Population phenomena inflate genetic associations of complex social traits

Tim T Morris [ab†], Neil M Davies [ab], Gibran Hemani [ab], George Davey Smith [ab]

[a] Medical Research Council Integrative Epidemiology Unit, University of Bristol, BS8 2BN, United Kingdom.

[b] Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom.

[†] Corresponding author: Tim Morris. Email: Tim.Morris@bristol.ac.uk

Key words: Genetics; education; socioeconomic; heritability; confounding; bias; ALSPAC.

Abstract

Heritability, genetic correlation and genetic associations estimated from samples of unrelated individuals are often perceived as confirmation that genotype causes the phenotype(s). However, these estimates can arise from indirect mechanisms due to population phenomena including population stratification, dynastic effects, and assortative mating. We introduce these, describe how they can bias or inflate genotype-phenotype associations, and demonstrate methods that can be used to assess their presence. Using data on educational achievement and parental socioeconomic position as an exemplar, we demonstrate that both heritability and genetic correlation may be biased estimates of the causal contribution of genotype. These results highlight the limitations of genotype-phenotype estimates obtained from samples of unrelated individuals. Use of these methods in combination with family-based designs may offer researchers greater opportunities to explore the mechanisms driving genotype-phenotype associations and identify the causes of estimate bias.

Introduction

Genotype-phenotype associations inferred from genetic data can be used to provide insight into the genetic architecture of complex traits, interrogate causal and non-causal associations between different phenotypes, and create phenotypic predictors.[1–3] In most situations, these applications depend upon a narrow definition of genotype-phenotype association; that there is a causal path from an individual's genotype to the same individual's phenotype. However, the common practice of using samples of unrelated individuals to estimate genotype-phenotype associations is liable to bias where other causal paths can confound these associations. For example, population dynamic phenomena such as population stratification, dynastic effects and assortative mating can induce correlations through confounding between genotypes and phenotypes. These processes do not reflect the causal pathways that are typically intended to be identified.

In this paper we first review the strategies for estimating genotype-phenotype associations in samples of unrelated individuals and then describe in detail the population dynamic phenomena that could bias these estimates away from the causal parameter. We then demonstrate empirical examples of these with a focus on socioeconomic phenotypes, suggest tools for detecting these biases, and discuss some potential consequences of, and solutions to them.

Estimation of genotype-phenotype associations

Fisher partitioned genotype-phenotype associations into two components[4], though the terms for these are seldom used consistently.[5] For simplicity we will refer to them as variant substitution effects, and confounding effects. Variant substitution effects can be thought of as the (counterfactual) change in an individual's phenotype that would occur as a result of changing that individual's genotype from conception (holding all else constant). In most cases this type of effect is the target of any genotype-phenotype association analysis. The mechanism that cascades from a variant substitution may be entirely molecular, for example altering gene expression that leads to disease, or it may be more complex and external, for example influencing behaviour which leads to environmental changes that in turn influence the phenotype. In both cases, there is a causal path from an individual's genotype to their phenotype that reflects a counterfactual model.[6] If this path of interest from genotype to phenotype is confounded, genotype-phenotype associations will not (solely) reflect underlying causal mechanisms but will be biased. Various population phenomena such as population stratification,[7] dynastic effects[8] and assortative mating[9,10] can introduce such confounding[2,11] (Figure 1 and Box 1). These population phenomena can be considered to inflate the true values of population estimates and represent the inaccuracy of the hypothetical counterfactual of substituting a sampled individual's genotype on their phenotype.

Genotype-phenotype associations are commonly estimated in three ways: SNP heritability, which represents the total genetic component of a trait estimated from variation in all measured SNPs; genetic correlation, which represents the correlation in effects of all measured SNPs on two or more phenotypes; and genetic associations, which represent how a phenotype is influenced by a specific SNP. SNP heritability and genetic correlation are estimated from whole genome data with methods such as GREML[12] and LD Score regression,[13,14] while genetic associations are estimated as per-SNP effects in genome-wide association studies (GWAS) using linear or logistic regression. Throughout, we refer to these collectively as genotype-phenotype associations. We focus on evaluating how various population level phenomena bias the parameters that can be estimated from whole genome-based approaches such as SNP heritability and genetic correlation, but the biases we describe can also inflate per-SNP estimates obtained from GWAS. We note that whole genome

methods are additionally susceptible to a separate set of biases which have been under considerable scrutiny, which arise when the observed SNPs follow different distributions to the unknown causal variants,[15] though they will not be discussed further here.

The population level phenomena that can contaminate genotype-phenotype associations

One mechanism which may bias genetic associations is population stratification (Figure 1, yellow paths), where confounding of genotype-phenotype associations is driven by population structure.[16] Population structure refers to systematic differences in allele frequencies between sub-populations (which often appears as geographical structure) due to ancestry.[7,17] Because phenotypes are often geographically patterned, spurious genotype-phenotype associations (both heritability and genetic correlation) can arise even when there does not exist a variant substitution effect on the phenotype. An oft-repeated example is that SNPs which have different frequencies in East Asian and European populations will be associated with chopstick use, even though the reasons underlying chopstick use are cultural rather than genetic.[18] Bias due to population stratification is commonly controlled for by restricting samples to a homogenous population and adjusting models for principal components of genotype, which capture common differences between sub-populations in allele frequencies. A recent study however demonstrated that geographical structure remains even after controlling for the first 100 principal components in large scale biobanks, far in excess of the 10 or 20 components commonly controlled for.[19] While it is not possible to prove that adjusting for principal components has controlled for all differences within the sample, one way to assess the impact of population stratification is to compare estimates obtained from unadjusted models and models that adjust for principal components. Attenuation in estimated effect sizes after principal component adjustment can provide evidence of population stratification, and the extent of this may be gauged by the extent of attenuation. However, in studies with a geographically homogenous sampling framework this may be insufficient.[19] Between-sibling study designs offer a robust solution because Mendel's first and second laws of independent segregation and assortment ensure that genetic differences between siblings are not correlated with environment.[20–23]

Genetic associations can also be biased by dynastic effects (Figure 1, red paths) whereby inherited SNPs operate indirectly on offspring phenotype via their effects in the parents' phenotype. For example, suppose that education associated SNPs at the parental generation contribute to the creation of education-enriching environments through the provision of books in the household. It follows that children of more educated parents will be more likely to inherit both education-associated SNPs (the biological path from offspring genotype to offspring education) and education-associated environments (the non-biological path from parental genotype to offspring education). This is a form of gene-environment correlation and can be thought of as a double contribution of genotype. Thus, social or environmental transmission effects can affect genotype-phenotype associations, leading to biased estimates of the causal effect of genotype on phenotype. It is important to note that under this model, the confounding effect is in fact due to a variant substitution effect, though the variant substitution occurred not in the individual being analysed but in their parents. It is possible that dynastic effects explain the relatively low estimates of the contribution of the shared environment from twin studies, which attribute these genetic nurture effects to the additive (heritable) effects of genetics. There is a large body of evidence suggesting that social phenotypes such as education and SEP are socially transmitted across generations,[24–26] and it is likely that genetic associations with these phenotypes will be affected by dynastic effects. The presence of dynastic effects can be tested and estimated with data on mother-father-offspring

trios or siblings.[20] Using polygenic scores, the raw association between offspring genotype and phenotype can be compared with its association when adjusted for maternal and paternal genotype. Attenuation of the raw association and direct (conditional) association between parental genotype and offspring phenotype supports an indirect effect of parental genotype on offspring phenotype and therefore the presence of dynastic effects. It is also possible to use non-transmitted parental SNPs to create a 'genetic nurture' polygenic score.[8] Because non-transmitted SNPs can only influence offspring phenotype indirectly, association between a non-transmitted score and offspring phenotype supports dynastic effects. Relatedness disequilibrium regression, which investigates changes in phenotypic similarity by relatedness amongst samples of siblings, can also be used to estimate bias in heritability estimates caused by environmental effects.[11] These methods all require data on genotyped mother-father-offspring trios and will be facilitated by large family based studies such as HUNT and MoBa.[27,28]

Assortative mating (Figure 1, green path) may also induce genetic associations between phenotypes. Assortative mating refers to the non-random pairing of spouses across the population and arises from mate selection based on phenotypic characteristics and social homogamy. There is evidence for assortative mating on a range of phenotypes including education and SEP.[9,10,29–31] Where phenotypes that are selected on have a genetic component, assortative mating will lead to spouses being more genetically similar to each other than to randomly selected individuals from a population. That is, phenotypic assortative mating across a population increases the likelihood of people mating with partners who are more genetically similar. While random mating would ensure even distribution of allele frequencies at the population level, assortative mating leads to systematic differences in allele frequencies (population stratification) and subsequent deviations from Hardy-Weinberg equilibrium over generations.[9] Assortative mating will lead to a disproportionate enrichment or depletion of education-associated alleles within spouse couples, and increased homozygosity, long-range linkage, and genetic variation in offspring across a population, biasing genotype-phenotype associations.[9,32,33] For example, offspring of parents with higher education are more likely to have a greater number of education-increasing alleles than offspring of parents with lower education. If spouses sort on different traits (i.e. cross trait assortative mating) then assortment can also induce genetic correlations between traits in offspring.[34] Assortative mating can lead to population stratification if it is sub-population specific,[35] and to disproportionate inheritance of the environment in addition to genotype if dynastic effects exist.

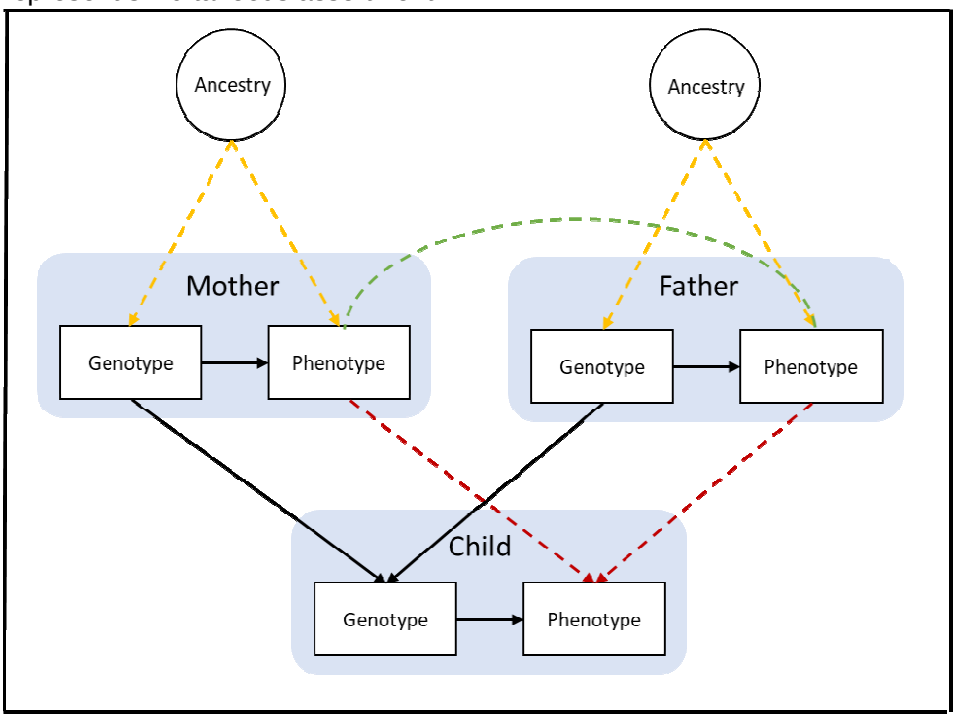
Determining the presence of population phenomena

Studies with a genetic focus that examine complex social phenomena such as education[36,37] may be particularly susceptible to bias arising due to population level phenomena. Education is one of many heavily studied social phenotypes in genetic studies.[38] Alongside parental occupational position it is one of the strongest determinants of health and social outcomes throughout the lifecourse.[39–43] Conceptually viewed as subcategories of broader socioeconomic position (SEP),[44,45] education and occupational position are strongly correlated phenotypically and genotypically,[43,46–50] highly heritable,[40,43,49–51] and have a complex genetic architecture characterised by high polygenicity.[38,49,52] The heritability of education has been estimated at 40% for years of education[40] and 60% for test score achievement.[53–56] Given this distinction we hereafter refer to years of education as 'attainment' and test score achievement as 'achievement'.[41] There is evidence of high genetic correlation (0.48-1) between educational attainment and other indicators of SEP such as social class,[43,49,50] but these may operate through an intermediate phenotype such as cognitive ability.[49,57,58] Cognitive ability is highly heritable[50,51,53] and correlates with many measures of SEP

phenotypically and genotypically.[39,46,50,51,53,59–63] The way in which complex social phenotypes such as education and occupation associate with genotype may have important implications for social policy to reduce inequalities throughout the lifecourse. It is therefore of paramount importance that results from studies investigating these phenotypes are interpreted correctly with an awareness of the mechanisms by which genotype-phenotype associations can arise.

Statistical methods to estimate genetic associations from unrelated individuals often assume no unmeasured population stratification, dynastic effects or assortative mating. Where these structures exist and are insufficiently controlled for, estimates of genetic associations will be biased due to hidden correlations in the data and incorrectly attributed to genetic effects.[29] To empirically explore the mechanisms described above, we performed a set of analyses using the example of educational achievement, SEP and cognitive ability in a UK birth cohort, The Avon Longitudinal Study of Parents and Children (ALSPAC). To demonstrate that results are not driven by genotyping errors or other biases, we also present results for C-reactive protein (CRP), a biomarker of inflammation that is associated with a range of complex diseases, as a negative control analysis.[64] CRP is a biomarker of inflammation that is associated with a range of complex diseases, is rarely observed in younger people and is unlikely to be influenced by assortative mating, dynastic effects or population stratification. Systematic population differences (population stratification) in CRP have been found to be insubstantial;[65,66] parental phenotypic effects of CRP are unlikely to influence offspring CRP (dynastic effects); and parents are very unlikely to selectively mate based upon CRP (assortative mating). First, we present univariate heritability and genetic correlation estimates for our phenotypes. Second, we use bivariate heritability as a measure of genetic influence on phenotypic similarity between phenotypes and estimate this for each phenotype-pair. Finally, we present results from a range of analyses designed to assess the presence of bias due to population stratification, dynastic effects and assortative mating.

Figure 1: Causal models of structures underlying genetic associations. Population stratification due to ancestral differences (yellow lines); Dynastic effects (red lines); and Assortative mating (green line). Arrows represent direction of effect; nondirected lines represent simultaneous assortment.



Box 1: Structures which can induce associations between an individual's genotype and their phenotype.

1. **Variant substitution effects**

Variant substitution effects can be thought of as the (counterfactual) change in an individual's phenotype that would occur as a result of changing that individual's genotype from conception (holding all else constant). They can be estimated for a single phenotype (univariate genetic association) or pairs of phenotypes (bivariate genetic association, genetic correlation or Mendelian randomization). In most cases this type of effect is the target of any genotype-phenotype association analysis.

2. **Population stratification**

Population stratification refers to confounding introduced to associations between population structure and phenotype by systematic differences in allele frequencies across sub-populations. This arises from ancestry differences due to non-random mating and subsequent genetic drift of allele frequencies between sub-population groups, historically caused by geographic and physical boundaries. If phenotypes also differ systematically between sub-populations, population stratification can lead to genotype-phenotype associations despite no causal relationship between the genotype and the phenotype.

3. **Dynastic effects**

Further to influencing offspring phenotype through genetic inheritance, parental genotype can indirectly influence offspring phenotype through its expression in the parental phenotype. Where this occurs, offspring may inherit both phenotype associated SNPs and phenotype associated environments from parents, leading to biased genetic associations.[8] For example, SNPs positively associated with education in the parent's generation may lead to the creation of educational rich environments (such as an increase in books in the household) which will have a positive impact upon the child's educational attainment. Here, a variant substitution effect in the parent is inducing confounding at the level of the individuals being studied. Dynastic effects refer to this 'inheritance' of environment in addition to genotype.

4. **Assortative mating**

Assortative mating refers to the process by which spouses select each other based upon certain phenotypic characteristics. If selected phenotypes have a genotypic component, then phenotypic selection induces greater genetic similarity between spouses than in the general population. The correlations that are induced between genotype and phenotype by phenotypic assortment will lead to biased estimates of the causal effect of genotype on phenotype in in subsequent generations.[32,34,33] While offspring inheritance of genotype is random conditional upon parent's genotype, assortative mating induces non-random inheritance patterns across groups based on phenotype.

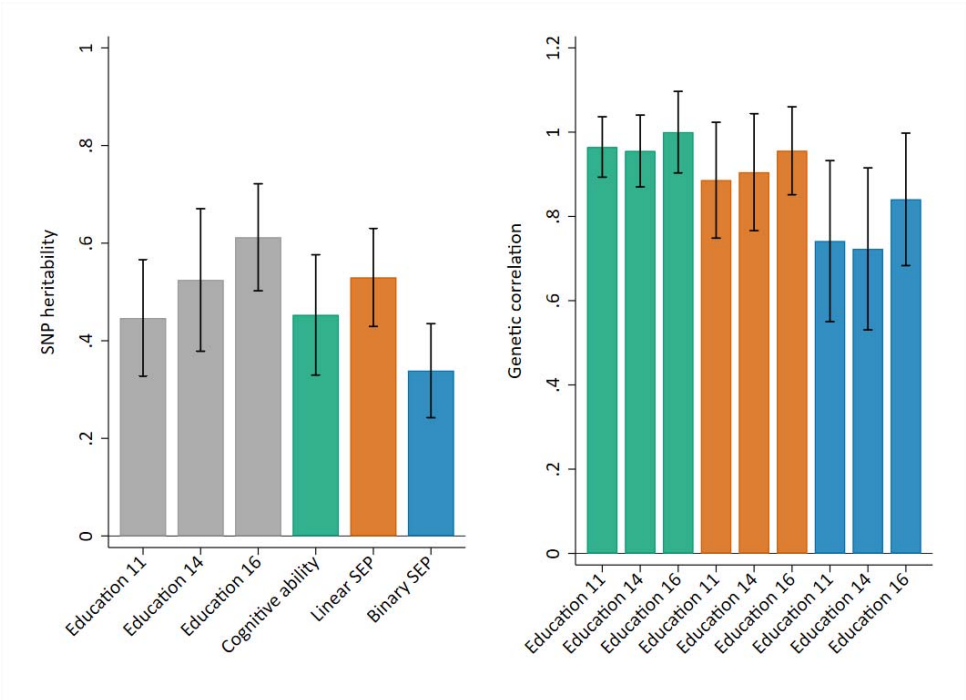
Results

Whole-genome estimates of genotype-phenotype associations for socioeconomic traits

Univariate phenotypic heritability

To investigate if and how genotype-phenotype associations may be biased we began by inferring the total contribution of all SNPs to the phenotypic variance, assuming an infinitesimal model of genetic architecture[12]. The SNP heritability of educational achievement increased with age from 44.7% (95%CI: 32.7 to 56.6) at age 11 to 52.5% (95%CI: 37.8 to 67.0) at age 14 and 61.2% (95%CI: 50.2 to 72.2) at age 16 (Figure 2a). The heritability of cognitive ability was estimated at 45.2% (95% CI: 33.0 to 57.6), and the heritability of SEP was estimated to be higher for a linear measure (53.0%; 95% CI: 42.9 to 63.0) than a binary measure (33.9%; 95% CI: 24.2 to 43.5).

Figure 2: SNP heritability and genetic correlations between phenotypes. Panel A: Grey bars represent educational achievement measured as exam point scores at ages 11, 14 and 16; green bar represents cognitive ability measured at age 8; orange bar represents a linear measure of socioeconomic position (SEP) measured as highest parental score on the Cambridge Social Stratification Score; blue bar represents a binary measure of SEP measured as "advantaged" for the highest two categories of Social Class based on Occupation and "disadvantaged" for the lower four categories. Panel B: Green bars represent genetic correlations between educational achievement at ages 11, 14 and 16 with cognitive ability measured at age 8; orange bars represent genetic correlations between educational achievement at ages 11, 14 and 16 with linear SEP; blue bars represent genetic correlations between educational achievement at ages 11, 14 and 16 with binary SEP. All analyses include adjustment for the first 20 principal components of population stratification. Parameter estimates in supplementary Tables S1 and S2.

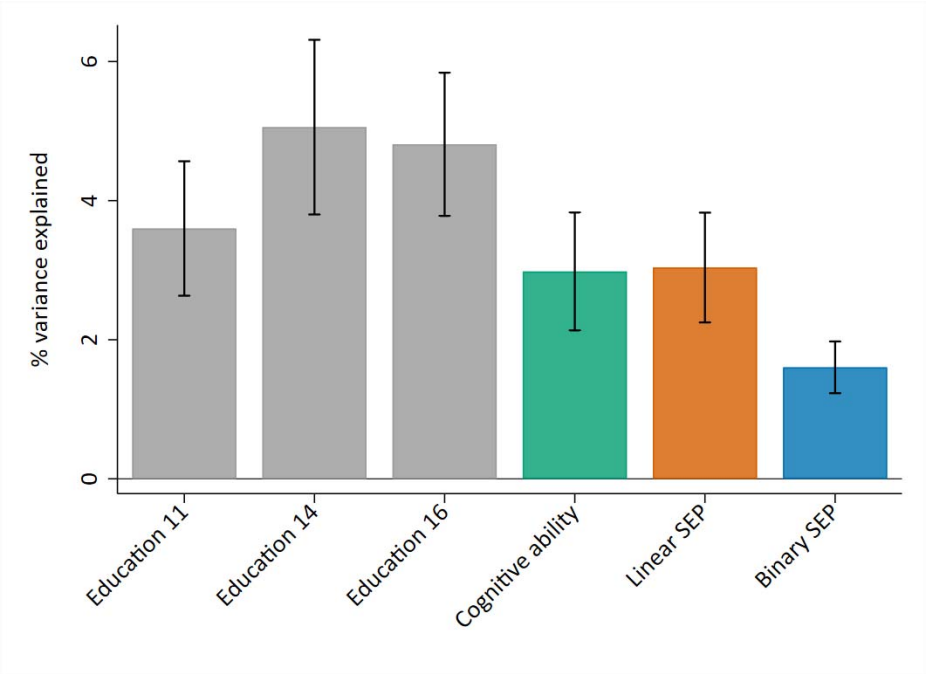


Genetic correlation

We next estimated genetic correlations between each phenotype-pair to infer the extent to which genetic effects were shared across phenotypes. Genetic correlations between educational achievement and cognitive ability were high and persisted throughout childhood within the range of 0.96 to 1 (Figure 2b). This suggests that the majority of SNPs which associate with educational achievement also associate with cognitive ability. Genetic correlations between educational achievement and SEP were also high: For the linear measure they ranged from 0.89 (95% CI: 0.75 to 1.02) to 0.96 (95% CI: 0.85 to 1.06), and for the binary measure from 0.76 (95% CI: 0.57 to 0.95) to 0.87 (95% CI: 0.71 to 1.04). The genetic correlations suggest that many SNPs which associate with educational achievement also associate with family SEP. These results were not driven by genotyping or imputation method (supplementary Tables S1-S4).

As a sensitivity analysis we estimated the amount of variance in each phenotype that could be explained by a polygenic score for educational achievement built from the largest GWAS of educational attainment to date (with the ALSPAC sample excluded).[38] This explained between 3.6% and 5.1% of the variation in educational achievement, 3.0% in cognitive ability and the linear measure of SEP, and 1.6% in the binary measure of SEP (Figure 3). That the polygenic score explains a similar amount of variation in the linear measure of SEP as educational achievement suggests a modest amount of pleiotropy in the SNPs used in the score, underscoring the high genetic correlations.

Figure 3: Variance explained in phenotypes by the educational achievement polygenic score. Polygenic score constructed from SNPs associated with education at $p<5x10^{-8}$. Grey bars represent educational achievement measured as exam point scores at ages 11, 14 and 16; green bar represents cognitive ability measured at age 8; orange bar represents a linear measure of family socioeconomic position (SEP) measured as highest parental score on the Cambridge Social Stratification Score; blue bar represents a binary measure of family SEP measured as "advantaged" for the highest two categories of Social Class based on Occupation and "disadvantaged" for the lower four categories. Standard errors were obtained through bootstrapping with 1000 replications. All analyses include adjustment for the first 20 principal components of population stratification. Parameter estimates in supplementary Table S5.

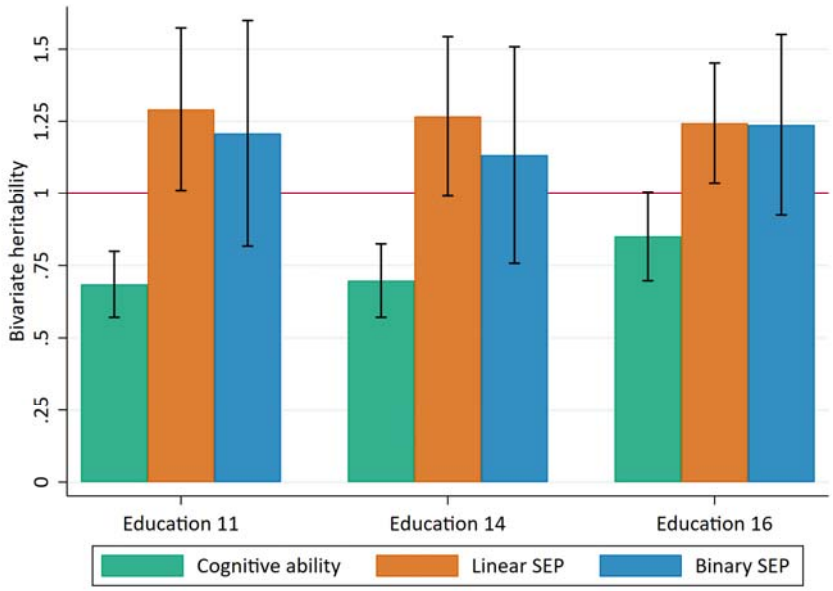


Bivariate SNP heritability

While genetic correlation estimates the correlation between the effects of SNPs on two phenotypes, it provides no information of how important genotype effects for one phenotype are for phenotypic differences in another. Bivariate heritability, which estimates the proportion of phenotypic correlation between two traits that can be attributed to genotype (calculated as $h^2_{AB} = \frac{r_g\sqrt{h^2_A}\sqrt{h^2_B}}{r_p}$), can be used to infer this. The bivariate heritabilities of educational achievement and cognitive ability range from 0.69 (SE: 0.06) at age 11 to 0.85 (SE: 0.08) at age 16 (Figure 4, supplementary Table S6). At face value this suggests that over two thirds of the phenotypic similarity between educational achievement and cognitive ability can be explained by shared common genetic variation in our sample. The bivariate heritabilities for educational achievement and SEP were estimated at greater than one for both the linear and binary measures (Figure 4, supplementary Table S6), and the standard errors suggest that this is not solely due to estimation imprecision. Bivariate heritability estimates greater than one are mathematically plausible because they are a ratio of two terms in which the numerator is not completely nested within the denominator. It is possible that bivariate heritability estimates above 1 may be an unbiased reflection of negative confounding caused by an

environmental factor, but this would require strong effects (see supplementary material). Bivariate heritabilities greater than one can therefore be taken as an indicator that estimates of univariate heritabilities or genetic correlation may have been biased, leading to overestimation of the genetic parameter ($r_g\sqrt{h_A^2}\sqrt{h_B^2}$). This information would not be obtained from genetic correlation estimates, demonstrating the usefulness of bivariate heritability for identifying the presence of bias due to population phenomena. We now investigate how these population phenomena may have biased our estimates.

Figure 4: Bivariate heritabilities between educational achievement, cognitive ability and socioeconomic position (SEP). Green bars represent cognitive ability measured at age 8; orange bars represent a linear measure of family SEP measured as highest parental score on the Cambridge Social Stratification Score; blue bars represent a binary measure of family SEP measured as "advantaged" for the highest two categories of Social Class based on Occupation and "disadvantaged" for the lower four categories. Educational achievement measured as exam point scores at ages 11, 14 and 16. Standard errors were obtained through simulations (see Data and methods). All analyses include adjustment for the first 20 principal components of population stratification. Parameter estimates in supplementary Table S6.



Exploring potential mechanisms of estimate inflation

Population stratification

Comparison of heritability estimates between models that omit and include the first 20 principal components indicates that bias due to population stratification as measured by the principal components is likely to be low (Table 1). The standard errors are relatively large and there is little evidence of differences in the heritability point estimates after additionally adjusting for the principal components. It is important to note that while adjustment for the first twenty principal components is unlikely to have removed all population stratification bias,[19] these results suggest that bias due to population stratification is likely to be low.

Table 1: SNP heritability estimates of phenotypes before and after adjusting for the first 20 principal components of ancestry. Standard errors in parentheses.

	Unadjusted	Population stratification adjusted
EA age 11	0.498 (0.056)	0.451 (0.064)
EA age 14	0.545 (0.069)	0.509 (0.079)
EA age 16	0.607 (0.052)	0.605 (0.059)
Cognitive ability	0.453 (0.058)	0.417 (0.066)
Linear SEP	0.568 (0.047)	0.547 (0.054)
Binary SEP	0.372 (0.046)	0.337 (0.052)

EA, educational achievement; SEP, socioeconomic position.

Dynastic effects

Table 2 shows associations between offspring education polygenic scores and educational achievement at age 16 before and after adjustment for parental polygenic scores, based on a sample of 1,095 mother-father-offspring trios. In the unadjusted model, a one standard deviation higher educational achievement polygenic score built from all SNPs is associated with a 0.340 (SE: 0.028) standard deviation higher achievement at age 16. After adjustment for parental polygenic scores this is attenuated to 0.223 (SE: 0.041), an attenuation of 34.4%. Using polygenic scores built only from SNPs that reached genomewide significance, the association of polygenic scores and educational achievement attenuated by 60.5% after adjustment for parental polygenic scores. Furthermore, parental genome-wide education polygenic scores remained associated with their child's education achievement conditional on the child's polygenic score, suggesting the presence of dynastic effects or assortative mating. Our negative control analyses of CRP based on 942 mother-father-offspring trios showed that a one standard deviation higher CRP polygenic score was associated with a 0.219 (SE: 0.030) standard deviation higher level of CRP. After adjustment for parental CRP polygenic scores this is attenuated to 0.192 (SE: 0.043), an attenuation of 12.4%. Neither the maternal nor paternal CRP polygenic were associated with offspring phenotypic CRP conditional on offspring CRP polygenic score, consistent with no dynastic effects for CRP as would be expected for such a biological phenotype.

Table 2: Associations between child and parent polygenic scores (PGS) with phenotypes (education n=1,095 trios; CRP n=942 trios). Standard errors in parentheses. Independent associations represent regression models with only a single PGS variable included, adjusted associations represent regression models with all 3 PGS variables. All models control for the first 20 principal components of ancestry. *P* values obtained from tests of seemingly unrelated regression on the child PGS coefficients between independent and co-adjusted models. Standard errors for attenuation were obtained through bootstrapping with 1000 replications.

	Independent associations	Adjusted associations	*p* value for difference	Attenuation in child PGS coefficient
Education				
Education PGS all SNPs				
Child PGS	0.340 (0.028)	0.223 (0.041)	<0.001	34.4% (9.3)

Mother PGS	0.261 (0.029)	0.110 (0.036)		
Father PGS	0.232 (0.030)	0.099 (0.034)		
Education PGS GWAS SNPs				
Child PGS	0.129 (0.029)	0.051 (0.042)	0.009	60.5% (35.0)
Mother PGS	0.072 (0.030)	0.035 (0.036)		
Father PGS	0.142 (0.029)	0.111 (0.036)		
CRP				
CRP GWAS SNPs				
Child PGS	0.219 (0.033)	0.192 (0.043)	0.330	12.4% (13.3)
Mother PGS	0.079 (0.033)	-0.002 (0.037)		
Father PGS	0.144 (0.032)	0.056 (0.037)		

PGS, polygenic score; CRP, C-Reactive protein.

Assortative mating

Table 3 demonstrates phenotypic and genotypic correlations for all available parental spouse pairs in the ALSPAC cohort. Phenotypic spousal correlations were positive for all phenotypes and similar to those estimated in other studies (cf. 0.41[10], 0.62[67] and 0.66[68]). This provides evidence of phenotypic assortative mating on both education and SEP between ALSPAC parents. To test if this phenotypic sorting induced genetic correlations between spouses, we examined genetic correlations between spouses based upon education polygenic scores. Positive correlations were observed between spouse-pairs for both polygenic scores, suggesting that the observed phenotypic assortment induced genetic assortment and that assortative mating likely contributed to bias in heritability estimates of educational achievement amongst offspring.[9] Turning to the negative control analysis, the spousal phenotypic correlation for CRP was 0.004 (0.030), and the spousal correlation of the CRP polygenic score was -0.009 (0.027). These results contrast to the spousal correlations on the social variables and imply no assortment on CRP.

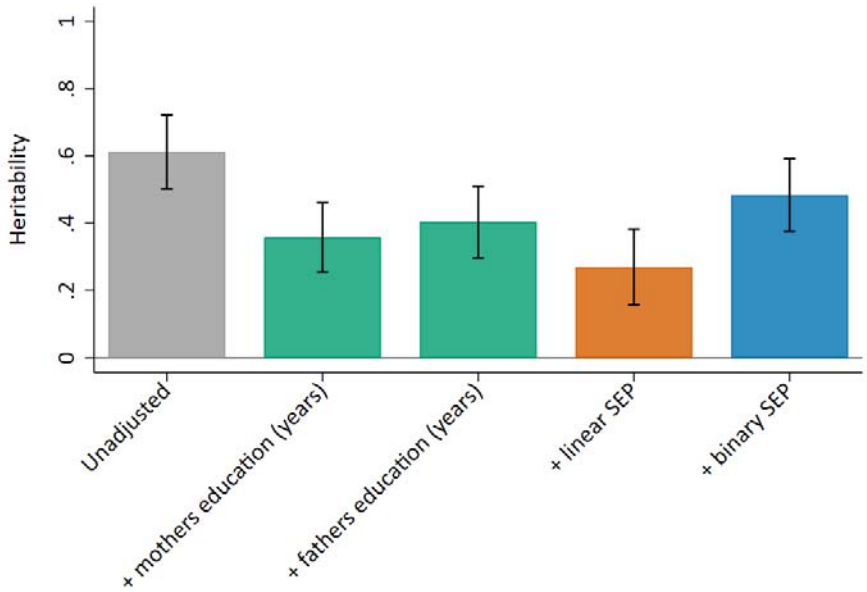
Table 3: Phenotypic and genotypic correlations between spouses. Standard errors in parentheses.

	Spouses	n
Phenotype		
Highest educational achievement	0.560 (0.011)	5,353
Linear socioeconomic position	0.434 (0.010)	6,858
Binary socioeconomic position	0.297 (0.008)	5,737
CRP	0.004 (0.030)	1,129
Genotype		
Education genetic score all SNPs	0.181 (0.025)	1,262
Education genetic score GWAS SNPs	0.080 (0.027)	1,262
CRP genetic score GWAS SNPs	-0.009 (0.027)	1,385

CRP, C-Reactive protein.

To further explore the potential impact of assortative mating and dynastic effects on our results we conducted additional sensitivity analyses controlling for parents' years of education and SEP. This approach still assumes no assortative mating or dynastic effects, but inconsistency between the main results and these sensitivity analyses provide an indication of bias in heritability due to these processes. The results of these analyses (Figure 4) demonstrate that the heritability of educational achievement at age 16 is greatly attenuated – by around half - when parental education or SEP is controlled for. This suggests that differences in educational achievement which are associated with common genetic variation can in part be explained by assortative mating, dynastic effects, or a combination of both. When these sensitivity analyses were applied to genetic correlation estimates between education and SEP, the impact of these biasing mechanisms was less clear, reflecting greater estimation imprecision (Figure S1 in supplementary material).

Figure 4: Heritability of educational achievement adjusting for parental socioeconomic variables. Grey bar represents the estimated heritability of educational achievement measured at age 16; green bars represent heritability adjusted for mothers and fathers' years of education; orange bar represents heritability adjusted for linear socioeconomic position (SEP); blue bar represents heritability adjusted for binary socioeconomic position.



Discussion

By analysing genetic contributions to socioeconomic phenotypes alongside a wide set of sensitivity analyses we have demonstrated how population phenomena can bias estimates of genetic contributions to complex social phenotypes from samples of unrelated individuals. The presence of genetic association does not necessarily imply a variant substitution effect solely giving rise to genotype-phenotype associations, but may reflect confounding by underlying population phenomena including population stratification, assortative mating and dynastic effects. These results importantly demonstrate that analyses using samples of unrelated individuals may not provide estimates of heritability or genetic correlation that are driven solely by causal genotype-phenotype relationships, and this likely reflects mechanisms influencing genome-wide association studies also.

Our results add to the growing body of evidence that estimates drawn from samples of unrelated individuals may overestimate heritability or genetic correlation[8,9,11] and bias Mendelian randomization studies.[20] Social phenotypes such as education and SEP, which are complex, highly assortative and dynastic, appear to be particularly susceptible to bias from population phenomena. It is therefore important that studies within the rapidly growing area of sociogenomic research[69] test for these phenomena using the methods we highlight and where possible draw upon data from family based studies. Estimating the attenuation of offspring polygenic scores from parental polygenic scores can help to identify dynastic effects; spousal correlations can provide information on the presence of assortative mating; and bivariate heritability can be used to identify overestimation in genetic parameters as a result of these phenomena.

Our SNP heritability estimates of educational achievement were higher than those previously estimated from a different UK cohort at around 25% at age 7 to 40% at age 16, though the confidence intervals between the estimates from the two studies overlap.[70] Differences in heritability estimates cannot be taken as evidence of misestimation though as they are relevant to a specific population at a specific time.[71] These SNP heritability estimates are higher than those for educational attainment,[38,72–74] which may reflect differences in the heritability of attainment and achievement. SNP heritabilities of attainment and achievement have not yet been estimated in the same sample but comparing different samples, achievement at the end of schooling has been estimated higher (SNP heritability: 0.4)[70] than lifetime attainment (SNP heritability: 0.2).[38] This discrepancy may reflect sample differences. and future research is required in samples with both attainment and achievement measured. Previous studies have highlighted that SNP heritability estimates are biased by family effects,[8,11] and these issues may have inflated our estimates. However, the strength of bias may be smaller for education test scores (achievement) which are likely to capture a more cognitive aspect of educational performance than the more social aspect of education that years of education (attainment) captures. As has been discussed previously,[75] the high heritabilities we observed may also reflect genuine differences due to the spatiotemporal homogeneity of the ALSPAC cohort. The mechanisms we investigated may also have larger effects in the ALSPAC study as a regional cohort than in other data samples; the impact of these mechanisms on more geographically dispersed studies such as UK Biobank is currently unknown.

Our estimates of the SNP heritability of cognitive ability (41.7%) and SEP (linear: 54.7%; binary: 33.7%) were broadly similar to educational achievement, and also exceeded those in previous studies of 29% for cognitive ability[50] and 20% for SEP.[49–51] That heritability was higher in the linear measure than the binary measure of SEP may reflect our cut-point in determining 'high' vs 'low' for the binary classification or genuine differences between the two measures. The estimates of proportion of variation in all phenotypes explained by the educational attainment polygenic score was broadly consistent with previous research[41] which used a polygenic score from an earlier GWAS.[52] Estimated genetic correlations between educational achievement, SEP and cognitive ability were consistent with findings from other cohorts[43,49,50] but with greater statistical precision due to larger sample sizes and the precision of GCTA over other methods.[76] Further research is required to investigate how these genetic associations persist into further and higher education.

Attenuation of genetic associations between children's polygenic score and educational achievement was between one and two thirds after controlling for both parents' polygenic scores, supporting the presence of dynastic effects whereby parental genotype indirectly affects offspring phenotype. Furthermore, both parents' scores remained robust predictors of children's achievement over and above the child's polygenic score. Phenotypic spousal correlations demonstrated strong evidence of parental assortative mating on educational attainment (r=0.56) and SEP (r=0.43), which

induced genetic correlations at education associated loci of r=0.18. Heritability estimates of educational achievement were attenuated by roughly half when parental education or SEP was controlled for. This supports bias in heritability estimates due to assortative mating and/or dynastic effects in ALSPAC. We found no strong evidence that our estimates were biased by population stratification as measured by the genetic principal components, but this may reflect the inability of genetic principal components to capture subtle population structure[77] rather than adequately control it.[13,19] It is also possible that our high estimates reflected the relatively homogenous educational environment experienced by the ALSPAC cohort when compared to previous studies. Environmental homogeneity increases the proportion of variation that can be attributed to genetic effects, and the ALSPAC children were all born within three years and mostly experienced the same school system within the same region of the UK. Our negative control analyses provided little evidence of dynastic effects or assortative mating for CRP in our sample. While this is expected, it strengthens confidence that the dynastic effects and assortative mating we observe for education are robust and do not arise from other issues such as genotyping errors.

Several limitations must be acknowledged in this study. First, measurement error on the phenotypes may have influenced our results. Genotyping accuracy and strict quality controls on the genetic data and educational achievement taken from administrative records should result in insufficient measurement error in these phenotypes to meaningfully bias our estimates. However, there may be some measurement inaccuracy in how well the education test scores capture underlying educational ability over and above test-retest reliability. Measurement error will be greater for SEP as these measures relied on self-reported data, but this would have to be differential and patterned to bias estimates (independent non-differential measurement error will only reduce statistical precision of the estimates, not bias them). Second, further residual population structure in the ALSPAC genetic relatedness matrix not captured by the principal components could bias our results.[19] We controlled for the first twenty principal components of population structure in our full analyses, but this is unlikely to account for all differences. Another possible source of bias in our study is that of shared environmental factors[78] due to schooling. Many children within our sample will attend the same schools and therefore share the same schooling environment. Because school choice in the UK is socioeconomically patterned,[79] correlations may be induced between parental SEP and school environment which would be attributed to additive genetic variation (i.e. genetic nurture effects). Recent research has demonstrated the importance of geography as a source of bias in heritability studies[80] and because we use a heavily geographically clustered cohort this may bias our heritability estimates. Third, the definition of educational attainment used in the GWAS to conduct the polygenic score was years of education, which is relatively crude and does not discriminate academic performance within each additional year of education. It is therefore possible that the score we use is capturing a social rather than performance aspect of education. Fourth, GREML assumes that causal SNPs have effects on phenotypes that are independent of linkage disequilibrium (LD) to other SNPs and minor allele frequency (MAF)[81]. Previous studies have demonstrated that violations to these assumptions can lead to biased SNP heritability and that multicomponent GREML methods (GREML-LDMS-R and GREML-LDMS-I) can obtain accurate SNP heritability estimates.[2,82] However, these extensions require much larger sample sizes to estimate than standard GREML approaches[82,83] and cannot be reliably estimated using our data. Further, the attenuation that we found due to population factors are in principle unrelated to these potential biases that arise due to genetic architecture assumptions. Therefore, while our revised estimates may be additionally biased due to modelling assumptions, it remains likely that that would occur in addition to the population level biases that we have described. Future studies on larger samples are required to test potential overestimation of SNP heritability for education and SEP using GREML-LDMS extensions. Finally, it is

possible that our estimates could have been biased by cryptic relatedness. To overcome this, we restricted our analytical sample to individuals with identity by descent less than 0.1, but it remains possible that some related participants will have been included. While data on mother-father-offspring trios provide opportunities to investigate the presence and strength of these mechanisms, mother-father-offspring-sibling quad approaches may offer further opportunities to test for heterogeneity in dynastic effects between siblings.

In conclusion, our results demonstrate some of the causal structures that may bias univariate and bivariate genetic estimates such as heritability and genetic correlations, particularly when applied to complex social phenotypes. Future studies may make use of the methodological tools that we highlight here to assess these alongside others.[8,9,11] Principally, family based study designs such as within-family[20], between-sibling[84], adoption[85] and half-sibling[86] will be better equipped to provide informative and accurate genetic associations given their robustness to population stratification, dynastic effects and assortative mating. Genetic studies investigating complex social relationships should be interpreted with care in light of these mechanisms, and results should be interpreted within a triangulation framework that considers the wider context of existing evidence.[87]

Data and methods

Study sample

Participants were children from the Avon Longitudinal Study of Parents and Children (ALSPAC). Pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992 were invited to take part in the study. The initial number of pregnancies enrolled was 14,541. When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. This additional recruitment resulted in a total sample of 15,454 pregnancies, resulting in 14,901 children who were alive at one year of age. From these there is genetic data available for 7,748 children on at least one of educational achievement, SEP and cognitive ability after quality control and removal of related individuals (see *Genetic data* below). For full details of the cohort profile and study design see [88,89]. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool at http://www.bristol.ac.uk/alspac/researchers/our-data/. The ALSPAC cohort is largely representative of the UK population when compared with 1991 Census data; there is under representation of some ethnic minorities, single parent families, and those living in rented accommodation.[88] Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004). We use the largest available samples in each of our analyses to increase precision of estimates, regardless of whether a child contributed data to the other analyses. Descriptive statistics for the raw variables used in the analyses and the phenotypic differences between ALSPAC participants included in our analyses and those who were excluded due to missing data genotype or phenotype data are in supplementary Table S7. Compared to participants who were excluded due to missing data, those included in the analyses had higher achievement at each stage of education, higher cognitive ability as measured at age 8, and came from higher SEP families as measured on both linear and binary.

Genetic data

DNA of the ALSPAC children was extracted from blood, cell line and mouthwash samples, then genotyped using references panels and subjected to standard quality control approaches. ALSPAC children were genotyped using the Illumina HumanHap550 quad chip genotyping platforms by 23andme subcontracting the Wellcome Trust Sanger Institute, Cambridge, UK and the Laboratory Corporation of America, Burlington, NC, US. ALSPAC mothers were genotyped using the Illumina human660W-quad array at Centre National de Génotypage (CNG) and genotypes were called with Illumina GenomeStudio. ALSPAC fathers and some additional mothers were genotyped using the Illumina HumanCoreExome chip genotyping platforms by the ALSPAC lab and called using GenomeStudio. All resulting raw genome-wide data were subjected to standard quality control methods in PLINK (v1.07). Individuals were excluded on the basis of gender mismatches; minimal or excessive heterozygosity; disproportionate levels of individual missingness (>3%) and insufficient sample replication (IBD < 0.8). Population stratification was assessed by multidimensional scaling analysis and compared with Hapmap II (release 22) European descent (CEU), Han Chinese, Japanese and Yoruba reference populations; all individuals with non-European ancestry were removed. SNPs with a minor allele frequency of < 1%, a call rate of < 95% or evidence for violations of Hardy-Weinberg equilibrium (P < 5E-7) were removed. Cryptic relatedness was assessed using an Identity by descent (IBD) estimate of more than 0.125 which is expected to correspond to roughly 12.5% alleles shared IBD or a relatedness at the first cousin level. Related participants that passed all other quality control thresholds were retained during subsequent phasing and imputation. For the

mother's, samples were removed where they had indeterminate X chromosome heterozygosity or extreme autosomal heterozygosity. After quality control, 9,115 participants and 500,527 SNPs for the children, 9,048 participants and 526,688 SNPs for the mothers, and 2,201 participants and 507,586 SNPs for the fathers (and additional mothers) passed these quality control filters.

We combined 477,482 SNP genotypes in common between the samples. We removed SNPs with genotype missingness above 1% due to poor quality and removed participants with potential ID mismatches. This resulted in a dataset of 20,043 participants containing 465,740 SNPs (112 were removed during liftover and 234 were out of HWE after combination). We estimated haplotypes using ShapeIT (v2.r644) which utilises relatedness during phasing. The phased haplotypes were then imputed to the Haplotype Reference Consortium (HRCr1.1, 2016) panel of approximately 31,000 phased whole genomes. The HRC panel was phased using ShapeIt v2, and the imputation was performed using the Michigan imputation server. After imputation and filtering on MAF >0.01 and info > 0.8 there were 7,191,388 SNPs. This gave 8,237 eligible children, 8,675 eligible mothers and 1,722 eligible fathers with available genotype data after exclusion of related participants using cryptic relatedness measures described previously.

Educational achievement

We use average fine graded point scores at the three major Key Stages of education in the UK at ages 11, 14 and 16. Point scores were obtained from the Key Stage 4 (age 16) database of the UK National Pupil Database (NPD) through data linkage to the ALSPAC cohort. The NPD represents the most accurate record of individual educational achievement available in the UK. The Key Stage 4 database provides a larger sample size than the earlier two Key Stage databases and contains data for each. Fine graded point scores provide a richer measure of a child's achievement than level bandings and were therefore chosen as the most accurate method of determining academic achievement during compulsory schooling.

Parental socioeconomic position

We use two measures of parental SEP; a binary classification based on the widely used Social Class based on Occupation (formerly Registrar Generals Social Class) of "high" [I and II] vs "low" [III-Non-manual, III-Manual, IV, and V] social classes and a continuous classification based upon the Cambridge Social Stratification Score (CAMSIS). Social Class based on Occupation assumes within-strata social homogeneity with clear boundaries, while CAMSIS provides a more flexible measure that accounts for social heterogeneity.[90]

Cognitive ability

Cognitive ability was measured during the direct assessment at age eight using the short form Wechsler Intelligence Scale for Children (WISC) from verbal, performance, and digit span tests[91] and administered by members of the ALSPAC psychology team overseen by an expert in psychometric testing. The short form tests have high reliability[92] and the ALSPAC measures utilise subtests with reliability ranging from 0.70 to 0.96. Raw scores were recalculated to be comparable to those that would have been obtained had the full test been administered and then age-scaled to give a total overall score combined from the performance and verbal subscales.

Educational attainment polygenic score

To test dynastic effects we used an educational attainment polygenic score with the 1,271 independent SNPs identified to associate with years of education at genome-wide levels of significance ($p<5e^{-8}$) in GWAS[38] using the software package PRSice.[93] PRSice was used to thin SNPs according to linkage disequilibrium through clumping, where the SNP with the smallest *P*-value in

each 250kb window was retained and all other SNPs in linkage disequilibrium with an r^2 of >0.1 were removed). This score was generated using GWAS results which had removed ALSPAC and 23andMe participants from the meta-analysis. In the GWAS the score using the 1,271 genome-wide significant SNPs explained 2.5-3.8% of the variation in educational attainment in the two prediction cohorts.

Negative control analyses

To ensure that our analyses were correctly identifying the biasing mechanisms that we outline and did not represent other biasing factors such as genotyping errors, we ran sensitivity analyses using C-reactive protein (CRP) as a negative control phenotype. CRP is a biomarker of inflammation that is associated with a range of complex diseases and is unlikely to be influenced by the biasing mechanisms that relate to the social phenotypes that we investigate. Offspring CRP was measured from non-fasting blood assays taken during direct assessment when the children were aged 9.

Statistical analysis

We estimate SNP heritability (hereafter referred to as heritability) using generalized restricted maximum likelihood (GREML) in the software package GCTA.[12] GCTA uses measured SNP level variation across the whole genome to estimate the proportion of variation in educational achievement, SEP, and cognitive ability that can be explained by common genetic variation. We use a series of univariate analyses of the form:

$$y = X\beta + g + \epsilon \#(1)$$

where y is the phenotype, X is a series of covariates, g is a normally distributed random effect with variance σ_g^2, and ϵ is residual error with variance σ_ϵ^2. Heritability is defined as the proportion of total phenotypic variance (genetic variance plus residual variance) explained by common genetic variation:

$$\frac{\sigma_g^2}{\sigma_g^2 + \sigma_\epsilon^2} \#(2)$$

Where genetically similar pairs are more phenotypically similar than genetically dissimilar pairs heritability estimates will be non-zero.

We estimate genetic correlations using bivariate GCTA, running nine sets of analyses between educational achievement at each age and each of linear SEP, binary SEP, and cognitive ability. Genetic correlation quantifies the extent to which SNPs that associate with one phenotype (i.e. educational achievement) also associate with another phenotype (i.e. cognitive ability). It therefore refers to the correlation of all genetic effects across the genome for phenotypes A and B, and is estimated as:

$$r_g = \frac{cov_g(A, B)}{\sqrt{var_g(A) var_g(B)}} \#(3)$$

Where r_g is the genetic correlation between phenotypes A and B, $var_g(A)$ is the genetic variance of phenotype A and $cov_g(A, B)$ is the genetic covariance between phenotypes A and B. Genetic correlations can indicate that two phenotypes are influenced by the same SNPs (i.e. have shared genetic architecture). In contrast, the bivariate heritability is the proportion of the phenotypic correlation that can be explained by the genotypes. Genetic correlations and bivariate heritability are likely to differ. For example, two phenotypes may be highly genetically correlated, but if they have low heritability then the bivariate heritability will be low. Bivariate heritability estimates the proportion of phenotypic correlations that can be explained by genetics. It is estimated as:

$$h_{AB}^2 = \frac{r_g\sqrt{h_A^2}\sqrt{h_B^2}}{r_p} \#(4)$$

Where r_g is the genetic correlation between phenotypes A and B, h_A^2 and h_B^2 are the heritabilities of phenotype A and B respectively, and r_p is the phenotypic correlation between phenotypes A and B. GCTA does not directly estimate coheritability or bivariate heritability terms and therefore cannot be used to estimate standard errors. Bivariate heritability estimates were derived from entering the GCTA estimated components into equation 4 above. Standard errors for the bivariate heritability estimates were estimated using simulations. First, we simulated one million observations for each input parameter (the two heritability terms, genetic correlation and phenotypic correlation) as normally distributed with a mean value corresponding to the point estimate and standard deviation corresponding to the standard error obtained from GCTA. Bivariate heritability was estimated for each observation and the standard deviation of all one million estimates taken as the estimated error for the point estimate. This approach to calculating standard errors assumes no covariance between the input parameters. Non-zero covariances would produce smaller standard errors and therefore this approach can be considered to provide conservative estimates.

We used data for unrelated participants, as indicated by the ALSPAC Genetic Relatedness Matrices (GRM). Population stratification is controlled for by using the first 20 principal components of inferred population structure as covariates in analyses. Continuous variables were inverse normally transformed to have a normal distribution, a requirement of GCTA.

References

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

2. Evans, L. M. *et al.* Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat. Genet.* (2018). doi:10.1038/s41588-018-0108-x

3. Belsky, D. W., Moffitt, T. E. & Caspi, A. Genetics in population health science: Strategies and opportunities. *Am. J. Public Health* (2013). doi:10.2105/AJPH.2012.301139

4. Fisher, R. *The Genetical Theory of Natural Selection*. (Oxford University Press, 1930).

5. Lee, J. J. & Chow, C. C. The causal meaning of Fisher's average effect. *Genet. Res. (Camb).* (2013). doi:10.1017/S0016672313000074

6. Pearl, J. *Causality: models, reasoning and inference*. **29**, (Springer, 2000).

7. Lao, O. *et al.* Correlation between Genetic and Geographic Structure in Europe. *Curr. Biol.* (2008). doi:10.1016/j.cub.2008.07.049

8. Kong, A. *et al.* The nature of nurture: Effects of parental genotypes. *Science (80-.).* **359**, 424–428 (2018).

9. Yengo, L. *et al.* Imprint of assortative mating on the human genome. *Nature Human Behaviour* (2018). doi:10.1038/s41562-018-0476-3

10. Robinson, M. R. *et al.* Genetic evidence of assortative mating in humans. *Nat. Hum. Behav.* **1**, 0016 (2017).

11. Young, A. I. *et al.* Relatedness disequilibrium regression estimates heritability without environmental bias. *Nat. Genet.* (2018). doi:10.1038/s41588-018-0178-9

12. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

13. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

14. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* (2015). doi:10.1038/ng.3406

15. Speed, D., Cai, N., Johnson, M. R., Nejentsev, S. & Balding, D. J. Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49**, 986–992 (2017).

16. Lawson, D. J. *et al.* Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Hum. Genet.* (2019). doi:10.1007/s00439-019-02014-8

17. Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* (2008). doi:10.1038/nature07331

18. Hamer, D. H. & Sirota, L. Beware the chopsticks gene. *Mol. Psychiatry* (2000). doi:10.1038/sj.mp.4000662

19. Haworth, S. *et al.* Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat. Commun.* **10**, 333 (2019).

20. Brumpton, B. *et al.* Within-family studies for Mendelian randomization: avoiding dynastic,

assortative mating, and population stratification biases. *bioRxiv* (2019). doi:10.1101/602516

21. Visscher, P. M. *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* (2006). doi:10.1371/journal.pgen.0020041

22. Hemani, G. *et al.* Inference of the genetic architecture underlying bmi and height with the use of 20,240 sibling pairs. *Am. J. Hum. Genet.* (2013). doi:10.1016/j.ajhg.2013.10.005

23. Davies, N. M. *et al.* Within family Mendelian randomization studies. *Hum. Mol. Genet.* (2019). doi:10.1093/hmg/ddz204

24. Black, S. E., Devereux, P. J. & Salvanes, K. G. Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *American Economic Review* (2005). doi:10.1257/0002828053828635

25. Hertz, T. *et al.* The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends. *B. E. J. Econom. Anal. Policy* (2008). doi:10.2202/1935-1682.1775

26. Bowles, S., Gintis, H. & Groves, M. O. *Unequal chances: Family background and economic success.* (Princeton University Press, 2009).

27. Magnus, P. *et al.* Cohort profile: The Norwegian Mother and Child Cohort Study (MoBa). *Int. J. Epidemiol.* (2006). doi:10.1093/ije/dyl170

28. Krokstad, S. *et al.* Cohort profile: The HUNT study, Norway. *Int. J. Epidemiol.* (2013). doi:10.1093/ije/dys095

29. Keller, M. C. *et al.* The Genetic Correlation between Height and IQ: Shared Genes or Assortative Mating? *PLoS Genet.* **9**, (2013).

30. Mare, R. D. Five Decades of Educational Assortative Mating. *Am. Sociol. Rev.* **56**, 15–32 (1991).

31. Vandenberg, S. G. Assortative mating, or who marries whom? *Behav. Genet.* **2**, 127–157 (1972).

32. Falconer, D. S. & Mackay, T. F. C. Introduction to quantitative genetics. *Introduction to quantitative genetics* (1996). doi:10.1016/j.bbabio.2008.04.029

33. Hartwig, F. P., Davies, N. M. & Davey Smith, G. Bias in Mendelian randomization due to assortative mating. *Genetic Epidemiology* (2018). doi:10.1002/gepi.22138

34. Buss, D. M. Human mate selection. *Am. Sci.* (1985).

35. Sebro, R. & Risch, N. J. A brief note on the resemblance between relatives in the presence of population stratification. *Heredity (Edinb).* (2012). doi:10.1038/hdy.2011.124

36. Belsky, D. W. *et al.* Genetic analysis of social-class mobility in five longitudinal studies. *Proc. Natl. Acad. Sci.* (2018). doi:10.1073/pnas.1801238115

37. Conley, D. *et al.* Is the Effect of Parental Education on Offspring Biased or Moderated by Genotype? *Sociol. Sci.* **2**, 82–105 (2015).

38. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* (2018). doi:10.1038/s41588-018-0147-3

39. Deary, I. J., Strand, S., Smith, P. & Fernandes, C. Intelligence and educational achievement.

Intelligence **35**, 13–21 (2007).

40. Branigan, A. R., Mccallum, K. J. & Freese, J. Variation in the heritability of educational attainment: An international meta-analysis. *Soc. Forces* **92**, 109–140 (2013).

41. Selzam, S. *et al.* Predicting educational achievement from DNA. *Mol. Psychiatry* (2017). doi:10.1038/mp.2016.107

42. Krapohl, E. *et al.* The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proc. Natl. Acad. Sci.* **111**, 15273–15278 (2014).

43. Davies, N. M., Hemani, G., Timpson, N. J., Windmeijer, F. & Davey Smith, G. The role of common genetic variation in educational attainment and income: evidence from the National Child Development Study. *Sci. Rep.* **5**, 16509 (2015).

44. Galobardes, B., Lynch, J. & Davey Smith, G. Measuring socioeconomic position in health research. *British Medical Bulletin* (2007). doi:10.1093/bmb/ldm001

45. Krieger, N., Williams, D. R. & Moss, N. E. Measuring Social Health Research: Concepts, Methodologies and Guidelines. *Annu. Rev. Public Health* (1997).

46. Morris, T., Dorling, D. & Davey Smith, G. How well can we predict educational outcomes? Examining the roles of cognitive ability and social position in educational attainment. *Contemp. Soc. Sci.* **11**, 1–15 (2016).

47. Bynner, J. & Joshi, H. Equality and opportunity in education: Evidence from the 1958 and 1970 birth cohort studies. *Oxford Rev. Educ.* **28**, 405–425 (2002).

48. Feinstein, L. Inequality in the early cognitive development of British children in the 1970 cohort. *Economica* **70**, 73–97 (2003).

49. Krapohl, E. & Plomin, R. Genetic link between family socioeconomic status and children/'s educational achievement estimated from genome-wide SNPs. *Mol Psychiatry* **21**, 437–443 (2016).

50. Marioni, R. E. *et al.* Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence* **44**, 26–32 (2014).

51. Trzaskowski, M. *et al.* Genetic influence on family socioeconomic status and children's intelligence. *Intelligence* **42**, 83–88 (2014).

52. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).

53. Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M. & Boomsma, D. I. Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Res.* **5**, 544–553 (2002).

54. Shakeshaft, N. G. *et al.* Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PLoS One* (2013). doi:10.1371/journal.pone.0080341

55. Miller, P., Mulvey, C. & Martin, N. Genetic and environmental contributions to educational attainment in Australia. *Econ. Educ. Rev.* (2002). doi:10.1016/s0272-7757(00)00018-2

56. Nielsen, F. Achievement and Ascription in Educational Attainment: Genetic and Environmental Influences on Adolescent Schooling. *Soc. Forces* (2007). doi:10.1353/sof.2006.0135

57. Tucker-Drob, E. M., Rhemtulla, M., Harden, K. P., Turkheimer, E. & Fask, D. Emergence of a gene × socioeconomic status interaction on infant mental ability between 10 months and 2 years. *Psychol. Sci.* **22**, 125–133 (2011).

58. Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B. & Gottesman, I. I. Socioeconomic status modifies heritability of IQ in young children. *Psychol. Sci.* **14**, 623–628 (2003).

59. Rindermann, H. & Neubauer, A. C. Processing speed, intelligence, creativity, and school performance: Testing of causal hypotheses using structural equation models. *Intelligence* **32**, 573–589 (2004).

60. Strenze, T. Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence* **35**, 401–426 (2007).

61. von Stumm, S. & Plomin, R. Intelligence Socioeconomic status and the growth of intelligence from infancy through adolescence. *Intelligence* **48**, 30–36 (2015).

62. Noble, K. G., McCandliss, B. D. & Farah, M. J. Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science* **10**, 464–480 (2007).

63. Duncan, G. J., Brooks-Gunn, J. & Klebanov, P. K. Economic Deprivation and Early Childhood Development. *Child Dev.* **65**, 296–318 (1994).

64. Lipsitch, M., Tchetgen Tchetgen, E. & Cohen, T. Negative Controls: A tool for detecting confounding and bias in observational studies. *Epidemiology* (2010). doi:10.1097/EDE.0b013e3181d61eeb

65. Okada, Y. *et al.* Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. *Hum. Mol. Genet.* (2011). doi:10.1093/hmg/ddq551

66. Ligthart, S. *et al.* Genome Analyses of >200,000 Individuals Identify 58 Loci for Chronic Inflammation and Highlight Pathways that Link Inflammation and Complex Disorders. *Am. J. Hum. Genet.* **103**, 691–706 (2018).

67. Heath, A. C., Eaves, L. J., Nance, W. E. & Corey, L. A. Social inequality and assortative mating: Cause or consequence? *Behav. Genet.* (1987). doi:10.1007/BF01066007

68. Abdellaoui, A. *et al.* Educational attainment influences levels of homozygosity through migration and assortative mating. *PLoS One* (2015). doi:10.1371/journal.pone.0118935

69. Freese, J. The Arrival of Social Science Genomics. *Contemp. Sociol.* **47**, 524–536 (2018).

70. Rimfeld, K. *et al.* The stability of educational achievement across school years is largely explained by genetic factors. *npj Sci. Learn.* **3**, 16 (2018).

71. Davey Smith, G. Epidemiology, epigenetics and the 'Gloomy Prospect': Embracing randomness in population health research and practice. *Int. J. Epidemiol.* **40**, 537–562 (2011).

72. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science (80-.).* **340**, 1467–1471 (2013).

73. de Vlaming, R. *et al.* Meta-GWAS Accuracy and Power (MetaGAP) Calculator Shows that Hiding Heritability Is Partially Due to Imperfect Genetic Correlations across Studies. *PLoS Genet.* (2017). doi:http://dx.doi.org/10.1371/journal.pgen.1006495

74. Benjamin, D. J. *et al.* The genetic architecture of economic and political preferences. *Proc. Natl. Acad. Sci.* (2012). doi:10.1073/pnas.1120666109

75. Morris, T. T., Davies, N. M., Dorling, D., Richmond, R. C. & Davey Smith, G. Testing the validity

of value-added measures of educational progress with genetic data. *Br. Educ. Res. J.* **0**, (2018).

76. Ni, G. *et al.* Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. *American Journal of Human Genetics* (2018). doi:10.1016/j.ajhg.2018.03.021

77. Leslie, S. *et al.* The fine-scale genetic structure of the British population. *Nature* (2015). doi:10.1038/nature14230

78. Tenesa, A. & Haley, C. S. The heritability of human disease: estimation, uses and abuses. *Nat. Rev. Genet.* **14**, 139–149 (2013).

79. Wilson, D. & Bridge, G. *School choice and equality of opportunity: an international systematic review*. (2019).

80. Heckerman, D. *et al.* Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc. Natl. Acad. Sci.* **113**, 7377 LP – 7382 (2016).

81. Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* **91**, 1011–1021 (2012).

82. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).

83. Visscher, P. M. *et al.* Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples. *PLoS Genet.* (2014). doi:10.1371/journal.pgen.1004269

84. Lawlor, D. A., Tilling, K. & Smith, G. D. Triangulation in aetiological epidemiology. *Int. J. Epidemiol.* **45**, 1866–1886 (2016).

85. Cheesman, R. *et al.* Comparison of adopted and non-adopted individuals reveals gene-environment interplay for education in the UK Biobank. *bioRxiv* (2019). doi:10.1101/707695

86. Kendler, K. S., Ohlsson, H., Sundquist, J. & Sundquist, K. Maternal half-sibling families with discordant fathers: A contrastive design assessing cross-generational paternal genetic transmission of alcohol use disorder, drug abuse and major depression. *Psychol. Med.* (2019). doi:10.1017/S0033291719000874

87. Munafò, M. R. & Davey Smith, G. Robust research needs many lines of evidence. *Nature* **553**, 399–401 (2018).

88. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* **42**, 111–127 (2013).

89. Fraser, A. *et al.* Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).

90. Prandy, K. Class, stratification and inequalities in health: a comparison of the Registrar-General's Social Classes and the Cambridge Scale. *Sociol. Heal. Illn.* **21**, 466–484 (1999).

91. Wechsler, D. *Wechsler Intelligence Scale for Children - Third Edition UK*. (The Psychological Corporation, 1992).

92. Crawford, J. R., Anderson, V., Rankin, P. M. & MacDonald, J. An index-based short-form of the WISC-IV with accompanying analysis of the reliability and abnormality of differences. *Br. J. Clin. Psychol.* **49**, 235–258 (2010).

93. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics*

31, 1466–1468 (2015).

Acknowledgements

We thank Sean Harrison for help with the simulations used to estimate bivariate heritability standard errors. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf); data used at age 23 was specifically funded by the Wellcome Trust and MRC [102215/2/13/2]. Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Bristol. REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies, providing 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. All ALSPAC data are available to bonafide researchers who obtain the necessary permissions from the ALSPAC executive committee. The Economics and Social Research Council (ESRC) support NMD via a Future Research Leaders grant [ES/N000757/1] and TTM via a postdoctoral research fellowship [ES/S011021/1]. The Medical Research Council (MRC) and the University of Bristol support the MRC Integrative Epidemiology Unit [MC_UU_12013/1, MC_UU_12013/9, MC_UU_00011/1]. This publication is the work of the authors and Tim Morris will serve as guarantor for the contents of this paper. No funding body has influenced data collection, analysis or its interpretations. This work was carried out using the computational facilities of the Advanced Computing Research Centre - http://www.bris.ac.uk/acrc/ and the Research Data Storage Facility of the University of Bristol - http://www.bris.ac.uk/acrc/storage/. TTM analysed and cleaned the data, interpreted results, wrote and revised the manuscript. NMD, GH and GDS interpreted the results, wrote and revised the manuscript. The authors declare that they have no competing interests.