General FAQ: Social-Science Genomics*

Daniel J. Benjamin
David I. Laibson
Michelle N. Meyer
Patrick A. Turley

25 June 2026

Abstract

Social-science genomics is an interdisciplinary field that uses genomic data to understand variation in social, behavioral, and economic outcomes. This FAQ provides an overview of the field. Our goal in providing this document is to create a reliable resource for the media and general public as well as for other researchers. All are welcome to draw on or reference this FAQ in their own writing to ensure consistent and accurate communication of genomic concepts.

* Edited with the assistance of Gemini. Benjamin: Anderson School of Management and David Geffen School of Medicine, University of California, Los Angeles, and NBER (email: daniel.benjamin@gmail.com); Laibson: Economics Department, Harvard University, and NBER (email: dlaibson@harvard.edu); Meyer: Department of Bioethics and Decision Sciences, Geisinger College of Health Sciences (e-mail: mmeyer@geisinger.edu); Turley: Center for Economics and Social Research and Economics Department, University of Southern California (email: pturley@usc.edu). We acknowledge support from NIH/NIA grant R24-AG089055 to the University of California Los Angeles.

1. Introduction and Background: What is Social-Science Genomics?

This FAQ provides an overview of the interdisciplinary field of social-science genomics. It was developed by several members and advisors of the Social Science Genetic Association Consortium (SSGAC), a team of social scientists that collaborates with medical researchers (https://www.thessgac.org/). The content of this document is synthesized from longer, paper-specific FAQs published by the SSGAC (https://www.thessgac.org/faqs).

Our goal in providing this document is to create a reliable resource for the media and general public as well as for other researchers. All are welcome to draw on or reference this FAQ in their own writing to ensure consistent and accurate communication of genomic concepts.

Social-science genomics is an interdisciplinary field that uses genomic data to understand variation in social, behavioral, and economic outcomes. A primary goal of research in this field is to study gene-environment interplay, which examines how genes and environments work together to shape human lives (e.g., Benjamin et al., 2026; Braudt, 2018; Conley, 2016; Conley & Fletcher, 2017; Dias Pereira et al., 2022; Freese, 2018; Martschenko, Trejo, & Domingue, 2019; Mills, Barban, & Tropf, 2020; Plomin et al., 2016).

2. The Tools of Social-Science Genomics

- **2.1 What is a Genome-Wide Association Study (GWAS)?** A GWAS is a "bottom-up" discovery method that scans the genome for correlations between millions of common genetic variants, called SNPs, and a phenotype (Visscher et al., 2017). A **phenotype** refers to any measurable characteristic of an individual—such as height, years of completed schooling, or health status—and the word "trait" is often used interchangeably with it.
    - **Population GWAS:** Most large GWAS analyze samples of unrelated individuals. These studies yield correlations, which may be negative, between each SNP and the phenotype. These correlations aggregate causal genetic effects, interpersonal effects (such as genetic nurture; see Section 3.1.2 below), and confounds related to genetic and environmental differences between groups that have arisen over time (Benjamin et al., 2026; Vilhjálmsson & Nordborg, 2013; see Section 3.1 for further discussion).
    - **Family-based GWAS:** By using samples of siblings or parent-offspring trios (the offspring and both biological parents), researchers can conduct a GWAS that accounts for the parental genotype. Because which genes a person inherits is random conditional on the parents, these studies yield estimates that have a causal interpretation (Benjamin et al., 2026; Young et al., 2022; see Section 3.2 for further discussion). That is, unlike population GWAS results, the results of family-based GWAS are not merely correlations but can be interpreted as estimates of causal genetic effects on the phenotype.
- **2.2 Polygenic Indices (PGIs)?** A PGI is a variable constructed using genetic data to predict a particular phenotype (Dudbridge, 2013; Burt, 2024). It is a weighted sum of the associations between each of many variants (typically more than one million) and the phenotype.
    - **Terminology:** Following Becker et al. (2021), we use the term "polygenic index" rather than

other common terms, such as “polygenic score” and “polygenic risk score,” because the words “risk” and “score” can unintentionally connote a value judgment about the desirability or consequences of a social, behavioral, or economic phenotype.

- **Prediction:** When researchers state that a PGI “predicts” an outcome, they merely mean the PGI is statistically correlated with the phenotype, even if that association is relatively weak. This scientific usage diverges from everyday language, where “prediction” connotes a forecast about future events. Instead, the scientific usage denotes out-of-sample generalizability, meaning researchers expect the statistical link to reappear in independent datasets—as long as environmental conditions in the independent datasets closely match those of the original sample—whether those data are historical or prospective.
- **The Nature of PGIs:** A PGI is a noisy and imperfect predictor of an individual’s phenotype. PGIs vary in their predictive power, depending on the phenotype and on the data and methodologies used for constructing and analyzing the PGI. Furthermore, the prediction provided by a PGI may be mediated by environmental pathways and may vary depending on the environmental context (Benjamin et al., 2026).
- **The “Portability Problem”:** All people share the overwhelming majority of their genetic makeup. However, when groups of people don’t mate randomly (e.g., due to geographic distance or cultural reasons), those groups can become genetically different over time, with some genetic variants becoming more or less common in different groups. As a result, PGIs developed in one group may not be as predictive for individuals from other groups if these groups have grown too genetically dissimilar over time. This so-called “portability problem” is due to differences in allele frequencies, linkage disequilibrium, and the environment (Duncan et al., 2019; Martin et al., 2019; Okbay et al., 2022; Mills, Barban, & Tropf, 2020).

3. Interpreting the Science: Correlation vs. Causation

- **3.1 Are GWAS variants “causal”?** In a population GWAS, a correlation can be driven by several distinct factors:
  1. **Causal Genetic Effects:** These are commonly called “direct genetic effects” and occur when the genetic variant itself influences the phenotype. A genetic causal effect corresponds to the hypothetical experiment of modifying an individual’s genotype at conception while holding both parental genotypes and the environment constant. A version of this experiment is naturally carried out by meiosis, the random process by which parental alleles are inherited by offspring. Causal genetic effects do not exist in a societal vacuum. For example, a genetic propensity to enjoy reading increases educational attainment in an environment in which reading is a common activity in the educational system. Accordingly, causal genetic effects include biological or behavioral pathways within the individual, as well as pathways mediated by others’ behavioral responses to the individual (Kong et al., 2018; Young et al., 2022) and societal structures and policies.
  2. **Interpersonal Genetic Effects:** These are commonly called “indirect genetic effects.” The most widely studied interpersonal genetic effects, commonly called “genetic nurture,” occur when a parent's genetic variant influences *parental* phenotypes that in turn shape the child's environment (Kong et al., 2018; Young et al., 2022). For example, parental genetic variants associated with educational attainment may cause parents to provide more books in the home. These effects manifest as a correlation with educational attainment in a population GWAS

because children inherit half of their genetic material from each parent; consequently, an individual's own genetic variants serve as a statistical proxy for the parental variants that shaped their upbringing.

3. **Population Stratification:** Population stratification refers to noncausal correlations caused by environmental factors that are correlated with genetic differences between groups of people (Price et al., 2006). A classic thought experiment illustrating this problem is a hypothetical GWAS of chopstick use: a researcher might find a genetic variant that is correlated with the use of chopsticks, but the correlation arises only because the variant is more frequent in groups where using chopsticks is a cultural norm (Hamer & Sirota, 2000).

- **3.2 The "Within-Family" Design:** By comparing siblings or controlling for parental genotypes, researchers ensure that parental genotypes are held constant. This approach allows scientists to estimate the causal genetic effect of an individual's own genetic variant, as opposed to the population coefficient, which includes interpersonal effects and confounds such as population stratification. The causal genetic effect is typically smaller than the population coefficient estimated in general population samples (Selzam et al., 2019; Benjamin et al., 2026; Lee et al., 2018).
- **3.3 Assortative Mating:** Because people tend to have children with partners who have similar phenotypes—a phenomenon known as assortative mating—genetic variants that influence those phenotypes become correlated across the genome (Robinson et al., 2017). For example, if parents assortatively mate on height, then offspring of a tall mother are also likely to inherit height-increasing genetic variants from the father. As a result, height-increasing genetic variants become correlated in the offspring across the genome, even if the genetic variants are on different chromosomes. In a population GWAS, these cross-genome correlations between genetic variants can inflate the GWAS correlation (between each variant and a phenotype) because the GWAS correlation for any single SNP captures not only its own causal effect but also the effects of other correlated genetic variants throughout the genome. This phenomenon is a major focus of modern social-science genomic methodology as it can lead population-based PGIs to overstate the predictive power of genetic variants (Yengo et al., 2018).
- **3.4 The "Variance Explained" (R^2) of a PGI:** R^2 measures how much of the differences between people can be predicted by a variable, such as a genetic variant or a PGI. For example, the best available PGI for educational attainment has an R^2 of roughly 12–15% in samples that look genetically similar to the "European genetic ancestry" sample of a well-known reference data set. Although low, this level of predictive power is comparable to, or higher than, that of traditional social variables frequently used in research, such as household income (Okbay et al., 2022; Lee et al., 2018). For perspective, a short-term precipitation forecast (e.g., 24-hour accumulation totals) has an R^2 typically exceeding 40%. This contrast underscores that while PGIs can provide valuable information for researchers, they do not currently offer high-precision individual prediction.

4. Applications in Research: Examples and Methodologies

- **4.1 Estimating the Causal Effects of Environmental Interventions:** Researchers often would like to estimate the effect of an environmental intervention on some phenotype—for example, the effect of attending preschool on later educational achievement. Although neither is a genetic variable, statistically controlling for a PGI can generate a more precise estimate of the effect of the environmental intervention. This is because the PGI may explain some of the variation in the phenotype among individuals. If so, then controlling for the PGI reduces the amount of unexplained

variation in the phenotype, facilitating detection of the causal effect of the environmental intervention (Benjamin et al., 2026; Rietveld et al., 2013). Hence, using a PGI can increase the statistical power to detect the effect of some other (non-genetic) variable for any given sample size.

- **4.2 Estimating the Causal Effects of Genetic Variants:** As discussed in Section 3.1, correlations identified in a population GWAS aggregate direct genetic effects, interpersonal genetic effects, and population stratification. To isolate the *causal* impact of genetic variants on a phenotype, researchers use family-based GWAS (see Section 2.1). This approach relies on the "Mendelian lottery"—the random process of meiosis that shuffles parental alleles during the formation of gametes. Because the specific alleles an offspring inherits are random conditional on the parental genotypes, the within-family design functions like a natural randomized controlled trial (see Section 3.2). By holding the parental genome constant, researchers can isolate causal genetic effects unaffected by confounding influences such as interpersonal genetic effects (see Section 3.1.2), or population stratification (see Section 3.1.3) (Benjamin et al., 2026). This methodology can also be extended to study the causal effects of PGIs, though the interpretation in that context requires additional nuance.
- **4.3 Improving Studies of the Environment By Partially Controlling for Genetics:** Genetic data can help researchers more accurately study the impact of social environments by reducing the impact of "genetic confounding." Consider a study on how parenting behaviors, such as reading to children, influence a child's later academic success. In an observational study, a positive correlation between reading with parents and success might not be causal; parents both affect the child's environment (reading the books) and pass their genetic variants to their children. If the same genetic factors that make a parent more likely to read to their child also contribute to the child's academic performance, the correlation between parenting behavior and child's academic performance will overestimate the effect of parenting alone. By including the child's PGI as a control variable, researchers can "soak up" some of this shared genetic influence, allowing for a clearer view of the environmental effect. However, it is important to note that this approach does not perfectly isolate causal effects. Because PGIs are noisy predictors and may correlate with other unmeasured environmental factors, the resulting estimates should be interpreted with caution (Benjamin et al., 2026; Belsky & Harden, 2019).
- **4.4 Parental Genetic Effects:** Social scientists want to understand how parents matter for their children's outcomes and in what ways. One approach involves identifying "parental genetic effects"—instances where a parent's genotype influences a child's phenotype not through direct genetic transmission, but by affecting the environment the parent provides (Kong et al., 2018; Young et al., 2022). This phenomenon is more commonly referred to as "indirect effects" or "genetic nurture" in the behavioral genetics literature (see Section 3.1). We use the terminology "parental genetic effects" here because "indirect" is easily misunderstood. Researchers can use parental genetic effects to generate hypotheses about which parental phenotypes matter for children's outcomes.
- **4.5 Identifying Gene-Environment ($G \times E$) Interaction:** In this type of research, researchers study how a person's genome changes how much the person is affected by the environment. This research includes evaluating policy interventions—such as mandatory schooling laws—to see if they benefit everyone equally or have disparate impacts depending on a person's genome (Belsky et al., 2016; Barcellos, Carvalho, & Turley, 2018). Furthermore, researchers examine how stable environmental contexts, such as parental socioeconomic status (SES), can moderate the influence of genetic predispositions on life outcomes (Bolyard & Savelyev, 2025).

- **4.6 Social Mobility and Intergenerational Transmission:** Genetic data can help researchers study how social environments constrain or facilitate opportunity across generations. Descriptive work has used PGIs to demonstrate that individuals with high predicted values for educational attainment achieve different economic and health outcomes depending on their family's initial socioeconomic status (Belsky et al., 2018; Bolyard & Savelyev, 2025). Moving beyond description, structural approaches such as that of Rustichini et al. (2023), enable researchers to identify and estimate causal pathways by statistically modeling how parental phenotypes, genetic inheritance, and phenotypes of children are interrelated. This so-called "structural identification" relies on key assumptions, for example that the model is correctly specified and that the biological and social parameters governing transmission are stable across the groups or time periods being compared.

5. Applications Outside of Research

- **5.1 Limitations for Individual-Level Prediction:** While PGIs for social and behavioral phenotypes are powerful tools for population-level research, they currently yield very imprecise predictions for individuals. Because these PGIs typically explain only a small fraction of the total variation across individuals for complex social outcomes, they usually do not provide actionable information regarding a specific person's future trajectory (see Section 3.4 for further discussion).
- **5.2 Limitations for Developing Interventions:** Results from a population GWAS, including PGIs, cannot immediately be directly applied to design interventions—such as altering pedagogy to improve graduation rates—because in almost all cases researchers have not yet identified the specific biological or social mechanisms that mediate the population-GWAS associations. For example, a genetic variant might influence educational attainment through any number of disparate pathways, such as individual differences in neurodevelopment, variations in personality phenotypes, or even the way social institutions respond to a child's physical characteristics. Without identifying these mediating pathways, there is no scientific basis for predicting how a particular intervention might interact with them to produce the desired outcome.
- **5.3 Research-Driven Benefits for Policy:** The main benefits of social-science genomics for policy are likely to come from research results that incorporate genetic data into social-science research—as detailed in Section 4—rather than from the direct use of genetic data in policy. For example, by using PGIs in large-scale studies, researchers can provide policymakers with better evidence about which environmental interventions work, for whom, and why (Benjamin et al., 2026).
- **5.4 Evaluating the Equity of Policy Impacts:** An example of a research-driven policy benefit is the evaluation of equity. Researchers can use genetic data to understand if social interventions help those with lower PGI-predicted values catch up or widen existing gaps. This evidence can inform whether a program should be universal or targeted toward those who face specific environmental or systemic hurdles (Harden, 2021).

6. Ethical and Social Implications

- **6.1 Potential Harms:** Without safeguards in place, social-science genomics research could lead to potential harms. For a comprehensive analysis of these and related risks, see Meyer et al. (2023).
    - **Fatalism** is the belief that a phenotype, whether observed or statistically predicted, is immutable, fixed, or predetermined. Fatalistic beliefs can lead individuals to give up trying to change their own phenotype. For policy makers with fatalistic beliefs, social-science genomic

research may be used to legitimize the status quo as unalterable and cast environmental interventions as futile.
    - **Stigmatization** is the belief that an individual or group is less capable or worthy based on genetic data. Stigmatization could be directed at oneself, at other particular individuals, or at groups.
    - **Discrimination and Inequitable Application** entail detrimental actions directed at individuals or groups in education, employment, health, insurance, immigration, reproduction, criminal justice, or other domains. This risk encompasses both the scientifically unsound or ethically improper applications of social-science genomics research and the unequal distribution of its beneficial applications.
    - **Biologization of Race and Ethnicity** is the interpretation of race and ethnicity as intrinsically biological (whereas concepts of race and ethnicity vary across cultures). When combined with other risks here, this biologizing can result in or facilitate racism.
    - **Resource Diversion** is the concern that devoting resources to social-science genomics research may distract from more effective environmental or social methods of understanding or addressing the same phenotypes.
- **6.2 Averting Potential Harms:** Several measures are currently in place to protect participants and the public.
    - **Legal Protections:** In many jurisdictions, laws—such as, in the U.S., the Genetic Information Nondiscrimination Act (GINA)—explicitly bar health insurers and employers from utilizing genetic data (Rothstein, 2008). We support the strengthening of legal safeguards that prevent genetic data from being used in ways that limit opportunity.
    - **Data Privacy and Security:** Research datasets have policies to protect research participants, and nearly all research is conducted with de-identified data. Much research in the field is conducted with biobanks, which have rules about who can access the data and what research can be conducted with it. For example, an insurance company could not use the data from the U.K. Biobank to learn the genetic profiles of its members (U.K. Biobank, 2024).
    - **Proactive Communication:** It has become a best practice in the field to create detailed FAQs (e.g., Martschenko et al., 2021) and supplementary materials for research papers to proactively correct common misconceptions and to communicate steps the researchers have taken to minimize risks such as those discussed above. Indeed, it is a goal of this FAQ to promote the production of such resources. Researchers have also recommended other responsible communication strategies; see Meyer et al. (2023).
- **6.3 Equity and Genetic Ancestry:** Diversity in genomic research is an ethical and scientific imperative in order to ensure that its benefits are equitably distributed and maximize the value of data that has been collected. For example, as discussed in Section 2.2, for research that uses PGIs, the current reliance primarily on data from individuals genetic similar to samples identified as having "European genetic ancestries" leads to a "portability problem": PGIs perform poorly when applied to individuals from a group with different linkage disequilibrium patterns, allele frequencies, and environmental contexts (Duncan et al., 2019; Martin et al., 2019). Without diversifying the collection of genomic data, the benefits of research using PGIs (as discussed in Section 4 and Sections 5.3–5.4) will remain largely restricted to the groups of people for whom more genetic data has been collected. Furthermore, existing methods of analyzing genomic data typically restrict the data to a group of individuals who are relatively similar genetically, omitting individuals outside that

group. Such methods are wasteful scientifically; more could be learned by studying as much of the available data as possible. Therefore another priority is developing methods that do not rely on such data restrictions.

7. Common Misconceptions

- **7.1 "Genetic" Means "Innate" or "Unchangeable":** One of the most prevalent misconceptions is that genetic associations imply immutability. In reality, most genetic effects on social outcomes are environmentally mediated and can be altered by social context, behavioral changes, or policy interventions (Goldberger, 1979; Visscher, 2008). Consider, for example, a child born with poor eyesight due to genetic factors. Without intervention, this child may struggle to read and consequently underperform in school. However, if the child is provided with eyeglasses, the "genetic" effect on their educational attainment may be reduced or even eliminated.
- **7.2 "The Gene for X":** There is no single gene for complex social outcomes like education or income. Such outcomes are highly polygenic, meaning that millions of variants each have a miniscule effect (Chabris et al., 2015).
- **7.3 "Nature vs. Nurture" and "Genetic Effects Are Purely Biological":** Many people assume genetic effects must operate through internal biological mechanisms. In reality, for social-science phenotypes, genetic effects most likely operate through environmental mechanisms "outside the skin"—such as an individual's personality or preferences or through the behavioral reactions of others (Jencks, 1980; see Sections 3.1 and 4.4 for further discussion of pathways mediated by the environment). For this reason, nature vs. nurture is a false dichotomy.
- **7.4 "Genetic Means Causal":** Correlations identified in a population GWAS aggregate causal genetic effects, interpersonal genetic effects, and confounds related to genetic ancestry (see Sections 3.1 and 3.2).

References

Barcellos, S. H., Carvalho, L. S., & Turley, P. (2018). Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences*, 115(42), E9765-E9772.

Becker, J., et al. (2021). Resource profile and user guide of the Polygenic Index Repository. *Nature Human Behaviour*, 5(12), 1744–1758.

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*, 28(1), 82–90.

Belsky, D. W., et al. (2016). The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychological Science*, 27(7), 957-972.

Belsky, D. W., et al. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences*, 115(31), E7273-E7282.

Benjamin, D. J., et al. (2026). Social-Science Genomics: Progress, Challenges, and Future Directions. *Journal of Economic Literature* (accepted).

Bolyard, A., & Savelyev, P. A. (2025). Understanding the educational attainment polygenic index and its interactions with SES in determining health in young adulthood. *Journal of Human Resources*. https://doi.org/10.3368/jhr.0521-11685R3

Braudt, D. B. (2018). Social genomics: An introduction to concepts, methods and data. *Sociology Compass*, 12(2), e12545.

Burt, C. H. (2024). Polygenic Indices (aka Polygenic Scores) in Social Science: A Guide for Interpretation and Evaluation. *Sociological Methodology*, 54(2), 300-350.

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science*, 24(4), 304-312.

Conley, D. (2016). Socio-genomic research using genome-wide molecular data. *Annual Review of Sociology*, 42, 275-299.

Conley, D., & Fletcher, J. (2017). *The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and Our Future*. Princeton University Press.

Dias Pereira, R., et al. (2022). Gene–Environment Interplay in the Social Sciences. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLOS Genetics*, 9(3), e1003348.

Duncan, L., et al. (2019). Analysis of polygenic risk score usage and performance in diverse human

populations. *Nature Communications*, 10, article 3328.

Freese, J. (2018). The arrival of social science genomics. *Contemporary Sociology*, 47(5), 524-536.

Goldberger, A. S. (1979). Heritability. *Economica*, 46(184), 327-347.

Hamer, D. H., & Sirota, L. (2000). Beware the chopsticks gene. *Molecular Psychiatry*, 5(1), 11-13.

Harden, K. P. (2021). *The Genetic Lottery: Why DNA Matters for Social Equality*. Princeton University Press.

Jencks, C. (1980). Heredity, environment, and public policy reconsidered. *American Sociological Review*, 45(5), 723-736.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., Benonisdottir, S., Oddsson, A., Halldorsson, B. V., Masson, G., Gudbjartsson, D. F., Helgason, A., Holm, H., & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, 359(6374), 424-428.

Lee, J. J., 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(8), 1112-1121.

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.

Martschenko, D. O., Domingue, B. W., Matthews, L. J., & Trejo, S. (2021). FoGS provides a public FAQ repository for social and behavioral genomic discoveries. *Nature Genetics*, 53(9), 1272-1274.

Martschenko, D. O., Trejo, S., & Domingue, B. W. (2019). Malleable and mutating: Educational attainment polygenic scores in health disparity research. *Social Science & Medicine*, 232, 227-236.

Meyer, M. N., et al. (2023). Wrestling with Social and Behavioral Genomics: Risks, Potential Benefits, and Ethical Responsibility. *Hastings Center Report*, 53(S1), S2–S49.

Mills, M. C., Barban, N., & Tropf, F. C. (2020). *An Introduction to Statistical Genetic Data Analysis*. MIT Press.

Okbay, A., et al. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*, 54(4), 437-449.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science*, 11(1), 3-23.

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

Rietveld, C. A., et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, 340(6139), 1467-1471.

Robinson, M. R., Kleinman, A., Graff, M., Vinkhuyzen, A. A., Couper, D., Miller, M. B., ... & Visscher, P. M. (2017). Genetic evidence of assortative mating in humans. *Nature Human Behaviour*, 1(1), 0016.

Rothstein, M. A. (2008). GINA, the ADA, and genetic discrimination in employment. *The Journal of Law, Medicine & Ethics*, 36(4), 837-840.

Rustichini, A., Iacono, W. G., Lee, J. J., & McGue, M. (2023). Educational Attainment and Intergenerational Mobility: A Polygenic Score Analysis. *Journal of Political Economy*, 131(10), 2724-2779.

Selzam, S., Ritchie, S. J., Pingault, J. B., Reynolds, C. A., O'Reilly, P. F., & Plomin, R. (2019). Comparing within- and between-family polygenic score prediction. *American Journal of Human Genetics*, 105(2), 351–363.

U.K. Biobank. (2024). *Access to UK Biobank data*. https://www.ukbiobank.ac.uk/about-us/how-we-work/access-to-uk-biobank-data/

Vilhjálmsson, B. J., & Nordborg, M. (2013). The nature of confounding in genome-wide association studies. *Nature Reviews Genetics*, 14(1), 1-2.

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.

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. *American Journal of Human Genetics*, 101(1), 5-22.

Yengo, L., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *Human Molecular Genetics*, 27(20), 3641-3649.

Young, A. I., Nehzati, S. M., Benonisdottir, S., Okbay, A., Jayashankar, H., Lee, C., Cesarini, D., Benjamin, D. J., Turley, P., & Kong, A. (2022). Mendelian imputation of parental genotypes improves estimates of direct genetic effects. *Nature Genetics*, 54(6), 897-905.