- rs333 resistance to HIV
- rs1805007 determines red hair and sensitivity to anesthetics
- rs9939609 triggers obesity and type-2 diabetes
- rs662799 prevents weight gain from high fat diets
- rs7495174 green eye color and rs12913832 for blue eye color
- rs7903146 in 3% of the population greatly increases the risk of type-2 diabetes
- rs12255372 linked to type-2 diabetes and breast cancer
- rs1799971 makes alcohol cravings stronger
- rs17822931 determines earwax, sweating and body odor
- rs1051730 and rs3750344 nicotine dependence
- rs4988235 lactose intolerance

- Explore the International HapMap project: https://www.genome.gov/10001688/international-hapmap-project

Required Readings:

Attia, J., Ioannidis, J. P., Thakkinstian, A., McEvoy, M., Scott, R. J., Minelli, C., ... & Guyatt, G. (2009). How to use an article about genetic association: A: Background concepts. *Jama*, *301*(1), 74-81.

Mills, M. C., Barban, N., & Tropf, F. C. (2020) Chapter 3 A Primer in Human Evolution. *An Introduction to Statistical Genetic Data Analysis*. The MIT Press, pp. 55-76. [Review Chapter 1]

Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. *Nature*, *470*(7333), 198.

Wall, J. D., & Pritchard, J. K. (2003). Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews Genetics*, *4*(8), 587-597.

Watson, J. D., Gann, A., Baker, T. A., Levine, M., Bell, S. B., Losick, R., & Harrison, S. C. (2014). Chapter 18. Transcriptional Regulation in Prokaryotes & Chapter 19. Transcriptional Regulation in Eukaryotes. *Molecular biology of the gene*. Boston: Benjamin-Cummings Publishing Company. 615-635 & 657-700. **[Hard reading - can start from videos on Khan academy about gene expression central dogma from DNA to protein - https://www.khanacademy.org/science/biology/gene-expression-central-dogma]**

Optional Readings:

Wiki: DNA profiling (https://en.wikipedia.org/wiki/DNA_profiling) and Human Genetic Variation (https://en.wikipedia.org/wiki/Human_genetic_variation)

Watson, J. D., Gann, A., Baker, T. A., Levine, M., Bell, S. B., Losick, R., & Harrison, S. C. (2014). Chapter 7. Techniques of Molecular Biology. *Molecular biology of the gene*. Boston: Benjamin-Cummings Publishing Company. 158-168.

Watson, J. D., Gann, A., Baker, T. A., Levine, M., Bell, S. B., Losick, R., & Harrison, S. C. (2014). Chapter 21. Gene Regulation in Development and Evolution. Transcriptional Regulation in Eukaryotes. *Molecular biology of the gene*. Boston: Benjamin-Cummings Publishing Company. 733-774.

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS genetics*, *4*(2), e1000008.

Check, E. (2006). Human evolution: how Africa learned to love the cow. *Nature,* 444: 994-996.

López, S., Van Dorp, L., & Hellenthal, G. (2015). Human dispersal out of Africa: a lasting debate. *Evolutionary Bioinformatics,* 11, EBO-S33489.

Lachance, J., Vernot, B., Elbers, C. C., Ferwerda, B., Froment, A., Bodo, J. M., ... & Zhang, K. (2012). Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell*, *150*(3), 457-469. (one of first studies in population genomics using whole-genome sequencing data)

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., ... & Ibrahim, M. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature genetics*, *39*(1), 31.

Ardlie, K. G., Kruglyak, L., & Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics*, *3*(4), 299.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., ... & Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature*, *411*(6834), 199.

Conrad, D. F., Jakobsson, M., Coop, G., Wen, X., Wall, J. D., Rosenberg, N. A., & Pritchard, J. K. (2006). A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nature genetics*, *38*(11), 1251.

Jakobsson, M., Scholz, S. W., Scheet, P., Gibbs, J. R., VanLiere, J. M., Fung, H. C., ... & Bras, J. M. (2008). Genotype, haplotype and copy-number variation in worldwide human populations. *Nature*, *451*(7181), 998-1003.

Hinds, D. A., Kloek, A. P., Jen, M., Chen, X., & Frazer, K. A. (2006). Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nature genetics*, *38*(1), 82.

Week 3 Topic: Classical Family studies

Main goals:

Students could

- Gain an overview of genetic recombination, linkage, and segregation analysis as molecular genetic approaches using family pedigree information
- Gain a basic knowledge of transmission test for linkage disequilibrium
- Grasp the study design of classical family studies (i.e., twin studies, adoption, and sibling studies) in behavioral genetics.
- Define and recognize the concept of heritability in family studies [The fundamental concept of heritability was introduced in Week 1]
- Understand and evaluate the equal environment assumption in twin studies
- Recognize the pros and cons of classical family studies.

Discussion of the readings:

We start exploring toolboxes of sociogenomics from family studies to unrelated individuals in association studies (Week 4). First, this week's readings introduce molecular genetic approaches using the information of family pedigree - linkage analysis and transmission tests for linkage disequilibrium. For many years, genetic linkage analysis can be used as a tool for estimating the genetic distance between two loci, and it was the primary tool used for the genetic mapping of Mendelian and complex traits with familial aggregation (Ott, Wang, & Leal, 2015). Second, the readings introduce behavioral genetic studies using the information of shared genes within families - classical twin study, sibling study, and adoption studies. Since monozygotic twins share 100% of their genes, and dizygotic twins and siblings share on average 50% of their genes, classical twin study starts from assessing the variance of a phenotype in a group and attempts to estimate how much of the variance is due to genetic effects (heritability), shared common environments, and non-shared environment (ACE model). Lastly, the heritability, the violation of the equal environment assumption of twin study, and other biases in family studies are evaluated.

Questions for students:

- Draw a pedigree of your family with known members and diseases
- Examine twin-heritability of two traits of your interest using the application MaTCH(Meta-Analysis of Twin Correlations and Heritability) accessible via http://match.ctglab.nl/#/home Gephi, https://gephi.org/

Required Readings:

Molecular Genetics:

Spielman, R. S., McGinnis, R. E., & Ewens, W. J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American journal of human genetics*, *52*(3), 506-516.

Easton, D. F., Bishop, D. T., Ford, D., & Crockford, G. P. (1993). Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. *American journal of human genetics*, *52*(4), 678-701.

Behavioral Genetics:

Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature reviews genetics*, *3*(11), 872 -882.

Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Archives of general psychiatry*, *60*(12), 1187-1192.

Van Dongen, J., Slagboom, P. E., Draisma, H. H., Martin, N. G., & Boomsma, D. I. (2012). The continuing value of twin studies in the omics era. *Nature Reviews Genetics*, *13*(9), 640-653.

Stunkard, A. J., Sørensen, T. I., Hanis, C., Teasdale, T. W., Chakraborty, R., Schull, W. J., & Schulsinger, F. (1986). An adoption study of human obesity. *New England Journal of Medicine*, *314*(4), 193-198.

Polderman, T.J.C., Benyamin, B, de Leeuw, C.A, Sullivan, P.F. et al. (2015). "Meta-analysis of the heritability of human traits based on fifty year of twin studies," *Nature Genetics* 47 (7), 702-709.

Felson, J. (2014). What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption. *Social Science Research*, *43*, 184-199.

Optional Readings:

Molecular Genetics:

Morton, N. E., & MacLean, C. J. (1974). Analysis of family resemblance. 3. Complex segregation of quantitative traits. *American journal of human genetics*, *26*(4), 489-503.

Spielman, R. S., & Ewens, W. J. (1998). A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *The American Journal of Human Genetics*, *62*(2), 450-458.

Martin, E. R., Monks, S. A., Warren, L. L., & Kaplan, N. L. (2000). A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *The American Journal of Human Genetics*, *67*(1), 146-154.

Ott, J., Wang, J., & Leal, S. M. (2015). Genetic linkage analysis in the age of whole-genome sequencing. *Nature Reviews Genetics*, *16*(5), 275.

Behavioral Genetics:

Pocock, N. A., Eisman, J. A., Hopper, J. L., Yeates, M. G., Sambrook, P. N., & Eberl, S. (1987). Genetic determinants of bone mass in adults. A twin study. *The Journal of clinical investigation*, *80*(3), 706-710.

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological medicine*, *25*(1), 63-77.

Levy, F., Hay, D. A., McStephen, M., Wood, C., & Waldman, I. (1997). Attention-deficit hyperactivity disorder: a category or a continuum? Genetic analysis of a large-scale twin study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(6), 737-744.

Keller, M. C., Coventry, W. L., Heath, A. C., & Martin, N. G. (2005). Widespread evidence for non-additive genetic variation in Cloninger's and Eysenck's personality dimensions using a twin plus sibling design. *Behavior genetics*, *35*(6), 707.

Frisell, T., Öberg, S., Kuja-Halkola, R., & Sjölander, A. (2012). Sibling Comparison Designs: Bios from Non-Shored Confounders and Measurement Error. *Epidemiology*, 713-720.

Bokhorst, C. L., Bakermans-kranenburg, M. J., Pasco Fearon, R. M., Van Ijzendoorn, M. H., Fonagy, P., & Schuengel, C. (2003). The importance of shared environment in mother–infant attachment security: A behavioral genetic study. *Child Development*, *74*(6), 1769-1782.

Cesarini, D., Johannesson, M, and Oskarsson, S. (2014). "Pre-Birth Factors, Post-Birth Factors, and Voting: Evidence from Swedish Adoption Data," *American Political Science Review* 108 (1).]

Conley, D., Rauscher, E., Dawes, C., Magnusson P.K.E. and Siegel, M.L. (2013). "Heritability and the Equal Environments Assumption: Evidence from Multiple Samples of Misclassified Twins," *Behavior Genetics* 43(4).

Week 4 Topic: Gene discovery: GWAS and statistical issues

Main goals:
Students could

- Understand research design, meta-analysis, and a data analysis plan of genome-wide association studies (GWAS).
- Know the fundamental aspects of statistical inference, methods, and heterogeneity for GWAS
- Compare GWAS to candidate gene studies to understand the importance of statistical power, replicability and robustness
- Understand the basics of imputation and quality control in GWAS
- Gain knowledge of the NHGRI-EBI GWAS Catalog and GIANT Consortium for an overview of GWAS
- Become aware of the current progress and future directions in the area of research

Discussion of the readings:
Association studies of unrelated individuals have been widely used for studying complex traits. There is a shift from the candidate gene approach to the genome-wide association study (GWAS), as more of a genetic basis is understood, and poor replicability of the candidate gene approach is criticized. GWAS with large sample size and stringent p-value threshold becomes the norm. Required readings include examples of large-scale GWAS studies and discussion of statistical issues, including power analysis, replicability, and robustness. The optional readings also include similar approach phenome-wide association studies in functional genomics, and more discussion about statistical issues, such as Winner's curse.

Questions for students:

- Choose a phenotype you are interested in and explore what GWAS of that phenotype has already been done. Check the GWAS catalog and examine the diagram (https://www.ebi.ac.uk/gwas/diagram) that contains various traits that have been studied to data at each chromosome.
- Explore GIANT Consortium. Use the data files and the LocusZoom to create a regional association plot of your interest. https://portals.broadinstitute.org/collaboration/giant/index.php/Main_Page
- Explore GWAS study on Ben Neale's site using the UK Biobank (http://www.nealelab.is/uk-biobank)

Required Readings:

Mills, M. C., Barban, N., & Tropf, F. C. (2020) Chapter 4 Genome-Wide Association Studies. *An Introduction to Statistical Genetic Data Analysis*. The MIT Press, pp. 77-100.

Hirschhorn, J., & Daly, M. (2005) Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics, 6, 95-108.

Lee, J. J., Wedow, R., Okbay, A. et al. (2018). "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals (& Supplementary Information, **Section 2.4-2.6, 23-26**)," *Nature Genetics*, 50, 1112-1121.

Tabor, H. K., Risch, N. J., & Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics*, *3*(5), 391-397.

de Bakker, P. I., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J., & Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nature genetics*, *37*(11), 1217-1223.

Rietveld, C. A., Conley, D., Eriksson, N, Esko, T. et al. (2014). "Replicability and Robustness of Genome-Wide Association Studies for Behavioral Traits," **(& Supplemental Material, 1-16**), *Psychological Science* 25 (11), 1975–1986.

Gelman, A., and Carlin, J. (2014). "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors," *Perspectives on Psychological Science*, 9(6), 641–651.

Sham, P. C., & Purcell, S. M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*, *15*(5), 335-346.

Evangelou, E., & Ioannidis, J. P. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*, *14*(6), 379-389.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics*, *101*(1), 5-22.

Optional Readings:

Cordell, H. J., & Clayton, D. G. (2005). Genetic association studies. *The Lancet*, *366*(9491), 1121-1131.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., ... & Pendlington, Z. M. (2016). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research*, *45*(D1), D896-D901.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., ... & Albrecht, E. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment (& Supplementary Information, **Section 1-5, 3-21)**. *science*, *340*(6139), 1467-1471.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., ... & Oskarsson, S. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539.

Okbay, A., Baselmans, B. M., De Neve, J. E., Turley, P., Nivard, M. G., Fontana, M. A., ... & Gratten, J. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature genetics*, *48*(6), 624.

Yang, J., Loos, R. J., Powell, J. E., Medland, S. E., Speliotes, E. K., Chasman, D. I., ... & Waite, L. (2012). FTO genotype is associated with phenotypic variability of body mass index. *Nature*, *490*(7419), 267.

Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benyamin, B., ... & De Leeuw, C. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method (& Supplemental **Section 8** "Correction of Effect Sizes for Winner's Curse") *Proceedings of the National Academy of Sciences*, *111*(38), 13790-13794.

Yang, J., Wray, N. R., & Visscher, P. M. (2010). Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, *34*(3), 254-257.

Fadista, J., Manning, A. K., Florez, J. C., & Groop, L. (2016). The (in) famous GWAS P-value threshold revisited and updated for low-frequency variants. *European Journal of Human Genetics*, *24*(8), 1202.

Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D., ... & Bitton, A. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature genetics*, *40*(8), 955.

Chabris, C. F., et al. (2012). "Most Reported Genetic Associations with General Intelligence Are Probably False Positives." *Psychological Science*, 23(11), 1314-1323.

Sebastiani, P., Solovieff, N., DeWan, A. T., Walsh, K. M., Puca, A., Hartley, S. W., ... & Myers, R. H. (2012). Genetic signatures of exceptional longevity in humans. *PloS one*, *7*(1), e29848.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., ... & Parkinson, H. (2013). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research*, *42*(D1), D1001-D1006.

Roden, D. M. (2017). Phenome-wide association studies: a new method for functional genomics in humans. *The Journal of physiology*, *595*(12), 4109-4115.

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., ... & Zhan, X. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature genetics*, *51*(2), 237.

Mills, M. C., & Rahal, C. (2019). A scientometric review of genome-wide association studies. *Communications biology*, *2*(1), 1-11.

Week 5 Topic: Population stratification

Main goals:

Students could:

- Grasp the concept and the common misnomers of population stratification
- Know common approaches to detect and control for population stratification
- Understand principal component analysis (PCA) and know how to interpret results from PCA
- Understand that ancestry does not equate to the socially constructed category of race, which is not a biological category
- Realize how genes mirror geography
- Know the basics of LD score regression

Discussion of the readings:
Population stratification refers to differences in allele frequencies between subpopulations in a population due to systematic differences in ancestry rather than association of genes with phenotypes (Freedman et al, 2004). Genomic control, structural association, and family-based associate tests are introduced as approaches to detect and control for population stratification in the required readings. LD score regression can separate contributions of polygenic effects and various confounding factors, such as population stratification, based on summary statistics from genome-wide association studies (GWASs) (Bulik-Sullivan et al, 2015). Currently, principal component analysis is widely used to control population stratification and decide individuals' ancestries. Examples include the widely-cited paper that used PCA to find genes mirror geography in Europe (Novembre et al., 2008). Further, the week's readings also include the discussion about the differences between racial and genetic ancestry, and recent evidence that previous analyses of polygenic adaptation of height were confounded by population stratification.

Questions for students:

- Go to the website that accompanies the Hellenthal et al. (2014) Science article: http://admixturemap.paintmychromosomes.com/the two tutorials under the "Historical Event" menu. Click on a labeled population on the map or selection from the "Target Population" drop-down menu. Examine some of the details of the past admixture events that they use to infer how that population has been formed. Pick one historical event and read the historical interpretation of the admixture signals.
- Explore ANCESTRYMAP 2.0 from https://reich.hms.harvard.edu/software (Patterson et al., 2004) finds skews in ancestry that is potentially associated with disease genes in recently mixed populations like African Americans.
- Go to the LD Hub(http://ldsc.broadinstitute.org/ldhub/), and explore Lookup Center and GWAShare Center. Then sing in with your Google account and go to Test Center following the three steps available on the automatic LD score regression platform (using input example data at Step 1 or other data that you have access to).

Required Readings:

Hamer, D. H. and Sirota, L. (2000). "Beware the chopsticks gene," *Molecular Psychiatry* 5, 11–13.

Cardon, L. R., & Palmer, L. J. (2003). Population stratification and spurious allelic association. *The Lancet*, *361*(9357), 598-604.

Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, *55*(4), 997-1004.

Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N., ... & Pato, M. T. (2004). Assessing the impact of population stratification on genetic association studies. *Nature genetics*, *36*(4), 388-393.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M.E. (2006). "Principal components analysis corrects for stratification in genome-wide association studies," *Nature Genetics* 38(8), 904–909.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., ... & Stephens, M. (2008). Genes mirror geography within Europe. *Nature*, *456*(7218), 98-101.

Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, *11*(7), 459-463.

Hellenthal, G., Busby, G. B., Band, G., Wilson, J. F., Capelli, C., Falush, D., & Myers, S. (2014). A genetic atlas of human admixture history. *Science*, 343(6172), 747-751.

Conley, D., & Fletcher, J. (2017). Chapter 5 IS RACE GENETIC? A NEW TAKE ON THE MOST FRAUGHT, DISTRACTING, AND NONSENSICAL QUESTION IN THE WORLD. *The Genome Factor: What the social genomics revolution reveals about ourselves, our history, and the future*. Princeton University Press, pp. 35-59.

Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., ... & Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*, *47*(3), 291-295.

Optional Readings:

Pritchard, J. K., & Rosenberg, N. A. (1999). Use of unlinked genetic markers to detect population stratification in association studies. *The American Journal of Human Genetics*, *65*(1), 220-228.

Wacholder, S., Rothman, N., & Caporaso, N. (2000). Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *Journal of the National Cancer Institute*, *92*(14), 1151-1158.

Reich, D. E., & Goldstein, D. B. (2001). Detecting association in a case-control study while correcting for population stratification. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, *20*(1), 4-16.

Wacholder, S., Rothman, N., & Caporaso, N. (2002). Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. *Cancer Epidemiology and Prevention Biomarkers*, *11*(6), 513-520.

Patterson, N., Hattangadi, N., Lane, B., Lohmueller, K. E., Hafler, D. A., Oksenberg, J. R., ... & Daly, M. J. (2004). Methods for high-density admixture mapping of disease genes. *The American Journal of Human Genetics*, *74*(5), 979-1000.

Campbell, C. D., Ogburn, E. L., Lunetta, K. L., Lyon, H. N., Freedman, M. L., Groop, L. C., ... & Hirschhorn, J. N. (2005). Demonstrating stratification in a European American population. *Nature genetics*, *37*(8), 868.

Witherspoon, D. J., Wooding, S., Rogers, A. R., Marchani, E. E., Watkins, W. S., Batzer, M. A., & Jorde, L. B. (2007). Genetic similarities within and between human populations. *Genetics*, *176*(1), 351-359.

Guo, G., Fu Y., Lee, H., Cai, T. et al. (2014). "Genetic Bio-Ancestry and Social Construction of Racial Classification in Social Surveys in the Contemporary United States," *Demography* 51(1), 141–172.

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., ... & Kenny, E. E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *The American Journal of Human Genetics*, *100*(4), 635-649.

Sohail, M., Maier, R.M., Ganna, A., Bloemendal, A., Martin, A.R., Turchin, M.C. et al. (2019). "Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies," *eLife,* 8. DOI: 10.7554/eLife.39702

Week 6 Topic: Estimating heritability

Main goals:
Students could:

- Define the meaning of heritability and become aware of common misconceptions of heritability
- Define and recognize the problem of missing heritability
- Know possible explanations for missing and hidden heritability
- Understand genetic restricted maximum likelihood model (GREML)
- Grasp the basics of relatedness disequilibrium regression

Discussion of the readings:
Classical twin studies have high heritability estimates that gene association studies cannot recover. To compare heritability estimates between different study designs and discuss the complicated matters of estimating heritability, this week's readings first discuss the concept, assumptions, and misconceptions of heritability, and why heritability is "missing" or "hidden", and then introduce approaches using whole genome data to estimate the heritability, including structural equation models, genetic restricted maximum likelihood (GREML) model (Yang et al, 2011), LD score regression (Bulik-Sullivan et al, 2015 in Week 5) and relatedness disequilibrium regression (RDR) model (Young, 2018).

Questions for students

- Find heritability estimates from twin studies and GWAS (and GREML estimates if possible) of two phenotypes of your interest. Compare the estimates and report whether heritability is "missing".
- Summarize the pros and cons of each heritability estimating method introduced this week.

Required Readings:

Conley, D., & Fletcher, J. (2017). Chapter 3 IF HERITABILITY IS SO HIGH, WHY CAN'T WE FIND IT? *The Genome Factor: What the social genomics revolution reveals about ourselves, our history, and the future*. Princeton University Press, pp. 35-59.

Goldberger, A.S. (2005). "Structural Equation Models in Human Behavior Genetics." *Identification and Inference for Econometric Models Essays in Honor of Thomas Rothenberg*, eds. DW Andrews and JS Stock. Cambridge: Cambridge University Press, 11–26.

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era—concepts and misconceptions. *Nature reviews genetics*, *9*(4), 255-266.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., ... & Cho, J. H. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747-753.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., ... & Goddard, M. E. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature genetics*, *42*(7), 565-569.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, *88*(1), 76-82.

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*, *109*(4), 1193-1198.

Yang, J., Zeng, J., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2017). Concepts, estimation and interpretation of SNP-based heritability. *Nature genetics*, *49*(9), 1304-1310.

Young, A. (2018). "Relatedness disequilibrium regression estimates heritability without environmental bias," *Nature Genetics* 50 (9), 1304-1310.

Evans, L. M., Tahmasbi, R., Vrieze, S. I. et al. (2018). "Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits," *Nature Genetics*, 50, 737-745. **Table 1**

Optional Readings:

Rijsdijk, F. V., & Sham, P. C. (2002). Analytic approaches to twin data using structural equation models. *Briefings in bioinformatics*, *3*(2), 119-133.

Visscher, P. M., Medland, S. E., Ferreira, M. A., Morley, K. I., Zhu, G., Cornes, B. K., ... & Martin, N. G. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS genetics*, *2*(3), e41.

Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature News*, *456*(7218), 18-21.

Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., & Wallace, B. (2008). Heritability of cooperative behavior in the trust game. *Proceedings of the National Academy of sciences*, *105*(10), 3721-3726.

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. Nature Reviews Genetics, 11(6), 446-450.

Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2011). Estimating missing heritability for disease from genome-wide association studies. *The American Journal of Human Genetics*, *88*(3), 294-305.

Lee, J. J., & Chow, C. C. (2014). Conditions for the validity of SNP-based heritability estimation. *Human genetics*, *133*(8), 1011-1022.

Conley, D., Siegal, M. L., Domingue, B. W., Harris, K. M., McQueen, M. B., & Boardman, J. D. (2014). Testing the key assumption of heritability estimates based on genome-wide genetic relatedness. *Journal of human genetics*, *59*(6), 342-345.

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A., Lee, S. H., ... & Snieder, H. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature genetics*, *47*(10), 1114.

Zhu, Z., Bakshi, A., Vinkhuyzen, A. A., Hemani, G., Lee, S. H., Nolte, I. M., ... & Mägi, R. (2015). Dominance genetic variation contributes little to the missing heritability for human complex traits. *The American Journal of Human Genetics*, *96*(3), 377-385.

Zheng, J., Erzurumluoglu, A. M., Elsworth, B. L., Kemp, J. P., Howe, L., Haycock, P. C., ... & Warrington, N. M. (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*, *33*(2), 272-279.

Tropf, F. C., Lee, S. H., Verweij, R. M., Stulp, G., Van Der Most, P. J., De Vlaming, R., ... & Iliadou, A. N. (2017). Hidden heritability due to heterogeneity across seven populations. *Nature human behaviour*, *1*(10), 757-765.

de Vlaming, R., Okbay, A. Rietveld, C.A. et al. (2017). "Meta-GWAS Accuracy and power calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies," *PLOS Genetics*, 13(1): e1006495.

Young, A. "Heritability estimation: the central problem" https://geneticvariance.wordpress.com/2017/10/09/heritability-estimation-the-central-problem-ii/

Young, A. "Relatedness disequilibrium regression explained" https://geneticvariance.wordpress.com/2018/08/13/relatedness-disequilibrium-regression-explained/

Young, A. "Missing heritability revisited" https://geneticvariance.wordpress.com/2019/04/10/missing-heritability-revisited/

Week 7 Topic: Working with Genetic Data I- GWAS

Main goals:

Students could

- Grasp the different formats of genomic data in the computer program PLINK
- Open and work with PLINK binary files and recode PLINK files into other formats
- Understand the basics of data management to select information on particular markers or a subsample of individuals
- Derive information on allele frequencies, phenotypes, and missing values and run simple quality control analysis
- Merge different genetic files
- Associate a phenotype to a PLINK file and run GWAS
- Know about the most prominently used data (e.g., UKBiobank, WLS, HRS, AddHealth, etc.) in sociogenomics for genome-wide association study discovery

Discussion of the readings:

The goal of this week is to give students a brief hands-on experience working with genetic data. Students will learn essential tools and knowledge to conduct GWAS. How to construct and work with polygenic scores will be introduced in Week 10. The focus of this week's reading is Part II of the book *An Introduction to Statistical Genetic Data Analysis* with sample codes and step-by-step instructions for using PLINK. Other readings include an overview of quality control, and imputation for GWAS. The optional readings include more information about PLINK and quality control, and introduction papers for widely used datasets in sociogenomics, such as UK Biobank, Add Health, WLS, HRS, etc.

Questions for students:

- How many columns and rows would a .ped file have for a sample of 100 individuals with 500,000 genotype variants?
- Download the HapMap-ceu set of PLINK binary files from the following link http://zzz.bwh.harvard.edu/plink/dist/example.zip. Explore the .fam and .bim, and describe the content of each column, and then do the following exercises:

1. From the raw SNP genotype file, create a compact binary fileset
2. Generate allele frequencies and missing values for all SNPs
3. Calculate Hardy-Weinberg equilibrium *P*-values for all SNPs
4. Create a QC-filter dataset with the following criteria:
    - Call rate > 95%
    - Minor Allele frequency > 1%
    - Hardy-Weinberg equilibrium test < 0.0001
5. Perform the basic association analysis of the QC-filtered dataset to produce a Manhattan plot

Required Readings:

Mills, M. C., Barban, N., & Tropf, F. C. (2020) Part II Working with Genetic Data (Chapter 7 Genetic Data and Analytical Challenges; Chapter 8 Data Management, Descriptive Statistics, and Quality Control; Chapter 9 Association Analysis, Population stratification, and genetic relatedness) *An Introduction to Statistical Genetic Data Analysis*. The MIT Press, pp. 151-242.

Marchini, J. and Howie, B. (2010). "Genotype imputation for genome-wide association studies," *Nature Reviews Genetics* 11, 499-511.

Winkler, T. W., et al. (2014). "Quality control and conduct of genome-wide association meta-analyses." *Nature Protocols*, 9(5): 1192–1212.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, et al. (2016). "Genome-wide association study identifies 74 loci associated with educational attainment," *Nature*. Supplemental Section 1.5 ("Quality Control").

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, *81*(3), 559-575.

1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68-74.

Optional Readings:

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), 7.

Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., & Zondervan, K. T. (2010). Data quality control in genetic case-control association studies. *Nature protocols*, *5*(9), 1564.

Marees, A. T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International journal of methods in psychiatric research*, *27*(2), e1608.

K. M. Harris et al., The National Longitudinal Study of Adolescent to Adult Health: Research design (2009) (available at https://www.cpc.unc.edu/projects/addhealth/design)

Herd, P., Carr, D., & Roan, C. (2014). Cohort profile: Wisconsin longitudinal study (WLS). *International journal of epidemiology*, *43*(1), 34-41.

Sonnega, A., Faul, J. D., Ofstedal, M. B., Langa, K. M., Phillips, J. W., & Weir, D. R. (2014). Cohort profile: the health and retirement study (HRS). *International journal of epidemiology*, *43*(2), 576-585.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., ... & Allen, N. E. (2017). Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American journal of epidemiology*, *186*(9), 1026-1034.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., ... & Cortes, A. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., ... & Liu, B. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, *12*(3), e1001779.

Week 8 Topic: Polygenic scores (PGS) I

Main goals:
Students could:

- Define and understand the origins of polygenic scores
- Understand how polygenic scores are constructed
- Know the basics of validation and prediction of polygenic scores
- Recognize the central challenges, why they are problematic, and potential solutions in working with polygenic scores
- Be introduced to applications of polygenic scores

Discussion of the readings:
This week's readings introduce how polygenic scores are constructed, used, and interpreted. The results from gene discovery are used to build PGS - a summary score of polygenic prediction of certain complex traits. the promise and pitfalls of PGS, and how PGSs are used and interpreted in sociogenomics. We will further discuss recent progress in improving polygenic scores and its application in gene-environment interaction studies in Week 10.

Questions for students:

- Think whether you can use PGS for the research questions you proposed in Week 1. If yes, what other information do you need? If not, state why and what genetic information do you need?
- Based on the readings, analyze the potential problems of using PGS in your research.

Required Readings:

Mills, M. C., Barban, N., & Tropf, F. C. (2020) Chapter 5 Introduction to Polygenic Scores and Genetic Architecture. *An Introduction to Statistical Genetic Data Analysis*. The MIT Press, pp. 101-128.

Lee, J.J., Wedow, R., Okbay, A. et al. (2018). "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals," *Nature Genetics*, 50, 1112-1121. **Supplementary Information, Section 6**

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*(7256), 748-752.

Dudbridge, F. (2013). "Power and predictive accuracy of polygenic risk scores." *PLoS Genet* 9(3), 1-17.

Wray, N. R., Yang, J., Hayes, B. J, Price, A. L. et al. (2013). "Pitfalls of predicting complex traits from SNPs," *Nature Reviews Genetics* 14, 507-515.

Vilhjálmsson, B.J., Yang J., Finucane, H.K., Gusev, A. et al. (2015). "Modeling linkage disequilibrium increases accuracy of polygenic risk scores." *The American Journal of Human Genetics* 97(4), 576 -592.

Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*, *19*(9), 581-590.

Optional Readings:

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, *273*(5281), 1516-1517.

Pharoah, P. D., Antoniou, A., Bobrow, M., Zimmern, R. L., Easton, D. F., & Ponder, B. A. (2002). Polygenic susceptibility to breast cancer and implications for prevention. *Nature genetics*, *31*(1), 33.

Daetwyler, H.D., Villaneuva, B., and Woolliams, J.A. (2008). "Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach," *PLoS One* 3(10).

Plomin, R., Haworth, C. M., & Davis, O. S. (2009). Common disorders are quantitative traits. *Nature Reviews Genetics*, *10*(12), 872.

Palla, L., & Dudbridge, F. (2015). A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *The American Journal of Human Genetics*, *97*(2), 250-259.

Krapohl, E., Euesden, J., Zabaneh, D., Pingault, J. B., Rimfeld, K., Von Stumm, S., ... & Plomin, R. (2016). Phenome-wide analysis of genome-wide polygenic scores. *Molecular psychiatry*, *21*(9), 1188

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., ... & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics*, *50*(9), 1219.

Francisco, M., & Bustamante, C. D. (2018). Polygenic risk scores: a biased prediction? *Genome medicine*, *10*(1), 100.

Week 9 Topic: Gene-environment interplay I - gene-environment interaction

Main goals:

Students could

- Understand and differentiate between types of the gene-environment interplay of gene-environment interaction (G×E) and gene-environment correlation (rGE)
- Understand the multiple ways of defining the environment, including by multi-levels, domains, and temporal aspects
- Recognize the history of G×E studies and common errors, from classic approaches to candidate gene and more recent genome-wide approaches
- Comprehend and differentiate between the central theoretical G*E models of diathesis-stress, differential susceptibility, bioecological (social compensation), and social control
- Grasp potential future directions in studying gene-environment interactions

Discussion of the readings:

The interplay between gene and environment, including gene-environment interaction (G×E) and gene-environment correlation (rGE), is of great interest to social scientists and has direct implications to personalized policy interventions. The majority of the research in this area examines G×E, which studies whether the effect of genotype on a phenotype varies across different environments. This effect could be causal or non-causal. Making any causal claim from G×E studies is still a big challenge, which will be further discussed in Week 12. Gene-environment correlation is the process by which an individuals' genotype influences or is associated with exposures to the environment. Week 11 will include more studies on rGE. Apart from the introduction and summary papers of gene-environment interplay and the problems of defining environment, this week's required readings also include the classical studies of G×E, such as genotype as moderator on individuals' risk for depression under life stress (Caspi, et al, 2003), and theoretical explanations of G×E models such as diathesis-stress and differential susceptibility hypotheses. The optional readings include more recent G*E studies.

Questions for students:

- Think of two research questions about gene-environment interaction. Review the literature and see whether these questions have been studied. If yes, what data and methods were used?
- Think of an example of the differential susceptibility model.

Required Readings:

Mills, M. C., Barban, N., & Tropf, F. C. (2020) Chapter 6 Gene-Environment Interplay. *An Introduction to Statistical Genetic Data Analysis*. The MIT Press, pp. 129-150.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological bulletin*, *84*(2), 309-322.

Boardman, J. D., Daw, J., & Freese, J. (2013). Defining the environment in gene–environment research: Lessons from social epidemiology. *American journal of public health*, *103*(S1), S64-S72.

Caspi, A, Sugden, K., Moffitt T.E., Taylor A., et al. (2003). "Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene," *Science* 301 (5631), 386-389.

Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry*, *168*(10), 1041-1049.

Thomas, D. (2010). Gene–environment-wide association studies: emerging approaches. *Nature Reviews Genetics*, *11*(4), 259-272.

Keller, M. C. (2014). Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biological psychiatry*, *75*(1), 18-24.

Shanahan, M. J., & Hofer, S. M. (2005). Social context in gene–environment interactions: Retrospect and prospect. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(Special_Issue_1), 65-76.

Belsky, J., Bakermans-Kranenburg, M. J., & Van IJzendoorn, M. H. (2007). For better and for worse: Differential susceptibility to environmental influences. *Current directions in psychological science*, *16*(6), 300-304.

Roisman, G. I., Newman, D. A., Fraley, R. C., Haltigan, J. D., Groh, A. M., & Haydon, K. C. (2012). Distinguishing differential susceptibility from diathesis–stress: Recommendations for evaluating interaction effects. *Development and psychopathology*, *24*(2), 389-409.

Optional Readings:

Via, S., & Lande, R. (1985). Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, *39*(3), 505-522.

Purcell, S. (2002). Variance components models for gene–environment interaction in twin analysis. *Twin Research and Human Genetics*, *5*(6), 554-5

Caspi A., McClay, J., Moffitt, T.E., Mill, J. et al. (2002). "Role of genotype in the cycle of violence in maltreated children," *Science* 297 (5582).

Eley, T. C., Sugden, K., Corsico, A., Gregory, A. M., Sham, P., McGuffin, P., ... & Craig, I. W. (2004). Gene–environment interaction analysis of serotonin system markers with adolescent depression. *Molecular psychiatry*, *9*(10), 908.

Kim-Cohen, J., Caspi, A., Taylor, A., Williams, B., Newcombe, R., Craig, I. W., & Moffitt, T. E. (2006). MAOA, maltreatment, and gene–environment interaction predicting children's mental health: new evidence and a meta-analysis. *Molecular psychiatry*, *11*(10), 903-913.

Caspi, A., Hariri, A. R., Holmes, A., Uher, R., & Moffitt, T. E. (2010). Genetic sensitivity to the environment: the case of the serotonin transporter gene and its implications for studying complex diseases and traits. *American journal of Psychiatry*, *167*(5), 509-527.

Karg, K., Burmeister, M., Shedden, K., & Sen, S. (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Archives of general psychiatry*, *68*(5), 444-454.

Reiss, D., Leve, L. D., & Neiderhiser, J. M. (2013). How genes and the social environment moderate each other. *American journal of public health*, *103*(S1), S111-S121.

Dick, D.M., Agrawal, A. Keller, M.C., Adkins A., et al. (2015). "Candidate gene–environment interaction research reflections and recommendations," *Perspectives on Psychological Science* 10 (1).

Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary–developmental theory of the origins and functions of stress reactivity. *Development and psychopathology, 17(2), 271-301.*

Mitchell, C., McLanahan, S., Brooks-Gunn, J., Garfinkel, I., Hobcraft, J., & Notterman, D. (2013). Genetic differential sensitivity to social environments: Implications for research. *American journal of public health, 103(S1), S102-S110.*

Belsky, J., Pluess, M., & Widaman, K. F. (2013). Confirmatory and competitive evaluation of alternative gene-environment interaction hypotheses. *Journal of Child Psychology and Psychiatry*, *54*(10), 1135-1143.

Duncan, L. E., Pollastri, A. R., & Smoller, J. W. (2014). Mind the gap: Why many geneticists and psychological scientists have discrepant views about gene–environment interaction (G× E) research. *American Psychologist*, *69*(3), 249.

Del Giudice, M. (2017). Statistical tests of differential susceptibility: Performance, limitations, and improvements. *Development and psychopathology*, *29*(4), 1267-1278.

Week 10 Topic: Working with genetic data II- PGS & G×E application

Main goals:
Students could:

- Recall the basics of polygenic scores(PGSs)
- Understand how to calculate a PGS using PRSice 2.0
- Know how to work with PGSs
- Become aware of the recent progress in improving polygenic scores
- Perform cross-trait prediction and estimate genetic covariation using PGSs
- Understand and estimate genetic confounding using PGSs
- Engage in an applied analysis of gene-environment interaction using PGSs

Discussion of the readings:
As polygenic scores are widely used in gene-environment interaction studies, students will have another hands-on experience with genetic data. This week's focus is constructing polygenic scores (PGSs) and using PGSs to study

gene-environment interactions (G×E). The readings continue the discussion of the pitfalls and caveats of using PGSs (in Week 8) and include the recent progress and updates of construction methods of PGS (e.g., Bayesian regression, MTAG, etc.), and the application of PGSs in G×E studies.

Questions for students:

- Follow the steps in this Basic Tutorial for Polygenic Risk Score analysis and replicate the scatter plot of polygenic score and height - https://choishingwan.github.io/PRS-Tutorial/.
- Explore HRS dataset and choose an available PGS of a phenotype of your interest (e.g., height) and try to answer following questions:
  1. What is the incremental R^2 of the PGS (e.g., height_PGS) predicting the phenotype (GWAS-heritability)?
  2. Does this PGS also predict other phenotypes in the same dataset? Refer to the cross-trait prediction discussion in section 11.2.2. in Chapter 11 of the book*An Introduction to Statistical Genetic Data Analysis*.
  3. Try to find another phenotype that has PGS, and explore whether the phenotypic association between these two phenotypes confounded by genes? Refer to the confounding discussion in section 11.2.3 as above.

Required Readings:

Mills, M. C., Barban, N., & Tropf, F. C. (2020) Part II Working with Genetic Data - Chapter 10 An Applied Guide to Creating and Validating Polygenic Scores & Part III Applications and Advanced Topics - Chapter 11 Polygenic Score and Gene-Environment Interaction Applications) *An Introduction to Statistical Genetic Data Analysis*. The MIT Press, pp. 243-313.

Choi, S. W., Mak, T. S. H., & O'reilly, P. (2018). A guide to performing Polygenic Risk Score analyses. *BioRxiv*, 416545, pp. 1-22.

Choi, S. W., & O'Reilly, P. F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *Giga Science*, *8*(7), giz082, pp. 1-6.

Mullins, N., Power, R. A., Fisher, H. L., Hanscombe, K. B., Euesden, J., Iniesta, R., ... & Uher, R. (2016). Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychological medicine*, *46*(4), 759-770.

Krapohl, E., Patel, H., Newhouse, S., Curtis, C. J., von Stumm, S., Dale, P. S., ... & Plomin, R. (2018). Multi-polygenic score approach to trait prediction. *Molecular psychiatry*, *23*(5), 1368-1374.

Martin, A.R., Daly, M.J. et al. (2019). "Predicting polygenic risk of psychiatric disorders." *Biological Psychiatry,* 97-109.

Optional Readings:

Loh, P. R., Tucker, G., Bulik-Sullivan, B. K., Vilhjalmsson, B. J., Finucane, H. K., Salem, R. M., ... & Patterson, N. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature genetics*, *47*(3), 284.

Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., ... & Magnusson, P. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature genetics*, *50*(2), 229-237.

Papageorge, N. and Thom, K. (2018). "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study." NBER Working paper. https://www.nber.org/papers/w25114]

Lee, J.J., Wedow, R., Okbay, A. et al. (2018). "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals," *Nature Genetics*, 50, 1112-1121. **Supplementary Information, Section 3.**

Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P., & Price, A. L. (2018). Mixed-model association for biobank-scale datasets. *Nature genetics*, *50*(7), 906.

Ge, T., Chen, C., Ni, Y. et al. (2019). "Polygenic prediction via Bayesian regression and continuous shrinkage priors," *Nature Communications* 10(1776).

Belsky, D.W., & Harden, K.P. (2019). "Phenotypic annotation: using polygenic scores to translate discoveries from genome-wide association studies from the top down," *Current Directions in Psychological Science*. Advance online publication. https://doi.org/10.1177%2F0963721418807729

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., ... & Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature communications*, *10*(1), 1-9.

Topic 11: Gene-Environment interplay II - Gene-Environment Correlation and Genetic Nurture

Main goals:
Students could

- Define and differentiate between different types of gene-environment correlation (rGE), including passive, evocative(reactive), and active.
- Comprehend the basic research designs of rGE models
- Recognize the challenges to identify gene-environment correlation and implications for G×E studies
- Understand the main arguments from both sides of the nature versus nurture debate
- Grasp potential future directions in this area of research

Discussion of the readings:
This week's readings focus on gene-environment correlation(rGE) and discuss the nature-nurture debate with examples of genetic nurture and more studies about gene-environment interaction. As introduced in Week 9, rGE is the process when genes and environment could work in tandem. Passive rGE refers to the association between the genotype a child inherits from her parents and the environment in which the child is raised. Evocative (or reactive) rGE refers to the association between an individual's genetically influenced behavior and the reaction of those in the individual's environment to that behavior. Finally, selective (or active) rGE refers to the association between an individual's (genetically influenced) traits or behaviors and the environmental niches selected by the individual (Jaffee & Price, 2007). This week's readings include sample articles of rGE, and the nature versus nurture debate, including genetic nurture. Genetic nurture is suggestive of passive rGE, which means that parents could have indirect effects on their children not by passing their alleles but through nurturing, which would be affected by their non-transmitted alleles.

Questions for students:

- Provide an example for each of the following types of passive, evocative, and selective gene-environment correlation.
- Draw the model of gene-environment interaction questions proposed last week and add possible gene-environment correlation pathways.

Required Readings:

Plomin, R., & Bergeman, C. S. (1991). The nature of nurture: Genetic influence on "environmental" measures. *Behavioral and Brain Sciences*, *14*(3), 373-386.

Collins, W. A., Maccoby, E. E., Steinberg, L., Hetherington, E. M., & Bornstein, M. H. (2000). Contemporary research on parenting: The case for nature and nurture. *American psychologist*, *55*(2), 218-232.

McCrae, R. R., Costa Jr, P. T., Ostendorf, F., Angleitner, A., Hřebíčková, M., Avia, M. D., ... & Saunders, P. R. (2000). Nature over nurture: temperament, personality, and life span development. *Journal of personality and social psychology*, *78*(1), 173-186.

Kendler, K. S., & Baker, J. H. (2007). Genetic influences on measures of the environment: a systematic review. *Psychological medicine*, *37*(5), 615-626.

Sacerdote, B. (2011). "Nature and Nurture Effects On Children's Outcomes: What Have We Learned From Studies of Twins And Adoptees?" *Handbook of Social Economics*, Chapter 1, 1-30.

Wagner, B., Li, J., Liu, H., & Guo, G. (2013). Gene–environment correlation: Difficulties and a natural experiment–based strategy. *American journal of public health*, *103*(S1), S167-S173.

Conley, D., Domingue, B. W., Cesarini, D., Dawes, C., Rietveld, C. A., & Boardman, J. D. (2015). Is the effect of parental education on offspring biased or moderated by genotype?. *Sociological Science*, *2*, 82-105.

Kong, A. Thorleifsson, G., Frigge, M.L. et al. (2018). "The nature of nurture: Effects of parental genotypes," *Science* 359, 424-428.

Optional Readings:

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological review*, *101*(4), 568-586.

Perry, B. D. (2002). Childhood experience and the expression of genetic potential: What childhood neglect tells us about nature and nurture. *Brain and mind, 3(1), 79-100*.

Chakravarti, A., & Little, P. (2003). Nature, nurture and human disease. *Nature, 421(6921), 412*.

Robinson, G. E. (2004). Beyond nature and nurture. *Science*, *304*(5669), 397-399.

Jaffee, S. R., & Price, T. S. (2007). Gene–environment correlations: A review of the evidence and implications for prevention of mental illness. *Molecular psychiatry*, *12*(5), 432.

Sacerdote, B. (2007). How large are the effects from changes in family environment? A study of Korean American adoptees. *The Quarterly Journal of Economics*, *122*(1), 119-157.

Conley, D. and Rauscher, E. (2012). "Genetic Interactions with Prenatal Social Environment: Effects on Academic and Behavioral Outcomes," *Journal of Health and Social Behavior* 54(1).

Lawson, H. A., Cheverud, J. M., & Wolf, J. B. (2013). Genomic imprinting and parent-of-origin effects on complex traits. *Nature Reviews Genetics*, *14*(9), 609-617.

Lee, D., Brooks-Gunn, J., McLanahan, S. S., Notterman, D., & Garfinkel, I. (2013). The Great Recession, genetic sensitivity, and maternal harsh parenting. Proceedings of the National Academy of Sciences, 110(34), 13780-13784.

Liu, H., & Guo, G. (2015). Lifetime socioeconomic status, historical context, and genetic inheritance in shaping body mass in middle and late adulthood. *American sociological review*, *80*(4), 705-737.

Liu, H. (2018). Social and genetic pathways in multigenerational transmission of educational attainment. *American Sociological Review*, *83*(2), 278-304.

Bates, T. C., Maher, B. S., Medland, S. E., McAloney, K., Wright, M. J., Hansell, N. K., ... & Gillespie, N. A. (2018). The nature of nurture: Using a virtual-parent design to test parenting effects on children's educational attainment in genotyped families. *Twin Research and Human Genetics*, *21*(2), 73-83.

Giuliani, C., Garagnani, P., & Franceschi, C. (2018). Genetics of human longevity within an eco-evolutionary nature-nurture framework. *Circulation Research*, *123*(7), 745-772.

Briley, D. A., Livengood, J., Derringer, J., Tucker-Drob, E. M., Fraley, R. C., & Roberts, B. W. (2019). Interpreting behavior genetic models: seven developmental processes to understand. *Behavior genetics*, *49*(2), 196-210.

Sotoudeh, R., Harris, K. M., & Conley, D. (2019). Effects of the peer metagenomic environment on smoking behavior. *Proceedings of the National Academy of Sciences*, *116*(33), 16302-16307.

Willoughby, E. A., McGue, M., Iacono, W. G., Rustichini, A., & Lee, J. J. (2019). The role of parental genotype in predicting offspring years of education: evidence for genetic nurture. *Molecular psychiatry*, 1-9.

Week 12 Topic: Pleiotropy and Genetic architecture

Main goals:
Students could

- Define and recognize pleiotropy and understand the shared genetic architecture of complex traits
- Understand the challenge of endogeneity in drawing causal inference from observational studies
- Understand the fundamentals of the Instrumental Variable (IV) approach in a Mendelian Randomization(MR) framework
- Grasp the main assumptions of the IV approach in an MR framework
- Be aware of extensions of MR and advances in the field (e.g., Egger regression, GIV model, CAUSE model)
- Be introduced to several applications of MR

Discussion of the readings:
To discuss genetic architecture and especially pleiotropy as a critical challenge for causal inference, Mendelian randomization, basically the instrumental variable analysis with genetic instruments, is introduced as a common approach for causal inference in sociogenomics. Because of pleiotropy, the exclusion restriction assumption of Mendelian randomization is often violated. New methods are proposed to adjust the bias, such as Egger regression in the required readings, and GIV regression and CAUSE model in the optional readings. This week's readings also include several applications of MR, mostly in genetic epidemiology.

Questions for students:

- Think of a phenotypic causal relationship that is important for you.
- Look up GWAS summary statistics for both the exposure and the outcome variable, and gather as much information as you can about the exposure variable, and think whether exclusion restriction assumption is violated.

- Analyze which of the following approaches would be the most suitable for your research question: using a single variant as an instrument, using summary results, or using polygenic scores.
- Try to think of a potential non-genetic IV for the same question (optional)

Required Readings:

Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M., and Smoller, J.W. (2013). "Pleiotropy in complex traits: challenges and strategies," *Nature Reviews Genetics* 14, 483-495.

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, *169*(7), 1177-1186.

Verbanck M., Chen C. Y., Neale B. and Do R. (2018). "Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases," *Nature Genetics* 50(5), 693-698.

Mills, M. C., Barban, N., & Tropf, F. C. (2020) Chapter 13 Mendelian Randomization and Instrumental Variables. *An Introduction to Statistical Genetic Data Analysis*. The MIT Press, pp. 339-358.

Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?. *International journal of epidemiology*, *32*(1), 1-22.

Pingault, J.B., O'Reilly, P.F., Schoeler, T., Ploubidis, G.B., Rijsdijk, F. and Dudbridge, F. (2018). "Using genetic data to strengthen causal inference in observational research," *Nature Reviews Genetics* 19, 566–580.

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., ... & Daly, M. J. (2015). An atlas of genetic correlations across human diseases and traits. *Nature genetics*, *47*(11), 1236-1243.

Bowden, J., Davey Smith, G., and Burgess, S. (2015). "Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression," *International Journal of Epidemiology* 44:2, 512-525.

Pickrell, J. K., Berisa, T., Liu, J. Z., Ségurel, L., Tung, J. Y., & Hinds, D. A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nature genetics*, *48*(7), 709-717.

Optional Readings:

Williams, G. C. (1957). Pleiotropy, natural selection, and the evolution of senescence. *evolution*, 398-411.

Stearns, F. W. (2010). One hundred years of pleiotropy: a retrospective. *Genetics*, *186*(3), 767-773.

Smith, G. D., & Ebrahim, S. (2004). Mendelian randomization: prospects, potentials, and limitations. *International journal of epidemiology*, *33*(1), 30-42.

Lawlor, Debbie A., Roger M. Harbord, Jonathan A. C. Sterne, Nic J. Timpson, and George Davey Smith. (2008). "Mendelian Randomization: Using genes as instruments for making causal inferences in epidemiology." Statistics in Medicine 27:1133-1163.

Rodriguez, S., Gaunt, T. R., & Day, I. N. (2009). Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies. *American journal of epidemiology*, *169*(4), 505-514.

Ference, B. A., Yoo, W., Alesh, I., Mahajan, N., Mirowska, K. K., Mewada, A., ... & Flack, J. M. (2012). Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. *Journal of the American College of Cardiology*, *60*(25), 2631-2639.

Vimaleswaran, K. S., Berry, D. J., Lu, C., Tikkanen, E., Pilz, S., Hiraki, L. T., ... & Wood, A. R. (2013). Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. *PLoS medicine*, *10*(2), e1001383.

Hemani, G., Yang, J., Vinkhuyzen, A., Powell, J. E., Willemsen, G., Hottenga, J. J., ... & Madden, P. A. (2013). Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs. *The American Journal of Human Genetics*, *93*(5), 865-875.

Tyrrell, Jessica, et al. (2016) "Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank," *British Medical Journal* 352: i582.

Bowden, J., Davey Smith, G., Haycock, P. C., & Burgess, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic epidemiology*, *40*(4), 304-314.

Davies, N. M., Holmes, M. V., & Smith, G. D. (2018). Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*, *362*, k601.

DiPrete, T.A., Burik, C.A.P. and Koellinger, P.D. (2018). "Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data," *PNAS,* 115 (22) E4970-E4979.

O'Connor, L.J. and Price, A.L. (2018). "Distinguishing genetic correlation from causation across 52 diseases and complex traits," *Nature Genetics* 50, 1728-1734.

Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., ... & Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature communications*, *9*(1), 224.

Morrison, J., Knoblauch, N., Marcus, J., Stephens, M., & He, X. (2019). Mendelian randomization accounting for horizontal and correlated pleiotropic effects using genome-wide summary statistics. *bioRxiv*, 682237.

Watanabe, K., Stringer, S., Frei, O., Mirkov, M. U., de Leeuw, C., Polderman, T. J., ... & Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nature genetics*, *51*(9), 1339-1348.

Week 13 Population genetics – Selection and Assortative Mating

Main goals:
Students could

- Understand and grasp the assumptions for genetic equilibrium and review the Hardy-Weinberg Principle
- Know the basics of population-genetics models of evolution – selection, mutation, genetic drift, migration, non-random mating, and linkage.
- Be aware of the assortative mating in today's society
- Understand how selection and assortative mating could confound causal inference

Discussion of the readings:
Population genetics is the study of genetic variation within populations over space and time. Assortative mating and natural selection are two major mechanisms to change the allele frequencies over generations through fertility and mortality. This week's readings include an overview of population genetics and examples for assortative mating and natural selection. The modeling details of population genetics are beyond the scope of this course.

More population genetics models and studies are available in optional readings.

Questions for students:

- Suppose you are a dictator of Planet X. You want to get rid of the recessive genetic disease that causes Planet X residents to be obsessed with money on the planet earth. There are about 25,000 people with this disease in the population, what is the allele frequency of the recessive allele in the population of the Planet X (about 300 million)?
- If you can set up a law to get rid of this scourge of society, what would that law be (thinking about the selection and assortative mating)?
- One extreme would be sterilizing all those people obsessed with human money on Planet X. How long will it take planet X population (in generations) to get rid of this genetic disease? (You can use computers to run simulations for you.)

Required Readings:

Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. Genetics Research, 8(3), 269-294.

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., ... & Lander, E. S. (2006). Positive natural selection in the human lineage. *science*, *312*(5780), 1614-1620.

Barreiro, L. B., Laval, G., Quach, H., Patin, E., & Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. *Nature genetics*, *40*(3), 340-345.

Christakis, N. A., & Fowler, J. H. (2014). Friendship and natural selection. *Proceedings of the National Academy of Sciences*, *111*(Supplement 3), 10796-10801.

Field, Y., Boyle, E. A., Telis, N., Gao, Z., Gaulton, K. J., Golan, D., ... & Pritchard, J. K. (2016). Detection of human adaptation during the past 2000 years. *Science*, *354*(6313), 760-764.

Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., ... & Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *ELife*, *8*, e39725, 1-46.

Thiessen, D., & Gregg, B. (1980). Human assortative mating and genetic equilibrium: An evolutionary perspective. *Ethology and Sociobiology*, *1*(2), 111-140.

Domingue, B. W., Fletcher, J., Conley, D., & Boardman, J. D. (2014). Genetic and educational assortative mating among US adults. *Proceedings of the National Academy of Sciences*, *111*(22), 7996-8000.

Conley, D., Laidley, T., Belsky, D. W., Fletcher, J. M., Boardman, J. D., & Domingue, B. W. (2016). Assortative mating and differential fertility by phenotype and genotype across the 20th century. *Proceedings of the National Academy of Sciences*, *113*(24), 6647-6652.

Hartwig, F. P., Davies, N. M., & Davey Smith, G. (2018). Bias in Mendelian randomization due to assortative mating. *Genetic epidemiology*, *42*(7), 608-620.

Optional Readings:

Chakravarti, A. (1999). Population genetics—making sense out of sequence. *Nature genetics*, *21*(1s), 56.

Schneider, S., Roessli, D., & Excoffier, L. (2000). Arlequin: a software for population genetics data analysis. *User manual ver*, *2*, 2496-2497.

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025-2035.

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in ecology & evolution*, *18*(4), 189-197.

Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., ... & Prchal, J. T. (2010). Genetic evidence for high-altitude adaptation in Tibet. *Science*, *329*(5987), 72-75.

Berg, J. J., & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS genetics*, *10*(8), e1004412, 1-25.

Beauchamp, J. P. (2016). Genetic evidence for natural selection in humans in the contemporary United States. *Proceedings of the National Academy of Sciences*, *113*(28), 7774-7779.

Domingue, B. W., Belsky, D. W., Harrati, A., Conley, D., Weir, D. R., & Boardman, J. D. (2017). Mortality selection in a genetic sample and implications for association studies. *International Journal of Epidemiology*, *46*(4), 1285-1294.

Kong, A., Frigge, M. L., Thorleifsson, G., Stefansson, H., Young, A. I., Zink, F., ... & Gudbjartsson, D. F. (2017). Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences*, *114*(5), E727-E732.

Gazal, S., Loh, P. R., Finucane, H. K., Ganna, A., Schoech, A., Sunyaev, S. et al. (2018). "Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations," *Nature Genetics*, 50.

Crow, J. F., & Felsenstein, J. (1968). The effect of assortative mating on the genetic composition of a population. *Eugenics quarterly*, *15*(2), 85-97.

Tropf, F. C., Stulp, G., Barban, N., Visscher, P. M., Yang, J., Snieder, H., & Mills, M. C. (2015). Human fertility, molecular genetics, and natural selection in modern societies. *PloS one*, *10*(6), e0126821.

Narasimhan, V. M., Hunt, K. A., Mason, D., Baker, C. L., Karczewski, K. J., Barnes, M. R., ... & Giorda, K. (2016). Health and population effects of rare gene knockouts in adult humans with related parents. *Science*, *352*(6284), 474-477.

Robinson, M. R., Kleinman, A., Graff, M., Vinkhuyzen, A. A., Couper, D., Miller, M. B., ... & van Vliet-Ostaptchouk, J. V. (2017). Genetic evidence of assortative mating in humans. *Nature Human Behaviour*, *1*(1), 0016.

Week 14 Topic: Ethical concerns and policy

Main goals:

- Raise awareness of the multiple ethical issues in genomics research
- Discuss genetic determinism and social policy
- Provide an overview of the clinical use of polygenic scores
- Reiterate the lack of diversity in current genomic research and implications for the wider applicability of results.
- Explore privacy, consent, legal issues, and insurance in relation to genomic data.

Discussion of the readings:

This week's required readings discuss ethical concerns of sociogenomic research, starting from three summary readings about the ethical and policy issues of genetic research. Specific topics are discussed in different readings

– informed consent, genetic information nondiscrimination (especially GINA 2008), the ethical issues of genetic determinism, insurance, social policy, and the clinical use of PGS. The optional readings also include summaries of eugenic history, genomic data sharing policy, specific case studies demonstrating complicated ethical and policy issues that could happen in the research and applications of genomic data (e.g., Myriad Genetics, Navajo Nation, etc.).

Questions for students:

- List and discuss three major ethical and policy concerns for the research question you raised at the beginning of the semester.

Required Readings:

Mills, M. C., Barban, N., & Tropf, F. C. (2020) Chapter 14 Ethical Issues in Genomics Research. *An Introduction to Statistical Genetic Data Analysis*. The MIT Press, pp. 359-373.

Nuffield Council on Bioethics, "Summary and Recommendations" *Genetics and Human Behavior: The Ethical Context* (2002) (pp. xix–xxxiii).

Berryessa, C. M., & Cho, M. K. (2013). Ethical, legal, social, and policy implications of behavioral genetics. *Annual Review of Genomics and Human Genetics*, *14*, 515-534.

McGuire, A. L., & Beskow, L. M. (2010). Informed consent in genomics and genetic research. *Annual review of genomics and human genetics*, *11*, 361-381.
https://www.genome.gov/about-genomics/policy-issues/Informed-Consent

Hudson, K. L., Holohan, M. K., & Collins, F. S. (2008). Keeping pace with the times—the Genetic Information Nondiscrimination Act of 2008. *New England Journal of Medicine*, *358*(25), 2661-2663.
https://www.eeoc.gov/laws/statutes/gina.cfm

Daykin, C. D., Akers, D. A., Macdonald, A. S., McGleenan, T., Paul, D., & Turvey, P. J. (2003). Genetics and insurance–some social policy issues. *British Actuarial Journal*, *9*(4), 787-830.

Yudell, M., Roberts, D., DeSalle, R., & Tishkoff, S. (2016). Taking race out of human genetics. *Science*, *351*(6273), 564-565.

Kim, M. S., Patel, K. P., Teng, A. K., Berens, A. J., & Lachance, J. (2018). Genetic disease risks can be misestimated across global populations. *Genome biology*, *19*(1), 179, 1-14.

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature genetics*, *51*(4), 584-591.

Optional Readings:

Juengst, E. T. (1991). Priorities in professional ethics and social policy for human genetics. *JAMA*, *266*(13), 1835-1836.

Alper, J. S., & Beckwith, J. (1993). Genetic fatalism and social policy: the implications of behavior genetics research. *The Yale journal of biology and medicine*, *66*(6), 511.

Harris, Y., Gorelick, P. B., Samuels, P., & Bempong, I. (1996). Why African Americans may not be participating in clinical trials. *Journal of the National Medical Association*, *88*(10),

Black, E. (2003). War against the weak: Eugenics and America's campaign to create a master race. *New York*.

Hall, W. D., Gartner, C. E., & Carter, A. (2008). The genetics of nicotine addiction liability: ethical and social policy implications. *Addiction*, *103*(3), 350-359.

Need, A. C., & Goldstein, D. B. (2009). Next generation disparities in human genomics: concerns and remedies. *Trends in Genetics*, *25*(11), 489-494.

Gold, E. R., & Carbone, J. (2010). Myriad Genetics: In the eye of the policy storm. *Genetics in Medicine*, *12*(1s), S39.

Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: on the deceptive determinism of DNA. *Psychological bulletin*, *137*(5), 800.

Garrison, N. A. (2013). Genomic justice for Native Americans: impact of the Havasupai case on genetic research. *Science, Technology, & Human Values*, *38*(2), 201-223.

Nature editorial, "Dangerous Work," *Nature* 502: 5-6 (Oct. 3, 2013), http://www.nature.com/polopoly_fs/1.13861!/menu/main/topColumns/topLeftColumn/pdf/502005b.pdf

Erika Check Hayden, "Ethics: Taboo Genetics," *Nature* 502: 26-28 (Oct. 3, 2013) http://www.nature.com/news/ethics-taboo-genetics-1.13858

Association of British Insurers. (2014). Concordat and Moratorium on Genetics and Insurance. https://www.abi.org.uk/globalassets/sitecore/files/documents/publications/public/2014/genetics/concordat-and-moratorium-on-genetics-and-insurance.pdf

Selections from Special Report, "The Genetics of Intelligence: Ethics and the Conduct of Trustworthy Research," *Hastings Center Report* 45(S1): S2-S72, (Sept.-Oct. 2015, Erik Parens & Paul S. Appelbaum eds.). https://onlinelibrary.wiley.com/toc/1552146x/45/S1

Tabery, J. (2015). Why Is Studying the Genetics of Intelligence So Controversial?. *Hastings Center Report*, *45*(S1), S9-S14.

Check Hayden, E. (2015). Genome researchers raise alarm over big data. *Nature News*.

Parens, E., & Appelbaum, P. S. (2015). An introduction to thinking about trustworthy research into the genetics of intelligence. *Hastings Center Report*, *45*(S1), S2-S8.

Reardon, S. (2017). Navajo Nation reconsiders ban on genetic research. *Nature News*, *550*(7675), 165.

Hann, K. E., Freeman, M., Fraser, L., Waller, J., Sanderson, S. C., Rahman, B., ... & Lanceley, A. (2017). Awareness, knowledge, perceptions, and attitudes towards genetic testing for cancer risk among ethnic minority groups: a systematic review. *BMC Public Health*, *17*(1), 503.

ASHG. (2018). ASHG denounces attempts to link genetics and racial supremacy. *Am. J. Hum. Genet.* 2018;103:636.

Knoppers, B. M., & Joly, Y. (2018). Introduction: the why and whither of genomic data sharing. *Human Genetics(2018):569-574*.

FAQs regarding Lee *et al*. (2018) "Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment" *Nature Genetics*, especially Section 3 (ethical and social implications.)

Majumder, M. A. (2018). United States: law and policy concerning transfer of genomic data to third countries. *Human genetics*, *137*(8), 647-655.

Phillips, M. (2018). International data-sharing norms: from the OECD to the General Data Protection Regulation (GDPR). *Human genetics*, *137*(8), 575-582.