

Implications of the genomic revolution for education research and policy

Tim T. Morris^{1,2}  | Stephanie von Hinke^{1,3,4} | Lindsey Pike^{1,2} |
 Neil R. Ingram⁵ | George Davey Smith^{1,2} | Marcus R. Munafò^{1,6} |
 Neil M. Davies^{1,2,7} 

¹Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK

²Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield Grove, Barley House, Bristol, UK

³School of Economics, University of Bristol, UK

⁴Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands

⁵School of Education, University of Bristol, Bristol, UK

⁶School of Psychological Science, University of Bristol, Bristol, UK

⁷K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence

Neil M. Davies, Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK.
 Email: neil.davies@bristol.ac.uk

Funding information

The Medical Research Council (MRC) and the University of Bristol support the MRC Integrative Epidemiology Unit, (MC_UU_00011/1, MC_UU_00011/7). NMD is supported via a Norwegian Research Council Grant number 295989. SvH, MM and GDS are supported via a NORFACE DIAL, grant (462-16-100); and SvH is supported via a European Research Council Starting Grant (851725)

Abstract

Research at the intersection of social science and genomics, ‘sociogenomics’, is transforming our understanding of the interplay between genomics, individual outcomes and society. It has interesting and maybe unexpected implications for education research and policy. Here we review the growing sociogenomics literature and discuss its implications for educational researchers and policymakers. We cover key concepts and methods in genomic research into educational outcomes, how genomic data can be used to investigate social or environmental effects, the methodological strengths and limitations of genomic data relative to other observational social data, the role of intergenerational transmission and potential policy implications. The increasing availability of genomic data in studies can produce a wealth of new evidence for education research. This may provide opportunities for disentangling the environmental and genomic factors that influence educational outcomes and identify potential mechanisms for intervention.

KEYWORDS

educational attainment, genetics, policy, social science, sociogenomics

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

Key insights

What is the main issue that the paper addresses?

This paper reviews the implications of recent discoveries in genomics for education research.

What are the main insights that the paper provides?

We describe how genetic data can be used to provide new insights into many research questions of interest, including understanding the causal mechanisms that link parents and offspring, and the influence of the environment on students' outcomes.

INTRODUCTION

A key aim of much social science research is identifying modifiable factors that can be used for interventions to improve people's life outcomes. Our genomes are fixed at conception and cannot be changed, so why should education researchers – and social scientists more broadly – be interested in an aspect of human nature that is immutable? Here we discuss how researchers can exploit genomic data to enrich research in the social sciences.

Human behaviour is impacted by and associated with both genetic variation (differences in DNA between people) and environmental circumstances. However, genetics and the environment are not simply two independent forces; they are inextricably linked and jointly influence human behaviour (Plomin et al., 2014). Most human traits are heritable; that is, statistical variation in those traits is driven at least partially by genetic variation between individuals (Turkheimer, 2000, for a glossary see Box S1). If our aim is a holistic understanding of individual differences in educational or other social outcomes, it is important to incorporate genomics. At a practical level, inference from quantitative statistical approaches that fail to account for genomic information may lead to spurious conclusions about relationships between social factors that are in fact influenced by genomic differences between individuals. For example, parenting behaviour is related to many offspring outcomes, but parents and their offspring share half of their genomes and also often have similar social environments. Are these observed correlations driven by parental behaviours, by genetic correlation between parents and their children, or by a combination of both? Genomic evidence may not only capture the causal effect of genomics on education, but also reflect the effects of other social or familial processes.

The use of genomic data may also allow researchers to make stronger causal inferences than from other observational data. Suppose an investigator is interested in whether myopia (shortsightedness) affects educational attainment (discussed in Box 1), an example we will come back to later. Here, a correlation between myopia and attainment is not sufficient for concluding that myopia causes educational attainment, given the potential impact of confounding factors and reverse causation. Confounding could occur if a third factor affected both risk of myopia and educational attainment, while reverse causation could occur if processes leading to educational attainment (i.e. some aspect of the school environment) affect short-sightedness. Under specific statistical assumptions, methods that exploit genetic variation can avoid these issues – which plague many observational studies – to reliably estimate causal effects. As with all empirical studies, however, a key consideration for interpreting genetic studies is whether the assumptions on which they depend are plausible.

Box 1 The definition of educational attainment

Many measures of the 'years of education' used in genome-wide association studies (GWASs) like that of Lee et al. were derived from qualifications (e.g. in UK Biobank), making them a noisy and imperfect measure. This has implications, particularly for Mendelian randomisation (MR) (Howe et al., 2020). Furthermore, these measures of educational attainment differ across countries and were harmonised to a common scale (the International Standard Classification of Education [ISCED]). However, clearly there is going to be heterogeneity in what a given number of years of education mean in one country at one specific point in time vs. another country or even within the same country at different points in time. For example, consider two siblings, one has 17 years of education and the other 18 (starting at age 4 and leaving at age 21/22). One has a humanities degree from a traditional university, the other has a vocational degree from a newer university. Under the International Standard Classification of Education (ISCED) both siblings would be coded as having the same amount of education, but clearly the educational experiences and the downstream consequences of that education may differ. Similar arguments can be made about individuals leaving state vs. private education at the same age, or even choosing different subjects while at secondary school.

However, despite the heterogeneity in the definition of educational attainment, the signal from educational GWASs has been very consistently replicated within European ancestry samples. Therefore, the definitions of educational attainment are clearly sufficiently consistent across these samples to produce a reliable signal that explains around 10% of the variation. Fascinatingly, there is evidence that these signals do not replicate well in non-European ancestry samples. For example, the association of educational attainment and the educational attainment polygenic score attenuates 85% in non-European sample samples (Lee et al., 2018). This is probably because the score was developed largely in samples of European ancestry. However to date we do not have particularly good understanding or evidence about this because there are relatively few non-European genetic samples compared with European samples (Mills & Rahal, 2019).

Researchers from education, economics, sociology and many other areas of social science have investigated the mechanisms which underlie the intergenerational transmission of education from one generation to the next. With the increasing availability of large datasets which have sampled mothers, fathers and offspring from the same families, researchers can use genetic variation to understand how educational advantage is transmitted within families.

While the genome is fixed at conception, the way that it affects educational outcomes may depend heavily on environmental circumstances. The opposite is also true: environmental impacts on individuals' outcomes may depend on one's genetic variation. Research in this area explores the interplay between genetic and environmental variation in shaping individual outcomes. Understanding whether the impact of different environments on educational outcomes differs across individuals with distinct genetic variation could help us to understand the existence of, causes of and development of educational inequalities within the population.

Below we provide an introduction to genetics for education researchers, discuss the relevance of recent findings in genomics to education researchers and how they can be used to understand the associations found in observational data, and finally explore the implications of these approaches for educational and social policies.

GENOMICS 101

The human genome consists of DNA which is stored in our cells as chromosomes. Most humans have 23 pairs of chromosomes, one of each pair being inherited from each parent. DNA consists of two coiled chains of nucleotides that are bonded together in pairs and stored in a double helix shape. There are four types of nucleotides, defined by a difference in their base molecule: (A) adenine; (C) cytosine; (G) guanine; (T) thymine. There are around 3.2 billion of these nucleotide pairs in the human genome, but only a minority of these differ between people (The, [1000 Genomes Project Consortium, 2015](#)). The points of nucleotide variation are referred to as single nucleotide polymorphisms (SNPs, pronounced 'snips'), and different copies of these SNPs are referred to as alleles. Because our chromosomes come in pairs, for any specific SNP, we normally would have zero, one or two copies of an allele. Genomes can be measured from biological material such as blood or saliva, and their base pairs can be determined via genotyping.

GENOME-WIDE ASSOCIATION STUDIES

Genome-wide association studies (GWASs) estimate associations between each SNP and a particular trait, or phenotype, such as educational attainment. Given the size of the human genome, a single GWAS can estimate millions of associations. Because so many associations are tested, GWASs use stringent 'genome-wide' levels of statistical significance ($p < 5 \times 10^{-8}$) to test whether there is strong evidence that a SNP associates with a phenotype ([Box 2](#)). To further reduce the likelihood of false positives, many GWASs replicate their initial results in external, independent datasets, and then use a Bonferroni corrected p -value for the number of SNPs exceeding genome-wide significance levels in a replication sample ([Box 2](#)). An important consideration of GWASs is that they may be based on highly selected samples such as the UK Biobank, in which only 5% of people contacted chose to participate. If genetic variation is associated with participation in these studies, then GWAS results may not be representative of the underlying population (Taylor et al., [2018](#)). Replication in more population-representative external cohorts helps to reduce the risk of non-representativeness, but it may still exist if replication cohorts are not fully population representative. GWASs can therefore be used to detect associations between specific traits and individual SNPs, but what relevance does this have for education and social scientific research?

The latest GWAS of educational attainment, defined as years of completed education (Lee et al., [2018](#)), included 1.1 million individuals. It identified 1271 education attainment-associated SNPs at genome-wide significance. These results have been replicated across multiple samples and demonstrate that common genetic variation associates with educational attainment. These associations may reflect biological effects of genetic variation on biological, social and psychological phenotypes that ultimately affect educational attainment. However, a substantial fraction of these associations are likely to reflect demographic and familial processes such as assortative mating (i.e. that spouses tend to be similar; see [Box S1](#) for a glossary) and effects of parents on their children (i.e. parenting; Brumpton et al., [2020](#); Davies, Howe, et al., [2019](#)). Thus, genetic data can provide compelling evidence not only about purely biological effects, but potentially about the impact of parents, the familial environment and society.

Box 2 Statistical significance in GWASs

In the context of GWASs, the genome-wide significance level has proven to be a useful indicator of association (Davey Smith, 2011). It is a relatively high bar, calculated on the basis of the number of independent variants in the genome, and the highest quality GWASs use a strict policy of replication in an independent sample. Thus, the threshold is 5×10^{-8} for discovery in the first sample, and then a Bonferroni corrected p -value is used for the number of SNPs exceeding genome-wide significance levels in a replication sample.

All that genome-wide significant signals from GWASs tell us is that there is evidence that a genetic variant is associated with a phenotype in the discovery and replication samples. This does not demonstrate that a genetic variant causes the phenotype in an individual, or is necessarily associated in the population of interest. There are a number of processes that could induce associations between SNPs and phenotypes, only one of which is a direct causal effect of the variant on the phenotype. Samples which are selected on the phenotype of interest and other traits (e.g. educational attainment and body mass index, BMI) may induce spurious correlations between SNPs causally associated with BMI and educational attainment and vice versa (i.e. selection/collider bias). Equally, population structure or familial effects can induce these associations. Therefore discovery that a genetic variant associates with a phenotype like educational attainment is really the first step in a scientific process; it is nevertheless an absolutely necessary (but not sufficient) step for elucidating the causal links between the genetic variant and a phenotype like educational attainment.

An important finding from GWASs of education and other social/behavioural phenotypes is that they are highly 'polygenic'. That is, these phenotypes are characterised not by a small number of SNPs with large effect sizes, but by a very large number of SNPs which each have tiny, but statistically detectable, effect sizes. Critically, *there is no such thing as a 'gene for' education*; of the 1271 education attainment-associated SNPs, the median SNP association was only 1.7 weeks of additional education throughout the lifecourse (1.1 and 2.6 weeks for the 5th and 95th centile SNPs; see [Box 1](#) for a discussion on measuring education). This pervasive collection of tiny effects – which has been discussed for over a century – has recently been labelled the omnigenic model (Boyle et al., 2017).

Such small effect sizes can only be detected in very large samples. More powerful predictive models can be constructed by aggregating individual SNPs into a polygenic score, which is a weighted average of the associations of the genetic variants and the trait. While individual SNP effects can be small, when combined into a polygenic score they can explain substantial amounts of trait variation. For example, the best performing polygenic score of the most recent education GWAS explains 11–13% of variation in educational attainment (Lee et al., 2018).

SNP HERITABILITY AND GENETIC CORRELATION

Researchers can use GWAS results to estimate the proportion of variation in a trait such as educational attainment that can be explained by measured genetic variation.¹ This estimate is termed the 'SNP heritability'. Studies may report the proportion of variation in

a trait explained by SNPs that were genome-wide significant, or all measured SNPs that were associated with the phenotype of interest, irrespective of statistical significance. The former approach tells us how much of the variation in a trait can be explained by variants with established effects, while the latter indicates how much variation can be explained by all measured genetic variation, even where it is noisily measured. Because there are fewer genome-wide significant SNPs, they typically explain far less of the variation in a trait than estimates using all SNPs; 2–3% of the variation in educational attainment vs. 11–13% in the most recent education GWAS (Lee et al., 2018). Heritability estimates from genomic studies are typically smaller than those from twin studies because they rely on measured rather than total genetic variation.

Genetic correlations, the correlations between genetic associations with one trait and a second trait, can also be estimated from genomic data. There are two commonly used methods for estimating genetic correlation, although they can also both estimate heritability as well. The first of these, called LD-score regression (Bulik-Sullivan et al., 2015), uses large-scale GWAS data. The second, called Genome-wide Complex Trait Analysis using Genome-based Restricted Maximum Likelihood (GCTA-GREML; Yang et al., 2011), uses genetic relatedness between individuals which is derived from individual-level genetic data. The intuition for GCTA-GREML is that if genetic variation explains (some of) phenotypic variation, individuals who are more genetically similar are expected to be more phenotypically similar (analogous to twin studies).

Studies have shown that higher educational attainment-associated genetic variation is also correlated with lower levels of smoking, lower body mass index (BMI), lower risk of coronary heart disease, lower triglycerides, higher high-density lipoprotein, being taller and being at higher risk of bipolar disorder (Figure 1). These results show that the genetic correlations between genetic variation related to one trait, such as BMI, and another trait, such as educational attainment, tend to be similar to their associations in other forms of observational studies. They also suggest that many genetic variants are likely to affect multiple phenotypes (referred to as pleiotropy).

The correlogram in Figure 1 illustrates the genetic correlations between a range of traits. This includes the genetic correlations between variants associated with attending college, from one of the first GWASs of educational attainment by Rietveld et al., and a range of other traits (Rietveld et al., 2013). Each of the squares in the figure represents a genetic correlation between two traits, e.g. attending college and being an ever/never smoker. This figure illustrates that genetic variants that associate with attending college also associate with a range of health traits. This therefore provides evidence of widespread pleiotropy (i.e. genetic variants that associate with multiple traits; Verbanck et al., 2017).

GENES AND ENVIRONMENTS

Researchers can use genetic data to explore the presence of gene–environment correlation (*rGE*). *rGE* refers to the fact that certain environments are more prevalent for carriers of certain genetic variants and can arise from three mechanisms (Plomin et al., 1977). First, *evocative* gene–environment correlation occurs when an individual's genetic variation evokes specific responses from others, creating a correlation between the individual's genetic variants and their environment. For example, having a high polygenic score for educational attainment (which may manifest in certain learning attitudes and behaviours) may invoke different responses from parents and teachers, whereby they create more learning opportunities.

Another form is *active* gene–environment correlation, which refers to the way that an individual can select their environment based upon their genetics. For example, are children

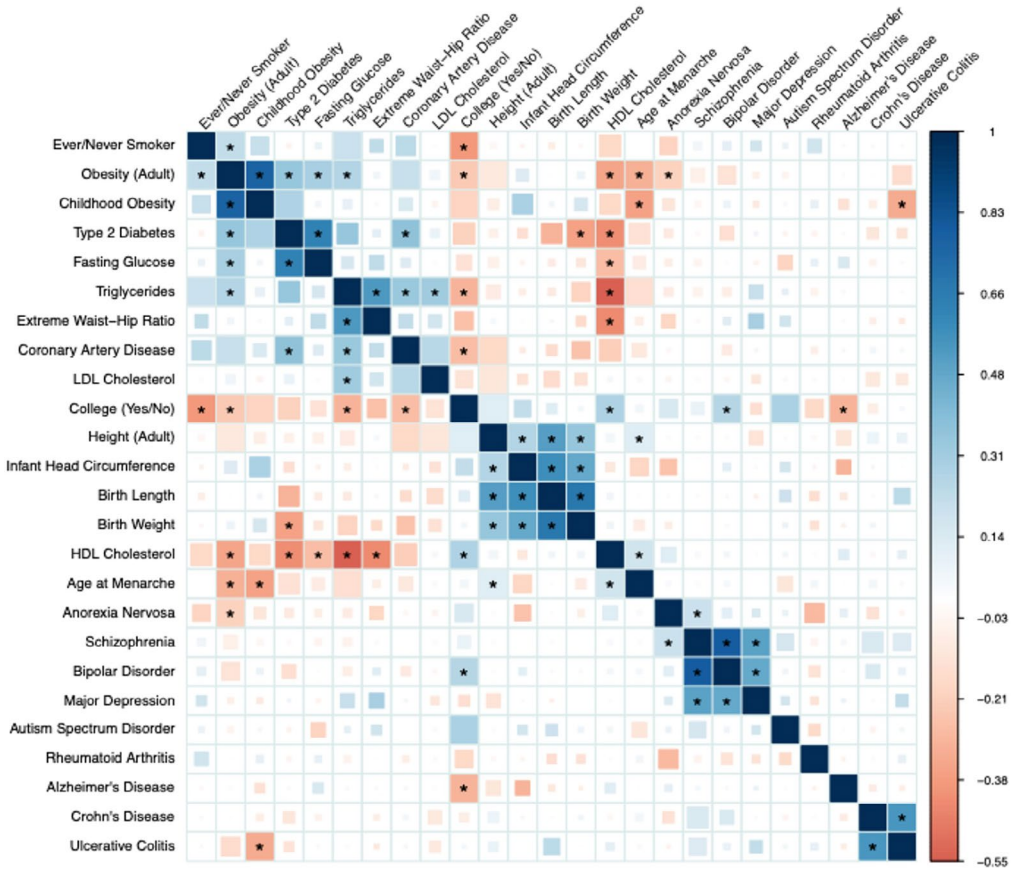


FIGURE 1 Genetic correlations from 25 genome-wide association studies. Squares represent the genetic correlation between two traits; blue represents positive genetic correlations while red represents negative genetic correlations. Coloured square size corresponds to p -values for tests of difference from zero (no genetic correlation), with larger squares denoting smaller p -values. Full-size squares denote values that pass a false discovery rate of 0.01, while squares with an asterisk denote values that pass Bonferroni correction for the 300 tests conducted. Educational attainment was measured in this study using the variable 'College (yes/no)'. This indicates that genetic variation associated with graduating from college is also associated with being less likely to be a smoker, to be obese, to have lower triglycerides and LDL cholesterol, to have higher high-density lipoprotein cholesterol and to have lower liability to bipolar disorder and Alzheimer's disease. Reproduced with permission from Bulik-Sullivan et al. (2015)

with a genetic predisposition for creative activities more likely to select certain subjects or activities? This is an active choice from the individual, which is correlated with their genetics.

Finally, there are *passive* gene–environment correlations, which often arise owing to the inheritance of both genetic variation and environments from one generation to the next. For example, the offspring of parents with high polygenic scores for education are also themselves likely to have high polygenic scores for education owing to direct genetic inheritance. If these parents create educational nurturing environments because of their genetics, the offspring are likely to also inherit environments that improve their learning and education. Hence, the genetic variation and environment that offspring inherit are correlated, a phenomenon known as 'genetic nurture' (see e.g. Bates et al., 2018; Kong et al., 2018; Plomin & Bergeman, 1991; Wang et al., 2021; Wertz et al., 2018).

One way social science research has used genomic data is via studies of gene-by-environment interplay which aim to quantify the way that genetic effects may vary with

Box 3 Gene–environment interactions and correlations

The latest GWAS by Lee et al. (2018) shows that genetic variation explains a substantial amount of variation of educational attainment. Environmental or family circumstances, such as parental time and monetary investments in children, are also important. A study of the gene–environment interplay can shed light on how these two interact in shaping educational outcomes. Indeed, Muslimova et al. (2020), Howe et al. (2021) and Demange et al. (2020) build on a well-known literature that shows that firstborns, on average, have more years of education than later-born siblings, probably owing to additional parental investments in the early years. Muslimova et al. (2020) used a sample of siblings from the UK Biobank, where both birth order and genetic variation are random. They show that siblings with a higher polygenic score benefit more from being firstborn than those with a lower polygenic score. Both one's birth order and one's genetic variation are random within families, meaning that the analysis explores the causal effect of such gene–environment interactions.

environmental context and vice versa (Box 3). Existing studies of gene–environment interactions have for example shown that the social environment moderates genetic vulnerability to poor socioeconomic outcomes (see e.g. Belsky et al., 2016, 2018; Meyers et al., 2013; Papageorge & Thom, 2020; Rutter, 2006; Shanahan et al., 2008). A challenge in these studies is having sufficient statistical power (i.e. large enough sample sizes) to reliably detect interactions. Many of the most celebrated (and cited) 'interactions' claimed in the behavioural and psychiatric genetic fields have failed to replicate (Duncan & Keller, 2011).

However, it is often unclear what may be driving the interplay between genes and environments: does one's genotype truly moderate the environmental effect, or is the interaction effect caused by gene–environment correlation? Suppose we are interested in the effect of the polygenic score for education on years of schooling. In the presence of genetic nurture, the polygenic score will partially capture environmental and familial effects shaped by the parental genotype (i.e. passive *rGE*). If the direct genetic effect and the familial effect operate in the same direction, this will cause upwards bias in the genetic effect, leading us to believe that the direct effect is larger than it truly is. Including an additional interaction effect of the polygenic score and the environmental factor may also overestimate the coefficient of interest owing to this *rGE*.

The use of family genetic data can address this problem directly as genetic variation is random between siblings and for offspring, conditional on parental genetic variation. This means that one can identify the causal direct genetic effect by exploiting within-family genetic designs that examine between-sibling differences or control for parental genotypes. Again, however, a challenge in these studies is statistical power.

CONFOUNDING, BIAS AND CAUSATION

Disentangling cause and effect from correlation is notoriously difficult. Bias owing to confounding factors or reverse causation impedes many studies, making it difficult to reliably draw causal inferences from observational data. Suppose a researcher is interested in the relationship between myopia (shortsightedness) and educational attainment. Correlations between these variables may arise because of a causal effect of myopia on education, a background factor that causes both of these variables (such as time spent reading) or a

reverse causal effect of education on myopia. Ordinary least squares regression techniques commonly applied to observational data are unable to distinguish directionality (causation vs. reverse causation), while causal effects can only be estimated if the regression model includes a sufficient set of confounder variables which are measured without error.

Under specific assumptions, researchers can exploit genomic data to provide more reliable causal evidence compared with methods that attempt to measure and adjust for confounders (e.g. multivariable adjustment or propensity score approaches). Mendelian randomisation uses the random inheritance of genetic variation from parents to offspring as a natural experiment to strengthen causal inference with respect to potentially modifiable influences on health and development (Davey Smith & Ebrahim, 2003; Davies et al., 2018; Richmond & Davey Smith, 2021, Sanderson et al., 2022). Mendelian randomisation can be implemented as an instrumental variable (IV) analysis, with the SNP (or SNPs) acting as an IV for the exposure under investigation. The IV methods depend on three instrumental variable assumptions: (a) that the instrument associates with the independent variable of interest (relevance); (b) that there are no factors that affect both the independent (exposure) and dependent (outcome) variables (no confounding/independence); and (c) that there are no direct paths from the instrument to the dependent variable except via the independent variable of interest (the exclusion restriction). Figure 2 illustrates these assumptions graphically.

Regarding the relationship between myopia and educational attainment, observational studies have found evidence that, on average, more short-sighted people remain in school for longer. However, it is not clear whether this is because learning-associated factors (such as reading books) affect people’s eyesight or whether shortsightedness affects people’s progress throughout education systems (Ip et al., 2008). The policy implications of these two hypotheses are very different. If short-sightedness affects educational attainment, then steps to address students’ eyesight could lead to a fairer education system. If something about the educational process affects the risk of myopia, then there might be ways to mitigate this risk (e.g. spending more time outdoors; Guggenheim et al., 2012). While observational studies will not provide evidence about the causal effect driving this association, genomic studies can provide more compelling evidence about the direction of the effect. For example, we could investigate the associations of myopia-associated SNPs with educational

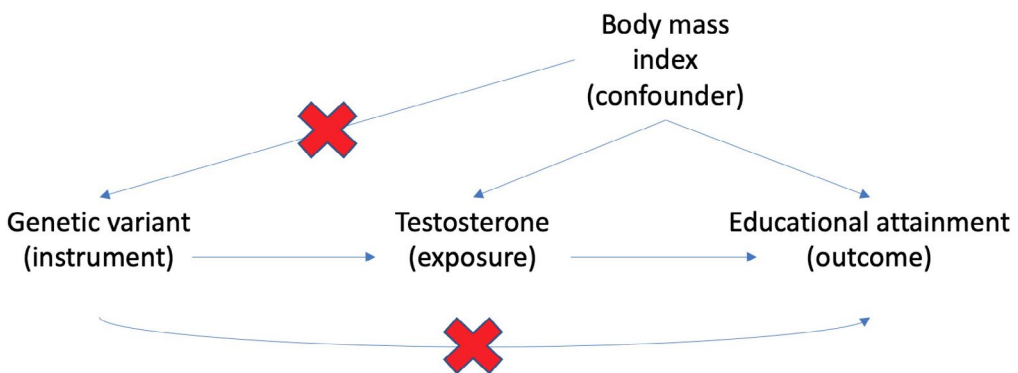


FIGURE 2 A directed acyclic graph illustrating the relationships between the independent variable (shortsightedness), dependent variable (educational attainment) and instrument (the genetic variant). The association of shortsightedness and educational attainment is potentially biased by socioeconomic position. Genetic variation in shortsightedness is unlikely to be related to socioeconomic position, is fixed at conception and cannot be affected by either shortsightedness or educational attainment. We can use these genetic variants to estimate the effect of shortsightedness on educational attainment (Mountjoy et al., 2018). One of the key strengths of MR is in demonstrating that some mechanisms are unlikely to be causal

attainment and the associations of education-associated SNPs with myopia to disentangle the direction of causation. If myopia affects educational attainment, we would expect individuals who inherit SNPs associated with myopia to have different educational attainment from individuals who do not inherit them. Conversely, if something about the educational process affects the risk of myopia, then we would expect that individuals who inherit education-associated SNPs to be more likely to develop myopia than individuals who do not inherit them (Figure 2). Mountjoy et al. (2018) investigated this hypothesis and found compelling evidence that myopia-associated SNPs did not strongly associate with educational attainment but that education-associated SNPs associated with myopia. These results suggest that the direction of causation runs from some process leading to educational attainment to myopia, rather than vice versa. Studies have been conducted for a wide range of other hypotheses and risk factors and can in theory be applied to a wide range of educational and social scientific traits where GWASs have been conducted.

Assessing the instrumental variable assumptions

The key to interpreting evidence from genetic studies is to determine whether assumptions on which they depend are plausible (Doidge & Dearden, 2017; Hemani et al., 2018a). GWASs have identified genetic variants that consistently associate with many traits of interest, so the first instrumental variable assumption (relevance) is likely to hold and can be assessed.

At conception, at each point of variation in the genome, offspring randomly inherit either their mother's or their father's DNA. DNA is set at conception, and so generally the environment does not affect germline genetic variation during the life course – a 90-year-old has the same DNA they were born with. This makes reverse causation unlikely (e.g. an outcome such as myopia, affecting genetic variation associated with educational attainment).

Furthermore, the random allocation of genetic variation at conception means that many, but not all, genetic variants are randomly distributed across the population. Hence the second instrumental variable assumption (independence) may hold, although this should always be evaluated.

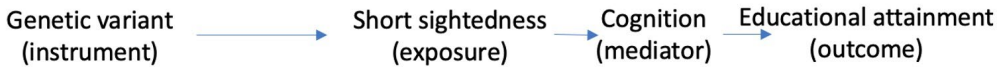
It is not possible to statistically prove that the genetic variants only affect the outcome via the proposed exposure, i.e. that the exclusion restriction holds. However, under the assumption that exposure has a constant effect on the outcome, it is possible to falsify this assumption by testing whether different genetic variants imply different effect sizes. If there is substantial heterogeneity in the estimates, then this suggests that a simple linear effect of the exposure on the outcome is unlikely, or that (one of) the instrumental variable assumptions do(es) not hold.

The exclusion restriction requires that instrumental variables only affect the outcome via the exposure of interest, i.e. that myopia-associated genetic variants only affect the outcome (educational attainment) via their effects on myopia. This may be plausible for genetic variants related to myopia, but it is more challenging for studies investigating the effects of educational attainment as an exposure. Educational attainment is generally considered to be a phenotype more distal from the genome than is myopia; therefore genetic variants that relate to educational attainment are likely to have more pleiotropic effects (i.e. the genetic variants are likely to be associated with other traits, such as cognition). Pleiotropy can be vertical, where a genetic variant affects one trait (e.g. cognition), which then affects an outcome (e.g. educational attainment). As long as the primary phenotype which the genetic variant relates to is correctly specified, then vertical pleiotropy is the essence of MR (Figure 3a; Davey Smith and Hemani 2014). Alternatively, pleiotropy can be horizontal, when a genetic variant affects both an exposure under investigation (e.g. educational attainment) and other variables through independent pathways. If one of these independent pathways

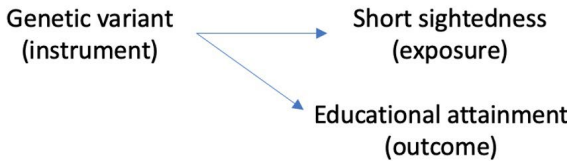
influences the outcome (e.g. short-sightedness), then this violates the exclusion restriction and causes bias in MR studies (Figure 3b).

If variants that primarily influence an upstream phenotype are mistakenly taken to be instruments for the downstream phenotype then inference will be incorrect (Davey Smith and Hemani 2014). For example, if a presumed genetic instrument for education is in fact primarily an instrument for cognition, then MR will estimate the causal effect of cognition – which includes any downstream effects acting through education – on the outcome. Methods exist that can improve the correct identification of the most proximal phenotype to a genetic variant (Hemani et al., 2017) (Figure 3c) Particularly problematic is when a genetic variant

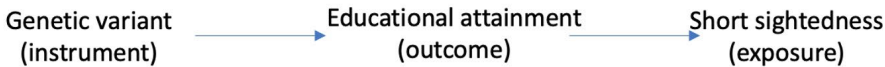
(a) Vertical pleiotropy



(b) Horizontal pleiotropy



(c) Mis-specifying the primary phenotype



(d) Correlated pleiotropy

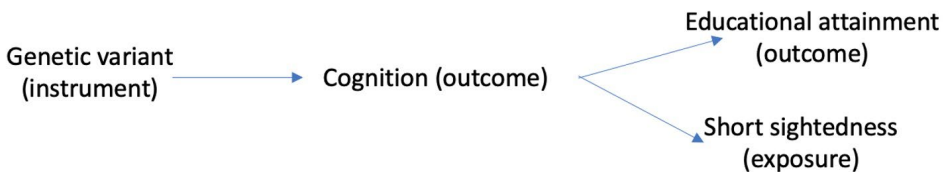


FIGURE 3 Possible structure of pleiotropic effects. In (a) the genetic variant affects short-sightedness, which in turn affects cognition, which in turn affects educational attainment. Vertical pleiotropy does not violate the assumptions of Mendelian randomisation (MR) here because educational attainment is purely a downstream factor from cognition. In contrast, in (b) if the genetic variant affects the exposure and outcome through distinct pathways then MR will be biased because horizontal pleiotropy violates the third instrumental variable assumption (exclusion restriction). Here, the variant can independently affect the outcome through the exposure or a third variable. In (c) a MR analysis which mistakenly identifies the primary phenotype influenced by the genetic variant used as an instrument is shown. If the instruments for education are considered to be instruments for a downstream trait influenced by education – such as short-sightedness – then inference will be in the wrong direction. This can be assessed using Stieger filtering (Hemani et al., 2018a). In this situation the pleiotropic effects will correlate with the effects through the exposure of interest, and such correlated pleiotropy will distort findings even when using many of the sensitivity analyses that have been developed for MR (Richmond & Davey Smith, 2021)

influences a confounder of the exposure to outcome association (Figure 3d). In this situation the pleiotropic effects will correlate with the effects through the exposure of interest, and such correlated pleiotropy will distort findings even when using many of the sensitivity analyses that have been developed for MR (Richmond & Davey Smith, 2021).

There are multiple methods which seek to evaluate the structure of pleiotropy and estimate the effects of exposure, even if some or all of the genetic variants have horizontally pleiotropic effects, or to evaluate the effects of multiple pathways, for example educational attainment and cognition, such as Anderson et al. (2020), Davies, Hill et al. (2019) and Sanderson et al. (2019). These use standard multivariable instrumental variable estimators to estimate the effects of multiple exposures using multiple instruments, e.g. see Greene (1993) and Wooldridge (2002).

Horizontal pleiotropy is probably less of an issue for more biologically proximal phenotypes, and may therefore have less of an impact on studies of some putative exposures that affect education (such as myopia) than studies of educational attainment as an exposure itself. This is doubly true if an exposure is found to have a precise null effect on educational attainment, such as in the example of shortsightedness. Horizontal and correlated pleiotropy are plausible explanations for why we might see associations between genetic variants known to associate with a trait like educational attainment and other outcomes such as myopia (i.e. false positive findings).

Mendelian randomisation requires a further fourth 'point identifying' instrumental variable assumption. This assumption can include that the exposure has a constant effect on the outcome or that the instrument (genetic variant) has a monotonic effect on the exposure; for more on this see Howe et al. (2021b).

Weighing evidence from genetic and randomised studies

The Educational Endowment Foundation (an education research funder in the UK) has funded a randomised trial of 700 students across 100 schools to estimate the impact of eyeglasses on educational outcomes for 4- and 5-year-olds who need them (Educational Endowment Foundation, 2021). This project will provide gold standard evidence for whether encouraging children to wear glasses can improve their reading achievement, mathematics achievement and visual acuity. The genetic evidence presented in Mountjoy et al. (2018) suggests that interventions to increase the use of glasses to overcome myopia (i.e. shortsightedness) are unlikely to substantially increase lifetime educational attainment (that is, the years of school attained).

However, there are crucial differences between the outcomes and exposures in the randomised trial and the genetic study. The exposure in the genetic study is myopia, whereas the intervention in the randomised trial covers a number of visual conditions, including short and long sightedness, amblyopia (lazy eye) and/or strabismus (a squint). The outcome in the genetic study relates to educational attainment, whereas the primary outcome for the trial is reading achievement measured at the end of the first year of school. Thus it is perfectly possible that the trial finds evidence of short-term increases in educational achievement in spite of the genetic evidence. Ultimately, the closest comparison will be whether this trial affects educational choices in 14 years when participants are aged 18+. This quite neatly highlights one of the strengths of genetically informed designs – that we can potentially provide some evidence about very long-term effects of interventions like glasses, that would take an enormously long time to discover using randomised controlled trials. However, it also highlights one of the limitations, that the results of genetic analysis estimate the effects of a specific exposure (myopia in the case of Mountjoy et al.) and a specific outcome (educational achievement/years of education). Both may differ from the hypothesis of interest, e.g. does encouraging 4- to 5-year-olds to wear glasses affect how well they learn to read?

Controlling for genetic variation

Genetic variables generally make very poor covariates. This is because they typically only explain a small proportion of the variation in a trait, and we know that there is almost certainly genetic variation that we have not identified (this is implied by the omnigenic model and limited statistical power of GWASs to identify all relevant genetic variation). Therefore, studies rarely control for genetic variables, and it is rarely a plausible or useful study design. There have been sophisticated attempts to combine instrumental variable analysis and genetic variation to, in effect, control for the mismeasured or unmeasured genetic variation contributing to a trait (most notably DiPrete et al., 2018; van Kippersluis et al., 2021) and repositories of polygenic indices (preprint: Becker et al., 2021; Richardson et al., 2019).

INTERGENERATIONAL TRANSMISSION

Given the predominance of data collection focussed on individuals rather than families, genomic data has historically only been available for unrelated individuals in large numbers. However, cohort and longitudinal studies are increasingly collecting genomic data on families, which offers new opportunities for studying the molecular and social mechanisms that mediate the intergenerational transmission of phenotypes such as educational attainment. For example, GWASs have suggested that the effects of SNPs that associate with educational attainment may be mediated via the effects of parents on their offspring. It is also possible to investigate the environmental mediation of these education-associated SNPs through familial and social factors, such as parenting behaviour. Recent research has demonstrated that SNP-educational attainment associations are at least partially mediated by environmental factors (Howe et al., 2021; Kong et al., 2018; Selzam et al., 2019). That is, estimated genetic associations do not represent just biological effects, but also social and demographic processes such as assortative mating and indirect effects such as genetic nurture. For example, if people who are more educated tend to have children with people who have lower BMI, then genetic variation that affects BMI and educational attainment will be correlated in their offspring. Similarly, if parents' BMI affects their offspring's educational attainment, then we would expect to see associations between genetic variation that affects BMI and the offspring's educational outcomes, even if BMI does not affect educational attainment at the individual level.

The half of parents' genomes that a child doesn't inherit, the 'non-inherited DNA', can also be used to investigate environmental mediation of genetic effects. Here, one can estimate parents' effects on their children through pathways other than direct genetic inheritance (Figure 4). Because these SNPs are not inherited, they cannot have a molecular effect within offspring and must therefore only impact offspring through the environment. These effects could include the impact of the family environment, such as parenting styles or parental personality, or population-level phenomena such as assortative mating. Hence, perhaps counter-intuitively, incorporating *genetic variation* in educational analysis can allow researchers to directly explore the role of the *family environment* in shaping offspring outcomes.

EDUCATION POLICY IMPLICATIONS

There has been a debate about the potential use of genetic data for educational policy (Grigorenko, 2007; Sabatello, 2018). The data and software resources are increasingly available (see Boxes 4 and 5). Some have argued that the presence of genetic associations with education and education-associated phenotypes indicates that educational policies

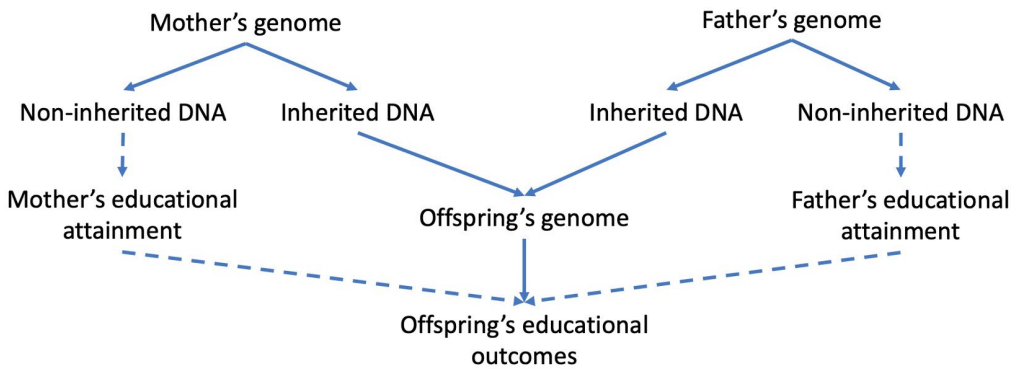


FIGURE 4 Relationship between parents' educational attainment and offspring educational attainment. The non-inherited genetic variation can be used to estimate the effects of parental education (or other parental traits) on offspring educational outcomes free from confounding via direct genetic inheritance

Box 4 UK genetic data available for education researchers

- Avon Longitudinal Study of Parents and Children – a longitudinal birth cohort study that sampled births between 1990 and 1992 in the Avon area. Large sample of mother and offspring genetic data, with a small number of samples from fathers. Questionnaire data and linkage available to the National Pupil Database, a census of exam results taken in UK state schools.
- Generation Scotland – large family based sample including genetic and epigenomic data. Extensive questionnaire, cognitive ability and linkage data.
- Millenium Cohort study – nationally representative birth cohort which also sampled some family members (i.e. parent–offspring trios). Data available on educational attainment and linkage to the National Pupil Database.
- National Child Development Study (1958) – representative cohort of individuals born in a single week in 1958 in the UK. Educational data collected during childhood and very long follow-up of health outcomes. Genetic data generated in a number of different samples, quality-controlled processed data will be available soon from the Centre for Longitudinal Studies.
- UK Biobank – a very large cohort study of nearly 500,000 individuals. The study has limited questionnaire data on educational attainment, how long participants remained in school and their qualifications. It has detailed health, biomarker, imaging data and linkages to medical records.
- Understanding Society (UKHLS) – household panel survey with genetic data on up to 12,000 individuals sampled by household.

should be tailored to individuals (Plomin, 2019; von Stumm & Plomin, 2021). This could be by identifying individuals with low genetic liabilities for education before entering the education system and targeting specialist education interventions to support their learning and address learning inequities.

However, there are two major issues with genetically informed policies aimed at individuals, such as using genomic information as a basis for school entry selection. First, there is growing evidence that observed SNP associations with education also reflect family and demographic factors as outlined above (Howe et al., 2021a; Kong et al., 2018; Morris, Davies,

Box 5 Resources for researchers

METHODS

There are a wide range of textbooks available that provide a detailed introduction to the use and analysis of genomic data (Asbury & Plomin, 2014; Mills et al., 2020) and reviews of the use of genetic data and educational outcomes (Freese, 2018; Plomin & von Stumm, 2018). The European Commission, Joint Research Centre (2019) report into social science genomics provides an excellent overview of recent developments in genomics and their implications for policy-makers. There are also a number of published reviews on MR and other genetic approaches, for example Davies et al. (2018) and Richmond & Davey Smith (2021) and specifically applied to familial data (Davies, Howe et al., 2019; Hwang et al., 2020; Thapar & Rice, 2021). An edited volume on a wide range of causal inference approaches involving human genetics is Davey Smith et al. (2021). Evans et al. (2021) published a special issue of *Behaviour Genetics* on describing methods for the genetic analysis of complex traits. Cesarini and Visscher (2017) provide an excellent summary of the potential uses of genetic evidence for educational research, building on earlier papers (Beauchamp et al., 2011; Benjamin et al., 2012). The Social Science Genetic Association Consortium (SSGAC) Summer Institute in Social Science Genomics provides an extensive reading list and lectures (SSGAC, 2021).

SOFTWARE

Genetic data can be used in educational research without specialist software. For example, if you have data on a specific genetic variation, i.e. how many alleles (zero, one or two) each participant has at a particular locus, these variables can be included in regression or other models in the same way as any other variable. However, many genetic datasets are provided in bespoke formats for genetic data. This is because there are a lot of variables (up to 80 million), and most standard formats (e.g. csv, Stata, SPSS, or R data files) struggle to handle this many variables. Common genetic data formats include .vcf (variant call format) and plink format (.bim,.bed and .fam for the summary information about the genetic variants, the genetic data for each participant and the ID information respectively). These files need to be processed with specialist software such as Plink (Purcell et al., 2007). Genetic correlations can be estimated using software such as GCTA (Yang et al., 2011). There are also specialist software packages and databases for running MR analysis such as the TwoSampleMR package for R and OpenGWAS (Hemani et al., 2018b).

Hemani et al., 2020; Wang et al., 2021). This evidence suggests that polygenic scores are likely to reflect not only the causal effects of SNPs on education, but also the influence of other social and familial processes. This may not matter if the goal is the prediction of general educational performance, but may matter if the goal is to identify specific causal pathways for support or to understand mechanisms, because non-causal relationships will not make good intervention targets. Second, genetic data currently add very little knowledge to individual-level predictions above other data that educators already have access to, such as age in year, sex and indicators of socioeconomic position such as free school meals (Morris, Davies & Davey Smith, 2020). Polygenic scores are likely to be useful for understanding

relationships at the population level, but it is less clear that they will be useful for providing individual-level predictions and personalised education (Davey Smith, 2011). Furthermore, genetically informing educational policy at the individual study level is unlikely to be politically acceptable or feasible. For further details see European Commission, Joint Research Centre (2019).

Genetically informed policies aimed at the group level, such as interventions designed to support pupils' progress, are probably more feasible. Genomic data also offers policy promise through how it can be used to assess structural differences within educational systems under differing confounding structures to traditional social scientific data. For example, genetic data can be used to investigate the benefits of studying at more advantaged schools net of genetic liability (Harden et al., 2020), selection differences between schools by socioeconomic position (Trejo et al., 2018), the differential performance of students with similar genetic liability for education from advantaged and disadvantaged backgrounds (Stumm et al., 2020), the potential role of genetics in fostering intergenerational social mobility (Belsky et al., 2018) and the robustness of value-added progress measures to pupil-level factors (Morris et al., 2018).

ETHICS

There are a number of ethical risks and concerns to the use of genetic data in educational research and policy. Schools currently use a wide array of data to select students for entry, group formation, targeted interventions or even exclusion. Children may respond to knowledge about their genomes (Van Wietmarschen et al., 2006) in ways that affect their learning, for example, by reducing their self-esteem (Kuther, 1994). There are also ethical issues surrounding informed consent from children and governing access to their genetic data. Velarde et al. (2021) argue that citizens should be able to decide 'who and how to trust' with their genetic information.

Many existing genetic studies have a Eurocentric approach (Mills & Rahal, 2019) and polygenic scores do not perform well across populations of different ancestries (Duncan et al., 2019). This means that inequalities between groups could be further entrenched by the systematically unequal benefit of genetic applications to European-ancestry populations. Furthermore, isolating genetic effects from social or environmental effects may be difficult given the social stratification that exists in many societies and the differential experience of environments between different groups (Martschenko et al., 2019). This is contextualised by an ugly history of eugenics that underlies some genetic research into intelligence in the nineteenth and twentieth centuries (Martschenko et al., 2019). Elements of this continue in the form of race pseudoscience that confuses or obfuscates genetic ancestry with race or ethnicity. Ancestry is a process-based concept that refers to the genetic similarity of individuals to one another that can be identified using genetic data (Yudell et al., 2016). Race is a pattern-based concept rooted in oppression that refers to the way that individuals have historically been ascribed to socially constructed or geographically defined groups and cannot be identified using genetic data. Ethnicity is a broad pattern-based approach rooted in identity that commonly refers to a broad range of shared factors such as traditions, cultural experiences or national origins, and cannot be identified using genetic data. For further information, see Birney et al. (2021), who provide a detailed discussion and guidance about the importance of the scientific language used to describe research using genetics.

SUMMARY

Findings from sociogenomics have the potential to improve our understanding of not only the genomic and biological factors, but also familial and social factors that combine across the life course to affect educational outcomes. Incorporation and uptake of genomic data may help to inform policymaking and the better design of interventions. This may be through identifying how nature and nurture interact to shape individual outcomes, elucidating the causal mechanisms underlying educational outcomes, and by providing novel insights into the transmission of educational advantage across generations. Many insights from genetics for education research are likely to be made at the population level (i.e. factors or processes that affect people on average), rather than for individual-level prediction (i.e. personalised education). With that, genetic data may be able to provide a wealth of evidence for education research.

ACKNOWLEDGEMENTS

We are grateful for comments from George Leckie, Danny Dorling and David Manley.

ETHICS STATEMENT

This review did not conduct any new empirical research and did not require ethical approval.

CONFLICT OF INTEREST

We receive, or have received funding for research into the use and application of genetics.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Tim T. Morris  <https://orcid.org/0000-0001-8178-6815>

Neil M. Davies  <https://orcid.org/0000-0002-2460-0508>

ENDNOTE

¹ For example, this is effectively the R^2 of genetic variation for the trait.

REFERENCES

- Anderson, E. L., Howe, L. D., Wade, K. H., Ben-Shlomo, Y., Hill, W. D., Deary, I. J., Sanderson, E. C., Zheng, J., Korologou-Linden, R., Stergiakouli, E., Davey Smith, G., Davies, N. M., & Hemani, G. (2020). Education, intelligence and Alzheimer's disease: Evidence from a multivariable two-sample Mendelian randomization study. *International Journal of Epidemiology*, 49, 1163–1172. <https://doi.org/10.1093/ije/dy280>
- Asbury, K., & Plomin, R. (2014). *G is for genes: the impact of genetics on education and achievement, Understanding children's worlds*. Wiley-Blackwell, .
- Bates, T. C., Maher, B. S., Medland, S. E., McAloney, K., Wright, M. J., Hansell, N. K., Kendler, K. S., Martin, N. G., & Gillespie, N. A. (2018). The nature of nurture: Using a virtual-parent design to test parenting effects on children's educational attainment in genotyped families. *Twin Research and Human Genetics*, 21, 73–83. <https://doi.org/10.1017/thg.2018.11>
- Beauchamp, J. P., Cesarini, D., Johannesson, M., van der Loos, M. J. H. M., Koellinger, P. D., Groenen, P. J. F., Fowler, J. H., Rosenquist, J. N., Thurik, A. R., & Christakis, N. A. (2011). Molecular genetics and economics. *Journal of Economic Perspectives*, 25, 57–82. <https://doi.org/10.1257/jep.25.4.57>
- Becker, J., Burik, C. A. P., Goldman, G., Wang, N., Jayashankar, H., Bennett, M., Belsky, D. W., Linnér, R. K., Ahlskog, R., Kleinman, A., Hinds, D. A., 23 and Me Research Group, Caspi, A., Corcoran, D. L., Moffitt, T. E.,

- Poulton, R., Sugden, K., Williams, B. S., Harris, K. M., ... Okbay, A. (2021). Resource profile and user guide of the polygenic index repository [Preprint]. *bioRxiv*. <https://doi.org/10.1101/2021.05.08.443158>
- Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., Conley, D., Fletcher, J. M., Freese, J., Herd, P., Moffitt, T. E., Poulton, R., Sicinski, K., Wertz, J., & Harris, K. M. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E7275–E7284. <https://doi.org/10.1073/pnas.1801238115>
- Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., Houts, R., Ramrakha, S., Sugden, K., Williams, B. S., Poulton, R., & Caspi, A. (2016). The genetics of success: how single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychological Science*, *27*(7), 957–972. <https://doi.org/10.1177/0956797616643070>
- Benjamin, D. J., Cesarini, D., Chabris, C. F., Glaeser, E. L., Laibson, D. I., Guðnason, V., Harris, T. B., Launer, L. J., Purcell, S., Smith, A. V., Johannesson, M., Magnusson, P. K. E., Beauchamp, J. P., Christakis, N. A., Atwood, C. S., Hebert, B., Freese, J., Hauser, R. M., Hauser, T. S., ... Lichtenstein, P. (2012). The promises and pitfalls of geno-economics. *Annual Review of Economics*, *4*, 627–662. <https://doi.org/10.1146/annurev-economics-080511-110939>
- Birney, E., Inouye, M., Raff, J., Rutherford, A., & Scally, A. (2021). The language of race, ethnicity, and ancestry in human genetic research. *arxiv*.
- Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded view of complex traits: From polygenic to omnigenic. *Cell*, *169*, 1177–1186. <https://doi.org/10.1016/j.cell.2017.05.038>
- Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. P., Harrison, S., Vie, G. Å., Cho, Y., Howe, L. D., Hughes, A., Boomsma, D. I., Havdahl, A., Hopper, J., Neale, M., Nivard, M. G., Pedersen, N. L., Reynolds, C. A., Tucker-Drob, E. M., Grotzinger, A., Howe, L., ... Davies, N. M. (2020). Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nature Communications*, *11*, 3519. <https://doi.org/10.1038/s41467-020-17117-4>
- Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., Duncan, L., Perry, J. R. B., 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*, 1236–1241. <https://doi.org/10.1038/ng.3406>
- Cesarini, D., & Visscher, P. M. (2017). Genetics and educational attainment. *Npj Science Learn*, *2*, 4. <https://doi.org/10.1038/s41539-017-0005-6>
- Davey Smith, G. (2011). Epidemiology, epigenetics and the 'Gloomy Prospect': Embracing randomness in population health research and practice. *International Journal of Epidemiology*, *40*, 537–562. <https://doi.org/10.1093/ije/dyr117>
- Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, *32*, 1–22. <https://doi.org/10.1093/ije/dyg070>
- Davey Smith, G., R. Richmond, & J.-B. Pingault (Eds.). (2021). *Combining human genetics and causal inference to understand human disease and development, Cold Spring Harbor perspectives in medicine*. Cold Spring Harbor Laboratory Press.
- Davies, N. M., Hill, W. D., Anderson, E. L., Sanderson, E., Deary, I. J., & Davey Smith, G. (2019). Multivariable two-sample Mendelian randomization estimates of the effects of intelligence and education on health. *eLife*, *8*. <https://doi.org/10.7554/eLife.43990>
- Davies, N. M., Holmes, M. V., & Davey Smith, G. (2018). Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*, *k601*. <https://doi.org/10.1136/bmj.k601>
- Davies, N. M., Howe, L. J., Brumpton, B., Havdahl, A., Evans, D. M., & Davey Smith, G. (2019). Within family Mendelian randomization studies. *Human Molecular Genetics*, *28*, R170–R179. <https://doi.org/10.1093/hmg/ddz204>
- Demange, P. A., Hottenga, J. J., Abdellaoui, A., Malanchini, M., Domingue, B. W., de Zeeuw, E. L., Rimfeld, K., Eley, T., Boomsma, D. I., van Bergen, E., Breen, G., Nivard, M. G., & Cheesman, R. (2020). Parental influences on offspring education: Indirect genetic effects of non-cognitive skills. *bioRxiv*. <https://doi.org/10.1101/2020.09.15.296236>
- DiPrete, T. A., Burik, C. A. P., & Koellinger, P. D. (2018). Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data. *Proceedings of the National Academy of Sciences*, *115*, E4970–E4979. <https://doi.org/10.1073/pnas.1707388115>
- Doidge, J. C., & Dearden, L. (2017). Beyond the limits of Mendelian randomisation [WWW Document]. *BMJ Rapid response*. <https://www.bmj.com/content/358/bmj.j3542/rr-2> (accessed 91421)
- Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *AJP*, *168*, 1041–1049. <https://doi.org/10.1176/appi.ajp.2011.11020191>
- 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*, 3328. <https://doi.org/10.1038/s41467-019-11112-0>
- Educational Endowment Foundation (2021). Glasses in Classes. <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/glasses-in-classes/>

- European Commission, Joint Research Centre (2019). Genome-wide association studies, polygenic scores and social science genetics: overview and policy implications.
- Evans, D. M., Medland, S. E., & Prom-Wormley, E. (2021). Introduction to the special issue on statistical genetic methods for human complex traits. *Behavior Genetics*, 51(3), 165–169. <https://doi.org/10.1007/s10519-021-10057-9>
- Freese, J. (2018). The arrival of social science genomics. *Contemporary Sociology: A Journal of Reviews*, 47, 524–536. <https://doi.org/10.1177/0094306118792214a>
- Greene, W. (1993). *Econometric analysis*, 2nd ed. Prentice Hall.
- Grigorenko, E. L. (2007). How can genomics inform education? *Mind, Brain, and Education*, 1, 20–27. <https://doi.org/10.1111/j.1751-228X.2007.00001.x>
- Guggenheim, J. A., Northstone, K., McMahon, G., Ness, A. R., Deere, K., Mattocks, C., Pourcain, B. S., & Williams, C. (2012). Time outdoors and physical activity as predictors of incident myopia in childhood: A prospective cohort study. *Investigative Ophthalmology & Visual Science*, 53, 2856. <https://doi.org/10.1167/iovs.11-9091>
- Harden, K. P., Domingue, B. W., Belsky, D. W., Boardman, J. D., Crosnoe, R., Malanchini, M., Nivard, M., Tucker-Drob, E. M., & Harris, K. M. (2020). Genetic associations with mathematics tracking and persistence in secondary school. *Npj Sci. Learn*, 5, 1. <https://doi.org/10.1038/s41539-020-0060-2>
- Hemani, G., Bowden, J., & Davey Smith, G. (2018a). Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human Molecular Genetics*, 27, R195–R208. <https://doi.org/10.1093/hmg/ddy163>
- Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., Tan, V. Y., Yarmolinsky, J., Shihab, H. A., Timpson, N. J., Evans, D. M., Relton, C., Martin, R. M., Davey Smith, G., Gaunt, T. R., & Haycock, P. C. (2018). The MR-Base platform supports systematic causal inference across the human genome. *eLife*, 7. <https://doi.org/10.7554/eLife.34408>
- Howe, L., Evans, D., Hemani, G., Smith, G. D., & Davies, N. M. (2021). Evaluating indirect genetic effects of siblings using singletons (preprint). *Open Science Framework*. <https://doi.org/10.31219/osf.io/j2rms>
- 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. (2021a). Within-sibship GWAS improve estimates of direct genetic effects (preprint). *Genetics*, <https://doi.org/10.1101/2021.03.05.433935>
- Howe, L. J., Tudball, M., Davey Smith, G., & Davies, N. M. (2021b). Interpreting Mendelian-randomization estimates of the effects of categorical exposures such as disease status and educational attainment. *International Journal of Epidemiology* dyab208, <https://doi.org/10.1093/ije/dyab208>
- Howe, L. J., Tudball, M., Smith, G. D., & Davies, N. M. (2020). Preprint. *Epidemiology*, <https://doi.org/10.1101/2020.12.14.20248168>
- Hwang, L.-D., Davies, N. M., Warrington, N. M., & Evans, D. M. (2020). Integrating family-based and mendelian randomization designs. *Cold Spring Harbor Perspectives in Medicine*, 11(3), a039503. <https://doi.org/10.1101/cshperspect.a039503>
- Ip, J. M., Saw, S.-M., Rose, K. A., Morgan, I. G., Kifley, A., Wang, J. J., & Mitchell, P. (2008). Role of near work in myopia: Findings in a sample of Australian school children. *Investigative Ophthalmology & Visual Science*, 49, 2903. <https://doi.org/10.1167/iovs.07-0804>
- Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsón, B. J., Young, A. I., Thorgeirsson, T. E., Benonisdóttir, S., Oddsson, A., Halldórsson, B. V., Masson, G., Gudbjartsson, D. F., Helgason, A., Björnsdóttir, G., Thorsteinsdóttir, U., & Stefánsson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, 359, 424–428. <https://doi.org/10.1126/science.aan6877>
- Kuther, T. L. (1994). Diagnostic classification of children within the educational system: Should it be eliminated? *ERIC*.
- Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghziyan, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, 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, 1112–1121. <https://doi.org/10.1038/s41588-018-0147-3>
- Martschenko, D., Trejo, S., & Domingue, B. W. (2019). Genetics and education: Recent developments in the context of an ugly history and an uncertain future. *AERA Open*, 5, 233285841881051. <https://doi.org/10.1177/2332858418810516>
- Meyers, J. L., Cerdá, M., Galea, S., Keyes, K. M., Aiello, A. E., Uddin, M., Wildman, D. E., & Koenen, K. C. (2013). Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit neighborhood health study. *Translational Psychiatry*, 3, e290. <https://doi.org/10.1038/tp.2013.63>
- Mills, M., Barban, N., & Tropf, F. C. (2020). *An introduction to statistical genetic data analysis*. The MIT Press.
- Mills, M. C., & Rahal, C. (2019). A scientometric review of genome-wide association studies. *Communications Biology*, 2, <https://doi.org/10.1038/s42003-018-0261-x>
- Morris, T. T., Davies, N. M., & Davey Smith, G. (2020). Can education be personalised using pupils' genetic data? *eLife*, 9, e49962. <https://doi.org/10.7554/eLife.49962>

- Morris, T. T., Davies, N. M., Dorling, D., Richmond, R. C., & Smith, G. D. (2018). Testing the validity of value-added measures of educational progress with genetic data. *British Educational Research Journal*, <https://doi.org/10.1002/berj.3466>
- Morris, T. T., Davies, N. M., Hemani, G., & Smith, G. D. (2020). Population phenomena inflate genetic associations of complex social traits. *Science Advances*, 6, eaay0328. <https://doi.org/10.1126/sciadv.aay0328>
- Mountjoy, E., Davies, N. M., Plotnikov, D., Smith, G. D., Rodriguez, S., Williams, C. E., Guggenheim, J. A., & Atan, D. (2018). Education and myopia: Assessing the direction of causality by mendelian randomisation. *BMJ*, k2022. <https://doi.org/10.1136/bmj.k2022>
- Muslimova, D., van Kippersluis, H., Rietveld, C. A., von Hinke, S., & Meddens, S. F. W. (2020). Dynamic complementarity in skill production: Evidence from genetic endowments and birth order. *SSRN Electronic Journal*, <https://doi.org/10.2139/ssrn.3748468>
- Papageorge, N. W., & Thom, K. (2020). Genes, education, and labor market outcomes: Evidence from the health and retirement study. *Journal of the European Economic Association*, 18, 1351–1399. <https://doi.org/10.1093/jeaa/jvz072>
- Plomin, R. (2019). Blueprint: How DNA makes us who we are.
- Plomin, R., & Bergeman, C. S. (1991). The nature of nurture: Genetic influence on 'environmental' measures. *Behavioral and Brain Sciences*, 14, 373–386. <https://doi.org/10.1017/S0140525X00070278>
- Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, 84, 309–322. <https://doi.org/10.1037/0033-2909.84.2.309>
- Plomin, R., Shakeshaft, N. G., McMillan, A., & Trzaskowski, M. (2014). Nature, nurture, and expertise. *Intelligence*, 45, 46–59. <https://doi.org/10.1016/j.intell.2013.06.008>
- Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics*, 19, 148–159. <https://doi.org/10.1038/nrg.2017.104>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81, 559–575. <https://doi.org/10.1086/519795>
- Richardson, T. G., Harrison, S., Hemani, G., & Davey Smith, G. (2019). An atlas of polygenic risk score associations to highlight putative causal relationships across the human genome. *eLife*, 8. <https://doi.org/10.7554/eLife.43657>
- Richmond, R. C., & Davey Smith, G. (2022). Mendelian randomization: Concepts and scope. *Cold Spring Harbor Perspectives in Medicine*, 12(1), a040501. <https://doi.org/10.1101/cshperspect.a040501>
- Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., Westra, H.-J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., Albrecht, E., Alizadeh, B. Z., Amin, N., Barnard, J., Baumeister, S. E., Benke, K. S., Bielak, L. F., Boatman, J. A., Boyle, P. A., ... Koellinger, P. D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, 340, 1467–1471. <https://doi.org/10.1126/science.1235488>
- Rutter, M. (2006). *Genes and behavior: nature – nurture interplay explained*. Blackwell Pub, Malden, MA.
- Sabatello, M. (2018). A genomically informed education system? Challenges for behavioral genetics. *Journal of Law, Medicine & Ethics*, 46, 130–144. <https://doi.org/10.1177/1073110518766027>
- Sanderson, E., Davey Smith, G., Bowden, J., & Munafò, M. R. (2019). Mendelian randomisation analysis of the effect of educational attainment and cognitive ability on smoking behaviour. *Nature Communications*, 10, 2949. <https://doi.org/10.1038/s41467-019-10679-y>
- Sanderson, E., Glymour, M. M., Holmes, M. V., Kang, H., Morrison, J., Munafò, M. R., Palmer, T., Mary Schooling, C., Wallace, C., Zhao, Q., & Davey Smith, G. (2022). Mendelian randomization. *Nature Reviews Methods Primers*, 2, 6. <https://doi.org/10.1038/s43586-021-00092-5>
- 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. *The American Journal of Human Genetics*, 105, 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>
- SSGAC (2021). MATERIALS [WWW Document]. <https://www.rsfgenicsschool.com/2019-materials> (accessed 52721)
- Stumm, S., Smith-Woolley, E., Ayorech, Z., McMillan, A., Rimfeld, K., Dale, P. S., & Plomin, R. (2020). Predicting educational achievement from genomic measures and socioeconomic status. *Dev Sci*, 23, <https://doi.org/10.1111/desc.12925>
- Taylor, A. E., Jones, H. J., Sallis, H., Euesden, J., Stergiakouli, E., Davies, N. M., Zammit, S., Lawlor, D. A., Munafò, M. R., Davey Smith, G., & Tilling, K. (2018). Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*, <https://doi.org/10.1093/ije/dyy060>

- Thapar, A., & Rice, F. (2021). Family-based designs that disentangle inherited factors from pre- and postnatal environmental exposures. *in vitro* fertilization, discordant sibling pairs, maternal versus paternal comparisons, and adoption designs. *Cold Spring Harbor Perspectives in Medicine*, 11, a038877. <https://doi.org/10.1101/cshperspect.a038877>
- The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature*, 526, 68–74. <https://doi.org/10.1038/nature15393>
- Trejo, S., Belsky, D. W., Boardman, J. D., Freese, J., Harris, K. M., Herd, P., Sicinski, K., & Domingue, B. W. (2018). Schools as moderators of genetic associations with life course attainments: evidence from the WLS and add health. *Sociological Science*, 5, 513–540. <https://doi.org/10.15195/v5.a22>
- Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, 9, 160–164. <https://doi.org/10.1111/1467-8721.00084>
- van Kippersluis, H., Biroli, P., Galama, T. J., von Hinke, S., Meddens, S. F. W., Muslimova, D., Pereira, R., & Rietveld, C. A. (2021). Using obviously-related instrumental variables to increase the predictive power of polygenic scores (preprint). *Genetics*, <https://doi.org/10.1101/2021.04.09.439157>
- Van Wietmarschen, H., Jongerden, J., & Ruivenkamp, G. (2006). The social construction of genetic identities: A research note. *Tailoring Biotechnologies*, 2, 109–120.
- Velarde, M. C. R., Tsantoulis, P., Burton-Jeangros, C., Aceti, M., Chappuis, P., & Hurst-Majno, S. (2021). Citizens' views on sharing their health data: The role of competence, reliability and pursuing the common good. *BMC Medical Ethics*, 22, 1–12. <https://doi.org/10.1186/s12910-021-00633-3>
- Verbanck, M., Chen, C.-Y., Neale, B., & Do, R. (2017). Widespread Pleiotropy Confounds Causal Relationships between Complex Traits and Diseases Inferred from Mendelian Randomization, <https://doi.org/10.1101/157552>
- von Stumm, S., & Plomin, R. (2021). Using DNA to predict intelligence. *Intelligence*, 86, 101530. <https://doi.org/10.1016/j.intell.2021.101530>
- Wang, B., Baldwin, J. R., Schoeler, T., Cheesman, R., Barkhuizen, W., Dudbridge, F., Bann, D., Morris, T. T., & Pingault, J.-B. (2021). Genetic nurture effects on education: A systematic review and meta-analysis, <https://doi.org/10.1101/2021.01.15.426782>
- Wertz, J., Caspi, A., Belsky, D. W., Beckley, A. L., Arseneault, L., Barnes, J. C., Corcoran, D. L., Hogan, S., Houts, R. M., Morgan, N., Odgers, C. L., Prinz, J. A., Sugden, K., Williams, B. S., Poulton, R., & Moffitt, T. E. (2018). Genetics and crime: Integrating new genomic discoveries into psychological research about antisocial behavior. *Psychological Science*, 29, 791–803. <https://doi.org/10.1177/0956797617744542>
- Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. The MIT press.
- Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88, 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>
- Yudell, M., Roberts, D., DeSalle, R., & Tishkoff, S. (2016). Taking race out of human genetics. *Science*, 351, 564–565. <https://doi.org/10.1126/science.aac4951>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Morris, T. T., von Hinke, S., Pike, L., Ingram, N. R., Davey Smith, G., Munafò, M. R., & Davies, N. M. (2022). Implications of the genomic revolution for education research and policy. *British Educational Research Journal*, 00, 1–21. <https://doi.org/10.1002/berj.3784>