

The “Golden Age” of Behavior Genetics?

Evan Charney 

The Samuel DuBois Cook Center on Social Equity, Duke University

Perspectives on Psychological Science
1–23

© The Author(s) 2022

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/17456916211041602

www.psychologicalscience.org/PPS



Abstract

The search for genetic risk factors underlying the presumed heritability of all human behavior has unfolded in two phases. The first phase, characterized by candidate-gene-association (CGA) studies, has fallen out of favor in the behavior-genetics community, so much so that it has been referred to as a “cautionary tale.” The second and current iteration is characterized by genome-wide association studies (GWASs), single-nucleotide polymorphism (SNP) heritability estimates, and polygenic risk scores. This research is guided by the resurrection of, or reemphasis on, Fisher’s “infinite infinitesimal allele” model of the heritability of complex phenotypes, first proposed over 100 years ago. Despite seemingly significant differences between the two iterations, they are united in viewing the discovery of risk alleles underlying heritability as a matter of finding differences in allele frequencies. Many of the infirmities that beset CGA studies persist in the era of GWASs, accompanied by a host of new difficulties due to the human genome’s underlying complexities and the limitations of Fisher’s model in the postgenomics era.

Keywords

behavioral genetics, environment, individual differences, behavioral methodology, quantitative methodology, scientific methodology

According to the so-called “first law” of behavior genetics, “all human behavioral traits are heritable,” no matter how socially and historically dependent and contingent they might seem (Turkheimer, 2000, p. 160). This “law,” construed as a “very robust empirical regularity (not a universal, mechanistic truth)” (Chabris et al., 2012, p. 305), has seemingly been confirmed by decades of twin and adoption studies. The search for genetic variants responsible for this heritability has proceeded in two phases.

The first phase was characterized by candidate-gene-association (CGA) studies. On the basis of such studies, researchers reported that polymorphisms of specific genes could predict, for example, likelihood or “risk” of performance on intelligence tests (Fiedorowicz et al., 2007), educational attainment/achievement (EA; Beaver et al., 2012; Shanahan et al., 2008), income/wealth (Frydman et al., 2011), and voting behavior (Fowler & Dawes, 2008). CGA studies have now mainly fallen out of favor in the behavior-genetics community (Chabris et al., 2012; Duncan & Keller, 2011; Flint et al., 2020). The second and current phase is characterized by genome-wide association studies (GWASs), polygenic risk scores (PRSs), and single-nucleotide polymorphism (SNP) heritability estimates. The claims are the same as in the era of CGA

studies: Polymorphisms of specific genes can predict, for example, the risk of performance on intelligence tests (Hill, Marioni, et al., 2019), educational attainment (J. J. Lee et al., 2018), income/wealth (Hill, Davies, et al., 2019), and voting behavior (Aaroe et al., 2021). The difference is that instead of a single polymorphism of relatively large effect size predicting a given behavior, millions of polymorphisms, each of minuscule effect size, are predictive. The current phase has recently been referred to as a “golden age”: “Genomic technology has ushered in a golden age of new tools to address enduring questions about how genes and environments combine to create unique human lives” (Harden, 2021, p. 37).

As in the CGA study era, current predictive claims, if correct, have profound social implications. For example, in a well-known study involving more than a million subjects (J. J. Lee et al., 2018), researchers reported that they could predict 11.4% of the population variance in years of schooling completed on the basis of a PRS derived from the effect sizes of millions of different

Corresponding Author:

Evan Charney, The Samuel DuBois Cook Center on Social Equity,
Duke University

Email: echar@duke.edu

alleles. According to Plomin and von Stumm (2018), direct-to-consumer PRSs will be used to predict people's genetic propensity to learn, reason, and solve problems. According to Dunkel et al. (2019), Jews have higher PRSs for general intelligence than Catholics and Lutherans. And according to Piffer (2019), PRSs predict supposed population-based differences in IQ.

In a blunt assessment of CGA studies, Keller commented, "This should be a real cautionary tale. How on Earth could we have spent 20 years and hundreds of millions of dollars studying pure noise?" (as quoted in Yong, 2019, para. 6). The rise of the new golden age from the ashes of CGA studies has occurred at a precipitous rate. Have the lessons of the "cautionary tale" been taken to heart? To a large extent, they have not. Some of the same problems that beset CGA studies have reappeared in new forms, together with a host of new difficulties related to new methods, new sources of data, and a unique focal point (sizable segments of the human genome, as opposed to individual genes).

CGA Studies

Basics

When variant forms of corresponding DNA sequences occur in more than 1% of a population, they are considered "common" in that population and are called polymorphisms (the expressions *allele* and *polymorphism* are used interchangeably). The search for the genetic basis of heritability, or genetic risk factors, is a search for differences in allele frequencies between case subjects and control subjects with dichotomous traits, or between those with different trait values in a quantitative trait (for the most part, for ease of exposition, attributes¹ are referred to as if they were dichotomous).

Before the dramatic advances in technology and reductions in cost that now enable the sequencing of millions of base pairs of the genomes of millions of individuals, the hunt for genetic variants focused on a small number of previously identified polymorphisms of a small number of genes. The most widely studied of these genes (about four or five) are involved in regulating a class of neurotransmitters known as monoamines and have long been the target of various neuropsychiatric medications. Polymorphisms of these genes were associated with differences in such things as transcriptional and translational efficiency. For example, researchers classified polymorphisms of the *MAOA* gene as having high (H) or low (L) transcriptional efficiency and correlated these polymorphisms with various behaviors via CGA studies (Caspi et al., 2002; Sabol et al., 1998). A standard CGA study is hypothesis-driven; that is, researchers propose, on the basis of the presumed biological effect of, for example, *MAOA-H* versus

MAOA-L, that those with *MAOA-L* are more likely to engage in a given behavior and test this hypothesis using a data set composed of cases and controls. If either group exhibits a statistically significant difference in *MAOA-L* frequencies (in the direction indicated by the hypothesis), this is taken as evidence for the hypothesis.

In little more than a decade, researchers published thousands of studies reporting statistically significant correlations between the same polymorphisms of the same handful of genes and every conceivable behavior.² And there was much speculation concerning the practical implications of these findings. For example, in a well-known study linking *MAOA-L*, childhood adversity, and risk of "antisocial" behavior, Caspi et al. noted that their findings could "inform the development of future pharmacological treatments" (2002, p. 853). Researchers suggested that early intervention services for families of *MAOA-L* children might be a means to reduce violent crime (Brooks-Crozier, 2011; Muller-Spahn, 2008). Legal scholars and courts debated whether *MAOA-L* could be a defense in a criminal trial (Denno, 2011), and medical ethicists suggested that we might have a moral obligation to avoid having children with the *MAOA-L* genotype (Savulescu, 2014).

Despite such hype, from their inception, CGA studies were plagued by persistent failures of replication (Charney & English, 2012). In 2012, the editor of *Behavior Genetics*, the premier behavior-genetics publication, noted, in setting forth strict criteria for the publication of CGA studies:

The literature on candidate gene associations is full of reports that have not stood up to rigorous replication [and] as a result, the psychiatric and behavior genetics literature has become confusing, and it now seems likely that many of the published findings of the last decade are wrong or misleading and have not contributed to real advances in knowledge. (Hewitt, 2012, p. 1)

Eight years later, Flint et al. commented: "There are literally thousands of papers reporting the results of [CGA studies], but it's not too harsh to say simply that these studies have taught us nothing useful about the genetic basis of psychiatric disease" (2020, p. 60).

Postmortem

In postmortem analyses of CGA studies, researchers have focused on phenomena that could lead to Type I errors (i.e., false positives). Prominent among these are multiple hypothesis testing (or "P-hacking") and insufficient power.

Multiple hypothesis testing. Null-hypothesis-significance testing is the most widely used data-analysis method in scientific disciplines. For testing a single hypothesis, the commonly employed value is $p \leq .05$. When more than one test is run without any correction in the form of a more stringent (i.e., lower) p -value threshold, the overall type I error rate (i.e., false positives) is much greater than 5%. For example, if we were to compare the frequencies of the MAOA-L genotype of case and control subjects for 1,000 different behavioral traits, each of these traits would constitute a hypothesis and 1,000 χ^2 tests (or linear regressions), each with its own null hypothesis. If the null hypothesis was true, an α level of .05 could theoretically produce 50 “significant” correlations by chance alone. The most straightforward way to deal with multiple hypothesis testing is the Bonferroni correction, in which the α level is divided by the number of tests performed. Dividing .05 by 1,000, for example, yields a p value of .00005.

Examples in the CGA study literature of multiple hypothesis testing without any statistical correction include data mining; searching for any correlation between a polymorphism and the (hundreds or thousands) of behavioral variables in a data set; adding interaction terms in the form of $G \times G$ and/or $G \times E$ when the polymorphism, by itself, fails to exhibit a statistically significant correlation with a given behavior; limiting the findings to various subgroups (e.g., specific genders, specific ethnicities, or specific genders and ethnicities) when the polymorphism fails to exhibit a statistically significant correlation with the general study population; or some combination of the above (Charney & English, 2012).

Insufficient power. Most researchers consider CGA studies to have been “underpowered” (i.e., to have had insufficient sample sizes), an assumption based in part on views concerning effect size (Border et al., 2019; Chabris et al., 2012; Farrell et al., 2015). CGA studies were characterized by claims that single alleles could have large effects (e.g., the MAOA-H polymorphism was reported to increase the likelihood of voting by 10%; Fowler & Dawes, 2008). As we shall see, the current view is that individual allelic effect sizes of complex traits are, by comparison, minuscule. Hence, from the present perspective, CGA studies were insufficiently powered to identify alleles of such small effect size.

GWASs

Basics

The main methodology used in the search for polymorphisms underlying the heritability of complex behaviors

has rapidly shifted from CGA studies to GWASs (Visscher et al., 2017). The focus of most current GWASs is the SNP, a type of genetic variant characterized by the substitution of a particular nucleotide at a given position or locus on a DNA molecule. SNPs are largely, but by no means exclusively, diallelic; that is, they come in two possible forms (e.g., A or G). SNPs can occur anywhere in human genomes—within genes or in intergenic regions—and are the most common form of genetic variation in human populations (numerically, although not in terms of the percentage of the genome implicated). SNPs occur once in every 1,000 nucleotides on average, resulting in roughly 4 to 5 million SNPs in a person’s genome. Typically, researchers focus on the search for SNPs that are common in a population, the assumption being that genetic risk factors for variation in “common” attributes will themselves be common.

Most GWASs today are said to be “hypothesis free”; that is, they do not involve a specific polymorphism or polymorphisms known in advance that researchers hypothesize to be associated with a specific phenotype on the basis of its presumed physiological effects. Rather, large numbers of SNPs (e.g., a million or more) in large numbers of case and control subjects are compared to ascertain whether there is any difference in SNP frequencies between the two. There are no *a priori* assumptions as to what these genetic variants might be. This form of GWAS is, in effect, a form of data mining. However, although GWASs do not test preexisting hypotheses, they are by no means “hypothesis-free.” A comparison of the frequencies of each of a million SNPs amounts to a million χ^2 tests (or linear regressions), each with its own null hypothesis. If the null hypothesis was true, an α level of .05 could theoretically result in 50,000 “statistically significant” associations by chance alone. If we employ a Bonferroni correction and divide .05 by 1 million (the number of tests performed), the resulting P value is 5.0×10^{-8} . This is the threshold of statistical significance commonly employed in GWASs. An SNP that achieves this significance level is said to have *genome-wide significance* and/or to be a *lead SNP*.

Infinite infinitesimal alleles

Initially, the identification of alleles with genome-wide significance in behavior genetics was slow going, but researchers were convinced that the problem was insufficient power to identify risk alleles of minuscule effect size. This conviction was based, in part, on GWASs of variance in height. Although height is considered a highly heritable trait, by 2009, only three alleles of genome-wide significance had been identified, and the most strongly associated allele was estimated to increase the chance of being taller (but the difference would be

only ~0.4 cm) and explained only 0.3% of the total phenotypic variation in height across the population (McEvoy & Visscher, 2009). McEvoy and Visscher noted that these findings represented the lowest hanging fruit and that, in what has become something of a mantra, “Even bigger sample sizes would be needed to find genes of smaller effect” (2009, p. 298).

Findings such as these, including the limited number of such findings, were given theoretical justification by Fisher’s “infinite infinitesimal allele” model (Fisher, 1930/1990). Fisher proposed that the heritability of complex traits involved the inheritance of an indefinitely large (“infinite”) number of alleles, each contributing a minuscule (“infinitesimal”) amount to trait heritability. In modern terms, this would be described as “massive polygenicity.” For Fisher, the contribution of alleles to heritability was primarily additive—the same assumption that underlies the additive model of heritability. The average mathematical effect of two or more alleles on trait heritability in a population is equivalent to the sum of their average individual effects in that population. Although the average effect of any particular allele might be insignificant, the combined average effect of alleles was not. Fisher’s model has become a “new old” dogma in behavior genetics. It explains the so-called “problem of missing heritability”: Even with massive sample sizes, researchers have been able to identify only a fraction of the heritability of complex behaviors as estimated by classical twin studies (Young, 2019). And it serves as a justification for ever-expanding sample sizes. Researchers have taken to combining data from multiple different GWASs, resulting in sample sizes of hundreds of thousands to more than a million and performing a metaanalysis on the combined data (Zeggini & Ioannidis, 2009).

Turkheimer (2016) has compared this approach to “high-tech p-hacking”:

In genome-wide-association studies, data on hundreds of thousands of individual bits of DNA are collected in large samples and then searched for significant results at highly stringent p levels. If (as usually happens) no significant results are discovered the first time around, the process is repeated with even larger samples, continuing until something significant finally emerges. “Hits,” as they have come to be known, are now being accumulated for many behavioral characteristics, but the effect sizes for individual SNPs or alleles are vanishingly small. But does this methodology sound familiar? Genome-wide association is unapologetic, high-tech p-hacking. (p. 27)

To defenders of this approach, what Turkheimer describes as “vanishingly small” effect sizes for individual

SNPs are those predicted by Fisher’s model, and they can be identified only by using massive sample sizes. In seeming confirmation of this, with growing sample sizes have come reports of ever more alleles of genome-wide significance. For example, consider several metaanalyses of EA, measured as the self-reported number of years of schooling and often treated as a proxy variable for performance on intelligence tests (Rietveld et al., 2014). In 2014, Rietveld et al. performed a metaanalysis of the combined data from 64 studies (or cohorts) across Europe and the United States that had a total of 126,559 persons of European ancestry (PoEA). Using a standard GWAS significance level of $p \leq 5 \times 10^{-8}$, they reported one SNP of genome-wide significance for EA. By 2016, using a sample of 293,723 PoEA, researchers reported 74 SNPs of genome-wide significance (Okbay et al., 2016). In 2018, with a combined sample of more than 1 million PoEA from 71 different GWASs and combined microarray data for approximately 10 million SNPs, researchers reported the identification of 1,271 SNPs of genome-wide significance (J. J. Lee et al., 2018). A similar pattern has been observed for other behavioral attributes (Plomin & von Stumm, 2018). And J. J. Lee et al. (2018) reported that, in line with Fisher’s model, the median effect size of the 1,271 lead SNPs correspond to risk of only 1.7 weeks of schooling per allele and, taken together, they accounted for only 3.86% of the variance in EA.

Linkage disequilibrium

Basics. According to Mendel’s law of independent assortment, any given allele will be inherited independently of any other allele. For example, if someone inherits the SNP allele with an A at a specific position on a particular chromosome, this tells us nothing about the allele they will inherit on the same chromosome 100 base pairs (bp) away. In reality, SNPs that are “close” to each other tend to be inherited together because the entire segment of DNA on which they are located is inherited as a single piece. This segment of DNA is referred to as a *haplotype*. The genotypes of different SNPs within a haplotype tend to be inherited together. Suppose that two SNPs are located 1000 bp (or 1 kilobase [kb]) apart in a haplotype. In SNP rs1000,³ the nucleotides G or T occur, and in SNP rs2000, the nucleotides C or G occur. Suppose we know that the genotype of rs1000 is T. On this basis, we can infer that the genotype of rs2000 is G (with varying degrees of probability). This nonrandom association of alleles at two different loci in a haplotype is known as *linkage disequilibrium* (LD), and it plays a critical role in GWASs.

The alleles of genome-wide significance that researchers associate with an attribute, including the

1,271 SNPs of genome-wide significance identified by J. J. Lee et al. (2018), are *not* “causal.” Instead, they are SNPs assumed to be in LD with causal-risk alleles.

A current genome assay typically checks a person’s DNA sequences for a million preidentified SNPs called *marker SNPs*. Haplotype maps have been constructed across the genomes of different populations that tell researchers which SNPs in a particular region are likely in LD. Given the genotype of the marker SNPs, researchers can infer the genotype of many more SNPs (or other forms of genetic variation) known (or presumed) to be in LD with them. This process, known as *genotype imputation*, is used to improve both coverage and power of a GWAS by inferring the alleles of ungenotyped SNPs according to the LD patterns derived from directly genotyped marker SNPs (Marchini & Howie, 2010). This allows researchers to examine large segments of the genome without having to genotype each of the three billion pairs of DNA nucleotides.

If researchers find a “hit” or correlation for a marker SNP, the assumption is generally that the marker SNP itself is not “causal” (Bush & Moore, 2012). Instead, an unknown genetic variant in LD with the marker SNP is deemed causal. The marker SNP serves as a proxy for the presumed causal genetic variant. To say that a genetic variant is *causal* means, in this context, that it has some physiological effect on phenotypic risk (i.e., it is a cause of phenotypic risk). Perhaps it causes a slight change in the transcription rate of a given gene or results in an altered form of a protein. How the causal variant influences risk is not known because the causal variant itself is not known. Therefore, a distinction can be drawn between marker SNPs and causal alleles or causal-risk alleles. Regarding risk, a marker SNP could also be referred to as a risk allele—not because it has any effect on phenotypic risk but because it is in LD with a causal-risk allele that does influence phenotypic risk.

Whether any two SNPs are in LD is a matter of probability, not certainty (and, as we shall see below, LD can vary by population). Researchers use as a measure of the strength of correlation between any two marker SNPs either a coefficient of LD (D') or, more commonly, an r^2 , which is equivalent to the Pearson correlation coefficient (a measure of the strength of the linear relationship between two variables). LD can range from 0 (no correlation) to 1 (perfect LD; Fig. 1).

Complexities. One difficulty regarding LD is that multiple marker SNPs that achieve genome-wide significance could be in LD with the same unknown causal allele and with each other. If all of them were counted as having genome-wide significance, this would result in overcounting. A common method for tackling this problem is

called *clumping*. Marker SNPs in LD are thinned, and only the marker SNP with the lowest p value is retained. After this is done for every region of LD, remaining marker SNPs are assumed to be *independent* (i.e., not in LD). A criticism of clumping is that researchers typically select an arbitrarily chosen threshold for determining whether marker SNPs are in LD in the first place (Choi et al., 2020). Other forms of clumping involve a second stage in which lead SNPs that are physically close to each other are merged and considered to be a single locus. A second common approach is to permit the inclusion of SNPs in LD but to adjust (or shrink) effect estimates on the basis of their correlation structure—for example, two correlated SNPs would both be preserved, but their respective effect estimates would be downsized commensurate with the degree of correlation. There is now a proliferation of computational tools designed to account for LD, including the software LDpred, a Bayesian method that weights each SNP by (an approximation to) the posterior mean of its conditional effect, given other SNPs, and incorporates estimations based on observed LD patterns in a population reference sample (Vilhjalmsson et al., 2015).

Despite computational tools, estimating LD remains a significant challenge

Measures of LD are particularly sensitive to population-based differences in allele frequencies and the choice of a cutoff for minor allele frequency—that is, the level at which the less frequently occurring [minor] allele in a population is deemed rare and excluded from the analysis (Linck & Battey, 2019). Furthermore, researchers generally assume that to be in LD lead SNPs must be located no more than, for example, 50 kb apart. However, in what is known as long-range LD (LRLD), LD spans regions more than 1 Mb (1,000 kb) apart on a chromosome. The most familiar example of LRLD concerns a region of chromosome 6 known as the major histocompatibility complex, which plays a key role in the immune system, in which thousands of SNPs exhibit LRLD across a region greater than 1.5 Mb. Analysis has shown that LRLD of more than 5 Mb is common across the human genome (Nelson et al., 2020; Park, 2019). If LRLD is not considered, the result can be a significant overcounting of marker SNPs presumed to be independent (i.e., not in LD). It is worth noting that LRLD can be a sign of epistasis ($G \times G$), a phenomenon that is hard to accommodate in an additive model (C. Zhang et al., 2017).

Further difficulties in estimating LD occur because they confound the assumption on which LD depends: that people have identical genomes in the somatic cells of their bodies. These difficulties include: De novo

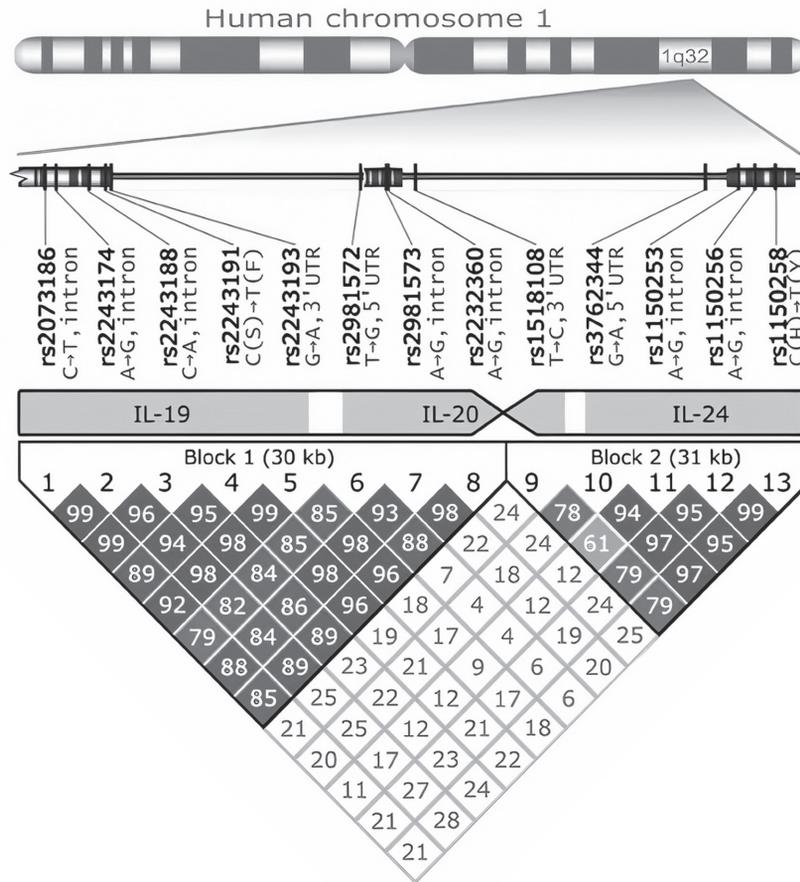


Fig. 1. Representation of two haplotypes on chromosome 1. The region 1q32 on the chromosome (to the right) is amplified below it. The location and the two forms of each single-nucleotide polymorphism (SNP) are shown. For example, SNP rs2073186 occurs as either a C or a T. It is located in an intron, a noncoding part of a gene. SNP rs2243191 occurs as a C or T. It is located in a protein-coding region of a gene and causes a single amino acid (the building block of proteins) to change from S (serine) to F (phenylalanine). Below this amplification are shown the three genes in which the SNPs are located—*IL-19*, *IL-20*, and *IL-24*—as well as intergenic regions (in white). Shown below that are two linkage disequilibrium (LD) blocks (or haplotypes) in region 1q32. The numbers 1 to 13 correspond to the SNPs shown above. At the bottom is a grid showing the degree of LD between each pair among the 13 SNPs. The numbers in the boxes are coefficients of LD (D'), which range from 0 to 1. For example, D' between SNP 1 (rs2073186) and SNP 3 (rs2243188) is .99, which indicates high LD. By contrast, D' between SNP 6 (rs2981572) and SNP 9 (rs1518108) is only .07. Reprinted by permission from Springer: *Genes & Immunity* (Possible relations between the polymorphisms of the cytokines *IL-19*, *IL-20* and *IL-24* and plaque-type psoriasis. Köks, S., Kingo, K., Vabrit, K., Rätsep, R., Karelson, M., Silm, H., & Vasar, E. *Genes & Immunity*, 6(5), 407–415. <https://doi.org/10.1038/sj.gene.6364216>), Copyright 2005.

copy-number variations (CNVs) and duplications or deletions of segments of DNA ranging from more than 50 bp to several megabases. Such variations are estimated to encompass up to 9.5% of the human genome (Iourov et al., 2019; Zarrei et al., 2015). They are de novo because they are not inherited and can be unique to every individual. De novo copy-number variations can result in the duplication or deletion of whole genes. Researchers have identified approximately 100 genes

for which both copies can be deleted in healthy persons without producing apparent phenotypic consequences. There is also extrachromosomal circular DNA, which are closed circles of DNA derived from chromosomal DNA that range in size from less than 100 bp to several kilobases and typically contain multiple active genes. Up to 10,000 extrachromosomal circular DNAs per cell have been identified in varying amounts and sizes and with different DNA sequences in all cell types (Ain et al.,

2020); polyploidy, which occurs when cells have more than two complete sets of chromosomes, is found, to varying degrees, in normal human cells and tissues. For example, almost all adult ventricular cardiomyocytes, which form the muscle wall of the heart, are polyploid, typically containing four complete sets of chromosomes (Gan et al., 2020). More than 50% of liver cells (hepatocytes) exhibit polyploidy (S. Zhang et al., 2019), and most large bone marrow cells (megakaryocytes) contain up to 32 sets of chromosomes (Mazzi et al., 2018). The result of these and other phenomena is widespread *somatic mosaicism*—the existence of different genomes in different cells of the body. Somatic mosaicism is now known to be the normal human condition and is prevalent in the brain (Breuss, 2020; Kaeser & Chun, 2020; B. Zhao, Wu, et al., 2019).

LD and causal inference. Given that postulated causal-risk alleles could be any alleles that are in LD with the marker SNPs, one would assume that nothing more could be said concerning biological causation. However, this has not stopped researchers from proposing biological pathways based on specific alleles. For example, J. J. Lee et al. (2018) observed that “relative to other genes, genes near our lead SNPs were overwhelmingly enriched for expression in the central nervous system [and] implicate genes involved in brain-development processes and neuron-to-neuron communication” (p. 1114). It is not clear what the authors mean here by “near,” “enriched,” or “implicated.” If by “near,” the authors mean within 50 to 500 kb of a marker SNP, they have failed to consider the possibility of LRLD. Furthermore, as far as being “overwhelmingly enriched” is concerned, there is nothing particularly distinctive about a gene being expressed in the brain. In-depth gene expression profiling has revealed that 84% of ~20,000 protein-coding genes are expressed in various regions of the human brain (Negi & Guda, 2017). This makes it likely that, by chance alone, SNPs that are “near” genes will be near genes whose protein products are expressed in the brain.

J. J. Lee et al. (2018) explained that they employed a “fine-mapping” statistical software program to identify 127 genes as being likely to contain causal SNPs in LD with some of their lead marker SNPs. They drew attention to one of these genes, *CACNA1H*, which, as the authors noted, is used to synthesize a calcium-channel protein that plays a role in neuronal excitability. The authors did not mention other genes “prioritized” by their software (Lee et al., Supplementary Tables) such as *PPIP5K2*, mutations of which have been associated with various forms of deafness; *PRKAG2*, transcribed to synthesize a protein involved in responding to energy demands in cardiac muscle and associated with various forms of heart disease; *EPB41L3*, transcribed to

synthesize erythrocyte membrane protein and implicated in several forms of cancer; as well as many genes of unknown function. And *CACNA1H* itself is also highly expressed in the kidney, liver, and heart and has been associated with certain forms of hypertension (Daniil et al., 2016). Applying the same or similar techniques and reasoning to GWASs of intelligence, Hill, Marioni, et al. noted (2019) that they “found evidence that neurogenesis and myelination—as well as genes expressed in the synapse, and those involved in the regulation of the nervous system—may explain some of the biological differences in intelligence” (p. 169), whereas Ye et al. (2020), drew attention to the brain-expressed genes in their “enrichment set” and the role of neurogenesis in risky behavior.

These researchers are cherry-picking genes “near” marker SNPs, the functions of which they believe accord with the nature of EA, intelligence, or risky behavior (i.e., all of these in some way involve the brain and neurons). This is a common practice, but it is not clear that any of this amounts to evidence. To get a sense of just how rudimentary such reasoning is, consider the growing body of evidence that the gut microbiome plays an essential role in brain development (Gao et al., 2020; Lu & Claud, 2019; C. R. Martin et al., 2018; Zhu et al., 2020). Perhaps relevant risk alleles are associated with genes (e.g., *FUT2*) that appear to affect microbiome composition in the gut (Kurilshikov et al., 2021).

Infinitesimal effects

What is perhaps most misleading about highlighting specific SNPs and using them to construct a schema of genetic neurodevelopmental etiology (of risk) is that according to researchers’ assumptions, the effect of any one of these SNPs on risk is *minuscule* (e.g., according to J. J. Lee et al. [2018], the marker SNP rs4860418 is associated with a .0071-unit decrease in genetic risk of EA. This effect size presumably matches that of the unknown causal-risk allele with which it is in LD). These effects are so small that it is inconceivable that the role of any of the unknown causal genetic variants could ever be demonstrated or analyzed experimentally. Of course, one could “knock out” the proposed gene in a mouse or find a family whose members lack the gene because of some rare condition and then connect the observed abnormal behavior with the phenotype under consideration. But this would simply be a way of distorting the findings by transforming what is insignificant and one effect among millions into the sole main effect.

To what extent is it biologically plausible, or even coherent, to assume that the cause of such a minuscule

Table 1. Lead SNPs for Intelligence and Educational Attainment Identified in Studies

Number of studies	Number of SNPs ($N = 6,272$)	SNPs (%)
1	5,432	93.44%
2	325	5.59%
3	43	0.74%
4	11	0.19%
5	2	0.03%
6	1	0.02%
7	0	0.00%
8	0	0.00%

effect can be traced back to specific inherited DNA sequences? The problem with such an assumption is that in a complex biological system, the signal or signature of a minuscule effect will be irretrievably lost in a sea of effects. As the effect size of variation in specific DNA sequences diminishes, the effect size of all the other processes normally involved in shaping gene expression grows, including the entire developmental process. Such diminutive allelic effect sizes make sense only in the statistical realm, not the biological realm.

Replication

Given the assumption of a single set or core set of genetic risk factors for a given phenotype (Maier et al., 2018), it is important to consider whether identification of lead marker SNPs for a given attribute is consistently replicated across studies. Let us assume that in this context, “lead” means a marker SNP of genome-wide significance (at $p \leq 5 \times 10^{-8}$) that has a p value lower than those of all the other marker SNPs with which it is assumed to be in LD (according to whatever statistical procedures are used in this determination).

Consider eight large meta-analyses of EA and/or intelligence/cognitive ability: (Davies et al., 2018; Hill, Marioni, et al., 2019; Kichaev et al., 2019; Lam et al., 2017, 2019; J. J. Lee et al., 2018; Okbay et al., 2016; Savage et al., 2018). Because EA is usually considered a proxy variable for intelligence, combining intelligence and EA should present no problems. Limiting ourselves to the unique marker SNPs found to have genome-wide significance in each study at $p \leq 5 \times 10^{-8}$ yields 6,272 SNPs total across the eight studies. Of these, 5,432, or 93.44%, occurred in only one study. (Table 1; see also the Supplemental Tables available online).⁴

This poor record of replication of purportedly independent and lead marker SNPs of genome-wide significance goes largely unnoticed, even though these same SNPs are often used to construct schemas of genetic causation of risk and PRSs.

PRSs

Basics

A PRS is intended to be a single-value estimate of a person’s genetic risk for an attribute (Choi et al., 2020). Take the example of constructing a PRS for EA. Suppose that a given GWAS (or combination of GWASs) contains information on 1 million marker SNPs for 100,000 persons as well as information on years of schooling completed. For each of the 100,000 members of the sample, the data would show the respondent’s EA and which SNP allele they possess for each of one million marker SNPs on a chip array. All these individual data could then be combined into “summary statistics,” which present the GWAS results in the form of population averages. For each SNP, the summary statistics identify the “effect” allele (A1; i.e., the marker SNP that shows a correlation with EA, which can be either positive or negative), the “no-effect” allele (A2), the frequency of A1 in the population, the average effect size of A1 on EA (an odds ratio for a dichotomous trait or a β for a quantitative trait), the standard error (SE), and the p value for the association. See Table 2 for an example.

Suppose we wanted to construct a newborn’s PRS for EA (specifically, her odds of completing high school), and suppose her genotype for marker SNP rs4686944 in Table 2 was GG, which is two “effect” or “risk” alleles. On the basis of our summary statistics, we would multiply the effect size of this allele (0.02041) by the number of risk alleles she has (2). We would repeat this process for each of her 1 million marker SNPs (after accounting for LD) that showed an association with completion of high school and sum the results. In theory, the resulting value would tell us the newborn’s “genetic risk” for higher or lower EA. In 2018, Plomin and von Stumm wrote as if this capability was imminent:

IQ GPSs [genetic polygenic score; equivalent to PRS] will be used to predict individuals’ genetic propensity to learn, reason and solve problems, not only in research but also in society, as direct to consumer genomic services provide GPS information that goes beyond single gene and ancestry information. We predict that IQ GPSs will become routinely available from direct-to-consumer companies along with hundreds of other medical and psychological GPSs that can be extracted from genome-wide genotyping on SNP chips. (p. 155)

At present, however, PRSs for behavior have no individual predictive value whatsoever (Morris et al., 2020), and whether any PRS for any phenotype has individual predictive value is an open question. Instead, their

Table 2. Simplified Summary Statistics for a Single SNP Allele From a Hypothetical Study of Educational Attainment

Marker name (SNP)	A1 (Effect allele)	A2	A1 Frequency	Effect size	SE	p
rs4686944	G	T	0.02041	0.029	0.01087	.0441

ostensible predictive value concerns populations: A PRS is said to be predictive if the average PRS in a case population is higher than the average PRS in a control population, or if the average PRS of those, for example, in the lowest decile of EA is lower than those in the highest decile.

Plomin and von Stumm (2018) made these points clear when discussing a PRS they constructed to predict students’ performance on a United Kingdom-wide examination, the General Certificate of Secondary Education (GCSE), administered at the end of compulsory education at age 16 years. According to the authors, a scatterplot between GCSE scores and the PRS (Fig. 2a) shows “the difficulty of predicting individual outcomes when the correlation is modest (0.30 in this example)” (Plomin & von Stumm, 2018, p. 156). Squaring this correlation, they estimated that the PRS predicted 9% of the variance in risk in their study population but noted that

Although higher [PRSs] can be seen to predict higher GCSE scores on average, there is great variability between individuals. . . . Individuals within the lowest and highest [PRS] deciles vary widely in school achievement. . . . The overlap in the two distributions is 61%. (p. 156)

However, on a more optimistic note, they comment regarding their study sample “EA2” (Fig. 2c):

Despite this variability, powerful predictions can be made at the extremes. . . . Specifically, the average school achievement of individuals in the lowest EA2 GPS decile is at the 28th percentile. For the highest EA2 GPS decile, the average school achievement is at the 68th percentile. (p. 156)

It is common practice to divide a PRS into lowest and highest deciles or quintiles and then note what seems like an impressive difference in the mean prevalence between the lowest and highest decile or quintile. But what may seem like an impressive difference is simply a general property of small correlations: Given large enough sample sizes, looking at the extremes will produce what appear to be large differences, even though the magnitude of the relationship is small. Finally, Plomin and von Stumm (2018), reflecting on

the poor individual predictive power of their PRS, have recourse to a familiar refrain: “As bigger and better GPSs emerge, the predictive power will increase” (p. 156).

Training the score. The construction of a PRS begins, as noted, with the summary statistics from various GWASs, referred to as the discovery sample. The score is then further developed on a training sample, which must be entirely separate from the discovery sample. (A good deal of confusion is generated by the lack of consistent terminology for these samples. What is referred to here as the training sample is also called, variously, the *target*, *validation*, *prediction*, or *replication* sample.) The objective of researchers, apparently, is to modify how they construct the PRS to achieve as high an R^2 in the training sample as possible. R^2 , or the incremental R^2 statistic, is a statistical measure representing the proportion of the variance for a dependent variable (in this case, risk of a phenotype of interest) that is explained by an independent variable or variables (marker SNPs) in a regression model. The higher the R^2 , the more of the variation in the data the PRS explains.

Researchers have an enormous amount of freedom in determining how to construct a PRS to achieve the largest possible R^2 . Consider the process of clumping, mentioned above. To achieve the highest R^2 , it is standard practice for researchers to modify both the LD r^2 thresholds (i.e., the level of correlation at which two SNPs are considered in LD), and the *clumping window*, the number of base pairs within which LD is considered possible (Prive et al., 2019). Thus, researchers might try out different combinations of r^2 thresholds (e.g., 0.01, 0.05, 0.1, 0.2, 0.5, 0.8, or 0.95) and clumping windows (e.g., 50, 100, 200, or 500 kb).

One might assume that only SNPs of genome-wide significance would be included in a PRS. However, it is now standard practice for researchers to try out different p -value thresholds for the inclusion of marker SNPs (A. R. Martin et al., 2019):

The standard PRS approach is to calculate several scores from SNPs meeting various p -value thresholds on a log scale ranging from genome-wide significance ($p < 5e-8$) to all independent SNPs ($p < 1$) [or $p = 1$], then compute and report accuracy for each PRS. In nearly all modern GWAS of

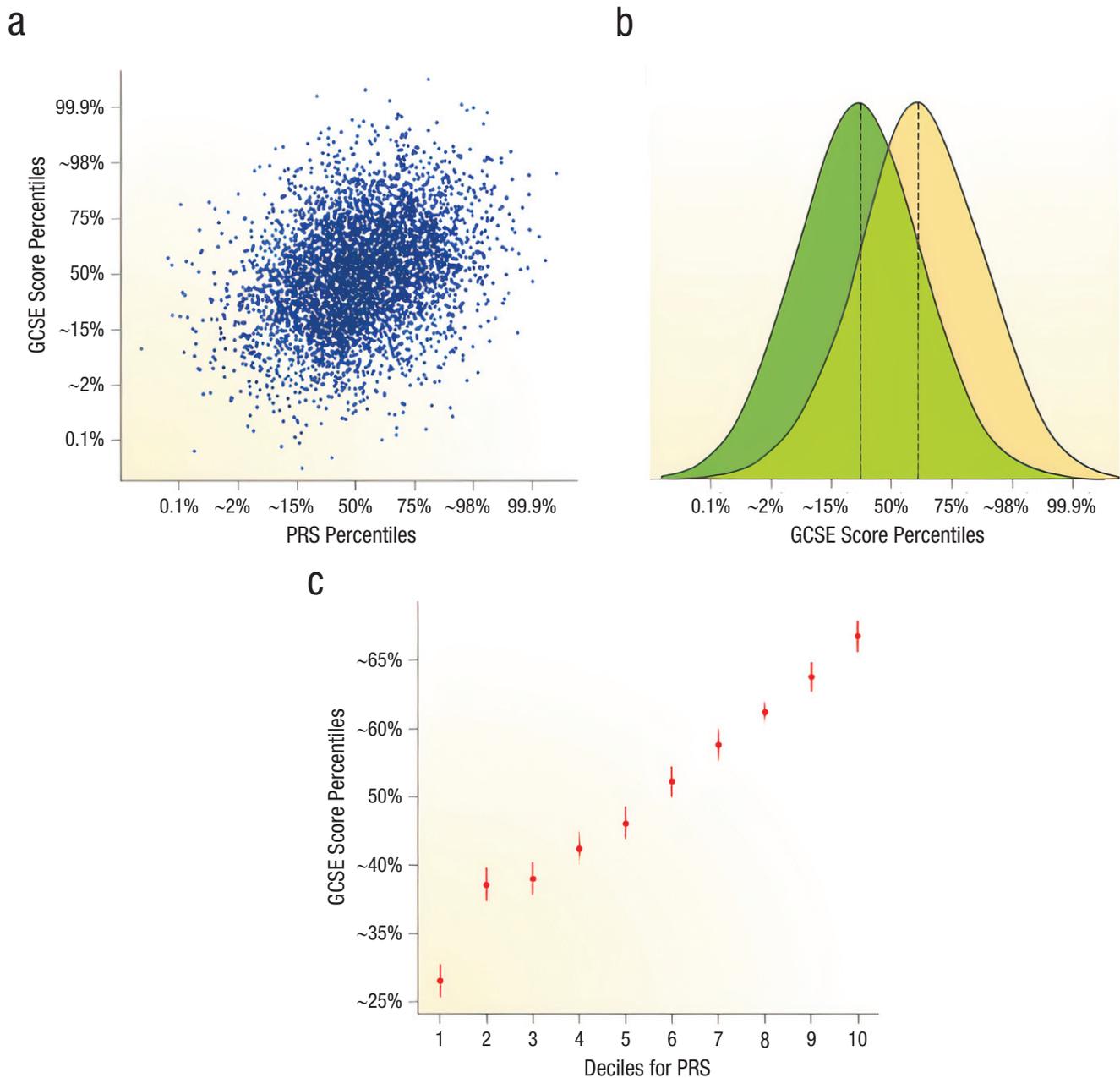


Fig. 2. Polygenic risk scores (PRSs) and individual versus population prediction. The scatterplot in (a) shows the relationship between educational attainment (EA) PRS percentiles and General Certificate of Secondary Education (GCSE) score percentiles. “EA2” refers to a specific PRS in Plomin and von Stumm (2018). The graph in (b) shows that the population distribution of PRS is normally distributed; the average PRS of those who score higher on the GCSE is slightly higher. On this basis, the authors predict that their PRS predicts 9% of the variance in genetic risk. The overlap in the two distributions is 61%. In the graph in (c), the sample was divided into 10 equal-sized groups (deciles) on the basis of EA2 PRS and shows the relationship between average EA2 PRSs and average GCSE score. Reprinted by permission from Springer: *Nature Reviews Genetics* (The new genetics of intelligence. Plomin, R., & von Stumm, S. *Nature Reviews Genetics*, 19(3), 148–159. <https://doi.org/10.1038/nrg.2017.104>), Copyright 2018.

complex traits, PRS computed using permissive p -value thresholds that aggregate the effects of 1,000s to 100,000s of independent SNPs typically explain more phenotypic variation than loci strictly meeting genome-wide significance. (p. 4)

In line with this, J. J. Lee et al. (2018, Supplementary Note, p. 128) noted that in constructing their PRS of EA, they tried out four different p -value thresholds: $p \leq 5 \times 10^{-8}$, 5×10^{-5} , 5×10^{-3} , and 1, and that their R^2 increased from .032 at $p \leq 5 \times 10^{-8}$ to 9.4% at $p \leq 1$.

They were able to achieve a yet higher R^2 of .114—the figure they ended up using in the end—when in addition to using $p \leq 1$, they switched from removing SNPs in LD with each other to using the software LDpred. (When the authors controlled for household income and the EA of the mother or father, the score’s incremental R^2 dropped to .046).

One might object that a significance threshold of $p = 1$ is nonsensical, but in the present context, it indicates (or entails) that after dealing with LD, *all* the remaining marker SNPs will be included in the PRS regardless of their statistical significance. Is this not an abandonment of the idea of statistical significance altogether while undertaking massive data mining? However, as the saying goes, the proof is in the pudding. Therefore, choosing a significance threshold of $p \leq 1$ (i.e., adopting no significance threshold) is justified, in the first instance, by the fact that it appears to work. It enables researchers to achieve a higher R^2 . But it also accords with the narrative of infinite infinitesimal alleles. Because the effects of individual SNPs are minuscule, the Bonferroni correction—or, as it turns out, any correction—is too stringent. Furthermore, what is statistically significant is not the effect of any individual SNP but their combined effect. Doubtless, any PRS constructed in this manner will contain many marker SNPs that do not affect phenotypic risk, but by including many more that do and otherwise would have been excluded, abandoning significance thresholds is justified.

There are many, many other decisions that researchers can make in their quest for the highest R^2 , including trying different algorithms embodied in various software programs to determine the weighting of individual SNPs, how to account for “winner’s curse” (the phenomenon in which estimates of the genetic effect based on new association findings tend to be upwardly biased), which cutoffs to use for minor allele frequency, or which programs to use to determine genotype imputation.

Model overfitting. All of this “freedom” on the part of researchers is a recipe for *model overfitting* (M. D. Lee et al., 2019; Mertens & Krypotos, 2019; Simmons et al., 2011). When researchers have so many “degrees of freedom”—that is, when they are free to try so many different analytic alternatives and value thresholds to achieve their preferred result, including setting significance thresholds—the likelihood of creating a manufactured rather than real statistical correlation is significant. All data sets have random quirks; an overfit model will incorporate these quirks to such an extent that the model explains the random error present in the data. Hence, an overfit model will not be generalizable because it describes the random error in the data rather than the relationships between variables. Ultimately, the regression coefficients represent noise rather than genuine relationships in the population. Inflated R^2 values are a sign

of overfit models, and overfit models are a common occurrence when researchers chase a high R^2 .

The problem of overfitting is exacerbated by the fact that in constructing PRSs, there are more predictors, in the form of individual SNPs, than the number of persons in the sample. For example, the study of J. J. Lee et al. (2018) has a sample size (N) of 1.1 million and 7.1 million predictors (i.e., SNPs). When there are more predictors than samples in the data set, you are confronted with a problem of “big-p, little-n” ($p > n$), also referred to as the “curse of dimensionality” (CoD; Altman & Krzywinski, 2018). Overfitting is an aspect of CoD. It occurs because the flexibility of prediction equations is in part determined by the number of variables involved (Lever et al., 2016). With increased flexibility, prediction and classification rules adapt to both the patterns in the population and the random idiosyncrasies of the training sample. Dimension-reduction methods such as principal component analysis can help to reduce dimensionality (see the Principal Components Analysis and LMM section, below) but may themselves be affected by CoD.

Estimating predictive performance is inferior to observing it in the real world

It is for these reasons that, according to the “gold standard,” before any claims that a PRS actually predicts anything can be made, a PRS developed on a training sample should be tested on a third sample entirely separate from both the discovery and training samples (Choi et al., 2020). Frequently, researchers do not do this. Sometimes, they will remove one or two study “cohorts” from their discovery sample (which consists of numerous cohorts), use these excluded cohorts as their training sample, and publish results derived from the training sample. To be sure, there are different statistical techniques designed to deal with the problem of model overfitting in the absence of a third independent sample and to estimate how accurately a predictive model will perform in practice. Cross-validation, for example, uses a single training sample that is then split into smaller “training” and “validation” subsets (Duncan et al., 2019). In discussing a recent study of the performance of cross-validation that was based on real-world data, Bauder et al. (2019) noted:

In real-world production applications, it is critical to establish a model’s usefulness by validating it on completely new input data and not just using the cross-validation results on a single historical data set. We present results for both evaluation methods, to include performance comparisons [and] find that using the separate training and [validation] sets generally outperforms cross-validation, indicating a better real-world model performance evaluation. (p. 19)

There is growing evidence that PRSs are overfit models. For example, Mostafavi et al. (2020) demonstrated that “the portability of a polygenic risk score can vary markedly depending on sample characteristics of both the original GWAS and the prediction set, and that this variation in prediction accuracy can be substantial” (p. 6). Variation in the samples for such things as the percentage of male versus female participants or the percentage of persons in various age or socioeconomic categories were all shown to substantially affect PRS accuracy.

Population Genetic Differences

Basics

In all the studies considered thus far, researchers employed data limited to persons of European ancestry. To be sure, PRSs have been constructed on the basis of discovery and training samples composed of persons of “Asian” or “African” descent. But care is taken to avoid “mixed” populations or, for example, using data derived from a GWAS of members of one ancestral group to construct a PRS for members of a different ancestral group. As J. J. Lee et al. (2018) note of their study:

Because the discovery sample used to construct the score consisted of people of European ancestry, we would not expect the predictive power of our score to be as high in other ancestry groups. Indeed, when . . . used to predict EA in a sample of African-Americans . . . the score only has an incremental R^2 of 1.6%, implying an attenuation of 85%. (p. 1115)

Researchers typically ascribe such R^2 attenuation to differences in *genetic population characteristics* between ancestral groups—differences in allele frequencies, haplotypes, degree of LD, and degree of genetic diversity (Duncan et al., 2019). These differences are the result of different population histories and are influenced by such phenomena as differences in migratory patterns, founder events, and population bottlenecks (loss of genetic variation that occurs when a new population is established by a very small number of people from a larger population), population expansions, relative population isolation, endogamy, inbreeding, adaptive pressures, and genetic drift (a random change in allele frequencies), to name a few. Genetic population characteristics are highly correlated with geography, although the sharing of certain allelic frequencies between populations need not entail a shared ancestry or geographic origin. Although researchers are aware of the existence of population

genetic differences *between* as well as *within* ancestral populations (as typically defined), they believe they can effectively handle the latter (but not the former).

A population is structured when it contains subpopulations often, but not exclusively, distinguished by geographic location that exhibit systematic differences in population genetic characteristics (or allele frequencies). Persons of European ancestry are a structured population at every level, and the more fine-grained one’s analysis, the more that nested levels of population structure appear. For example, the ancestral population of the United Kingdom is composed of 17 distinct population “clusters” that are highly localized and differ in their population genetic characteristics (Leslie et al., 2015). Structured populations, which are most populations, are an omnipresent threat to the validity of genetic association studies because of *population stratification*.

Population stratification

Population stratification occurs when researchers mistake population genetic differences in allele frequencies for differences between case subjects and control subjects in risk alleles. Consider a simplified example—simplified in that it involves only one allele when a PRS can involve millions. Because of population genetic differences, the Roma have a higher percentage of the *CYP2C19*2* allele (Petrovic et al., 2020). Suppose our phenotype was EA and the Roma were overrepresented in our study population. The Roma have, on average, much lower EA than other Europeans (Eurostat, n.d.). Hence, we might conclude that *CYP2C19*2* is a risk allele for *low* EA. However, any association between *CYP2C19*2* and EA would be the result of population stratification. Were we, for example, to exclude the Roma altogether from our sample, any correlation between *CYP2C19*2* and EA would disappear (assuming, that is, that there were no other ancestral populations that exhibited similar population differences in *CYP2C19*2* frequencies). The EA of the Roma are so low not because they have a high frequency of the *CYP2C19*2* allele but because of their history as a persecuted and excluded minority (Van den Bogaert, 2018).

The two most widely used statistical methods for dealing with population structure are principal components analysis (PCA) and linear mixed modeling (LMM; Y. Zhang & Pan, 2015). PCA is used to identify patterns or “axes of variation” that explain the greatest amount of variance in allele (i.e., SNP) frequency in the sample. In a GWAS, what typically accounts for the greatest amount of variation in allelic frequency (the first principal component) is not the attribute of interest but the

ancestries of the participants, which often correspond to geographical regions. Researchers adjust genotypes and phenotypes by amounts attributable to ancestry along each axis. It is common for researchers to do a PCA on the entire set of genotype data, then use the first 5, 10, 20, or more principal components as covariates in the association model. An alternative widely used approach is LMM, which incorporates both fixed and random effects, population structure being treated as a random effect.

There is strong evidence that neither of these approaches solves the problem of population stratification. For example, in several different studies, researchers reported that the average PRS for height increased from south-to-north across Europe (i.e., exhibited a “latitudinal cline”), paralleling average population differences in height from Italy to the Netherlands (Berg & Coop, 2014; Racimo et al., 2018; Robinson et al., 2015; Turchin et al., 2012; Zoledziewska et al., 2015). The results of these studies were highly touted not only as multiply replicated PRSs for height but also as an example of polygenic adaptation. Most of these studies were based on data assembled by an international collaborative effort known as the Genetic Investigation of ANthropometric Traits (GIANT) Consortium, consisting of the combined summary statistics from 79 individual GWASs totaling 253,288 persons of European ancestry from across Europe. In subsequent studies (Berg et al., 2019; Sohail et al., 2019), researchers attempted to replicate these findings using a larger sample from the U.K. Biobank.

The U.K. Biobank contains GWAS and “health and well-being” data of 500,000 volunteer participants from the United Kingdom. Researchers limited themselves to participants who self-identified as being of “white British ancestry” ($N = 336,474$). This study population was larger and more homogeneous, in terms of ancestry, than the population that constituted the GIANT Consortium from which the PRSs for height had been originally derived. Researchers failed to replicate the original findings. As Berg et al. (2019) noted of their results, “what once appeared an ironclad example of population genetic evidence for polygenic adaptation now lacks any strong support” (p. 14). Both Berg et al. (2019) and Sohail et al. (2019) concluded that the differences in the PRSs for height were picking up ancestral population differences in allele frequencies between, for example, the Italians and the Swedes, that had nothing to do with height. And they knew this because these scores did not identify differences in height (or at the very least, were significantly attenuated) in a more homogeneous population that contained neither Italians nor Swedes. Of their findings, Berg et al. (2019) commented that “methods for

correcting for population stratification in GWAS may not always be sufficient for polygenic trait analyses” (abstract).

Doubts concerning the effectiveness of current methodologies in dealing with population stratification have continued. For example, the U.K. Biobank, a large segment of which has been assumed to represent a single relatively homogeneous population of persons of “white British ancestry,” itself exhibits population structure (Cook et al., 2020). This is not surprising given that, as noted above, “white British ancestry” does not constitute a structure-free population. Both Cook et al. (2020) and Haworth et al. (2019), showed that PRSs based on GWAS data from the U.K. Biobank for traits including education, income, body mass index (BMI), hypertension, smoking, and alcohol consumption were correlated with *birth location*. These associations with geography persisted even after they used PCA involving up to 100 principal components (Zaidi & Mathieson, 2020) and LMM. Differences in all these attributes—education, income, BMI, hypertension, smoking, and alcohol consumption—and many more are known to vary by geographic region throughout Great Britain. At the same time, allele frequencies are known to differ throughout Great Britain by geographic region because of differences in ancestral populations, which provides a source of covariance between genotype and attribute that will lead to population stratification. As Haworth et al. note (2019), “this phenomenon is important, both as a source of ecological-level covariance between genotypes and geographically heterogeneous complex traits and because of its apparent persistence across different analytical contexts and modes of statistical adjustment” (p. 6),

Researchers uncovered the same problem in analyzing PRSs for coronary artery disease, rheumatoid arthritis, schizophrenia, waist–hip ratio, body-mass index, and height in a Finnish population (Kerminen et al., 2019). Ancestral Finns are a population that, given the nature of their migratory history and relative isolation from the rest of Europe, one would expect to be more genetically homogeneous than “whites of British ancestry” (Kaariainen et al., 2017). Nonetheless, even after employing PCA and LLM, the authors observed strong correlations between the geographic distribution of the PRSs and Finnish population structure that runs from southwest to northeast.

Sibship studies

Researchers generally assume that studies of siblings from the same parents are immune to population stratification because genetic differences between siblings are due to the random partitioning of parental alleles

(Selzam et al., 2019). Hence, it is common for researchers to attempt to replicate GWASs and PRSs using family studies. Many studies have shown that when a PRS score derived from a general-population GWAS discovery sample is applied to a sibling population, the R^2 significantly declines (J. J. Lee et al., 2018; Mostafavi et al., 2020; Trejo & Domingue, 2018). For example, J. J. Lee et al. (2018) noted a decline of 40% in their PRS when applied to a sibling study population. There have been numerous and at times convoluted attempts to explain this phenomenon, but the most straightforward is, in the words of Lello et al. (2020), that “at least some of the observed power in polygenic prediction among non-sibling people comes from effects such as subtle population stratification (perhaps correlated to environmental conditions or family socio-economic status)” (p. 9).

Zaidi and Mathieson (2020) demonstrated that if researchers use data from a general population GWAS discovery sample and develop a PRS using a sibling training sample, population stratification persists. Researchers have started to use sibling discovery samples to overcome this problem, and such studies have shown just how inflated population-based R^2 estimates are. For example, in a recent study of 159,701 siblings using sibling GWAS data (Howe et al., 2021), researchers observed that the within-sibships SNP heritability estimate for EA attenuated by 71% from the population estimate (population $b^2 = 0.14$; within-sibships $b^2 = 0.04$), while the estimate for cognition attenuated by 46% (population $b^2 = 0.24$; within-sibships $b^2 = 0.13$). Reported genetic correlations (for explanation, see Pleiotropy section below) between EA and height (Valge et al., 2019) and EA and BMI (Zeng et al., 2019), when reestimated with sibship data, were close to 0. The authors of this study (Howe et al., 2021) noted that “population estimates are likely to be driven by demography [i.e., population stratification] and indirect genetic effects” (p. 7).

Are sibling and family studies completely immune from population stratification, as is commonly assumed? They are not. It is known, for example, that in so-called admixed populations, stratification can persist in sibship studies (Mersha et al., 2015; Thornton et al., 2012; Wang et al., 2016). An admixed population is one in which members have recent ancestry from two or more separate sources. For example, as an “ancestral population,” African Americans arose within the past 400 years, and genome-wide ancestry estimates show averages of 73.2% African, 24.0% European, and 0.8% Native American ancestry (Bryc et al., 2015). Although persons of European ancestry are not an admixed population in this sense, the rise of multiracial and multiethnic children in the U.S. and Europe introduces admixture by degrees.

Genetic Correlations

Proxy phenotypes

What kind of phenotypes are “income” and “EA”? According to Hill, Davies, et al. (2019), there are no risk alleles for income *per se*. Instead, there are risk alleles for attributes that are correlated with, and have a causal effect on, income:

Genetic variants do not act directly on income; instead, genetic variants are associated with partly heritable traits (such as intelligence, conscientiousness, health, etc.), which have their own complex gene-to-phenotype paths (including neural variables) and are ultimately associated with income. (p. 19)

Indeed, the idea of a phenotype “income,” apart from all behaviors and features of persons that might influence it, and apart from the particular social, political, and economic institutions in which income exists and its distribution occurs, is nonsensical. Furthermore, as Hill, Davies, et al. (2019) note, any correlation between a given attribute (e.g., intelligence or health) and income is socially and historically contingent: Income could just as well depend on service to the party, and continually shifting political policies can alter, if not eliminate, the degree of correlation between income and health. The same is true of EA. The grade one completes in a particular educational system (i.e., EA), in addition to being correlated with various socioeconomic factors, is correlated with different behavioral and health-related attributes, some of which exert a causal influence on EA and may also be heritable. But “the grade one completes” is not itself a phenotype that influences the grade one completes in addition to these behavioral and physical attributes. It is for such reasons, presumably, that EA is treated as a proxy phenotype (Rietveld et al., 2014): “Educational attainment is a good proxy phenotype for cognitive performance, because cognitive performance is strongly genetically influenced and causally affects educational attainment” (p. 13791).

PRSs for proxy phenotypes such as income and EA can, presumably, predict heritable attributes with which income and EA are correlated and that exert a causal influence on them. However, income and EA are also commonly said to be “genetically correlated,” not only with each other and with intelligence but also with a host of other attributes. How is this possible? Inasmuch as there is now an explosion of studies exploring genetic correlations between phenotypes, this is a question of some importance.

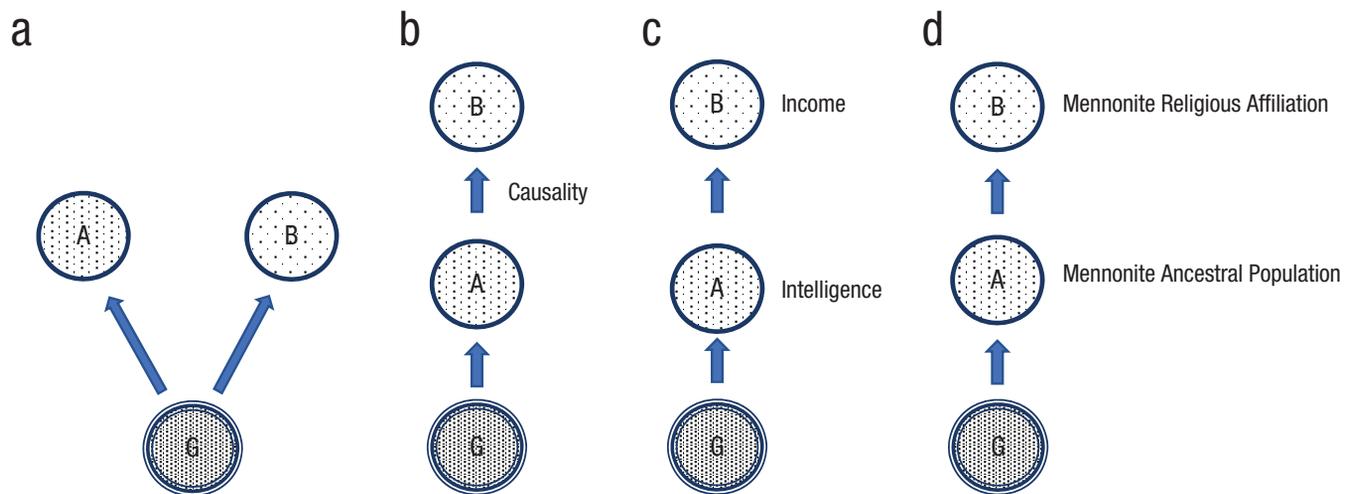


Fig. 3. Horizontal and vertical pleiotropy. In horizontal pleiotropy (a), two attributes, A and B, are correlated because they share a certain percentage of genetic risk alleles. In vertical pleiotropy (b), two attributes, A and B, are correlated not because they share genetic risk alleles, but because A has a causal influence on B. B could be heritable with its own risk alleles that are distinct from A’s, or it could be nonheritable and without any risk alleles, as in the case of income. In (c), the relationship between intelligence and income is depicted as an example of vertical pleiotropy. In (d), the relationship between being a member of a specific Central European subpopulation and having a Mennonite religious affiliation is depicted as an instance of vertical pleiotropy.

Pleiotropy

Genetic correlation, or “genetic overlap,” is a measure of the proportion of variance that two traits share as a result of shared risk alleles (Hackinger & Zeggini, 2017). It is taken as an indication of *pleiotropy*, the phenomenon in which the same allele serves as a risk factor for two different phenotypes. Researchers typically divide pleiotropy into horizontal pleiotropy and vertical (or mediated) pleiotropy.

Horizontal pleiotropy. Horizontal pleiotropy (Fig. 3a) occurs when two or more traits share a certain percentage of risk alleles. It can serve as an explanation for their phenotypic correlation or co-occurrence: They co-occur because their risk alleles are shared and hence inherited together. If genetic correlations are sufficiently high, a PRS for one trait can predict a certain percentage of the variance of the correlated trait. So called cross-trait PRSs have become a popular technique to measure the genetic similarity of polygenic traits (Leppert et al., 2020; Pouget et al., 2019). In the current context, claims of genetic correlation do not concern the correlation of known risk alleles (which, it will be recalled, are not identified in GWASs). Rather, it is a correlation between marker alleles and a given trait. The marker SNPs the two traits have in common are assumed to be in LD with the same risk alleles (van Rheenen et al., 2019), but there is no way that researchers can know this to be the case.

Vertical pleiotropy and heritability. In contrast to horizontal pleiotropy, vertical pleiotropy (Fig. 3b) occurs

when two phenotypes, A and B, are correlated not because they share risk alleles but rather because A has a causal effect on B. B may be heritable with causal-risk alleles of its own, or it may not be heritable. The latter is the type of relationship assumed, for example, by Hill, Davies, et al. (2019) between intelligence (and health) and income. Intelligence (A), which possesses (according to Hill and most others) genetic risk alleles, stands in a causal relation to income (B), which does not (Fig. 3c).⁵ However, calling vertical pleiotropy a form of pleiotropy is an egregious misnomer. The correlation between intelligence and income has nothing to do with shared alleles because “genetic variants do not act directly on income” (Hill, Davies, et al., 2019, p. 19). Any genetic correlation between intelligence—or anything else—and income, is *spurious*, and it is for this reason that, until recently, vertical pleiotropy was referred to as *spurious pleiotropy* (van Rheenen et al., 2019; Wagner & Zhang, 2011).

Equally misleading are claims that income and EA are *heritable*. One might object that nothing in the concept of heritability requires that a trait deemed heritable be influenced by the transmission of parental risk alleles *for that trait*. It is sufficient that a heritable trait A stands in a causal relationship to another trait B. If we are going to accept this, then it is not clear on what grounds population stratification can be considered confounding. Take the following example. The Mennonites represent a branch of the Anabaptist movement that began in northern and central Europe in the 16th century and have a well-documented migration and genealogical history. They are an ancestral population with a history of endogamy, and contemporary Mennonites exhibit distinct

population genetic characteristics, including differences in allele frequencies (Melton, 2012). Consider a PRS that predicts (B) Mennonite religious affiliation, which, presumably, is nonheritable (Fig. 3d). It predicts this because it predicts (A) being a member of the Mennonite population. Being a member of this population is heritable and has genetic risk factors. The relationship between A and B is causal. Why would we not say that being a member of the Mennonite ancestral population (A) exhibits “vertical pleiotropy” with Mennonite religious affiliation (B) and that as a result, Mennonite religious affiliation is heritable? “Religious affiliation” is often referred to as an example of one of the few human “behaviors” that is not heritable (D’Onofrio et al., 1999). Given the claim that income and EA are heritable, it is not clear why religious affiliation is not heritable.

An abundance of correlations

With more genetic correlation studies have come ever more reports of genetic correlations (whether horizontal or vertical). Intelligence, for example, has been reported to be significantly positively genetically correlated with, among other things, EA, income, autism, short-sightedness, cheese intake, eating Muesli cereal, grip strength in the right hand, age at first birth, longevity, parental longevity, voting behavior, anorexia nervosa, wearing glasses or contact lenses, subjective well-being, intracranial volume, height, drinking ground coffee, belonging to a religious group as a leisure activity, and iron and magnesium levels, and to be negatively correlated⁶ with type 2 diabetes, schizophrenia, hyperopia, bipolar disorder, Alzheimer’s disease, obesity/BMI, number of children, neuroticism, major depressive disorder, coronary artery disease, heart attack, ischemic and large vessel disease stroke, bipolar disorder, attention-deficit/hyperactivity disorder, osteoarthritis, hypertension, levels of C-reactive protein (a protein made in response to inflammation), and lung cancer (Aaroe et al., 2021; Abbott et al., 2021; Davies et al., 2018; Engen et al., 2020; Hagenaars et al., 2016; Hill, Davies, et al., 2019; Hill, Marioni, et al., 2019; Ligthart et al., 2018; Savage et al., 2018; Valge et al., 2019; Y. Zhao, Ning, et al., 2019). Given the number of correlations, when researchers claim that a PRS predicts intelligence, how do they know they are predicting *intelligence* (however precisely this is measured) and not one or more of the above?

Many findings of genetic correlations are contradictory and/or have not been consistently replicated. For example, Engen et al. (2020) report no negative genetic correlation between intelligence and schizophrenia; Hill, Davies, et al. (2019) report no genetic correlation between intelligence and subjective well-being; and

Hill, Marioni, et al. (2019) report no negative genetic correlation between intelligence and bipolar disorder. In a study using sibship data, Howe et al. (2021), report no genetic correlation between intelligence and height, BMI, and levels of C-reactive protein. It is often assumed that high genetic correlations entail high phenotypic correlations (an assumption known as “Cheverud’s conjecture” [Cheverud, 1988]). But such need not be the case. For example, in a study of assorted psychiatric conditions, Roelfs et al. (2021) noted that a large portion of significant genetic correlation occurred between attributes that were phenotypically uncorrelated. This might help explain seemingly contradictory findings. For example, researchers have reported a positive genetic correlation between autism and EA (Grove et al., 2019), although phenotypically, the correlation is negative (Toft et al., 2021). Autism has also been genetically correlated, in different studies, with both higher intelligence (Bulik-Sullivan et al., 2015; Clarke et al., 2016) and intellectual disability (Jensen et al., 2020). A phenotype may show a positive genetic correlation with one of two genetically correlated traits and a negative correlation with the other. For example, schizophrenia is frequently reported to be genetically correlated with lower general cognitive function and higher EA (Lam et al., 2019; Ohi et al., 2018; Savage et al., 2018), even though higher general cognitive function and higher EA are reported to be highly genetically and phenotypically correlated.

Conclusion

It has been argued that many of the same problematic research practices that undermined CGA studies persist in the era of GWASs. Some of them, such as data mining without statistical correction and unrestrained researcher degrees of freedom, are, for the most part, no longer seen as problematic. Replication studies have failed to consistently find the same results, a problem compounded by a lack of clarity as to what counts as replication in the first place. Population stratification is often cited as one of the reasons for the failure of CGA studies (Cardon & Palmer, 2003), and there is overwhelming evidence that bias due to population stratification persists as well. It should not be surprising that techniques designed to deal with correlations involving a single allele are ineffective when dealing with tens of millions of alleles. The CGA study era was characterized by claims that transcended the available knowledge. Currently, however, researchers do not hesitate to propose genetic neuronal pathways underlying variation in complex behaviors without knowing what the genetic risk factors are; or they assume that two genetically correlated traits share not only the same marker SNPs

but also the same risk alleles in LD with them. As new sources of bias are uncovered and researchers attempt to deal with them, the fraction of heritability explained by SNPs goes down, not up. The “excessive pleiotropy” of the CGA study era concerned only a handful of polymorphisms. But the proliferating numbers of genetic correlations reported in the GWAS era do not seem any more biologically plausible, despite involving thousands to millions of alleles. It also raises the question of how researchers know that their PRSs are, in fact, predicting what they claim they are. Finally, it matters whether a social aspect of a person such as income or EA is heritable or is correlated with—because it is influenced by—a heritable characteristic (e.g., chronic illness). Heritability by association is no more a scientific principle than guilt by association is a moral one.

Researchers employ the tools of modern genetic research in service of Fisher’s century-old model. This model, however, is not up to the task of serving as a guide for identifying (for the most part still hypothetical) risk alleles. Using this model is like trying to use Newtonian physics to study objects traveling at the speed of light. Current GWASs are focused almost exclusively on finding additive effects when recent molecular—as opposed to statistical—evidence indicates that biological systems are anything but additive; that is, they are characterized by widespread epistasis or $G \times G$ interactions (Miton et al., 2021; Reddy & Desai, 2021). Nothing points to the limitations of Fisher’s model more clearly than supposed “genetic correlations.” To assume that $X\%$ of alleles shared between two phenotypes entails $X\%$ of shared risk reduces biological causation to a matter of counting. To take the presence of an allele as equal to its assumed fractional additive effect is to ignore the genetic, epigenetic, and environmental “background” and the interactive biological system within which allelic effects come to be in the first place. Most of the statistical algorithms employed in the GWAS era depend on the assumption that potentially highly disruptive phenomena such as LRLD, de novo CNVs, extrachromosomal circular DNA, polyploidy, somatic mosaicism, and cryptic population structure are absent, even though we know they are widespread. Of course, much more needs to be said about these matters than can be said here.

In describing what he calls the “gloomy prospect” for using GWAS to find alleles underlying heritability, Turkheimer commented (2016):

EA and divorce are not discernible entities at a genetic level of analysis. What we will see instead is a proliferation of small, diverse, contingent findings that do not accumulate into coherent scientific theories. These will not be robust findings

with large effect sizes; they will be the signature of a complex problem being addressed at the wrong level of analysis. They will be the keyless sidewalk under the genomic streetlight. (p. 27)

Turkheimer was right. There is little evidence that current approaches have either advanced our knowledge of how genes contribute to complex behaviors or given us new tools to predict them.

Transparency

Action Editor: Laura A. King

Editor: Laura A. King

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

ORCID iD

Evan Charney  <https://orcid.org/0000-0001-7884-7410>

Notes

1. The terms *attributes*, *traits*, and *phenotypes* will be used interchangeably. The use of the term *phenotype* is not intended, for the most part, to imply an assumption of the existence of a corresponding genotype.
2. For a list of some of the phenotypes reported to be associated with polymorphisms of *MAOA*, *5HTT*, *DRD2*, and *DRD4*, see Charney (2012).
3. The rs number is an accession number used by researchers and databases to refer to specific SNPs. It stands for Reference SNP cluster ID.
4. See the “SNPs per Study” and “Table SNPs > 1 studies” tab in the Supplemental Material. The data for all these studies are available online at GWAS Catalog (<https://www.ebi.ac.uk/gwas/>).
5. Researchers generally treat the causal relationship between intelligence and income/EA in the same manner: (Heritable) intelligence can be a cause of variation in EA, but variation in EA is not considered a possible cause of variation in intelligence, despite a good deal of evidence to the contrary (Hegelund et al., 2020; Kremen et al., 2019; Leist et al., 2021). If both intelligence and EA influenced each other, they would stand in a relationship of reciprocal or bidirectional causation.
6. A negative genetic correlation indicates that individuals with high genotypic values for trait A tend to have low genotypic values for trait B.

References

- Aaroe, L., Appadurai, V., Hansen, K. M., Schork, A. J., Werge, T., Mors, O., Borglum, A. D., Hougaard, D. M., Nordentoft, M., Mortensen, P. B., Thompson, W. K., Buil, A., Agerbo, E., & Petersen, M. B. (2021). Genetic predictors of educational attainment and intelligence test performance predict voter turnout. *Nature Human Behavior*, *5*, 281–291. <https://doi.org/10.1038/s41562-020-00952-2>

- Abbott, L., Neale, B., & Palmer, D. (2021). *Genetic correlation between traits and disorders in the UK Biobank*. <https://ukbb-rg.hail.is/>
- Ain, Q., Schmeer, C., Wengerodt, D., Witte, O. W., & Kretz, A. (2020). Extrachromosomal circular DNA: Current knowledge and implications for CNS aging and neurodegeneration. *International Journal of Molecular Sciences*, *21*(7), Article 2477. <https://doi.org/10.3390/ijms21072477>
- Altman, N., & Krzywinski, M. (2018). The curse(s) of dimensionality. *Nature Methods*, *15*(6), 399–400. <https://doi.org/10.1038/s41592-018-0019-x>
- Bauder, R. A., Herland, M., & Khoshgoftaar, T. M. (2019). Evaluating model predictive performance: A Medicare fraud detection case study 2019. In *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*. Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/IRI.2019.00016>
- Beaver, K. M., Wright, J. P., DeLisi, M., & Vaughn, M. G. (2012). Dopaminergic polymorphisms and educational achievement: Results from a longitudinal sample of Americans. *Developmental Psychology*, *48*(4), 932–938. <https://doi.org/10.1037/a0026313>
- Berg, J. J., & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLOS Genetics*, *10*(8), Article e1004412. <https://doi.org/10.1371/journal.pgen.1004412>
- Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., Boyle, E. A., Zhang, X., Racimo, F., Pritchard, J. K., & Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *eLife*, *2019*(8), Article e39725. <https://doi.org/10.7554/eLife.39725>
- Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *American Journal of Psychiatry*, *176*(5), 376–387. <https://doi.org/10.1176/appi.ajp.2018.18070881>
- Breuss, M. A. (2020). *Somatic mosaicism in the mature brain reveals clonal cellular distributions during cortical development*. bioRxiv. <https://www.biorxiv.org/content/10.1101/2020.08.10.244814v1>
- Brooks-Crozier, J. (2011). The nature and nurture of violence—early intervention services for the families of MAOA-low children as a means to reduce violent crime and the costs of violent crime. *Connecticut Law Review*, *140*, 531–573. https://opencommons.uconn.edu/law_review/140/
- Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & Mountain, J. L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics*, *96*(1), 37–53. <https://doi.org/10.1016/j.ajhg.2014.11.010>
- Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R. ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan, L., Perry, J. R., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L., & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. <https://doi.org/10.1038/ng.3406>
- Bush, W. S., & Moore, J. H. (2012). Genome-wide association studies. *PLOS Computational Biology*, *8*(12), Article e1002822. <https://doi.org/10.1371/journal.pcbi.1002822>
- Cardon, L. R., & Palmer, L. J. (2003). Population stratification and spurious allelic association. *Lancet*, *361*(9357), 598–604. [https://doi.org/10.1016/s0140-6736\(03\)12520-2](https://doi.org/10.1016/s0140-6736(03)12520-2)
- Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A., & Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, *297*(5582), 851–854. <https://doi.org/10.1126/science.1072290>
- Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J., Cesarini, D., van der Loos, M., Johannesson, M., Magnusson, P. K., Lichtenstein, P., Atwood, C. S., Freese, J., Hauser, T. S., Hauser, R. M., Christakis, N., & Laibson, D. (2012). Most reported genetic associations with general intelligence are probably false positives. *Psychological Science*, *23*(11), 1314–1323. <https://doi.org/10.1177/0956797611435528>
- Charney, E. (2012). *4 genes predict everything*. <https://rb.gy/zx9fd7>
- Charney, E., & English, W. (2012). Candidate genes and political behavior. *American Political Science Review*, *106*(1), 1–34. <https://doi.org/10.1017/s0003055411000554>
- Cheverud, J. M. (1988). A comparison of genetic and phenotypic correlations. *Evolution*, *42*(5), 958–968. <https://doi.org/10.1111/j.1558-5646.1988.tb02514.x>
- Choi, S. W., Mak, T. S., & O'Reilly, P. F. (2020). Tutorial: A guide to performing polygenic risk score analyses. *Nature Protocols*, *15*(9), 2759–2772. <https://doi.org/10.1038/s41596-020-0353-1>
- Clarke, T.-K., Lupton, M. K., Fernandez-Pujals, A. M., Starr, J., Davies, G., Cox, S., Pattie, A., Liewald, D. C., Hall, L. S., MacIntyre, D. J., Smith, B. H., Hocking, L. J., Padmanabhan, S., Thomson, P. A., Hayward, C., Hansell, N. K., Montgomery, G. W., Medland, S. E., Martin, N. G., . . . McIntosh, A. M. (2016). Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Molecular Psychiatry*, *21*(3), 419–425. <https://doi.org/10.1038/mp.2015.12>
- Cook, J. P., Mahajan, A., & Morris, A. P. (2020). Fine-scale population structure in the UK Biobank: Implications for genome-wide association studies. *Human Molecular Genetics*, *29*(16), 2803–2811. <https://doi.org/10.1093/hmg/ddaa157>
- Daniil, G., Fernandes-Rosa, F. L., Chemin, J., Blesneac, I., Beltrand, J., Polak, M., Jeunemaitre, X., Boulkroun, S., Amar, L., Strom, T. M., Lory, P., & Zennaro, M. C. (2016). CACNA1H mutations are associated with different forms of primary aldosteronism. *EBioMedicine*, *13*, 225–236. <https://doi.org/10.1016/j.ebiom.2016.10.002>
- Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., Hagenaars, S. P., Ritchie, S. J., Marioni, R. E., Fawns-Ritchie, C., Liewald, D. C. M., Okely, J. A., Ahola-Olli, A. V., Barnes, C. L. K., Bertram, L., Bis, J. C., Burdick, K. E., Christoforou, A., DeRosse, P., . . . Deary, I. J. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications*, *9*(1), Article 2098. <https://doi.org/10.1038/s41467-018-04362-x>

- Denno, D. W. (2011). Courts' increasing consideration of behavioral genetics evidence in criminal cases: Results of a longitudinal study. *Michigan State Law Review*, 2011, 967–1047. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2065523
- D'Onofrio, B. M., Eaves, L. J., Murrelle, L., Maes, H. H., & Spilka, B. (1999). Understanding biological and social influences on religious affiliation, attitudes, and behaviors: A behavior genetic perspective. *Journal of Personality*, 67(6), 953–984. <https://doi.org/10.1111/1467-6494.00079>
- Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R., & Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications*, 10(1), Article 3328. <https://doi.org/10.1038/s41467-019-11112-0>
- 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. <https://doi.org/10.1176/appi.ajp.2011.11020191>
- Dunkel, C. S., Woodley of Menie, M. A., Pallesen, J., & Kirkegaard, E. O. W. (2019). Polygenic scores mediate the Jewish phenotypic advantage in educational attainment and cognitive ability compared with Catholics and Lutherans. *Evolutionary Behavioral Sciences*, 13(4), 366–375. <https://doi.org/10.1037/ebs0000158>
- Engen, M. J., Lyngstad, S. H., Ueland, T., Simonsen, C. E., Vaskinn, A., Smeland, O., Bettella, F., Lagerberg, T. V., Djurovic, S., Andreassen, O. A., & Melle, I. (2020). Polygenic scores for schizophrenia and general cognitive ability: Associations with six cognitive domains, premorbid intelligence, and cognitive composite score in individuals with a psychotic disorder and in healthy controls. *Translational Psychiatry*, 10(1), Article 416. <https://doi.org/10.1038/s41398-020-01094-9>
- Eurostat. (n.d.). *Population by educational attainment level, sex and age (%) - main indicators*. Retrieved on August 1, 2021, from https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=edat_lfse_03&lang=en
- Farrell, M. S., Werge, T., Sklar, P., Owen, M. J., Ophoff, R. A., O'Donovan, M. C., Corvin, A., Cichon, S., & Sullivan, P. F. (2015). Evaluating historical candidate genes for schizophrenia. *Molecular Psychiatry*, 20(5), 555–562. <https://doi.org/10.1038/mp.2015.16>
- Fiedorowicz, J. G., Moser, D. J., Hynes, S. M., Beglinger, L. J., Schultz, S. K., & Ellingrod, V. L. (2007). LA allelic heterozygosity of the 5HTTLPR polymorphism is associated with higher cognitive function and lower interpersonal sensitivity. *Psychiatric Genetics*, 17(1), 3–4. <https://doi.org/10.1097/YPG.0b013e328010f498>
- Fisher, R. A. S. (1990). *The genetical theory of natural selection*. The Clarendon Press. (Original work published 1930)
- Flint, J., Greenspan, R. J., & Kendler, K. S. (2020). *How genes influence behavior*. Oxford University Press.
- Fowler, J. H., & Dawes, C. T. (2008). Two genes predict voter turnout. *The Journal of Politics*, 70(3), 579–594. <https://doi.org/10.1017/s0022381608080638>
- Frydman, C., Camerer, C., Bossaerts, P., & Rangel, A. (2011). MAOA-L carriers are better at making optimal financial decisions under risk. *Proceedings of the Royal Society B: Biological Science*, 278(1714), 2053–2059. <https://doi.org/10.1098/rspb.2010.2304>
- Gan, P., Patterson, M., & Sucov, H. M. (2020). Cardiomyocyte polyploidy and implications for heart regeneration. *Annual Review of Physiology*, 82, 45–61. <https://doi.org/10.1146/annurev-physiol-021119-034618>
- Gao, W., Baumgartel, K. L., & Alexander, S. A. (2020). The gut microbiome as a component of the gut-brain axis in cognitive health. *Biological Research for Nursing*, 22(4), 485–494. <https://doi.org/10.1177/1099800420941923>
- Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Baekvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., . . . Borglum, A. D. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, 51(3), 431–444. <https://doi.org/10.1038/s41588-019-0344-8>
- Hackinger, S., & Zeggini, E. (2017). Statistical methods to detect pleiotropy in human complex traits. *Open Biology*, 7(11), Article 170125. <https://doi.org/10.1098/rsob.170125>
- Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C., Ritchie, S. J., Marioni, R. E., Fawns-Ritchie, C., Cullen, B., Malik, R., Metastroke Consortium, I. C., f, B. P. G., SpiroMeta, C. Charge Consortium Pulmonary Group, C. C. A., Longevity, G., Worrall, B. B., Sudlow, C. L., Wardlaw, J. M., Gallacher, J., Pell, J., . . . Deary, I. J. (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank ($N=112\,151$) and 24 GWAS consortia. *Molecular Psychiatry*, 21(11), 1624–1632. <https://doi.org/10.1038/mp.2015.225>
- Harden, K. P. (2021). “Reports of my death were greatly exaggerated”: Behavior genetics in the postgenomic era. *Annual Review of Psychology*, 72, 37–60. <https://doi.org/10.1146/annurev-psych-052220-103822>
- Haworth, S., Mitchell, R., Corbin, L., Wade, K. H., Dudding, T., Budu-Aggrey, A., Carslake, D., Hemani, G., Paternoster, L., Smith, G. D., Davies, N., Lawson, D. J., & Timpson, N. J. (2019). Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nature Communications*, 10(1), Article 333. <https://doi.org/10.1038/s41467-018-08219-1>
- Hegelund, E. R., Grønkvær, M., Osler, M., Dammeyer, J., Flensburg-Madsen, T., & Mortensen, E. L. (2020). The influence of educational attainment on intelligence. *Intelligence*, 78, Article 101419. <https://doi.org/10.1016/j.intell.2019.101419>
- Hewitt, J. K. (2012). Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behavior Genetics*, 42(1), 1–2. <https://doi.org/10.1007/s10519-011-9504-z>
- Hill, W. D., Davies, N. M., Ritchie, S. J., Skene, N. G., Bryois, J., Bell, S., Di Angelantonio, E., Roberts, D. J., Xueyi, S., Davies, G., Liewald, D. C. M., Porteous, D. J., Hayward, C., Butterworth, A. S., McIntosh, A. M., Gale, C. R., & Deary, I. J.

- (2019). Genome-wide analysis identifies molecular systems and 149 genetic loci associated with income. *Nature Communications*, *10*(1), Article 5741. <https://doi.org/10.1038/s41467-019-13585-5>
- Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., McIntosh, A. M., Gale, C. R., Davies, G., & Deary, I. J. (2019). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry*, *24*(2), 169–181. <https://doi.org/10.1038/s41380-017-0001-5>
- Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., Chittoor, G., Lind, P. A., Palviainen, T., van der Zee, M. D., Cheesman, R., Mangino, M., Wang, Y., Li, S., Klaric, L., Ratliff, S. M., Bielak, L. F., Nygaard, M., Reynolds, C. A., . . . Davies, N. M. (2021). *Within-sibship GWAS improve estimates of direct genetic effects*. bioRxiv 2021.2003.2005.433935. <https://doi.org/10.1101/2021.03.05.433935>
- Iourov, I. Y., Vorsanova, S. G., & Yurov, Y. B. (2019). The variome concept: Focus on CNVariome. *Molecular Cytogenetics*, *12*, Article 52. <https://doi.org/10.1186/s13039-019-0467-8>
- Jensen, M., Smolen, C., & Girirajan, S. (2020). Gene discoveries in autism are biased towards comorbidity with intellectual disability. *Journal of Medical Genetics*, *57*(9), 647–652. <https://doi.org/10.1136/jmedgenet-2019-106476>
- Kaariainen, H., Muilu, J., Perola, M., & Kristiansson, K. (2017). Genetics in an isolated population like Finland: A different basis for genomic medicine? *Journal of Community Genetics*, *8*(4), 319–326. <https://doi.org/10.1007/s12687-017-0318-4>
- Kaesler, G., & Chun, J. (2020). Brain cell somatic gene recombination and its phylogenetic foundations. *Journal of Biological Chemistry*, *295*(36), 12786–12795. <https://doi.org/10.1074/jbc.REV120.009192>
- Kerminen, S., Martin, A. R., Koskela, J., Ruotsalainen, S. E., Havulinna, A. S., Surakka, I., Palotie, A., Perola, M., Salomaa, V., Daly, M. J., Ripatti, S., & Pirinen, M. (2019). Geographic variation and bias in the polygenic scores of complex diseases and traits in Finland. *American Journal of Human Genetics*, *104*(6), 1169–1181. <https://doi.org/10.1016/j.ajhg.2019.05.001>
- Kichaev, G., Bhatia, G., Loh, P. R., Gazal, S., Burch, K., Freund, M. K., Schoech, A., Pasaniuc, B., & Price, A. L. (2019). Leveraging polygenic functional enrichment to improve GWAS power. *American Journal of Human Genetics*, *104*(1), 65–75. <https://doi.org/10.1016/j.ajhg.2018.11.008>
- Köks, S., Kingo, K., Vabrit, K., Rätsep, R., Karelson, M., Silm, H., & Vasar, E. (2005). Possible relations between the polymorphisms of the cytokines IL-19, IL-20 and IL-24 and plaque-type psoriasis. *Genes & Immunity*, *6*(5), 407–415. <https://doi.org/10.1038/sj.gene.6364216>
- Kremen, W. S., Beck, A., Elman, J. A., Gustavson, D. E., Reynolds, C. A., Tu, X. M., Sanderson-Cimino, M. E., Panizzon, M. S., Vuoksima, E., Toomey, R., Fennema-Notestine, C., Hagler, D. J., Jr., Fang, B., Dale, A. M., Lyons, M. J., & Franz, C. E. (2019). Influence of young adult cognitive ability and additional education on later-life cognition. *Proceedings of the National Academy of Sciences, USA*, *116*(6), 2021–2026. <https://doi.org/10.1073/pnas.1811537116>
- Kurilshikov, A., Medina-Gomez, C., Bacigalupe, R., Radjabzadeh, D., Wang, J., Demirkan, A., Le Roy, C. I., Raygoza Garay, J. A., Finnicum, C. T., Liu, X., Zhernakova, D. V., Bonder, M. J., Hansen, T. H., Frost, F., Ruhlemann, M. C., Turpin, W., Moon, J. Y., Kim, H. N., Lull, K., . . . Zhernakova, A. (2021). Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nature Genetics*, *53*(2), 156–165. <https://doi.org/10.1038/s41588-020-00763-1>
- Lam, M., Hill, W. D., Trampush, J. W., Yu, J., Knowles, E., Davies, G., Stahl, E., Huckins, L., Liewald, D. C., Djurovic, S., Melle, I., Sundet, K., Christoforou, A., Reinvang, I., DeRosse, P., Lundervold, A. J., Steen, V. M., Espeseth, T., Raikkonen, K., . . . Lencz, T. (2019). Pleiotropic meta-analysis of cognition, education, and schizophrenia differentiates roles of early neurodevelopmental and adult synaptic pathways. *American Journal of Human Genetics*, *105*(2), 334–350. <https://doi.org/10.1016/j.ajhg.2019.06.012>
- Lam, M., Trampush, J. W., Yu, J., Knowles, E., Davies, G., Liewald, D. C., Starr, J. M., Djurovic, S., Melle, I., Sundet, K., Christoforou, A., Reinvang, I., DeRosse, P., Lundervold, A. J., Steen, V. M., Espeseth, T., Raikkonen, K., Widen, E., Palotie, A., . . . Lencz, T. (2017). Large-scale cognitive GWAS meta-analysis reveals tissue-specific neural expression and potential nootropic drug targets. *Cell Reports*, *21*(9), 2597–2613. <https://doi.org/10.1016/j.celrep.2017.11.028>
- Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linner, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., . . . Cesarini, D. (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. <https://doi.org/10.1038/s41588-018-0147-3>
- Lee, M. D., Criss, A. H., Devezer, B., Donkin, C., Etz, A., Leite, F. P., Matzke, D., Rouder, J. N., Trueblood, J. S., White, C. N., & Vandekerckhove, J. (2019). Robust modeling in cognitive science. *Computational Brain & Behavior*, *2*(3), 141–153. <https://doi.org/10.1007/s42113-019-00029-y>
- Leist, A. K., Bar-Haim, E., & Chauvel, L. (2021). Inequality of educational opportunity at time of schooling predicts cognitive functioning in later adulthood. *SSM - Population Health*, *15*, Article 100837. <https://doi.org/10.1016/j.ssmph.2021.100837>
- Lello, L., Raben, T. G., & Hsu, S. D. H. (2020). Sibling validation of polygenic risk scores and complex trait prediction. *Scientific Reports*, *10*(1), Article 13190. <https://doi.org/10.1038/s41598-020-69927-7>
- Leppert, B., Millard, L. A. C., Riglin, L., Davey Smith, G., Thapar, A., Tilling, K., Walton, E., & Stergiakouli, E. (2020). A cross-disorder PRS-pheWAS of 5 major psychiatric disorders in UK Biobank. *PLOS Genetics*, *16*(5), Article e1008185. <https://doi.org/10.1371/journal.pgen.1008185>

- Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E. C., Cunliffe, B., Wellcome Trust Case Control, C., International Multiple Sclerosis Genetics, C., Lawson, D. J., Falush, D., Freeman, C., Pirinen, M., Myers, S., Robinson, M., Donnelly, P., & Bodmer, W. (2015). The fine-scale genetic structure of the British population. *Nature*, *519*(7543), 309–314. <https://doi.org/10.1038/nature14230>
- Lever, J., Krzywinski, M., & Altman, N. (2016). Model selection and overfitting. *Nature Methods*, *13*(9), 703–704. <https://doi.org/10.1038/nmeth.3968>
- Lighthart, S., Vaez, A., Vosa, U., Stathopoulou, M. G., de Vries, P. S., Prins, B. P., Van der Most, P. J., Tanaka, T., Naderi, E., Rose, L. M., Wu, Y., Karlsson, R., Barbalic, M., Lin, H., Pool, R., Zhu, G., Mace, A., Sidore, C., Trompet, S., . . . Alizadeh, B. Z. (2018). Genome analyses of >200,000 individuals identify 58 loci for chronic inflammation and highlight pathways that link inflammation and complex disorders. *American Journal of Human Genetics*, *103*(5), 691–706. <https://doi.org/10.1016/j.ajhg.2018.09.009>
- Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639–647. <https://doi.org/10.1111/1755-0998.12995>
- Lu, J., & Claud, E. C. (2019). Connection between gut microbiome and brain development in preterm infants. *Developmental Psychobiology*, *61*(5), 739–751. <https://doi.org/10.1002/dev.21806>
- Maier, R. M., Visscher, P. M., Robinson, M. R., & Wray, N. R. (2018). Embracing polygenicity: A review of methods and tools for psychiatric genetics research. *Psychological Medicine*, *48*(7), 1055–1067. <https://doi.org/10.1017/S0033291717002318>
- Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, *11*(7), 499–511. <https://doi.org/10.1038/nrg2796>
- Martin, A. R., Daly, M. J., Robinson, E. B., Hyman, S. E., & Neale, B. M. (2019). Predicting polygenic risk of psychiatric disorders. *Biological Psychiatry*, *86*(2), 97–109. <https://doi.org/10.1016/j.biopsych.2018.12.015>
- Martin, C. R., Osadchiy, V., Kalani, A., & Mayer, E. A. (2018). The brain-gut-microbiome axis. *Cellular and Molecular Gastroenterology and Hepatology*, *6*(2), 133–148. <https://doi.org/10.1016/j.jcmgh.2018.04.003>
- Mazzi, S., Lordier, L., Debili, N., Raslova, H., & Vainchenker, W. (2018). Megakaryocyte and polyploidization. *Experimental Hematology*, *57*, 1–13. <https://doi.org/10.1016/j.exphem.2017.10.001>
- McEvoy, B. P., & Visscher, P. M. (2009). Genetics of human height. *Economics & Human Biology*, *7*(3), 294–306. <https://doi.org/10.1016/j.ehb.2009.09.005>
- Melton, P. (2012). Mennonite migrations: Genetic and demographic consequences. In M. H. Crawford (Ed.), *Causes and consequences of human migration: An evolutionary perspective* (pp. 299–316). Cambridge University Press.
- Mersha, T. B., Ding, L., He, H., Alexander, E. S., Zhang, X., Kurowski, B. G., Pilipenko, V., Kottyan, L., Martin, L. J., & Fardo, D. W. (2015). Impact of population stratification on family-based association in an admixed population. *International Journal of Genomics*, *2015*, Article 501617. <https://doi.org/10.1155/2015/501617>
- Mertens, G., & Kryptos, A. M. (2019). Preregistration of analyses of preexisting data. *Psychologica Belgica*, *59*(1), 338–352. <https://doi.org/10.5334/pb.493>
- Miton, C. M., Buda, K., & Tokuriki, N. (2021). Epistasis and intramolecular networks in protein evolution. *Current Opinion in Structural Biology*, *69*, 160–168. <https://doi.org/10.1016/j.sbi.2021.04.007>
- Morris, T. T., Davies, N. M., & Davey Smith, G. (2020). Can education be personalised using pupils’ genetic data? *eLife*, *2020*(9), Article e49962. <https://doi.org/10.7554/eLife.49962>
- Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K., & Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *eLife*, *2020*(9), Article e48376. <https://doi.org/10.7554/eLife.48376>
- Muller-Spahn, F. (2008). Individualized preventive psychiatry: Syndrome and vulnerability diagnostics. *European Archives of Psychiatry and Clinical Neuroscience*, *258*, 92–97. <https://doi.org/10.1007/s00406-008-5020-4>
- Negi, S. K., & Guda, C. (2017). Global gene expression profiling of healthy human brain and its application in studying neurological disorders. *Scientific Reports*, *7*(1), Article 897. <https://doi.org/10.1038/s41598-017-00952-9>
- Nelson, D., Kelleher, J., Ragsdale, A. P., Moreau, C., McVean, G., & Gravel, S. (2020). Accounting for long-range correlations in genome-wide simulations of large cohorts. *PLoS Genetics*, *16*(5), Article e1008619. <https://doi.org/10.1371/journal.pgen.1008619>
- Ohi, K., Sumiyoshi, C., Fujino, H., Yasuda, Y., Yamamori, H., Fujimoto, M., Shiino, T., Sumiyoshi, T., & Hashimoto, R. (2018). genetic overlap between general cognitive function and schizophrenia: A review of cognitive GWASs. *International Journal of Molecular Science*, *19*(12), Article 3822. <https://doi.org/10.3390/ijms19123822>
- Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., Turley, P., Chen, G. B., Emilsson, V., Meddens, S. F., Oskarsson, S., Pickrell, J. K., Thom, K., Timshel, P., de Vlaming, R., Abdellaoui, A., Ahluwalia, T. S., Bacelis, J., Baumbach, C., . . . Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539–542. <https://doi.org/10.1038/nature17671>
- Park, L. (2019). Population-specific long-range linkage disequilibrium in the human genome and its influence on identifying common disease variants. *Scientific Reports*, *9*(1), Article 11380. <https://doi.org/10.1038/s41598-019-47832-y>
- Petrovic, J., Pesic, V., & Lauschke, V. M. (2020). Frequencies of clinically important CYP2C19 and CYP2D6 alleles are graded across Europe. *European Journal of Human Genetics*, *28*(1), 88–94. <https://doi.org/10.1038/s41431-019-0480-8>
- Piffer, D. (2019). Evidence for recent polygenic selection on educational attainment and intelligence inferred from GWAS hits: A replication of previous findings using

- recent data. *Psych*, 1(1), 55–75. <https://doi.org/10.3390/psych1010005>
- Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics*, 19(3), 148–159. <https://doi.org/10.1038/nrg.2017.104>
- Pouget, J. G., Schizophrenia Working Group of the Psychiatric Genomics, C., Han, B., Wu, Y., Mignot, E., Ollila, H. M., Barker, J., Spain, S., Dand, N., Trembath, R., Martin, J., Mayes, M. D., Bossini-Castillo, L., Lopez-Isac, E., Jin, Y., Santorico, S. A., Spritz, R. A., Hakonarson, H., Polychronakos, C., . . . Knight, J. (2019). Cross-disorder analysis of schizophrenia and 19 immune-mediated diseases identifies shared genetic risk. *Human Molecular Genetics*, 28(20), 3498–3513. <https://doi.org/10.1093/hmg/ddz145>
- Prive, F., Vilhjalmsdottir, B. J., Aschard, H., & Blum, M. G. B. (2019). Making the most of clumping and thresholding for polygenic scores. *American Journal of Human Genetics*, 105(6), 1213–1221. <https://doi.org/10.1016/j.ajhg.2019.11.001>
- Racimo, F., Berg, J. J., & Pickrell, J. K. (2018). Detecting polygenic adaptation in admixture graphs. *Genetics*, 208(4), 1565–1584. <https://doi.org/10.1534/genetics.117.300489>
- Reddy, G., & Desai, M. M. (2021). Global epistasis emerges from a generic model of a complex trait. *eLife*, 2021(10), Article e64740. <https://doi.org/10.7554/eLife.64740>
- Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benjamin, B., Chabris, C. F., Emilsson, V., Johnson, A. D., Lee, J. J., de Leeuw, C., Marioni, R. E., Medland, S. E., Miller, M. B., Rostapshova, O., van der Lee, S. J., Vinkhuyzen, A. A., Amin, N., Conley, D., . . . Koellinger, P. D. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences, USA*, 111(38), 13790–13794. <https://doi.org/10.1073/pnas.1404623111>
- Robinson, M. R., Hemani, G., Medina-Gomez, C., Mezzavilla, M., Esko, T., Shakhbazov, K., Powell, J. E., Vinkhuyzen, A., Berndt, S. I., Gustafsson, S., Justice, A. E., Kahali, B., Locke, A. E., Pers, T. H., Vedantam, S., Wood, A. R., van Rheenen, W., Andreassen, O. A., Gasparini, P., . . . Visscher, P. M. (2015). Population genetic differentiation of height and body mass index across Europe. *Nature Genetics*, 47(11), 1357–1362. <https://doi.org/10.1038/ng.3401>
- Roelfs, D., Alnaes, D., Frei, O., van der Meer, D., Smeland, O. B., Andreassen, O. A., Westlye, L. T., & Kaufmann, T. (2021). Phenotypically independent profiles relevant to mental health are genetically correlated. *Translational Psychiatry*, 11(1), Article 202. <https://doi.org/10.1038/s41398-021-01313-x>
- Sabol, S. Z., Hu, S., & Hamer, D. (1998). A functional polymorphism in the monoamine oxidase A gene promoter. *Human Genetics*, 103(3), 273–279.
- Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., . . . Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, 50(7), 912–919. <https://doi.org/10.1038/s41588-018-0152-6>
- Savulescu, J. (2014). The nature of the moral obligation to choose the best children. In A. Akabayashi (Ed.), *The future of bioethics* (pp. 170–182). Oxford University Press.
- 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. <https://doi.org/10.1016/j.ajhg.2019.06.006>
- Shanahan, M. J., Vaisey, S., Erickson, L. D., & Smolen, A. (2008). Environmental contingencies and genetic propensities: Social capital, educational continuation, and dopamine receptor gene DRD2. *American Journal of Sociology*, 114, S260–S286. <https://doi.org/10.1086/592204>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Sohail, M., Maier, R. M., Ganna, A., Bloemendal, A., Martin, A. R., Turchin, M. C., Chiang, C. W., Hirschhorn, J., Daly, M. J., Patterson, N., Neale, B., Mathieson, I., Reich, D., & Sunyaev, S. R. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife*, 2019(8), Article 39702. <https://doi.org/10.7554/eLife.39702>
- Thornton, T., Tang, H., Hoffmann, T. J., Ochs-Balcom, H. M., Caan, B. J., & Risch, N. (2012). Estimating kinship in admixed populations. *American Journal of Human Genetics*, 91(1), 122–138. <https://doi.org/10.1016/j.ajhg.2012.05.024>
- Toft, G., Liu, C., Menon, J., Schendel, D., Loss, G., & Ehrenstein, V. (2021). Assessment of educational attainment and employment among individuals with autism spectrum disorder in Denmark. *JAMA Pediatrics*, 175(6), 601–608. <https://doi.org/10.1001/jamapediatrics.2021.0124>
- Trejo, S., & Domingue, B. W. (2018). Genetic nature or genetic nurture? Introducing social genetic parameters to quantify bias in polygenic score analyses. *Biodemography and Social Biology*, 64(3–4), 187–215. <https://doi.org/10.1080/19485565.2019.1681257>
- Turchin, M. C., Chiang, C. W., Palmer, C. D., Sankaraman, S., Reich, D., Genetic Investigation of, A. T. C., & Hirschhorn, J. N. (2012). Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics*, 44(9), 1015–1019. <https://doi.org/10.1038/ng.2368>
- Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, 9(5), 160–164. <https://doi.org/10.1111/1467-8721.00084>
- Turkheimer, E. (2016). Weak genetic explanation 20 years later: Reply to Plomin et al. (2016). *Perspectives on Psychological Science*, 11(1), 24–28. <https://doi.org/10.1177/1745691615617442>
- Valge, M., Meitern, R., & Horak, P. (2019). Morphometric traits predict educational attainment independently of

- socioeconomic background. *BMC Public Health*, 19(1), Article 1696. <https://doi.org/10.1186/s12889-019-8072-7>
- Van den Bogaert, S. (2018). *Segregation of Roma Children in Education*. Nijhoff.
- van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. (2019). Genetic correlations of polygenic disease traits: From theory to practice. *Nature Reviews Genetics*, 20(10), 567–581. <https://doi.org/10.1038/s41576-019-0137-z>
- Vilhjalmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindstrom, S., Ripke, S., Genovese, G., Loh, P.-R., Bhatia, G., Do, R., Hayeck, T., & Won, H.-H. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) study, Kathiresan, S., Pato, M., Pato, C., Tamimi, R., Stahl, E., . . . Price, A. L. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *American Journal of Human Genetics*, 97(4), 576–592. <https://doi.org/10.1016/j.ajhg.2015.09.001>
- 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. <https://doi.org/10.1016/j.ajhg.2017.06.005>
- Wagner, G. P., & Zhang, J. (2011). The pleiotropic structure of the genotype-phenotype map: The evolvability of complex organisms. *Nature Reviews Genetics*, 12(3), 204–213. <https://doi.org/10.1038/nrg2949>
- Wang, X., Xiao, R., Zhu, X., & Li, M. (2016). Gene mapping in admixed families: A cautionary note on the interpretation of the transmission disequilibrium test and a possible solution. *Human Heredity*, 81(2), 106–116. <https://doi.org/10.1159/000446956>
- Ye, J., Liu, L., Xu, X., Wen, Y., Li, P., Cheng, B., Cheng, S., Zhang, L., Ma, M., Qi, X., Liang, C., Kafle, O. P., Wu, C., Wang, S., Wang, X., Ning, Y., Chu, X., Niu, L., & Zhang, F. (2020). A genome-wide multiphenotypic association analysis identified candidate genes and gene ontology shared by four common risky behaviors. *Aging*, 12(4), 3287–3297. <https://doi.org/10.18632/aging.102812>
- Yong, E. (2019, May 17). A waste of 1,000 research papers. *The Atlantic*. <https://www.theatlantic.com/science/archive/2019/05/waste-1000-studies/589684/>
- Young, A. I. (2019). Solving the missing heritability problem. *PLOS Genetics*, 15(6), Article e1008222. <https://doi.org/10.1371/journal.pgen.1008222>
- Zaidi, A. A., & Mathieson, I. (2020). *Demographic history impacts stratification in polygenic scores*. bioRxiv. <https://doi.org/10.1101/2020.07.20.212530>
- Zarrei, M., MacDonald, J. R., Merico, D., & Scherer, S. W. (2015). A copy number variation map of the human genome. *Nature Reviews Genetics*, 16(3), 172–183. <https://doi.org/10.1038/nrg3871>
- Zeggini, E., & Ioannidis, J. P. (2009). Meta-analysis in genome-wide association studies. *Pharmacogenomics*, 10(2), 191–201. <https://doi.org/10.2217/14622416.10.2.191>
- Zeng, L., Ntalla, I., Kessler, T., Kastrati, A., & Erdmann, J. Group, U. K. B. C. C. C. W., Danesh, J., Watkins, H., Samani, N. J., Deloukas, P., & Schunkert, H. (2019). Genetically modulated educational attainment and coronary disease risk. *European Heart Journal*, 40(29), 2413–2420. <https://doi.org/10.1093/eurheartj/ehz328>
- Zhang, C., de Smith, A. J., Smirnov, I. V., Wiencke, J. K., Wiemels, J. L., Witte, J. S., & Walsh, K. M. (2017). Non-additive and epistatic effects of HLA polymorphisms contributing to risk of adult glioma. *Journal of Neuro-Oncology*, 135(2), 237–244. <https://doi.org/10.1007/s11060-017-2569-7>
- Zhang, S., Lin, Y. H., Tarlow, B., & Zhu, H. (2019). The origins and functions of hepatic polyploidy. *Cell Cycle*, 18(12), 1302–1315. <https://doi.org/10.1080/15384101.2019.1618123>
- Zhang, Y., & Pan, W. (2015). Principal component regression and linear mixed model in association analysis of structured samples: Competitors or complements? *Genetics Epidemiology*, 39(3), 149–155. <https://doi.org/10.1002/gepi.21879>
- Zhao, B., Wu, Q., Ye, A. Y., Guo, J., Zheng, X., Yang, X., Yan, L., Liu, Q. R., Hyde, T. M., Wei, L., & Huang, A. Y. (2019). Somatic LINE-1 retrotransposition in cortical neurons and non-brain tissues of Rett patients and healthy individuals. *PLOS Genetics*, 15(4), Article e1008043. <https://doi.org/10.1371/journal.pgen.1008043>
- Zhao, Y., Ning, Y., Zhang, F., Ding, M., Wen, Y., Shi, L., Wang, K., Lu, M., Sun, J., Wu, M., Cheng, B., Ma, M., Zhang, L., Cheng, S., Shen, H., Tian, Q., Guo, X., & Deng, H. W. (2019). PCA-based GRS analysis enhances the effectiveness for genetic correlation detection. *Briefings in Bioinformatics*, 20(6), 2291–2298. <https://doi.org/10.1093/bib/bby075>
- Zhu, S., Jiang, Y., Xu, K., Cui, M., Ye, W., Zhao, G., Jin, L., & Chen, X. (2020). The progress of gut microbiome research related to brain disorders. *Journal of Neuroinflammation*, 17(1), Article 25. <https://doi.org/10.1186/s12974-020-1705-z>
- Zoledziewska, M., Sidore, C., Chiang, C. W. K., Sanna, S., Mulas, A., Steri, M., Busonero, F., Marcus, J. H., Marongiu, M., Maschio, A., Ortega Del Vecchio, D., Floris, M., Meloni, A., Delitala, A., Concas, M. P., Murgia, F., Biino, G., Vaccargiu, S., Nagaraja, R., . . . Cucca, F. (2015). Height-reducing variants and selection for short stature in Sardinia. *Nature Genetics*, 47(11), 1352–1356. <https://doi.org/10.1038/ng.3403>