A Litmus Test for Confounding in Polygenic Scores

Samuel Pattillo Smith[1,2], Olivia S. Smith[1,2], Hakhamanesh Mostafavi[3], Dandan Peng[4], Jeremy J. Berg[5], Michael D. Edge[4,+], and Arbel Harpak[1,2,+]

[1]*Department of Population Health, University of Texas at Austin, Austin, TX*
[2]*Department of Integrative Biology, University of Texas at Austin, Austin, TX*
[3]*Department of Population Health, New York University, New York, NY*
[4]*Department of Computational Biology, University of Southern California, Los Angeles, CA*
[5]*Department of Human Genetics, University of Chicago, Chicago, IL*
[+]*Correspondence should be addressed to M.D.E (edgem@usc.edu) and A.H. (arbelharpak@utexas.edu)*

Abstract

Polygenic scores (PGSs) are being rapidly adopted for trait prediction in the clinic and beyond. PGSs are often thought of as capturing the direct genetic effect of one's genotype on their phenotype. However, because PGSs are constructed from population-level associations, they are influenced by factors other than direct genetic effects, including stratification, assortative mating, and dynastic effects ("SAD effects"). Our interpretation and application of PGSs may hinge on the relative impact of SAD effects, since they may often be environmentally or culturally mediated. We developed a method that estimates the proportion of variance in a PGS (in a given sample) that is driven by direct effects, SAD effects, and their covariance. We leverage a comparison of a PGS of interest based on a standard GWAS with a PGS based on a sibling GWAS—which is largely immune to SAD effects—to quantify the relative contribution of each type of effect to variance in the PGS of interest. Our method, Partitioning Genetic Scores Using Siblings (PGSUS, pron. "Pegasus"), breaks down variance components further by axes of genetic ancestry, allowing for a nuanced interpretation of SAD effects. In particular, PGSUS can detect stratification along major axes of ancestry as well as SAD variance that is "isotropic" with respect to axes of ancestry. Applying PGSUS, we found evidence of stratification in PGSs constructed using large meta-analyses of height and educational attainment as well as in a range of PGSs constructed using the UK Biobank. In some instances, a given PGS appears to be stratified along a major axis of ancestry in one prediction sample but not in another (for example, in comparisons of prediction in samples from different countries, or in ancient DNA vs. contemporary samples). Finally, we show that different approaches for adjustment for population structure in GWASs have distinct advantages with respect to mitigation of ancestry-axis-specific and isotropic SAD variance in PGS. Our study illustrates how family-based designs can be combined with standard population-based designs to guide the interpretation and application of genomic predictors.

Introduction

Genome-wide association studies (GWASs) in humans have revealed that the genetic basis of variation for many health conditions and other traits is highly polygenic. That is, a substantial proportion of the variation derives from a large number of common genetic variants with marginal effects that are individually small. GWASs further revealed that the joint effect of these variants is often well-captured by a simple linear combination, consistent with longstanding theoretical predictions. Following from these observations, attention has turned toward the construction of genomic predictors of traits, so-called "polygenic scores" (PGSs). A PGS for a trait is typically a sum over a set of loci of an individual's alleles, each weighted by its GWAS-estimated effect.

PGSs have many potential uses, including in the clinic[1,2], as instruments to account for genetic variation when studying environmental effects[3], and for studying human evolution[4,5]. However, GWAS and polygenic scores in humans suffer from a general weakness: a lack of experimental control. Neither environments nor genetic backgrounds can be randomized among individuals within GWAS samples. When a specific locus's alleles are correlated with other causal factors (genetic, societal, or environmental), they become spuriously associated with the phenotype. This can bias allelic effect estimates with respect to the causal genetic effect of the focal locus (and linked causal variants)[6–8].

What factors contribute to population-based GWAS allelic effect estimates? It has long been established that, in addition to the causal effects of one's alleles on one's own phenotype ("direct" genetic effects), allelic effect estimates can be affected by confounding due to population stratification[6,9]. More recently, other phenomena that affect GWAS allelic effect estimates have received increased attention[10,11], such as assortative mating[11–15] and "dynastic" (or indirect parental) genetic effects. Dynastic effects arise by virtue of the correlation between an individual's genotype and that of their parent, when the parent's genotype influences the focal individual's phenotype[16–19]. For example, if maternal genotypes affect the focal individual's uterine environment during early embryonic development, then there will be an association between the focal individual's genotype and trait, even if the genotype only operates via the mother. We henceforth refer to these influences as "SAD" (Stratification, Assortment, and Dynastic) effects, although there may be additional systematic influences on GWAS allelic effect estimates beyond these three factors (e.g., indirect genetic effects from other related individuals). Although GWAS analysis pipelines include steps aimed at adjusting for environmental and genetic confounding, even residual SAD effects that are small on a per-SNP basis can combine to have large effects on the variance of a PGS across individuals[8,20–23].

When considering the application of a PGS to individual genomes, it is important to know the extent to which variation among individuals' PGS is due to SAD effects. Yet, to our knowledge, there are no methods that address this question. To illustrate the importance of this task, consider an example in which variation among individuals in a PGS for cancer is largely due to correlates of SNPs included in the PGS (index SNPs) in the GWAS sample, such as exposure to pollutants. The PGS may nonetheless be predictive—this would depend on whether the genotype-exposure correlation in the GWAS sample exists in the prediction sample. Even if a PGS is predictive, but largely because of SAD effects, should we use the PGS as a predictor of genetic risk? The answer may vary across different applications of PGSs. Regardless, a diagnostic tool measuring the impact of SAD factors could provide grounds for an

informed decision.

The fraction of PGS variation in a sample that is attributable to SAD effects may also depend on the sample's dissimilarity to the GWAS sample—in terms of genetic ancestry, environment and social context [22,24,25]. Therefore, understanding how SAD effects in the GWAS sample affect PGS variation in distinct prediction samples may help us understand PGS portability.

One promising way to evaluate the impact of SAD effects on a PGS is family-based designs, such as sibling GWAS, in which trait differences among pairs of full siblings are regressed on differences in their genotypes. Sibling GWAS are largely robust to SAD effects [22,26–28], though they are not strictly immune [15] and complexities can arise in the presence of genetic interactions or other factors [11]. Family-based designs are typically underpowered compared with population-based GWAS, and are therefore rarely used to build PGS for anything but research purposes. Nevertheless, family studies can, in principle, help us distinguish direct genetic effects from SAD effects contributing to population-based GWAS estimates.

Here, we develop a method, Partitioning Genomic Scores Using Siblings (PGSUS, pron. "Pegasus"), that partitions the variance of a PGS derived from a population-based GWAS ("standard-GWAS" below) on the basis of a comparison with the corresponding allelic effect estimates from a sibling-based GWAS ("sib-GWAS" below). PGSUS goes beyond previous approaches by quantifying the contribution of direct genetic effects versus SAD effects (and their covariance) to a PGS as applied in a given prediction sample. PGSUS can be viewed as a litmus test for the presence of confounding and can therefore inform specific applications and interpretations of PGSs.

Results and Discussion

Model

We consider a PGS applied to a "prediction sample"—a sample of genotypes for which the PGS is calculated. Our main aim is to identify instances in which substantial PGS variance is due to SAD effects. In particular, components of SAD variance aligning with major axes of ancestry are suggestive of stratification along that ancestry axis (see **A litmus test for confounding**). We begin by outlining our model, which motivates the partitioning of PGS variance that PGSUS performs.

PCA partitioning of polygenic score variance. The variance of a PGS can be partitioned into orthogonal components attributable to principal components of the genotype matrix of the prediction sample. In **Text S1**, we show that

$$\mathrm{Var}[\mathbf{Z}] = \sum_{i=1}^{\min(n,\ell)} \lambda_i (\mathbf{U_i} \cdot \mathbf{b})^2, \tag{1}$$

where Z is a length-n vector of PGS values, λ_i is the eigenvalue for the i-th principal component of the prediction sample's genotype matrix, $\mathbf{U_i}$ is a vector holding the loadings of index SNPs on principal component i, b is a vector of allelic effect estimates, and $\cdot$ represents the dot product. n is the number of individuals in the sample used to perform the PCA and ℓ is the number of index SNPs. This variance partitioning is equivalent to the partitioning that would be obtained by regressing individual PGS values on individual-level principal-component coordinates but calculated via a projection of the SNP allelic

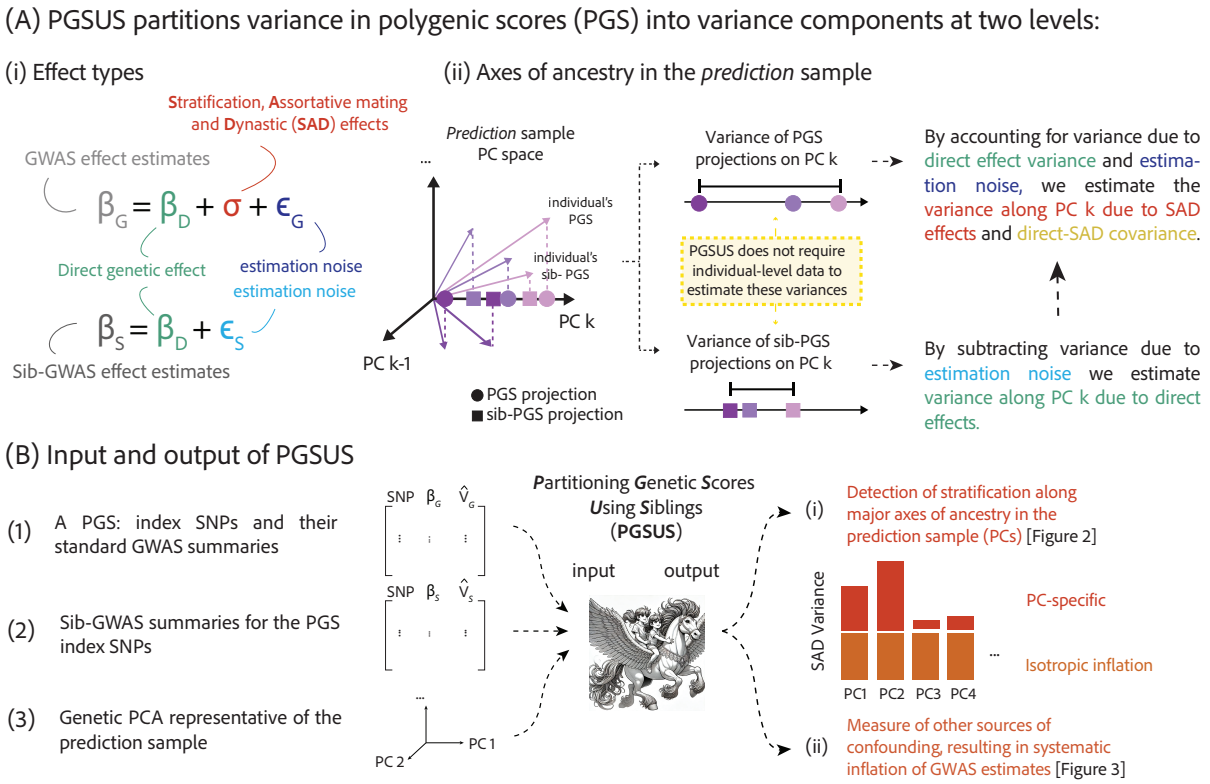

Figure 1. Partitioning Genetic Scores Using Siblings (PGSUS). (A) PGSUS partitions the variance of a PGS among individuals in a prediction sample at two levels. Namely, variance components are attributable to both (i) effect type: direct effects, SAD effects, and their covariance, as well as estimation noise; and (ii) a principal component of the genotype matrix of the prediction sample. The partitioning by effect types derives from the generative model shown in (i) for standard GWAS effect estimates and the corresponding sib-based estimates. **(B)** PGSUS does not require individual-level data. It only requires an input of GWAS summary statistics, corresponding sib-based summary statistics, and a PCA of a genotype matrix representative of the sample to which the PGS is to be applied. The output is the full set of variance components and corresponding null acceptance regions. SAD variance components significantly larger than their corresponding eigenvalues would predict are suggestive of stratification along the corresponding PC. In addition, "isotropic" SAD variance, i.e. large SAD variance that is not specific to a PC, may suggest other modes of confounding such as assortative mating, dynastic effects and population structure confounding that is not specific to a PC. However, other factors such as ascertainment biases can substantially influence isotropic inflation as well.

effect estimates onto the PCs. Thanks to this equivalence, PGSUS does not require any individual-level data, but only GWAS summary statistics and a PCA of the prediction sample (or a representative sample from the population to which the researcher wishes to apply the PGS; **Fig. 1A**). In addition, expressing the variance partitioning in this way allows us to link it to a generative model for the allelic effects (see below). The quantity $\mathbf{U_i} \cdot \mathbf{b}$ is the scalar projection of the allelic effect estimates $\mathbf{b}$ on vector $\mathbf{U_i}$ and can be thought of as the effect estimated for principal component i. It is also closely related to the covariance of the vector of allelic effect estimates with $\mathbf{U_i}$.

Generative model of allelic effect estimates. We model the allelic effect estimates from a

standard GWAS, β_G, as the sum

$$\beta_G = \beta_D + \sigma + \epsilon_G, \tag{2}$$

where β_D is the true direct genetic effect, σ represents SAD effects, and ϵ_G is a measurement error. At each locus, we assume that ϵ_G has expectation zero and standard deviation equal to the standard error of β_G as estimated in the GWAS. In a sib-GWAS, we model allelic effect estimates as

$$\beta_S^* = \alpha^{-1}(\beta_D + \epsilon_S). \tag{3}$$

Here, ϵ_S is the measurement error, again with expectation zero and standard deviation given by the standard error of β_S as estimated in the sib-GWAS. The parameter α, the isotropic inflation factor, captures systematic differences in the magnitude of allelic effect estimates between the standard-GWAS and sib-GWAS. In the **Isotropic inflation** section, we discuss both SAD factors and technical issues that may contribute to systematic inflation. We adopted a strategy of estimating α (see Section **Estimation of the isotropic inflation factor**), and then adjusting for isotropic inflation. Namely, we define

$$\beta_S = \alpha\beta_S^* = \beta_D + \epsilon_S \tag{4}$$

and do not consider isotropic inflation further in the two-level partitioning that follows.

We highlight the assumption that the direct effect β_D is the same in standard- and sib-GWAS based allelic effect estimates. This may not be true in general, especially in the presence of differences in ancestry, environment and social context between the standard and the sibling samples[11,22]. (Even without such differences, and no SAD effects, there are differences in the expectations of β_D and β_S[11]—but they are likely small, see **Text S2**).

Partitioning of PGS variance by both effect types and PCs. In combination, the expression for the variance of a PGS in **Eq. 1** and the models for the allelic effect estimates contributing to a PGS in **Eqs. 2,4** suggest a partitioning of variance in a PGS due to the projections of direct effects, SAD effects, and measurement error on each principal component. Specifically, suppose a vector of the allelic effect estimates from GWAS at index SNPs, $\overrightarrow{\beta_G}$, replaces $\mathbf{b}$ in **Eq. 1**, as in

$$\mathrm{Var}[\mathbf{Z}] = \sum_{i=1}^{\min(n,\ell)} \lambda_i \left(\mathbf{U_i} \cdot \overrightarrow{\beta_G}\right)^2 = \sum_{i=1}^{\min(n,\ell)} \lambda_i \left(\mathbf{U_i} \cdot \left[\overrightarrow{\beta_D} + \overrightarrow{\sigma} + \overrightarrow{\epsilon_G}\right]\right)^2. \tag{5}$$

Distributing the $\mathbf{U_i}$ and expanding the expression on the right gives

$$\begin{aligned}\mathrm{Var}[\mathbf{Z}] = \sum_{i=1}^{\min(n,\ell)} \lambda_i \Bigg(&\left[\mathbf{U_i} \cdot \overrightarrow{\beta_D}\right]^2 + \left[\mathbf{U_i} \cdot \overrightarrow{\sigma}\right]^2 + \left[\mathbf{U_i} \cdot \overrightarrow{\epsilon_G}\right]^2 + 2\left[\mathbf{U_i} \cdot \overrightarrow{\beta_D}\right]\left[\mathbf{U_i} \cdot \overrightarrow{\sigma}\right] + \\ &2\left[\mathbf{U_i} \cdot \overrightarrow{\beta_D}\right]\left[\mathbf{U_i} \cdot \overrightarrow{\epsilon_G}\right] + 2\left[\mathbf{U_i} \cdot \overrightarrow{\sigma}\right]\left[\mathbf{U_i} \cdot \overrightarrow{\epsilon_G}\right]\Bigg).\end{aligned} \tag{6}$$

We are principally interested in estimating three of the terms that arise from the expansion in **Eq. 6**, which can be interpreted as variance components attributable to direct effects, to SAD effects, and to

the covariance of direct effects and SAD effects. In particular,

$$c_{Di} = \lambda_i \left(\mathbf{U_i} \cdot \overrightarrow{\beta_D} \right)^2 \tag{7}$$

is the component of variance in the PGS due to direct effects specific to principal component i of the genotype matrix, and

$$c_{\sigma i} = \lambda_i \left(\mathbf{U_i} \cdot \overrightarrow{\sigma} \right)^2 \tag{8}$$

is the component of variance due to SAD effects specific to principal component i. We also consider

$$c_{(D \cdot \sigma) i} = 2\lambda_i \left(\mathbf{U_i} \cdot \overrightarrow{\beta_D} \right) \left(\mathbf{U_i} \cdot \overrightarrow{\sigma} \right), \tag{9}$$

which is a contribution to variance due to covariance of direct and SAD effects specific to principal component i. This last term can be negative, unlike the previous two, which are constrained to be non-negative. In the **Estimation of variance components** section, we derive estimators. Our basic strategy is to isolate $(\mathbf{U_i} \cdot \overrightarrow{\sigma})$ using the projection of the difference between allelic effect estimates from the standard GWAS and those from the sib-GWAS, and proceed with moment-based estimation.

We developed PGSUS as freely available software that performs this partitioning. The software takes three inputs: (i) the set of index sites used in the polygenic score, their estimated effects and standard errors from the standard-GWAS used to construct the PGS; (ii) corresponding estimates from a sib-GWAS of full siblings assumed to be sampled from the same population as the standard-GWAS, such that the samples are subject to the same direct effects; and (iii) either a matrix of genotypes from the prediction sample for the corresponding loci or corresponding, precomputed eigenvalues and eigenvectors from the genotype matrix of the prediction sample (**Fig. 1B**). Using this set of inputs, PGSUS first estimates and adjusts for isotropic inflation. Then, PGSUS partitions the PGS variance into PC-specific, effect-type-specific subcomponents.

A litmus test for confounding

To illustrate how PGSUS can be used as a diagnostic tool, we begin by examining a case in which stratification is already known to contribute to PGS variance. If an axis of stratification largely overlaps with a PC of the prediction sample, we expect to see a large PC-specific SAD variance component.

In 2019, researchers discovered that the GIANT consortium's height GWAS[29] was confounded along a north-south ancestry axis in Europe[20,21]. We constructed a PGS using this GWAS and partitioned its variance in the 1000 Genomes European superpopulation as the prediction sample[30]. As expected, the SAD variance component on PC2 (which best tags the north-south ancestry axis[31]; **Fig. 2A** inset) is large—37% of the size of the total variance of the PGS (**Fig. 2A**). This variance component is significantly larger than expected under a null model with no PC-specific SAD variance (95% null acceptance regions are shown in **Fig. 2**; see **Text S3** for details on how the empirical null distribution was derived).

This result confirms that PGSUS can detect a "positive control", a known example of confounding along a major axis of genetic ancestry. An important point about the interpretation of results produced by PGSUS is that they tell us about variation in a specific PGS, not about a trait or even about a specific

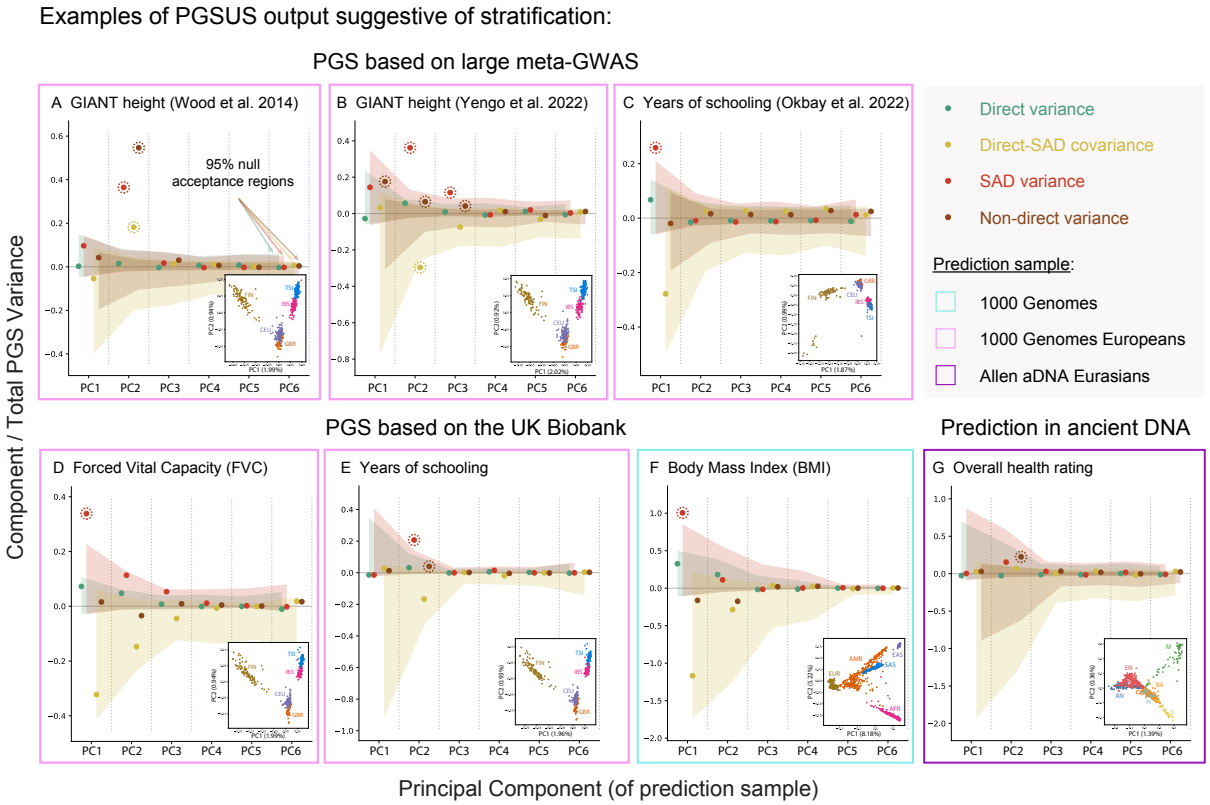

Figure 2. PGSUS partitioning uncovers stratification in polygenic scores. For seven examples, we show variance components along the first six PCs divided by the total variance in PGS in the prediction sample. Shaded regions show empirical permutation-based null acceptance regions. Components deviating from this expectation are highlighted with dashed circles. Insets show the prediction-sample individual coordinates along some of the PCs on which we observed significantly large variance components; labels are shorthand for subsamples in the 1000 Genomes (1KG) data or the Allen Ancient DNA Resource Eurasian subset. Note that insets will vary slightly between the same prediction cohort as they are constructed using the set of SNPs selected through clumping using trait-specific summary statistics. The 1KG subsample labels in panels **(A)-(E)** correspond to the five European subpopulations: Finnish in Finland (FIN), Utah residents with northern and western European ancestry (CEU), Iberian populations in Spain (IBS), British from England and Scotland (GBR), and Tuscan in Italy (TSI). The 1KG subsample labels in panel **(F)** and correspond to the five 1KG superpopulations: European (EUR), admixed Americans (AMR), East Asian (EAS), South Asian (SAS), and African (AFR). Finally, labels in panel **(I)** correspond to six ancient population groupings as defined in Mallick and Reich [32]: Anatolian Neolithic (AN), European Neolithic (EN), Historical Europe (H), Bronze Age (BA), Steppe pastoralists (S), and Mesolithic (M). Percentages of variance explained by each PC are given in the parentheses of each axis label in the insets. All PGSs shown are constructed using clumping and thresholding with a marginal association thresholds as follows: $p < 1$ for panels **(A,B,E,G)**; $p < 10^{-3}$ for panels **(C,D)**; $p < 10^{-5}$ for panel **(F)**. PGS partitioned in panels **(D-G)** are based on GWASs with 20 GWAS sample PCs as covariates as the only adjustment for population structure.

GWAS for that trait. To demonstrate this, we considered partitionings of other height PGSs with respect to the same prediction sample. For a similarly constructed PGS using a GWAS conducted in the UK Biobank (UKB), the SAD variance component on PC2 is large but does not significantly deviate from the null expectation (**Fig. S2A**). Furthermore, using the GIANT GWAS again, but constructing the PGS

with a more stringent threshold for the significance of the index SNP, the SAD variance component on PC2 is only 2.9% of the total variance of the PGS (**Fig. S1**; marginal GWAS p-value $< 10^{-3}$ instead of p-value < 1 used in **Fig. 2A**).

We examined PGSs based on two other GWAS meta-analyses. A PGS based on the 2022 version of the GIANT height GWAS[33] also shows significantly large SAD variance suggestive of stratification along the same north-south European ancestry axis (**Fig. 2B**). A partitioning of a PGS based on the years of schooling GWAS of 2022 ("EA4"[34]), which has also been previously suggested to be affected by SAD factors[19,34,35], shows evidence of stratification along PC1 of the same European prediction sample (**Fig. 2A**), which most prominently differentiates Finnish from non-Finnish individuals in this sample (**Fig. 2C** inset).

Confounding in PGS built using modern biobanks

The GIANT 2014 study[29] was, until recently, the largest GWAS of height, the complex trait most studied via GWAS, and so evidence of confounding sparked concern for complex trait research as a whole[8,36]. Our partitionings of GIANT 2014, GIANT 2022 and EA4 suggest the concern could extend to other recent GWAS meta-analyses as well. Some modern biobanks were designed to recruit large cohorts that are relatively ancestrally homogeneous with the intention of minimizing stratification[37–40]. Nonetheless, the extent to which single biobank-based PGSs are affected by confounding remains an open question.

PGSUS can address this question, as we illustrate next by partitioning the variance of a suite of PGSs constructed using GWASs we performed in the UKB for 11 highly heritable physiological traits and 6 social or behavioral traits (**Table S1**). We constructed PGSs by clumping and thresholding with different p-value thresholds, based on GWASs performed with various adjustments for population structure. All GWASs were performed using the self-identified White British subset of the UKB, which is relatively ancestrally homogeneous. Variances in the PGSs were then partitioned with respect to either the entire 1000 Genomes Phase 3 sample (1KG)[30], its European subset (1KG Europeans) or to a Eurasian subsample of the the Allen Ancient DNA Resource (AADR)[32] as the prediction sample.

Results from partitionings of this suite of PGSs can be found in **Files S1-S8**, and general conclusions are discussed below. Although we have performed many tests, we do not consider corrections for multiple hypothesis testing. For computational reasons, we have also limited our permutation-based hypothesis testing to a significance level of 0.05 (**Text S3**). Our discussion of specific cases in this manuscript should only be considered as illustrations of the results PGSUS can provide and their implications when they are indeed statistically significant. They are not statements about omnibus hypotheses about SAD effects in the PGSs or datasets analyzed.

For concreteness, we first focus on a set of GWASs performed with a standard adjustment for 20 principal components (PCs) of the GWAS-sample genotype matrix and PGSs constructed through clumping and thresholding[41–43] with a marginal association threshold of $p < 10^{-5}$ with respect to 1KG Europeans. In 9 of the 17 PGSs examined, we detected a significant SAD variance component in at least one of the top 6 PCs (**File S1**). For example, the SAD variance component on PC1 is approximately one third (33.8%) of the total variance in the PGS for forced vital capacity (FVC; **Fig. 2D**). One possible contributor to PC-specific SAD variance is environmental and social confounding: lower FVC is associated with cigarette

smoking[44], which in turn is associated with educational attainment[45]. Among our GWAS-sample individuals, both pack-years of smoking and years of schooling are significantly associated with position on PC1 (**Fig. S3**). Another example is in a UKB-based PGS for years of schooling (**Fig. 2E**) that, like the EA4-based PGS, showed significant SAD variance components with respect to 1KG Europeans, but on PC2 in this case.

PGSUS informs PGS portability

PGSUS sometimes detected significant PC-specific SAD effects in a PGS with respect to one prediction sample and no such evidence with respect to another sample. For example, a PGS constructed for body mass index (BMI) showed no evidence for PC-specific SAD effects when applied to 1KG Europeans (**File S1**). Yet in the full 1KG sample, we detected significant SAD variance on PC1, which in part tags differentiation between 1KG European and non-European samples (**Fig. 2F**).

Portability to ancient DNA samples is also of interest. Several studies have used polygenic summaries based on a similar UKB GWAS to predict past phenotypes[46–49] or infer natural selection in recent European history[50–52]. We also examined partitionings of PGSs derived from UKB-based GWAS with respect to a sample of ancient West Eurasians who lived between 13,000 and 1,000 years ago (Allen Ancient DNA Resource[32]). In several PGSs for household income, overall health rating, skin color, and years of schooling, we found either significant SAD variance or total non-direct variance component (i.e., the sum of the SAD variance component and the direct-SAD covariance component) among the top six PCs (**Figs. S4-S5**). For example, there is significantly large non-direct variance along PC2 in a PGS for overall health rating, a trait argued to be under directional selection in ancient Eurasia during this time[50] (**Fig. 2G**). We note that these top PCs are also significantly correlated with a number of sample variables, such as date (as determined by conventional radiocarbon dating) and sequencing coverage (**Fig. S6**). This raises the concern that confounding between these variables and UKB-based effect estimates of index SNPs may spuriously contribute to polygenic signals of selection.

The examples of BMI in 1KG data and traits argued to be under directional selection in ancient Eurasia help illustrate that PGSUS, as diagnostic of confounding, can inform us on the lack of portability of a PGS to an intended prediction sample. There are many factors that can influence PGS portability, including differences in linkage disequilibrium patterns, genetic variance, and environmental variance between the GWAS sample and the prediction sample, all of which may change over time[22,24,25,53,54]. The effects of confounding on PGS portability are not well understood. Still, our results should raise caution with respect to evidence for natural selection acting in the past—especially since SAD variance could reflect more recent environmental or social factors, such as contemporary stratification or dynastic effects.

Isotropic inflation

Often, standard-GWAS allelic effect estimates are larger than sib-GWAS effect estimates (after accounting for the degree of estimation error in each design) in a way that is not explained by any small set of principal components. Such "isotropic inflation" in standard GWAS could be due to assortative mating[10–14,55] or indirect parental effects[10,11,18,19,56] if they act isotropically with respect to principal component space;

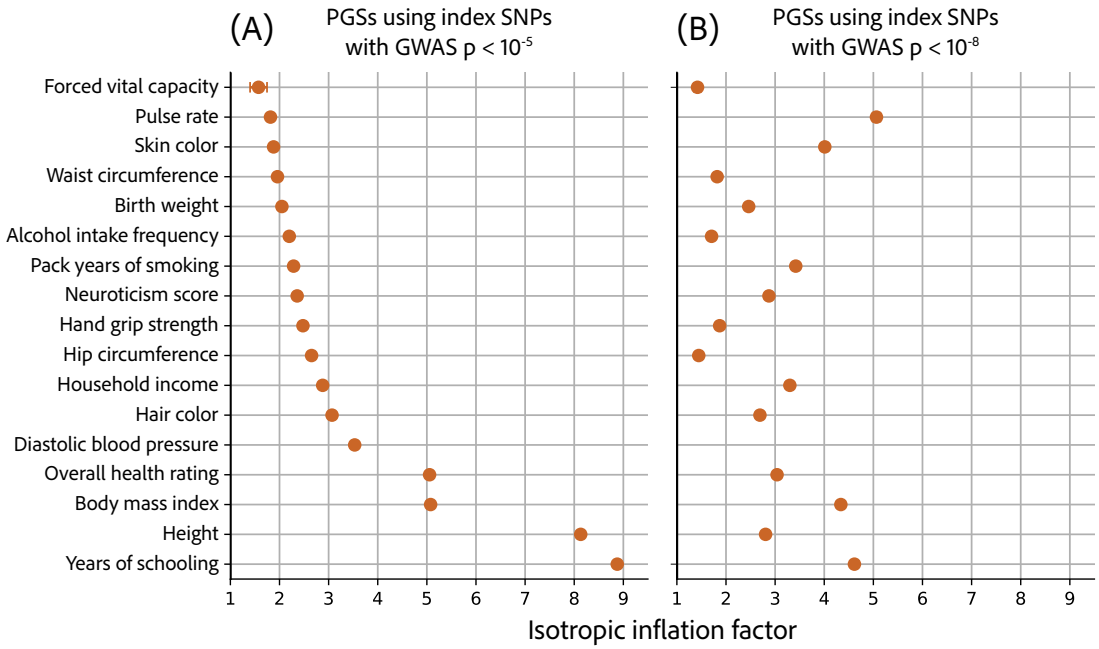

Figure 3. The isotropic inflation factor. Isotropic inflation refers to SAD variance that is not PC-specific, and manifests as uniformly larger magnitude effects in standard-GWAS compared with the magnitude of direct effects estimated from sib-GWAS. A factor of 1 indicates no isotropic inflation. Error bars show +/- bootstrap standard error, and are often too small to be visible. Isotropic inflation may be driven by population stratification, assortative mating, indirect effects, or other issues, such as different measurement units in the standard-GWAS and sib-GWAS or ascertainment bias. Shown are isotropic inflation factor estimates for PGSs constructed from a GWAS in the White British subset of UKB using clumping and thresholding, with a marginal association p-value threshold of **(A)** 10^{-5} or **(B)** 10^{-8} partitioned with respect to the 1KG prediction sample.

similarly, it could be due to some modes of population stratification [11,15,34,57,58]. For example, confounding between genetic similarity and similarity in environmental exposures or social context such as access to healthcare or education.

There are also technical drivers of isotropic inflation, for example, differences in units of measurement among studies (e.g., centimeters as opposed to standard deviations of height). Another technical factor driving isotropic inflation may be in biases related to the ascertainment of index SNPs, such as the "winner's curse" [59,60], as PGS index SNPs are typically ascertained in the same GWAS sample in which their effects are estimated. Empirically, across PGS marginal association thresholds, we observed that more stringent ascertainment tended to lead to more isotropic inflation (**Figs. 3,S7,S9**). Estimates of the isotopic inflation factor also depended on the prediction sample (e.g., when comparing **Fig. 3** and **Fig. S8A,B**). This is in part because of a trade-off between the two types of SAD variance: large PC-specific SAD variance components imply smaller isotropic inflation, all else being equal.

Our analyses cannot identify the specific drivers of isotropic inflation. As discussed, a plausible driver of isotropic inflation is SAD effects. Despite this possibility, we do not consider variance due to isotropic inflation as a contributor to SAD variance when reporting PGS variance partitionings in this manuscript. Instead, the PGS partitionings we report are with respect to the isotropic-inflation-adjusted allelic effect

estimates of **Eq. 4**. Therefore, our estimates of variance components (e.g. components shown in **Fig. 2**) can be interpreted as either (i) assuming isotropic inflation to be due to technical factors or (ii) as estimated variance components after adjusting for the "global" effects of isotropic inflation, which may in part be caused by SAD factors. In the section **Estimation of the isotropic inflation factor**, we detail our estimation approach.

Among a set of PGSs (UKB-based, with clumping and thresholding on a GWAS association p-value $< 10^{-5}$) applied to the 1KG prediction sample, all PGSs show some isotropic inflation (**Fig. 3A**). The highest isotropic inflation factors are for years of schooling (8.88) and height (8.13). When we constructed PGSs differently, using a GWAS association p-value $< 10^{-8}$ for index SNPs, the rankings of isotropic inflation factors across traits often varied widely, illustrating again that PGSUS characterizes the variance of a specific PGS in a given prediction sample, and not a trait or a GWAS (**Text S6; Figs. 3B, S7-S9**).

Do population structure adjustments in GWASs mitigate SAD variance?

The analyses of PGS variance discussed up to this point have been based on PGSs computed from GWASs in which we (or others) adjusted allelic effect estimates by including the top PCs of the GWAS sample's genotype matrix as covariates in the GWAS. In our in-house GWAS, the top 20 PCs were included. This is a widely-used approach[20,21,61,62]. One question is whether other approaches to adjusting for confounding in GWAS produce PGSs with less SAD variance.

We addressed this question by comparing the number of significantly large SAD variance components among the top six PCs and the isotropic inflation factor when using various methods to adjust for population structure. The GWASs were again performed in the UKB White British subsample, and PGSs based on these GWASs were partitioned with respect to the 1KG cohort as the prediction sample. We first compared the adjustment for 20 GWAS sample PCs with no adjustment at all. Our expectation was that no PC adjustment in the GWAS should lead to an equal or larger number of significant PC-specific SAD components. This was indeed the case for 16 of 17 PGSs (**Figs. 4, S10A**). This expectation was also met across the majority of PGSs (but not all) in sensitivity analyses, such as when we applied more lenient association p-value thresholding (resulting in PGSs including hundreds of thousands of SNPs; **Table S2**, **Fig S11**) or considered the top 20 instead of 6 top PCs (**Fig. S12**).

Still, PC adjustments did not remove all PC-specific SAD variance and occasionally even had the counterintuitive effect of increasing PC-specific SAD variance (panels A in **Figs. S10-S13**). One reason may be that PGS index SNPs are distinct in several ways from the SNPs that influence PCA the most. They are variants strongly associated with traits and therefore plausibly subject to stronger selection[63,64], rarer, and otherwise distinct from weakly associated variants in their distribution across individuals[25,65–68]. In contrast, GWAS sample PCs capture the structure of common, less constrained variation.

Relatedly, the adjustment for GWAS sample PCs is aimed at capturing the main axes of population structure in the GWAS sample. However, when a PGS based on the GWAS is applied to a distinct prediction sample, other axes of population structure in the GWAS sample shared with the prediction sample may contribute to confounding[5,69]. Considering axes of population structure shared between

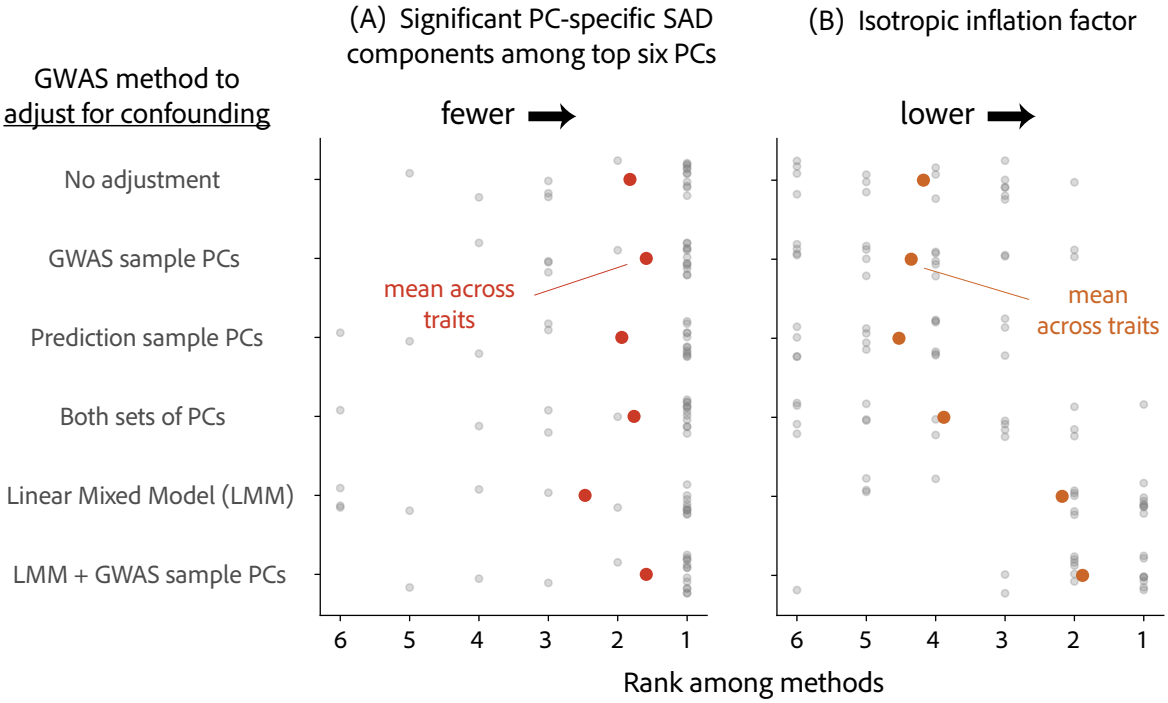

Figure 4. The utility of GWAS population structure adjustments in mitigating SAD variance. We constructed PGSs using clumping and thresholding in a UKB-based GWASs with an ascertainment threshold of p-value $< 10^{-5}$. We then partitioned the PGS variance in the 1KG prediction sample and examined SAD variance signals. The PGSs were based on GWASs with various methods to adjust for population structure. "GWAS sample PCs" refers to a GWAS adjustment for the first 20 principal components of the genotype matrix of the UKB White British cohort in which the GWAS is performed. "Prediction sample PCs" refers to the inclusion of 20 1KG genotype matrix PCs. "LMM" refers to the the *BOLT-LMM* implementation of a linear mixed model. For each trait-method pair (one grey point), **(A)** shows the rank of the adjustment method in terms of the number of significantly large PC-specific SAD components (p-value< 0.05), and **(B)** shows its rank in terms of isotropic inflation factor. Rank 1 is given to the GWAS adjustment method(s) that yielded the fewest significant PC-specific SAD variance components in **(A)** and lowest isotropic inflation factor in **(B)**; Rank 2 is given to the second fewest / smallest, and so forth.

GWAS samples and prediction samples may be a fruitful direction in mitigating SAD variance[69]. We tested this idea by including the top 20 PCs of the *prediction sample*'s genotype matrix in GWAS. Adjustments for the GWAS-sample PCs, the prediction-sample PCs, or both did comparably well at mitigating PC-specific SAD variance (**Fig. 4A**; also see **Fig. S14** for the correlations between GWAS and prediction sample PCs). In addition to the lack of noticeable improvement over the adjustment for GWAS-sample PCs, we note that, in practice, it is not always known *a priori* to what samples the PGS might be applied when it is constructed.

Lastly, we considered the performance of GWAS adjustment methods in mitigating isotropic inflation. All four adjustments considered thus far resulted in isotropic inflation factors that were large and similar on average across traits (ranging between 3.9 and 4.5; **Figs. 4B, S10**). We considered an additional approach, a linear mixed model (LMM, specifically with *BOLT-LMM*[70]), that aims to adjust

for confounding of genome-wide genetic similarity with phenotypic similarity[6,70]. LMM-based GWASs can be thought of as adjusting for relatedness along all PCs[71,72] and have been shown to perform well in adjusting for assortment, admixture, and cryptic relatedness that are not well captured by individual PCs[73]. Indeed, PGSs based on LMM GWASs resulted in the smallest isotropic inflation factors (**Figs. 4B, S10B**). However, the improved mitigation of isotropic SAD variance was accompanied by the worst performance in mitigating PC-specific SAD variance (in the majority of cases, and on average, across traits; **Figs. 4A, S10A**; also see a specific example for a household income PGS in **Fig. 2C**). This apparent trade-off between isotropic inflation and PC-specific SAD variance is somewhat intuitive, given that LMM down-weights allelic effect estimates evenly across PCs, ignoring PC-specific patterns of confounding[71,72]. When we used an LMM in addition to adjusting for GWAS sample PCs, we observed the best mitigation of both isotropic and PC-specific SAD variance signals across all adjustment methods considered (in the vast majority of cases, and on average, across traits; bottom row in **Fig. 4**; right column in **Fig. S10A,B**; see related discussion in Zhang and Pan[74]).

Conclusion

Through their uses in the clinic, research, and beyond, PGSs are sometimes assumed to represent direct, causal effects of an individual's genotype on their phenotype. Thus far, to our knowledge, most attempts to measure the extent to which this is true have been qualitative or heuristic in nature. The challenge of measuring the degree to which a PGS (and its variation in a prediction sample) instead reflects SAD factors—including environmentally and socially mediated factors—is critical to its interpretation and application. PGSUS is a proposed step toward addressing this challenge.

Acknowledgements

We thank Graham Coop for his major contributions in conceptualizing PGSUS and guidance throughout its development. We thank Ipsita Agrawal, Yi Ding, Elliot Tucker-Drob, Jonathan Pritchard, Molly Przeworski and members of the Harpak Lab for feedback on the manuscript. The work was funded by NIH grant RF1AG073593 to Elliot Tucker-Drob, NIH grants F32GM130050 and R35GM137758 to M.D.E., and NIH grant R35GM151108, a Pew Scholarship, and a fellowship from the Simons Foundation's Society of Fellows (#633313) to A.H. This study was conducted using the UK Biobank resource under application 92741, as approved by the University of Texas at Austin institutional review board (study protocol 00003287). The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing computational resources that have contributed to the research results reported within this paper.

Methods

Estimation of variance components

We take a moment-based approach to estimating the variance components in **Eqs. 7-9**. We first describe an estimation procedure that applies if the isotropic inflation factor α is known. In practice, we multiply sib-GWAS effect estimates in **Eq. 3** by an estimate of α to estimate β_S, which we then use as input to the procedures described in this section. We return to the estimation of α later. In **Text S3**, we describe our empirical permutation strategy for testing hypotheses about the variance components.

Estimating the variance components given the isotropic inflation factor α. Our approach assumes that the difference between standard-GWAS based and sib-GWAS-based allelic effect estimates, other than differences in estimation noise, is that the sib-based estimates are free of SAD effects (**Eqs. 2-4**). We note that, even in the absence of SAD effects, the estimated effects are expected to differ in the two study designs[11]—but the difference is often small (**Text S2**). Applying **Eq. 1**, we then reason that the projection of the differences between standard-GWAS and sib-GWAS allelic effect estimates on the principal components will provide information about the proportion of variance in a PGS attributable to SAD.

Applying the model in **Eqs. 2-3** to index SNPs $j \in \{1, ..., \ell\}$, the variance component of the difference between standard-GWAS and sib-GWAS allelic effect estimates that is due to the ith principal component is

$$\lambda_i \left(\sum_{j=1}^{\ell} [\beta_{Gj} - \beta_{Sj}] U_{ij} \right)^2 = \lambda_i \left(\sum_{j=1}^{\ell} [\sigma_j + \epsilon_{Gj} - \epsilon_{Sj}] U_{ij} \right)^2 . \tag{10}$$

To derive the expectation of the expression in **Eq. 10**, we assume that the λ (eigenvalue), σ (SAD effect), and U (locus loadings on eigenvector) terms are fixed and that all the measurement error terms all have expectation 0 (i.e. $\mathrm{E}[\epsilon_{G_j}] = \mathrm{E}[\epsilon_{S_j}] = 0$ for all j). We also assume that the error terms from standard-GWAS and the sib-GWAS are uncorrelated with each other (i.e. $\mathrm{E}[\epsilon_{G_j} \epsilon_{S_k}] = 0$ for all j and k, including $j = k$). The assumption that the measurement errors from the standard-GWAS and sib-GWAS studies are uncorrelated might be problematic if, e.g., the sib-GWAS is performed in a subset of the standard-GWAS sample. Nonetheless, we adopt it here. Under these assumptions,

$$\mathrm{E}\left[\lambda_i \left(\sum_{j=1}^{\ell} [\beta_{Gj} - \beta_{Sj}] U_{ij} \right)^2 \right] = \lambda_i \left(\sum_{j=1}^{\ell} \sigma_j U_{ij} \right)^2 + \mathrm{E}\left[\left(\lambda_i \sum_{j=1}^{\ell} \epsilon_{Gj} U_{ij} \right)^2 \right] + \mathrm{E}\left[\left(\lambda_i \sum_{j=1}^{\ell} \epsilon_{Sj} U_{ij} \right)^2 \right]. \tag{11}$$

The first term on the right is c_{σ_i}, the variance component of the PGS attributable to SAD effects along principal component i, and one of our main estimands of interest. The latter two terms are variance components attributable to measurement errors in the standard-GWAS allelic effect estimates and the sib-GWAS allelic effect estimates, respectively. If we can assume that measurement errors at distinct loci are uncorrelated, or, slightly less restrictively, that

$$\sum_{j=1}^{\ell}\sum_{k\neq j}U_{ij}U_{ik}\mathrm{E}[\epsilon_{Gj}\epsilon_{Gk}]=0, \tag{12}$$

then the expected variance component due to GWAS errors is

$$\mathrm{E}\left[\lambda_i\left(\sum_{j=1}^{\ell}\epsilon_{Gj}U_{ij}\right)^2\right]=\lambda_i\sum_{j=1}^{\ell}U_{ij}^2v_{Gj}^2, \tag{13}$$

where v_{Gj}^2 is the variance of the measurement error of the standard-GWAS allelic effect estimate at locus j. Making an analogous assumption for the measurement errors in the sib-GWAS allelic effect estimates, we arrive at an estimator for the variance component of the PGS due to SAD effects along principal component i, $c_{\sigma i}=\lambda_i(\sum_{j=1}^{\ell}\sigma_jU_{ij})^2$,

$$\hat{c}_{\sigma i}=\lambda_i\left(\sum_{j=1}^{\ell}[\beta_{Gj}-\beta_{Sj}]U_{ij}\right)^2-\lambda_i\sum_{j=1}^{\ell}U_{ij}^2\hat{v}_{Gj}^2-\lambda_i\sum_{j=1}^{\ell}U_{ij}^2\hat{v}_{Sj}^2, \tag{14}$$

where $\hat{v}_{Gj}$ is an estimated standard error of the standard-GWAS allelic effect estimate and $\hat{v}_{Sj}$ is an estimated standard error of the sib-GWAS allelic effect estimate at locus j.

Similarly, we can estimate the variance component of the PGS attributable to direct effects along the ith principal component of the genotype matrix of the prediction sample, $c_{Di}=\lambda_i(\sum_{j=1}^{\ell}\beta_{Dj}U_{ij})^2$, by starting with the projection of the sib-GWAS allelic effect estimates on principal component i. Under the same assumptions about the measurement errors in the sib-GWAS allelic effect estimates used immediately above (i.e. that the measurement errors have expectation 0 and $\sum_{j=1}^{\ell}\sum_{k\neq j}U_{ij}U_{ik}\mathrm{E}[\epsilon_{S_j}\epsilon_{S_k}]=0$), we have

$$\mathrm{E}\left[\lambda_i\left(\sum_{j=1}^{\ell}\beta_{Sj}U_{ij}\right)^2\right]=\lambda_i\mathrm{E}\left[\left(\sum_{j=1}^{\ell}[\beta_{Dj}+\epsilon_{Sj}]U_{ij}\right)^2\right]=\lambda_i\left(\sum_{j=1}^{\ell}\beta_{Dj}U_{ij}\right)^2+\lambda_i\sum_{j=1}^{\ell}U_{ij}^2v_{Sj}^2. \tag{15}$$

Thus, to estimate c_{Di}, we use

$$\hat{c}_{Di}=\lambda_i\left(\sum_{j=1}^{\ell}\beta_{Sj}U_{ij}\right)^2-\lambda_i\sum_{j=1}^{\ell}U_{ij}^2\hat{v}_{Sj}^2. \tag{16}$$

Finally, adapting **Eq. 6**, the variance along the ith principal component of the GWAS-based PGS can be written as

$$\begin{aligned}\lambda_i\left(\sum_{j=1}^{\ell}\beta_{Gj}U_{ij}\right)^2=&c_{Di}+c_{\sigma i}+c_{(D\cdot\sigma)i}+\lambda_i\left(\sum_{j=1}^{\ell}\epsilon_{Gj}U_{ij}\right)^2+\\&2\lambda_i\left(\sum_{j=1}^{\ell}\beta_{Dj}U_{ij}\right)\left(\sum_{j=1}^{\ell}\epsilon_{Gj}U_{ij}\right)+2\lambda_i\left(\sum_{j=1}^{\ell}\sigma_jU_{ij}\right)\left(\sum_{j=1}^{\ell}\epsilon_{Gj}U_{ij}\right).\end{aligned} \tag{17}$$

The final two terms have expectation zero because of the assumption in **Eq. 12** and because all the other variables in the final two terms are treated as fixed. We thus estimate $c_{(D\cdot\sigma)i}$ by plugging in estimators for the other variance components and rearranging, giving

$$\begin{aligned}\hat{c}_{(D\cdot\sigma)i} &= \lambda_i\left(\sum_{j=1}^{\ell}\beta_{Gj}U_{ij}\right)^2 - \hat{c}_{Di} - \hat{c}_{\sigma i} - \lambda_i\sum_{j=1}^{\ell}U_{ij}^2\hat{v}_{Gj}^2 \\ &= \lambda_i\left(\sum_{j=1}^{\ell}\beta_{Gj}U_{ij}\right)^2 - \lambda_i\left(\sum_{j=1}^{\ell}\beta_{Sj}U_{ij}\right)^2 - \lambda_i\left(\sum_{j=1}^{\ell}[\beta_{Gj}-\beta_{Sj}]U_{ij}\right)^2 + 2\lambda_i\sum_{j=1}^{\ell}U_{ij}^2\hat{v}_{Sj}^2 \\ &= 2\lambda_i\left(\sum_{j=1}^{\ell}\beta_{Sj}U_{ij}\right)\left(\left(\sum_{j=1}^{\ell}\beta_{Gj}U_{ij}\right) - \left(\sum_{j=1}^{\ell}\beta_{Sj}U_{ij}\right)\right) + 2\lambda_i\sum_{j=1}^{\ell}U_{ij}^2\hat{v}_{Sj}^2. \end{aligned} \quad (18)$$

As mentioned previously, the estimands c_{Di} and $c_{\sigma i}$ are constrained to be non-negative, whereas $c_{(D\cdot\sigma)i}$ may take positive or negative values. However, the estimators $\hat{c}_{Di}$, $\hat{c}_{\sigma i}$, and $\hat{c}_{(D\cdot\sigma)i}$ presented here can all take positive or negative values.

Large direct variance components as evidence of natural selection

Direct-effect variance components that deviate from the null acceptance regions (green shaded regions in **Fig. 2**) are potentially suggestive of selection: these regions include the center of the distribution of the component under a model in which trait-associated alleles (namely, PGS index SNPs) evolve neutrally. We observed several such cases among top PCs in the PGSs we partitioned (**Files S1-S8**). For example, PC1 of 1KG in a PGS for waist circumference (**Fig. S16D**), consistent with previous reports on selection for pelvic morphology to minimize obstetric obstruction [75] and its impact on ilium morphological development [76], and hip-width proportions [77]. In many instances this signal was accompanied by significantly large SAD variance along the same PC. Such an observation might be expected if both selection and stratification are at play, or if selection acts on both direct and dynastic effects. However, a simulation study we performed suggested that such co-occurrence of large SAD and direct components on a PC may arise in the presence of stratification along that PC, even if there is no selection. (See further discussion in **Text S6** and Zaidi and Mathieson [67] for a similar observation.) We therefore cannot interpret a large direct variance component as necessarily pointing to evidence of natural selection in these cases.

Estimation of the isotropic inflation factor

The presence of isotropic inflation, namely $\alpha \neq 1$, renders the approach in the previous subsection problematic. With α as a free parameter, there are four estimands per principal component (α and the three variance components) and only three pieces of information per principal component (the projections of the sib-GWAS allelic effect estimates, standard-GWAS allelic effect estimates, and their difference on the principal component). However, α should be shared across all principal components, which gives other possibilities for estimation.

Because we are mostly interested in principal-component-specific variance components for the "top"

principal components—i.e. the principal components associated with the largest eigenvalues—and given evidence that other PCs often do not capture meaningful axes of population structure[7]—we adopt a strategy in which we use the principal components with smaller eigenvalues to estimate α, meaning that the σ terms in the top principal components can be interpreted as departures from the relationship between sib-GWAS and standard-GWAS allelic effect estimates on the other components.

In particular, we assume that on the $(k+1)$th to final principal components (where PCs are ranked by eigenvalue), the PC-specific SAD effects σ are 0, yielding

$$\mathrm{Var}[\sum_{i=k+1}^{\min(n,\ell)} \lambda_i(\mathbf{U_i}\cdot\mathbf{b})^2] = \sum_{i=k+1}^{\min(n,\ell)} \lambda_i\left(\left[\mathbf{U_i}\cdot\overrightarrow{\beta_D}\right]^2 + \left[\mathbf{U_i}\cdot\overrightarrow{\epsilon_G}\right]^2 + 2\left[\mathbf{U_i}\cdot\overrightarrow{\beta_D}\right]\left[\mathbf{U_i}\cdot\overrightarrow{\epsilon_G}\right]\right). \tag{19}$$

As before, the third term in parentheses is assumed to have expectation 0; the second one has an expectation that we can estimate using the standard error of the standard-GWAS allelic effect estimates; and by a similar estimation as in the estimator of c_{Di} in **Eq. 16**, we can get an alternative estimate of the variance due to direct effects along PC $i > k$,

$$\hat{c}'_{Di} = \lambda_i\left(\sum_{j=1}^{\ell}\beta_{Gj}U_{ij}\right)^2 - \lambda_i\sum_{j=1}^{\ell}U_{ij}^2\hat{v}_{Gj}^2. \tag{20}$$

Setting the right hand side equal to the other estimator of c_{Di} (**Eq. 16**) gives

$$\lambda_i\left(\sum_{j=1}^{\ell}\beta_{Gj}U_{ij}\right)^2 - \lambda_i\sum_{j=1}^{\ell}U_{ij}^2\hat{v}_{Gj}^2 = \lambda_i\left(\sum_{j=1}^{\ell}\beta_{Sj}U_{ij}\right)^2 - \lambda_i\sum_{j=1}^{\ell}U_{ij}^2\hat{v}_{Sj}^2. \tag{21}$$

Since, instead of observing β_{Si}, we observe $\beta^*_{Si} = \alpha\beta_{Si}$, and obtain estimated squared standard errors for the observed sib-GWAS allelic effect estimates $\hat{v}^{2*}_{Sj} = \alpha^2\hat{v}^2_{Sj}$, then we can write

$$\lambda_i\left(\sum_{j=1}^{\ell}\beta_{Gj}U_{ij}\right)^2 - \lambda_i\sum_{j=1}^{\ell}U_{ij}^2\hat{v}_{Gj}^2 = \alpha^2\left[\lambda_i\left(\sum_{j=1}^{\ell}\beta^*_{Sj}U_{ij}\right)^2 - \lambda_i\sum_{j=1}^{\ell}U_{ij}^2\hat{v}^{2*}_{Sj}\right]. \tag{22}$$

or

$$\hat{c}_{Di} = \alpha^2\hat{c}'_{Di} \qquad \forall i \geq k. \tag{23}$$

This expression suggests that we might estimate α as the square root of the coefficient relating $\hat{c}_{Di}$ to $\hat{c}'_{Di}$ in a no-intercept regression across PCs $i >= k$. Note that both the terms in **Eq. 23** have measurement error— given in **Eq. 42** for $\hat{c}_{Di}$ and **Eq. 43** for $\hat{c}'_{Di}$. We estimate α^2 setting $k = 101$ and using Deming Regression, an errors-in-variables method which accounts for uncertainty in each of the two axes that are considered[78].

What drives isotropic inflation? A major factor may be in forms of population structure confounding that act isotropically or at least along many PCs. For instance, confounding due to stratification can act isotropically when the environmental resemblance among individuals is proportional to their genetic relatedness, i.e. the genetic relatedness matrix. Such covariance between environmental and genetic

features is unlikely to act along the axis of stratification captured by a single PC, instead leading to inflation of standard-GWAS summary statistics with respect to many PCs.

A second source of confounding that may act isotropically is assortative mating, or the tendency toward the formation of mating pairs with similar phenotypes, which plays a role in complex trait variation in human populations[55,79–84]. Similar to stratification, the effects of assortative mating on genetic variance may not localize to an individual principal component. In fact, some forms of assortative mating produce patterns that conform to the isotropic inflation factor assumed here, in which the allelic effect estimate is inflated by a scalar with size dependent on the strength of assortative mating[35,85]

Third, indirect parental, or dynastic, effects are also often thought of as acting isotropically, at least in part[10,16,18,45]. However, the extent to which parental genetic effects impact GWAS allelic effect estimates is still unclear. Recent research indicates that signals presumed to reflect dynastic effects may in fact reflect stratification as well[19].

Fourth, a technical driver of isotropic inflation, discussed above, is index-SNP ascertainment biases. Winner's curse[59] may lead allelic effect estimates to be systematically larger in the sample in which the effects were ascertained as significant. Throughout our analyses here, ascertainment is based on significance of marginal SNP association in the standard GWAS, and so intuition dictates that allelic effect estimates would be larger in the standard GWAS and the isotropic inflation factor would be driven to be larger than 1. However, there are likely additional ascertainment biasing effects at play[15,58]. Empirically, our estimates of α were highly sensitive to the choice of a marginal GWAS association p-value (**Figs. 3, S8-S9, S15-S18**).

One example is clumping, a strategy of accounting for LD in the ascertainment of index SNPs[42,86]. Clumps are sets of contiguous SNPs (in terms of chormosomal position) from which at most one SNP is used as an index SNP, with the rationale being that SNPs of the same clump are in high LD. Typically, in the clumping and thresholding technique, the clumps are chosen, through a greedy algorithm, to be centered around the most signficiant SNPs that are not yet included in other clumps. We hypothesized that because SNPs in the same clump are in LD, using even random SNPs as index SNPs would induce ascertainment bias and increase isotropic inflation, as long as these SNPs were chosen from the "best" clumps, i.e. clumps centered around a highly significant SNP. Indeed, choosing index SNPs in this manner (second column in **Fig. S15**) resulted in comparable isotropic inflation factors to traditional clumping and thresholding (i.e., choosing the most significant SNPs from the clumps containing the most significant SNPs; first column in **Fig. S15**) in two traits we examined.

Our results suggest that ascertainment-related issues are plausibly a major driver of variation in isotropic inflation factor across methods of PGS construction. We discuss the effects of SNP ascertainment in more depth in **Text S6** and **Fig. S17**.

As for variation in isotropic inflation factors across traits, we were not able to identify a strong predictor for it. Isotropic inflation estimates in 1KG Europeans were not significantly correlated with LD Score regression[87]-based SNP heritability estimates across traits (**Fig. S19**). Further, we found that the correlation coefficient between individual trait values and Townsend deprivation index, a composite metric of socioeconomic status, were not correlated with the isotropic inflation factor estimates of the same traits (**Fig. S20**).

Data

UK Biobank. The UK Biobank is a large database with extensive phenotypic and genetic information for over half a million individuals in the UK who were between 40 and 69 years of age at the time of recruitment[37]. The individuals included in this study were those who passed checks excluding individuals who: were identified by the UK Biobank to have sex chromosome aneuploidy, self-reported sex differing from sex determined by genotype data, and participants who withdrew from the study as of the last withdrawal notification delivery (April 24, 2024). We excluded individuals in the standard GWAS cohort who were third-degree relatives or closer as identified by the UK Biobank in data field 22021. We further limited our sample to those individuals who were classified as "White British" (data field 22006) by the UKB. These individuals self identified as both White and British. In addition, these individuals are very tightly clustered in the genetic principal component space[37,88]. Finally, individuals who we identified as having full siblings in the cohort were removed to ensure no sample overlap between the standard-GWAS and sib-GWAS cohort. For each phenotype, we also omitted individuals who had missing data for the phenotype of interest, resulting in between 96,942 and 321,718 individuals across the 17 continuous traits we analyzed in the standard-GWAS. In the sibling cohort, we only included pairs of siblings who both had phenotypic measurements, resulting in between 2,163 and 17,328 sibling pairs across phenotypes. A full account of the traits studied, corresponding UK Biobank data field identifiers, and GWAS sample sizes can be found in **Table S1**.

A few specific phenotypes are worth noting for the transformations and filtering steps. We calculated the phenotype 'years of schooling' by converting the maximum educational attainment of the participants to years following Okbay et al.[89]. For the 'hand grip strength' phenotype, we took the average of the measurement across both hands. Finally for diastolic blood pressure levels, we adjusted the measurements of individuals who were taking blood pressure medication upward by 10 mmHg[90].

From the relatedness information provided by the UK Biobank in Resource 531 and data field 22021, we identified 17,353 White British full-sibling pairs using the following protocol. Groups of siblings were identified as those pairs of individuals who had a kinship coefficient between 0.1768 and 0.3536, as well as an IBS0 value greater than 0.0012, as estimated using the software KING[91]. From each of the resulting family groups we then selected two individuals, resulting in the aforementioned 17,353 sibling pairs where every individual is only a part of a single sibling pair. Each individual identified as a part of a sibling pair were omitted from the standard-GWAS analysis of each trait.

We identified a set of 9,607,691 genetic variants that passed the following quality control filters using the *PLINK 2* software[86] in the standard cohort to include in both the standard-GWASs and sib-GWASs for all traits. Using the set of genotyped and imputed variants, we first filtered out all variants with and INFO score lower than 0.8. We then removed all SNPs with genotype missingness greater than 0.05 using the *PLINK 2* flag `--geno 0.05` and minor allele count greater than 5, *PLINK 2* flag `--mac 5`. Finally, we removed all variants that were not biallelic SNPs (`--snps-only`) and all variants which are far from Hardy-Weinberg equilibrium (`--hwe 1e-10`).

1000 Genomes. In order to estimate the SAD variance in the 17 complex traits of interest in the UK Biobank, we downloaded the whole genome sequences from phase three of the 1000 Genomes (1KG) project at `https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/`. From the initial set

of roughly 80 million SNPs, we extracted the set of overlapping SNPs between those remaining in the UK Biobank which resulted in 6,131,234 SNPs to be included in our analysis. We used these variants for all future analyses of both the entire 1KG cohort ($N = 2,504$ sometimes referenced as the 1KG All) and those individuals in the European superpopulation ($N = 504$, referenced as the 1KG Europeans).

Software availability

All of the summary statistics generated in this study are available on the Harpak Lab website's data tab `https://www.harpaklab.com/data`. The software for SAD variance estimation and a working example are available at `https://github.com/harpak-lab/PGSUS`. GIANT summary statistics were downloaded from `https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files`. EA4 Summary statistics were downloaded from the SSGAC data repository (`https://thessgac.com/`). Finally, AADR data were downloaded from `https://reich.hms.harvard.edu/datasets`. See further details on external data processing in **Text S7**.

References

1. Amit V Khera, Mark Chaffin, Krishna G Aragam, Mary E Haas, Carolina Roselli, Seung Hoan Choi, Pradeep Natarajan, Eric S Lander, Steven A Lubitz, Patrick T Ellinor, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, 50(9):1219–1224, 2018.

2. Cathryn M Lewis and Evangelos Vassos. Polygenic risk scores: from research tools to clinical instruments. *Genome medicine*, 12(1):44, 2020.

3. K Paige Harden, Benjamin W Domingue, Daniel W Belsky, Jason D Boardman, Robert Crosnoe, Margherita Malanchini, Michel Nivard, Elliot M Tucker-Drob, and Kathleen Mullan Harris. Genetic associations with mathematics tracking and persistence in secondary school. *NPJ Science of Learning*, 5(1):1, 2020.

4. Jeremy J Berg and Graham Coop. A population genetic signal of polygenic adaptation. *PLoS Genetics*, 10(8):e1004412, 2014.

5. Megan K Le, Olivia S Smith, Ali Akbari, Arbel Harpak, David Reich, and Vagheesh M Narasimhan. 1,000 ancient genomes uncover 10,000 years of natural selection in Europe. *bioRxiv*, 2023.

6. Bjarni J Vilhjálmsson and Magnus Nordborg. The nature of confounding in genome-wide association studies. *Nature Reviews Genetics*, 14(1):1–2, 2013.

7. Alexander I Young, Stefania Benonisdottir, Molly Przeworski, and Augustine Kong. Deconstructing the sources of genotype-phenotype associations in humans. *Science*, 365(6460):1396–1400, 2019.

8. John W Benning, Jedidiah Carlson, Ruth G Shaw, and Arbel Harpak. Confounding fuels hereditarian fallacies. *bioRxiv*, 2024.

9. Eric S Lander and Nicholas J Schork. Genetic dissection of complex traits. *Focus*, 265(3):2037–458, 2006.

10. Alexander I Young, Seyed Moeen Nehzati, Stefania Benonisdottir, Aysu Okbay, Hariharan Jayashankar, Chanwook Lee, David Cesarini, Daniel J Benjamin, Patrick Turley, and Augustine Kong. Mendelian imputation of parental genotypes improves estimates of direct genetic effects. *Nature Genetics*, 54(6):897–905, 2022.

11. Carl Veller, Molly Przeworski, and Graham Coop. Causal interpretations of family gwas in the presence of heterogeneous effects. *Proceedings of the National Academy of Sciences*, 121(38): e2401379121, 2024. doi: 10.1073/pnas.2401379121.

12. Benjamin W Domingue, Jason Fletcher, Dalton Conley, and Jason D Boardman. Genetic and educational assortative mating among us adults. *Proceedings of the National Academy of Sciences*, 111(22):7996–8000, 2014.

13. Matthew R Robinson, Aaron Kleinman, Mariaelisa Graff, Anna AE Vinkhuyzen, David Couper, Michael B Miller, Wouter J Peyrot, Abdel Abdellaoui, Brendan P Zietsch, Ilja M Nolte, et al. Genetic evidence of assortative mating in humans. *Nature Human Behaviour*, 1(1):0016, 2017.

14. J Graham Ruby, Kevin M Wright, Kristin A Rand, Amir Kermany, Keith Noto, Don Curtis, Neal Varner, Daniel Garrigan, Dmitri Slinkov, Ilya Dorfman, et al. Estimates of the heritability of human longevity are substantially inflated due to assortative mating. *Genetics*, 210(3):1109–1124, 2018.

15. Carl Veller and Graham M Coop. Interpreting population-and family-based genome-wide association studies in the presence of confounding. *PLoS Biology*, 22(4):e3002511, 2024.

16. Augustine Kong, Gudmar Thorleifsson, Michael L Frigge, Bjarni J Vilhjalmsson, Alexander I Young, Thorgeir E Thorgeirsson, Stefania Benonisdottir, Asmundur Oddsson, Bjarni V Halldorsson, Gisli Masson, et al. The nature of nurture: Effects of parental genotypes. *Science*, 359(6374): 424–428, 2018.

17. Daniel W Belsky and K Paige Harden. Phenotypic annotation: using polygenic scores to translate discoveries from genome-wide association studies from the top down. *Current Directions in Psychological Science*, 28(1):82–90, 2019.

18. Arbel Harpak and Michael D Edge. Gwas deems parents guilty by association. *Proceedings of the National Academy of Sciences*, 118(27):e2109433118, 2021.

19. Michel G Nivard, Daniel W Belsky, K Paige Harden, Tina Baier, Ole A Andreassen, Eivind Ystrøm, Elsje van Bergen, and Torkild H Lyngstad. More than nature and nurture, indirect genetic effects on children's academic achievement are consequences of dynastic social processes. *Nature Human Behaviour*, 8(4):771–778, 2024.

20. Jeremy J Berg, Arbel Harpak, Nasa Sinnott-Armstrong, Anja Moltke Joergensen, Hakhamanesh Mostafavi, Yair Field, Evan August Boyle, Xinjun Zhang, Fernando Racimo, Jonathan K Pritchard, et al. Reduced signal for polygenic adaptation of height in uk biobank. *eLife*, 8:e39725, 2019.

21. Mashaal Sohail, Robert M Maier, Andrea Ganna, Alex Bloemendal, Alicia R Martin, Michael C Turchin, Charleston WK Chiang, Joel Hirschhorn, Mark J Daly, Nick Patterson, et al. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife*, 8:e39702, 2019.

22. Hakhamanesh Mostafavi, Arbel Harpak, Ipsita Agarwal, Dalton Conley, Jonathan K Pritchard, and Molly Przeworski. Variable prediction accuracy of polygenic scores within an ancestry group. *eLife*, 9:e48376, 2020.

23. Minhui Chen, Carlo Sidore, Masato Akiyama, Kazuyoshi Ishigaki, Yoichiro Kamatani, David Schlessinger, Francesco Cucca, Yukinori Okada, and Charleston WK Chiang. Evidence of polygenic adaptation in sardinia at height-associated loci ascertained from the biobank japan. *The American Journal of Human Genetics*, 107(1):60–71, 2020.

24. Kangcheng Hou, Ziqi Xu, Yi Ding, Ravi Mandla, Zhuozheng Shi, Kristin Boulier, Arbel Harpak, and Bogdan Pasaniuc. Calibrated prediction intervals for polygenic scores across diverse contexts. *Nature Genetics*, pages 1–11, 2024.

25. Joyce Y Wang, Neeka Lin, Michael Zietz, Jason Mares, Vagheesh M Narasimhan, Paul J Rathouz, and Arbel Harpak. Three open questions in polygenic score portability. *bioRxiv*, pages 2024–08, 2024.

26. Gonçalo R Abecasis, William OC Cookson, and Lon R Cardon. Pedigree tests of transmission disequilibrium. *European Journal of Human Genetics*, 8(7):545–551, 2000.

27. David B Allison and Michael C Neale. Joint tests of linkage and association for quantitative traits. *Theoretical Population Biology*, 60(3):239–251, 2001.

28. Wei-Min Chen and Gonçalo R Abecasis. Family-based association tests for genomewide association scans. *The American Journal of Human Genetics*, 81(5):913–926, 2007.

29. Andrew R Wood, Tonu Esko, Jian Yang, Sailaja Vedantam, Tune H Pers, Stefan Gustafsson, Audrey Y Chu, Karol Estrada, Jian'an Luan, Zoltán Kutalik, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, 46 (11):1173–1186, 2014.

30. 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422):56, 2012.

31. John Novembre and Matthew Stephens. Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, 40(5):646–649, 2008.

32. Swapan Mallick and David Reich. The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes. `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FFIDCW`, 2023. Accessed: 2024-09-17.

33. Loïc Yengo, Sailaja Vedantam, Eirini Marouli, Julia Sidorenko, Eric Bartell, Saori Sakaue, Marielisa Graff, Anders U Eliasen, Yunxuan Jiang, Sridharan Raghavan, et al. A saturated map of common genetic variants associated with human height. *Nature*, 610(7933):704–712, 2022.

34. Aysu Okbay, Yeda Wu, Nancy Wang, Hariharan Jayashankar, Michael Bennett, Seyed Moeen Nehzati, Julia Sidorenko, Hyeokmoon Kweon, Grant Goldman, Tamara Gjorgjieva, et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*, 54(4):437–449, 2022.

35. Tammy Tan, Hariharan Jayashankar, Junming Guan, Seyed Moeen Nehzati, Mahdi Mir, Michael Bennett, Esben Agerbo, Rafael Ahlskog, Ville Pinto de Andrade Anapaz, Bjørn Olav Åsvold, Stefania Benonisdottir, Laxmi Bhatta, Dorret I. Boomsma, Ben Brumpton, Archie Campbell, Christopher F. Chabris, Rosa Cheesman, Zhengming Chen, China Kadoorie Biobank Collaborative Group, Eco de Geus, Erik A. Ehli, Abdelrahman G. Elnahas, Estonian Biobank Research Team, Finngen, Andrea Ganna, Alexandros Giannelis, Liisa Hakaste, Ailin Falkmo Hansen, Alexandra Havdahl, Caroline Hayward, Jouke-Jan Hottenga, Mikkel Aagaard Houmark, Kristian Hveem, Jaakko Kaprio, Arnulf Langhammer, Antti Latvala, James J. Lee, Mikko Lehtovirta, Liming Li, LifeLines Cohort Study, Kuang Lin, Richard Karlsson Linnér, Stefano Lombardi, Nicholas G. Martin, Matt McGue, Sarah E. Medland, Andres Metspalu, Brittany L. Mitchell, Guiyan Ni, Ilja M. Nolte, Matthew T. Oetjens, Sven Oskarsson, Teemu Palviainen, Rashmi B. Prasad, Anu Reigo, Kadri Reis, Julia Sidorenko, Karri Silventoinen, Harold Snieder, Tiinamaija Tuomi, Bjarni J. Vilhjálmsson, Robin G. Walters, Emily A. Willoughby, Bendik S. Winsvold, Eivind Ystrom, Jonathan Flint, Loic Yengo, Peter M. Visscher, Augustine Kong, Elliot M. Tucker-Drob, Richard Border, David Cesarini, Patrick Turley, Aysu Okbay, Daniel J. Benjamin, and Alexander Strudwick Young. Family-gwas reveals effects of environment and mating on genetic associations. *medRxiv*, 2024. doi: 10.1101/2024.10.01.24314703. URL `https://www.medrxiv.org/content/early/2024/10/04/2024.10.01.24314703`.

36. Nick Barton, Joachim Hermisson, and Magnus Nordborg. Why structure matters. *eLife*, 8:e45380, 2019.

37. Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O'Connell, et al. The uk biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209, 2018.

38. Akiko Nagai, Makoto Hirata, Yoichiro Kamatani, Kaori Muto, Koichi Matsuda, Yutaka Kiyohara, Toshiharu Ninomiya, Akiko Tamakoshi, Zentaro Yamagata, Taisei Mushiroda, et al. Overview of the biobank japan project: Study design and profile. *Journal of epidemiology*, 27(Supplement_III): S2–S8, 2017.

39. Mitja I Kurki, Juha Karjalainen, Priit Palta, Timo P Sipilä, Kati Kristiansson, Kati M Donner, Mary P Reeve, Hannele Laivuori, Mervi Aavikko, Mari A Kaunisto, et al. Finngen provides genetic insights from a well-phenotyped isolated population. *Nature*, 613(7944):508–518, 2023.

40. Liis Leitsalu, Toomas Haller, Tõnu Esko, Mari-Liis Tammesoo, Helene Alavere, Harold Snieder, Markus Perola, Pauline C Ng, Reedik Mägi, Lili Milani, et al. Cohort profile: Estonian biobank of the estonian genome center, university of tartu. *International journal of epidemiology*, 44(4): 1137–1147, 2015.

41. Andries T Marees, Hilde De Kluiver, Sven Stringer, Florence Vorspan, Emmanuel Curis, Cynthia Marie-Claire, and Eske M Derks. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*, 27 (2):e1608, 2018.

42. Shing Wan Choi, Timothy Shin-Heng Mak, and Paul F O'Reilly. Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, 15(9):2759–2772, 2020.

43. Scott Hazelhurst, Noah Zaitlen, Bogdan Pasaniuc, Bjarni Vilhjálmsson, and Shaun Aron. Medical population genetics and GWAS for complex diseases 19th april - 24th april 2015 training material archive, March 2022.

44. Benjamin Burrows, Ronald J Knudson, Martha G Cline, and Michael D Lebowitz. Quantitative relationships between cigarette smoking and ventilatory function. *American Review of Respiratory Disease*, 115(2):195–205, 1977.

45. Yuchang Wu, Xiaoyuan Zhong, Yunong Lin, Zijie Zhao, Jiawen Chen, Boyan Zheng, James J Li, Jason M Fletcher, and Qiongshi Lu. Estimating genetic nurture with summary statistics of multigenerational genome-wide association studies. *Proceedings of the National Academy of Sciences*, 118(25):e2023184118, 2021.

46. Manuel Ferrando-Bernal, Colin M Brand, and John A Capra. Inferring human phenotypes using ancient dna: from molecules to populations. *Current Opinion in Genetics & Development*, 90: 102283, 2025.

47. Samantha L Cox, Christopher B Ruff, Robert M Maier, and Iain Mathieson. Genetic contributions to variation in human stature in prehistoric europe. *Proceedings of the National Academy of Sciences*, 116(43):21484–21492, 2019.

48. Samantha L Cox, Hannah M Moots, Jay T Stock, Andrej Shbat, Bárbara D Bitarello, Nicole Nicklisch, Kurt W Alt, Wolfgang Haak, Eva Rosenstock, Christopher B Ruff, et al. Predicting skeletal stature using ancient dna. *American Journal of Biological Anthropology*, 177(1):162–174, 2022.

49. David Gokhman, Keith D Harris, Shai Carmi, and Gili Greenbaum. Predicting the direction of phenotypic difference. *bioRxiv*, pages 2024–02, 2024.

50. Ali Akbari, Alison R Barton, Steven Gazal, Zheng Li, Mohammadreza Kariminejad, Annabel Perry, Yating Zeng, Alissa Mittnik, Nick Patterson, Matthew Mah, et al. Pervasive findings of directional selection realize the promise of ancient dna to elucidate human adaptation. *bioRxiv*, 2024.

51. Evan K. Irving-Pease, Alba Refoyo-Martínez, William Barrie, Andrés Ingason, Alice Pearson, Anders Fischer, Karl-Göran Sjögren, Alma S. Halgren, Ruairidh Macleod, Fabrice Demeter, Rasmus A. Henriksen, Tharsika Vimala, Hugh McColl, Andrew H. Vaughn, Leo Speidel, Aaron J. Stern, Gabriele Scorrano, Abigail Ramsøe, Andrew J. Schork, Anders Rosengren, Lei Zhao, Kristian Kristiansen, Astrid K. N. Iversen, Lars Fugger, Peter H. Sudmant, Daniel J. Lawson, Richard Durbin, Thorfinn Korneliussen, Thomas Werge, Morten E. Allentoft, Martin Sikora, Rasmus Nielsen, Fernando Racimo, and Eske Willerslev. The selection landscape and genetic legacy of ancient Eurasians. *Nature*, 625:312–320, 2024. doi: 10.1038/s41586-023-06705-1. URL `https://doi.org/10.1038/s41586-023-06705-1`.

52. William Barrie, Yaoling Yang, Evan K. Irving-Pease, Kathrine E. Attfield, Gabriele Scorrano, Lise Torp Jensen, Angelos P. Armen, Evangelos Antonios Dimopoulos, Aaron Stern, Alba Refoyo-Martinez, Alice Pearson, Abigail Ramsøe, Charleen Gaunitz, Fabrice Demeter, Marie Louise S. Jørkov, Stig Bermann Møller, Bente Springborg, Lutz Klassen, Inger Marie Hyldgård, Niels Wickmann, Lasse Vinner, Thorfinn Sand Korneliussen, Morten E. Allentoft, Martin Sikora, Kristian Kristiansen, Santiago Rodriguez, Rasmus Nielsen, Astrid K. N. Iversen, Daniel J. Lawson, Lars Fugger, and Eske Willerslev. Elevated genetic risk for multiple sclerosis emerged in Steppe pastoralist populations. *Nature*, 625:321–328, 2024. doi: 10.1038/s41586-023-06618-z. URL `https://doi.org/10.1038/s41586-023-06618-z`.

53. Maryn O Carlson, Daniel P Rice, Jeremy J Berg, and Matthias Steinrücken. Polygenic score accuracy in ancient samples: Quantifying the effects of allelic turnover. *PLoS Genetics*, 18(5): e1010170, 2022.

54. Yi Ding, Kangcheng Hou, Ziqi Xu, Aditya Pimplaskar, Ella Petter, Kristin Boulier, Florian Privé, Bjarni J Vilhjálmsson, Loes M Olde Loohuis, and Bogdan Pasaniuc. Polygenic scoring accuracy varies across the genetic ancestry continuum. *Nature*, 618(7966):774–781, 2023.

55. Tanya B Horwitz, Jared V Balbona, Katie N Paulich, and Matthew C Keller. Evidence of correlations between human partners based on systematic reviews and meta-analyses of 22 traits and uk biobank analysis of 133 traits. *Nature Human Behaviour*, 7(9):1568–1583, 2023.

56. James J Lee, Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, Tuan Anh Nguyen-Viet, Peter Bowers, Julia Sidorenko, Richard Karlsson Linnér, et al. Gene discovery and polygenic prediction from a 1.1-million-person gwas of educational attainment. *Nature Genetics*, 50(8):1112, 2018.

57. Laurence J Howe, Michel G Nivard, Tim T Morris, Ailin F Hansen, Humaira Rasheed, Yoonsu Cho, Geetha Chittoor, Rafael Ahlskog, Penelope A Lind, Teemu Palviainen, et al. Within-sibship

738 genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Ge-*
739 *netics*, 54(5):581–592, 2022.

740 58. Alan J Aw, Jeremy McRae, Elior Rahmani, and Yun S Song. Highly parameterized polygenic
741 scores tend to overfit to population stratification via random effects. *bioRxiv*, pages 2024–01, 2024.

742 59. Rui Xiao and Michael Boehnke. Quantifying and correcting for the winner's curse in genetic
743 association studies. *Genetic Epidemiology: The Official Publication of the International Genetic*
744 *Epidemiology Society*, 33(5):453–462, 2009.

745 60. Cameron Palmer and Itsik Pe'er. Statistical correction of the winner's curse explains replication
746 variability in quantitative trait genome-wide association studies. *PLoS Genetics*, 13(7):e1006916,
747 2017.

748 61. Nick Patterson, Alkes L Price, and David Reich. Population structure and eigenanalysis. *PLoS*
749 *Genetics*, 2(12):e190, 2006.

750 62. Alkes L Price, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and
751 David Reich. Principal components analysis corrects for stratification in genome-wide association
752 studies. *Nature Genetics*, 38(8):904–909, 2006.

753 63. Guy Sella and Nicholas H Barton. Thinking about the evolution of complex traits in the era of
754 genome-wide association studies. *Annual review of genomics and human genetics*, 20(1):461–493,
755 2019.

756 64. Roshni A Patel, Clemens L Weiss, Huisheng Zhu, Hakhamanesh Mostafavi, Yuval B Simons,
757 Jeffrey P Spence, and Jonathan K Pritchard. Conditional frequency spectra as a tool for studying
758 selection on complex traits in biobanks. *bioRxiv*, pages 2024–06, 2024.

759 65. Iain Mathieson and Gil McVean. Differential confounding of rare and common variants in spatially
760 structured populations. *Nature Genetics*, 44(3):243–246, 2012.

761 66. Iain Mathieson and Gil McVean. Demography and the age of rare variants. *PLoS Genetics*, 10(8):
762 e1004528, 2014.

763 67. Arslan A Zaidi and Iain Mathieson. Demographic history mediates the effect of stratification on
764 polygenic scores. *eLife*, 9:e61548, 2020.

765 68. E Koch, NJ Connally, N Baya, MP Reeve, M Daly, B Neale, ES Lander, A Bloemendal, and
766 S Sunyaev. Genetic association data are broadly consistent with stabilizing selection shaping
767 human common diseases and traits. *bioRxiv*, pages 2024–06, 2024.

768 69. Jennifer Blanc and Jeremy J Berg. Testing for differences in polygenic scores in the presence of
769 confounding. *bioRxiv*, 2023.

770 70. Po-Ru Loh, George Tucker, Brendan K Bulik-Sullivan, Bjarni J Vilhjálmsson, Hilary K Finucane,
771 Rany M Salem, Daniel I Chasman, Paul M Ridker, Benjamin M Neale, Bonnie Berger, et al. Effi-
772 cient bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics*,
773 47(3):284–290, 2015.

774 71. Gabriel E Hoffman. Correcting for population structure and kinship using the linear mixed model:
775 theory and extensions. *PloS One*, 8(10):e75707, 2013.

776 72. Joshua G Schraiber, Michael D Edge, and Matt Pennell. Unifying approaches from statistical
777 genetics and phylogenetics for mapping phenotypes in structured populations. *bioRxiv*, pages
778 2024–02, 2024.

779 73. Yiqi Yao and Alejandro Ochoa. Limitations of principal components in quantitative genetic asso-
780 ciation models for human studies. *eLife*, 12:e79238, 2023.

781 74. Yiwei Zhang and Wei Pan. Principal component regression and linear mixed model in association
782 analysis of structured samples: competitors or complements? *Genetic epidemiology*, 39(3):149–155,
783 2015.

784 75. Mihaela Pavličev, Roberto Romero, and Philipp Mitteroecker. Evolution of the human pelvis and
785 obstructed labor: new explanations of an old obstetrical dilemma. *American Journal of Obstetrics*
786 *and Gynecology*, 222(1):3–16, 2020.

787 76. Mariel Young, Daniel Richard, Mark Grabowski, Benjamin M Auerbach, Bernadette S De Bakker,
788 Jaco Hagoort, Pushpanathan Muthuirulan, Vismaya Kharkar, Helen K Kurki, Lia Betti, et al. The
789 developmental impacts of natural selection on human pelvic morphology. *Science Advances*, 8(33):
790 eabq4884, 2022.

791 77. Eucharist Kun, Emily M Javan, Olivia Smith, Faris Gulamali, Javier de la Fuente, Brianna I
792 Flynn, Kushal Vajrala, Zoe Trutner, Prakash Jayakumar, Elliot M Tucker-Drob, et al. The genetic
793 architecture and evolution of the human skeletal form. *Science*, 381(6655):eadf8009, 2023.

794 78. Robert James Adcock. A problem in least squares. *The Analyst*, 5(2):53–54, 1878.

795 79. Karl Pearson and Alice Lee. On the laws of inheritance in man: I. inheritance of physical characters.
796 *Biometrika*, 2(4):357–462, 1903.

797 80. Ronald A Fisher. Xv.—the correlation between relatives on the supposition of mendelian inher-
798 itance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2):
799 399–433, 1919.

800 81. Sewall Wright. Systems of mating. iii. assortative mating based on somatic resemblance. *Genetics*,
801 6(2):144, 1921.

802 82. Dalton Conley, Thomas Laidley, Daniel W Belsky, Jason M Fletcher, Jason D Boardman, and
803 Benjamin W Domingue. Assortative mating and differential fertility by phenotype and genotype

804 across the 20th century. *Proceedings of the National Academy of Sciences*, 113(24):6647–6652,
805 2016.

806 83. Xiaoyin Li, Susan Redline, Xiang Zhang, Scott Williams, and Xiaofeng Zhu. Height associated
807 variants demonstrate assortative mating in human populations. *Scientific Reports*, 7(1):15689,
808 2017.

809 84. Loic Yengo, Matthew R Robinson, Matthew C Keller, Kathryn E Kemper, Yuanhao Yang, Maciej
810 Trzaskowski, Jacob Gratten, Patrick Turley, David Cesarini, Daniel J Benjamin, et al. Imprint of
811 assortative mating on the human genome. *Nature Human Behaviour*, 2(12):948–954, 2018.

812 85. Alexander Strudwick Young. Estimation of indirect genetic effects and heritability under assorta-
813 tive mating. *bioRxiv*, 2023. doi: 10.1101/2023.07.10.548458. URL `https://www.biorxiv.org/`
814 `content/early/2023/07/11/2023.07.10.548458`.

815 86. Christopher C Chang, Carson C Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M
816 Purcell, and James J Lee. Second-generation PLINK: rising to the challenge of larger and richer
817 datasets. *Gigascience*, 4(1):s13742–015, 2015.

818 87. Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Schizophre-
819 nia Working Group of the Psychiatric Genomics Consortium, Nick Patterson, Mark J Daly, Alkes L
820 Price, and Benjamin M Neale. Ld score regression distinguishes confounding from polygenicity in
821 genome-wide association studies. *Nature Genetics*, 47(3):291–295, 2015.

822 88. Simon Haworth, Ruth Mitchell, Laura Corbin, Kaitlin H Wade, Tom Dudding, Ashley Budu-
823 Aggrey, David Carslake, Gibran Hemani, Lavinia Paternoster, George Davey Smith, et al. Apparent
824 latent structure within the uk biobank sample has implications for epidemiological analysis. *Nature*
825 *Communications*, 10(1):333, 2019.

826 89. Aysu Okbay, Jonathan P Beauchamp, Mark Alan Fontana, James J Lee, Tune H Pers, Cornelius A
827 Rietveld, Patrick Turley, Guo-Bo Chen, Valur Emilsson, S Fleur W Meddens, et al. Genome-wide
828 association study identifies 74 loci associated with educational attainment. *Nature*, 533(7604):
829 539–542, 2016.

830 90. Evangelos Evangelou, Helen R Warren, David Mosen-Ansorena, Borbala Mifsud, Raha Pazoki,
831 He Gao, Georgios Ntritsos, Niki Dimou, Claudia P Cabrera, Ibrahim Karaman, et al. Genetic
832 analysis of over 1 million people identifies 535 new loci associated with blood pressure traits.
833 *Nature Genetics*, 50(10):1412–1425, 2018.

834 91. Ani Manichaikul, Josyf C Mychaleckyj, Stephen S Rich, Kathy Daly, Michèle Sale, and Wei-Min
835 Chen. Robust relationship inference in genome-wide association studies. *Bioinformatics*, 26(22):
836 2867–2873, 2010.