ARTICLE

Negative selection on complex traits limits phenotype prediction accuracy between populations

Arun Durvasula[1] and Kirk E. Lohmueller[1,2,3,*]

Summary

Phenotype prediction is a key goal for medical genetics. Unfortunately, most genome-wide association studies are done in European populations, which reduces the accuracy of predictions via polygenic scores in non-European populations. Here, we use population genetic models to show that human demographic history and negative selection on complex traits can result in population-specific genetic architectures. For traits where alleles with the largest effect on the trait are under the strongest negative selection, approximately half of the heritability can be accounted for by variants in Europe that are absent from Africa, leading to poor performance in phenotype prediction across these populations. Further, under such a model, individuals in the tails of the genetic risk distribution may not be identified via polygenic scores generated in another population. We empirically test these predictions by building a model to stratify heritability between European-specific and shared variants and applied it to 37 traits and diseases in the UK Biobank. Across these phenotypes, ∼30% of the heritability comes from European-specific variants. We conclude that genetic association studies need to include more diverse populations to enable the utility of phenotype prediction in all populations.

Introduction

The past decade of genome wide association studies (GWASs) has uncovered a plethora of trait-associated loci scattered across the genome.[1–4] Geneticists have devoted many resources to turning these associations into phenotype prediction models that aggregate variants across the genome into a polygenic score. Such scores can be used to guide healthcare decisions for a variety of traits and diseases,[5] and recent work has suggested these polygenic scores may be ready for clinical use.[6,7] While individuals with high polygenic risk for diseases have been found via these scores, for example in atherosclerosis[8] and breast cancer,[9] challenges remain in applying these polygenic scores uniformly across populations. Recent analyses have suggested that because many of the largest studies are concentrated on European populations, polygenic scores may be biased and less informative in non-European populations.[10–15] There are several reasons why polygenic scores may not transfer well across populations. One possibility is that alleles have different effect sizes in different populations, owing to differences in interactions with the environment.[16] Another possibility is that differences in linkage disequilibrium (LD) between variants across populations means that causal variants may be tagged differently in non-European populations, leading to differences in effect sizes.[11,17] Finally, the original polygenic score performance in Europeans may be inflated because of population stratification.[18,19]

Here, we propose that an additional reason for the lack of transferability of polygenic scores is that each population has its own genetic architecture, owing to the evolutionary processes that give rise to traits. Under this reasoning, a population's demographic history influences the number of causal variants and their frequencies, resulting in some phenotypic variance coming from causal variants that are population specific. For example, work on the genetic architecture of skin color in African populations has uncovered distinct loci affecting the trait in each population, suggesting that populations with independent demographic histories can end up with different genetic architectures and causal variants for the same traits.[20] Indeed, modeling work suggests that genetic architecture is an outcome of the evolutionary process rather than a trait-specific property.[21]

Recent exponential growth in human populations has created an excess of new variants that tend to be low frequency and population specific (private variation[22–24]). Population genetic models of genetic architecture that include negative selection suggest that, in aggregate, low-frequency variants could contribute substantially to traits.[25–27] Application of these models to large-scale genetic datasets has discovered that many traits are under apparent negative selection, ranging from anthropometric traits to molecular phenotypes.[28–33] Depending on the interplay between allele frequency and effect size, these variants could make up a large portion of the heritability for many traits, as demonstrated by a recent GWAS on height and BMI using whole-genome sequencing data.[34,35] Because narrow-sense heritability is the proportion of variance explained by additive genetic factors, it is directly related to the accuracy of phenotypic prediction as the variance explained by the polygenic score.[36] If these private variants contribute substantially to heritability, it

[1]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA 90095, USA; [3]Interdepartmental Program in Bioinformatics, University of California, Los Angeles, Los Angeles, CA 90095, USA
*Correspondence: klohmueller@ucla.edu
https://doi.org/10.1016/j.ajhg.2021.02.013.

follows that the variants will not be useful for phenotype prediction between populations because they are not present in other populations. The proportion of narrow-sense heritability that private variants explain places an upper bound on the accuracy of polygenic scores between populations.

In this study, we use simulations under demographic scenarios of recent explosive population growth with varying amounts of negative selection as well as analyses of empirical data to test the role of private variants in complex traits.

Material and methods

Population genetic modeling and simulations

We performed forward simulations by using SLiM v.3.[37] We simulated a demographic history for a European and an African population according to the demographic model fit by Gravel et al.[38] (including migration). The African population size expanded to 14,474 individuals and the European population began at a size of 1,032 individuals after splitting from Africa and grew exponentially at a rate of 0.38% per generation for 920 generations. We simulated a mutational target size of 5 Mb with a mutation rate of 1.2×10^{-8} per base pair (bp) and a recombination rate of 1×10^{-8} per bp. To simulate selection across the entire region, we drew selection coefficients for new mutations from a gamma distribution with parameters fit by Kim et al.[39] (mean = −0.01026, α = 0.186). We sampled 10,000 haploid genomes from each population. To simulate a quantitative trait, we followed the model described by Eyre-Walker[25] and the framework set by Lohmueller[21] where a SNP's effect on a trait, β, is given by

$$\beta = \delta s^\tau (1 + \varepsilon) C,$$

where $\delta \in \{-1, 1\}$ with equal probability, $\varepsilon \sim N(0, 0.5)$, and s is the selection coefficient of a variant segregating in the population at the end of the simulation. C is a scaling factor for effect sizes and controls the heritability for a given mutational target size. In these simulations, C was set to obtain a heritability of $\sim$0.4 (see Table S1). Finally, τ reflects the relationship between a SNP's effect on fitness and the trait. $\tau = 0$ indicates no relationship between fitness and the trait, while $\tau > 0$ indicates that mutations that are more evolutionarily deleterious are those that have larger effects on the trait. In this model, when $\tau > 0$, the trait itself may be under direct selection or it may be correlated with a trait under selection. We call variants private and shared on the basis of their allele frequency in a sample of 10,000 chromosomes from both populations (see below).

To compare our simulation results to the empirical data from the Exome Aggregation Consortium (ExAC), which includes African American individuals, we computed the expected allele frequency for SNP i in simulated admixed African American individuals $(p_{AA, i})$ as

$$p_{AA,i} = \alpha_i p_{EUR,i} + (1 - \alpha_i) p_{AFR,i}$$

where $p_{EUR, i}$ and $p_{AFR, i}$ denote the allele frequencies in Europe and Africa, respectively. For each SNP, we drew an admixture proportion $\alpha_i \sim Beta(2, 8)$ in order to incorporate variance in the admixture proportion along the genome. The parameters of the beta distribution were chosen to match the observed variation in admixture proportion in African American individuals[40] and result in a mean proportion of African ancestry of 80%.

Defining the proportion of heritability from private variants: $h^2_{private}$

We begin by describing a model in which an individual, i, in a population, ϕ, has a phenotype, y_i, that is a linear combination of genotypes ($\boldsymbol{x_i}$, $x_{ij} \in \{x_{i1}, ..., x_{iM}\}$), effect sizes ($\boldsymbol{\beta}$, $\beta_j \in \{\beta_1, ..., \beta_M\}$), and a normally distributed term describing the effect of the environment, $e_i \sim N(0, V_E)$:

$$y_i = \boldsymbol{x_i^T} \boldsymbol{\beta} + e_i.$$

The narrow-sense heritability, h^2, of the phenotype, $\boldsymbol{y}$, in the population is given by

$$h^2 = \frac{V_A}{Var(\boldsymbol{y})}$$

where the variance of the phenotype can be decomposed into additive, dominance, interacting, and environmental terms: $Var(\boldsymbol{y}) = V_A + V_D + V_I + V_E$. The additive genetic variance is $V_A = 2 \sum_{j=1}^{M} p_j(1 - p_j)\beta_j^2$ when there are M variants, where p_j is the allele frequency for variant j and β_j is the effect size of variant j.

We wish to examine the proportion of heritability that comes from a particular class of variants. Consider a sister population, ψ, that diverged from the population described above (ϕ). Variants in population ϕ can be partitioned into those that appear only in ϕ (private variants) or those that appear in both populations (shared variants). The total number of variants is the sum of the number of shared and number of private variants, $M = M_p + M_s$. We wish to partition the heritability into these two classes, h_p^2 and h_s^2, which make up the total heritability: $h^2 = h_p^2 + h_s^2$. Define $h^2_{private}$ to be the proportion of the heritability accounted for by the private variants.

The quantity of interest, then, is

$$h^2_{private} = \frac{h_p^2}{h^2} = \frac{V_{A,p}}{V_A}.$$

The additive genetic variance from private variants is $V_{A,p} = 2 \sum_{j=1}^{M} p_j(1 - p_j)\beta_j^2 z_j$, where z_j is an indicator function that is 1 when the variant j is private (with probability $P(\omega_j)$) to the population and 0 otherwise. We describe how z_j is estimated below when analyzing empirical data (see model to identify private variants).

Polygenic score calculation

We compute three sets of polygenic scores on the simulated individuals: (1) using all variants, (2) using variants private to the simulated population of interest, and (3) using variants shared between the simulated European and African populations. For each haploid genome, we sum the effect sizes, β, for each class of variants, resulting in three scores for each genome. We standardize the scores by subtracting the mean of the true polygenic score (class 1) and dividing by the standard deviation of the true polygenic score (class 1). We compute the Pearson correlation between classes 1 and 2 as well as classes 1 and 3 and report the r^2 value as a percentage.

Model to identify private variants

When analyzing the empirical UK Biobank data, it is challenging to assess whether a particular variant is private or shared. If a variant is seen only in one population, it is possible that it is truly private to that population, or instead, it is shared but at too low a frequency to have been discovered with the number of individuals

samples from the other population. To address this issue, we built a probabilistic model to evaluate the probability that a variant is private to a population given the number of copies of the allele in that population (that is, the allele frequency).

We begin with the intuition that rare alleles tend to be private and common alleles tend to be shared between populations, even in the presence of migration. Migration can be thought of as sampling alleles from one population and placing them in the other population. Under this model, rare alleles will tend to stay within a population and not transfer between populations. This suggests that allele frequency is informative in determining whether an allele is private or not.

Wakeley and Hey[41] use coalescent theory to determine the frequency spectrum of private variants. An application of Bayes' rule allows us to calculate the following probability:

$$P(\omega\big|i) = \frac{P(i|\omega)P(\omega)}{P(i)},$$

where $i \in \{1, ..., n\}$ is the number of copies of the allele in the sample (n) and $\omega \in \{0, 1\}$ is 1 if the allele is private and 0 if not. $P(i|\omega = 1)$ is the site frequency spectrum of private variants, and $P(i)$ is given by the full site frequency spectrum. For example, in a constant-sized equilibrium population, $P(i) = (\theta/i)/(\sum_{i}^{n} \theta/i)$. $P(\omega)$ is the probability of a variant's being private to a population.

Wakeley and Hey[41] provide expressions to obtain these quantities in a constant-sized equilibrium population without natural selection. However, here we are concerned with populations that are not in equilibrium and with variants under negative selection, so we obtain these probabilities via simulation under a particular demographic model and distribution of fitness effects.

In the results presented here, we use the demographic model from Gravel et al.[38] that relates European and African populations. We use a distribution of fitness effects from Kim et al.,[39] assuming that mutations are additive (that is, $h = 0.5$) and that selection coefficients, s, are drawn from a gamma distribution with mean = -0.01026 and shape = 0.186. Using these parameters, we simulate data for 10,000 European chromosomes by using SLiM[37] and compute (1) the proportional site frequency spectrum for private variants ($P(i|\omega)$), (2) the proportional site frequency spectrum for all variants ($P(i)$), and (3) the proportion of private variants ($P(\omega)$). We defined private variants in the simulation as those that appear in the simulated European population but not the simulated African population.

Next, we store these quantities in a lookup table and use them to compute the probability that a variant is private given the number of copies of the allele in the empirical data. In the UK Biobank dataset, alleles are present at frequency 1×10^{-6} and higher. However, in simulations, the lowest allele frequency is 1×10^{-4}. For alleles below this frequency, we set the probability equal to the probability for alleles at a frequency of 1 in 10,000.

Testing our probabilistic model to infer private variants

We evaluated the ability of our model to distinguish between private and shared variants by simulating new data and performing binary classification, calling a variant private if the $P(\omega|i)$ exceeded some threshold, t. We varied this threshold and computed the number of true positive (private variants that are truly private), false positives (private variants that are truly shared), false negatives (shared variants that are truly private), and true negatives (shared variants that are truly shared). We summarized this by us-

ing receiver operator characteristic and precision recall curves (Figure S1; Tables S2 and S3).

We also validated our model by using data from ExAC.[42] For each variant in ExAC, we used our model to compute the probability that the variant is private to the non-Finnish European population on the basis of the allele frequency in that population. Then, we checked whether variants were observed in a sample of 10,406 African and African American samples.

Partitioning heritability

We applied our Bayesian model to predict which variants are private to GWAS summary statistics from 37 traits in the UK Biobank released by the Neale lab (see web resources). We computed the additive genetic variance for variants with a high posterior probability of being private to the British cohort and divided that by the total amount of additive genetic variance explained by SNPs to obtain our estimate of $h^2_{private}$ (Note S1). We also performed the inference by using a randomized algorithm to correct for the effects of LD and misestimated effect sizes as well as population stratification (Notes S2, S3, S4, and S5; Figures S3, S4, S5, S6, S7, S8, and S9). Finally, we also independently replicated the results on BMI by using data from the GIANT consortium[43] (Note S1). Importantly, this partitioning of the heritability into shared and private components does not make use of the τ-model[25] that relates a mutation's effect on fitness to its effect on the trait.

Results

The distribution of European-specific variants in data and models

We begin by precisely defining private variants in the datasets and models that we consider. Studies of genomic variation point to the out-of-Africa bottleneck and subsequent explosive growth in population size as a key driver of the distribution of genomic variation. We focus on a simplified model of this history (Figure 1A; Gravel et al.[38]). We define private variants as those that are found in Europe but are absent from Africa and shared variants as those that are found in both populations. Note that by our definition, private variants may be shared between other out-of-Africa populations (e.g., between Europe and East Asia) because of shared recent history.

One potential concern with this definition of whether a variant is private to Europe is that it may depend on the sample size of the African population used in the comparison. We examined this possibility by computing the probability of not observing an allele present in a sample of African individuals across a range of minor allele frequencies (MAFs) with a sample size of 10,000 chromosomes. This sample size is approximately similar to the sample size of the ExAC dataset (Lek et al.[42]). We find that variants with a frequency as low as 10^{-3} in the African population have a nearly 100% probability of being sampled in ExAC (Figure S2). Thus, we would correctly classify variants segregating at low frequency in Africa as being shared.

Next, we examined the number of private variants in European populations compared to African populations in
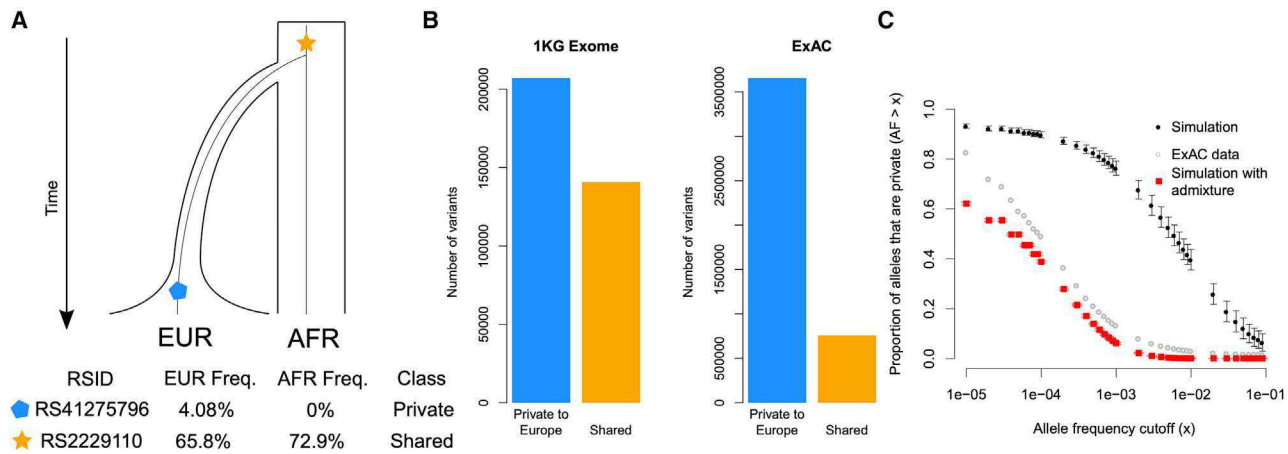
Figure 1. Human population history generates population-specific variants

(A) Model for variants that are shared (common to Europe [EUR] and Africa [AFR]) and private (occurring only in EUR and absent from AFR). Bottom, examples of private and shared variants from ExAC.[42]

(B) The number of non-synonymous variants that are private to European populations and absent from African populations (blue bars) and the number of non-synonymous variants that are shared between the two populations in the 1KG exome dataset and the ExAC dataset (orange bars).

(C) The proportion of non-synonymous alleles above a given frequency that are private to Europe and absent from Africa in the ExAC dataset and in simulations based on human history. Note that because the ExAC dataset contains admixed African American individuals, the proportion of private variants is reduced compared with the original simulation (black dots). Modeling this admixture (red dots) shows a better fit to this dataset. Error bars denote standard deviation across simulation replicates.

two datasets: the 1000 Genomes (1KG) data and the ExAC data. In order to meaningfully compare the two datasets, we focused on variants contained in the exome. For both datasets, there are many more private variants in the European population compared to shared variants (Figure 1B). This is expected under models of human history where many shared alleles were lost during the out-of-Africa bottleneck and new mutations accumulated independently in the out-of-Africa population. Because of the small population size, some of these mutations could drift to a higher frequency than they would have in a larger population.

We next conducted simulations under this model of human evolution, where an ancestral population splits into a group that underwent a genetic bottleneck out of Africa (representing a European population) and a group that stayed within Africa without a bottleneck (representing an African population; Figure 1A [38]), coupled with varying levels of negative selection on traits (including no negative selection). We include negative selection by modifying the relationship between a mutation's effect on the trait and its effect on reproductive fitness by using the model put forth by Eyre-Walker in 2010[25] (see material and methods). This model includes a parameter, τ, which ties the selection coefficient of a mutation to its effect on a trait.[25] Larger values of τ imply that more evolutionarily deleterious mutations have larger effects on the trait. Importantly, our model includes exponential growth in the out-of-Africa population, which creates an excess of private variants, as well as low levels of migration between the European and African populations, which can turn some private variants into shared variants. We compared our simulations to data from ExAC and found that our simulations predicted

more higher-frequency private alleles than are observed in the data (Figure 1C). However, the ExAC data contains admixed African American individuals. Admixture can introduce variants that are private to Europe into the sample labeled "African." We simulated this admixture process (see material and methods) and found that the resulting simulation matches the data closely, suggesting that our model is a reasonable approximation of human demography and selection (Figure 1C).

Population genetic models predict population-specific variants account for heritability and impact polygenic scores

We reasoned that since there are many private causal variants in our simulations, they may account for a substantial proportion of the heritability in aggregate. We examined the contribution of private variants to heritability and found that when traits are not tied to fitness ($\tau = 0$), private variants account for ~30% of the heritability (Figure 2A). However, when the coupling between trait effects and fitness effects is moderate ($\tau = 0.25$) or strong ($\tau = 0.5$), private variants account for over half of the heritability, and there is a maximum of ~79% under strong coupling (Figures 2B and 2C). These results suggest that many causal variants, which jointly explain much of the heritability, tend to be population specific. This effect is a consequence of how the trait relates to fitness as well as the demographic history of the population.

The fact that many of the variants that affect the trait are not shared across populations may limit the applicability of polygenic scores derived from European populations to other populations. This effect would be distinct from
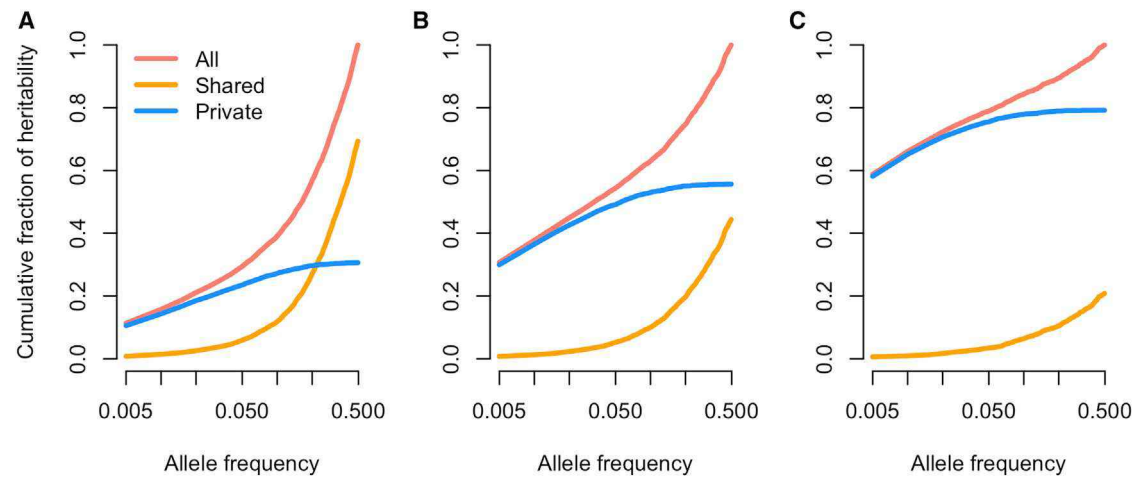
Figure 2. The effect of natural selection on the relationship between heritability and allele frequency
(A–C) Cumulative fraction of heritability explained by private and shared variants under (A) no relation between a mutation's effect on fitness and the trait ($\tau = 0$), (B) moderate coupling between a mutation's effect on fitness and the trait ($\tau = 0.25$), and (C) strong coupling between a mutation's effect on fitness and the trait($\tau = 0.5$). Note that the x axis is on a log scale. As τ increases, a greater fraction of heritability comes from variation that is found only within Europe.

imperfect tagging of causal variants due to differences in LD patterns between populations. To test for this effect in simulated data, we calculated true polygenic scores for individuals in the simulated European and African populations and asked how well polygenic scores derived from only private variants and only shared variants correlated with the true polygenic scores. Polygenic scores derived from only shared variants represent the case where a polygenic score can be transferred from Europe to another population. If shared variant effect sizes correlate well between populations, despite not contributing to a majority of additive genetic variance, polygenic scores may still be accurate across populations. We note that these simulations include identification of the true causal SNPs and, as such, are much higher than polygenic score accuracies reported elsewhere.[13] These simulations represent the best-case scenario for polygenic scores. We found that when traits are independent of fitness, the shared polygenic score has a 91% correlation in Europe and 96% correlation in Africa with the true polygenic score, suggesting that polygenic scores can be applied between populations (Figure 3A). However, we found that when trait effects are tied to fitness effects, the correlation between shared polygenic scores and the true polygenic scores decreases (Figures 3B and 3C) and the correlation between private polygenic scores and true polygenic scores increases (Figures 3D, 3E, and 3F). Note that in the analysis with private polygenic scores, each population uses variants private to that population but not from the other population. That is, the African private polygenic score uses variants private to Africa. This suggests that the reduction in accuracy does not depend on the population's specific demography, as the same pattern is present in European and African populations. For traits with strong coupling between trait effects and fitness effects ($\tau = 0.5$), the correlation between the

true polygenic scores and the polygenic scores derived from shared variants drops to 62% in Europe and 57% in Africa (Table S4). These findings suggest that polygenic scores based solely on shared variants may be substantially less accurate than polygenic scores using all variants and may not transfer between populations well when the variants with the greatest effects on the trait are those under the most negative selection.

While shared variants do not capture the full distribution of polygenic scores, we asked whether individuals in the tail of the true polygenic score distribution remained in the tail when examining shared variants only. When there is no coupling between fitness and trait effects ($\tau = 0$), shared variants capture 35% of the tail correctly in Europe and 28% of the tail correctly in Africa (Table 1). However, when there is moderate coupling ($\tau = 0.25$), this number drops to 11% in Europe and 7% in Africa. When there is strong coupling, the polygenic score based on shared variants identifies none of the individuals in the tails of the distribution. If the trait under consideration is a disease, this analysis suggests that a polygenic score based on shared variation cannot identify individuals at the highest risk for that disease. In contrast, when considering only private variants, the polygenic score correctly identifies 44%–46% of individuals who are at the extremes of the distribution. These results suggest that when using scores derived from European populations, individuals who are truly in the tails of the polygenic score distribution will not be identified via shared variants alone, corresponding to a high false-negative error rate. In addition, the low recall for both of these polygenic scores suggests many individuals that are in the tails of the distribution will be missed.

While our simulations suggest private variants may be an important component of the heritability and may limit
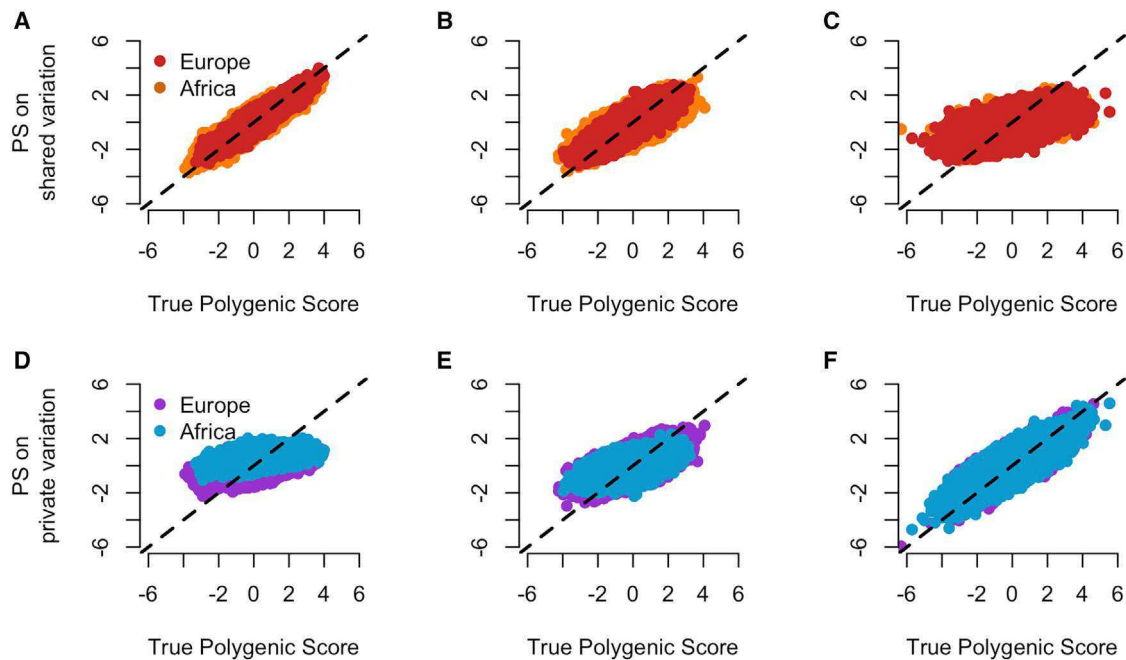
Figure 3. The relationship between polygenic scores and natural selection
(A–F) Polygenic score accuracy for shared variants only (top row) and private variants only (bottom row) in Europe and Africa on simulated data with different degrees of negative selection. In the bottom row, each score uses private variants from within the population being considered (e.g., for Africa, we use variants private to Africa) but not from the other population. The black line shows the 1:1 line. (A and D) No relationship between a mutation's effect on fitness and its effect on the trait ($\tau = 0$). (B and E) Moderate coupling between fitness and trait effects ($\tau = 0.25$). (C and F) Strong coupling ($\tau = 0.5$). As the strength of coupling increases, polygenic scores computed from shared variation become less correlated with the true polygenic score. However, at the same time, polygenic scores computed from private variation become more correlated with the true polygenic scores.

phenotype prediction across populations, their precise role depends on the extent of negative selection acting on traits (either directly or through pleiotropy), which remains an open question.[28–30,32,33] Thus, we next tested how much of the heritability private variants account for in real GWAS data in European populations, where GWAS data is abundant.

A model for private variation

We built a Bayesian model to classify variants segregating in the UK Biobank as private or shared by using the allele frequency conditional on a demographic model and distribution of fitness effects inferred for a European population (see material and methods). To validate our model, we simulated a new dataset under the same European demographic model and recorded whether each allele was observed in both populations. Then, we calculated the probability of each allele's being private to the European population. We classified variants as private if the probability $P(\omega|i) \geq t$, $\omega \in \{0, 1\}$ is 1 if the allele is private and 0 if not, $i \in \{1, ..., n\}$ is the number of copies of the allele in the sample, and t is some probability cutoff. For each cutoff, we calculated (1) the number of variants that we predict are private and are truly private (true positives), (2) the number of variants that we predict are private and are truly not private (false positives), (3) the number of variants that we predict are not private and are truly private (false negatives), and

(4) the number of variants that we predict are not private and are truly not private (true negatives).

We summarize these numbers by using two curves: a precision-recall curve (Figure S1A) and a receiver operator characteristic (ROC) curve (Figure S1B). We find that at a precision of 94%, we have a recall of 99% and that the area under the ROC curve is 0.80, suggesting that our model is able to distinguish between private and shared variants on the basis of allele frequency alone (Table S3). We also tested the model on a simulated dataset including five times more individuals than the 10,000 individuals used in the initial simulation. Importantly, for this comparison, we used the same lookup table, based on 10,000 individuals, as before. This allows us to test how sample size affects our inferences. We find that the precision-recall curve is largely the same, but there is a decrease in the ROC curve (AUROC = 0.70).

In addition, examining $P(\omega|i)$ versus the allele frequency in the simulated independent dataset (Figure S1C), we find that alleles higher than ~10% frequency have a negligible probability of being private. This is consistent with the intuition that common alleles are unlikely to be private.

We also examined several posterior probability thresholds in detail ($t \in \{0.1, 0.23, 0.4\}$; Table S3). Across these thresholds, we find that the false discovery rate (FDR) from simulations is ~5%, suggesting that the model is relatively robust to the threshold used.

Table 1. The effect of natural selection on identifying high-risk individuals

τ	Shared (Europe)	Private (Europe)	Shared (Africa)	Private (Africa)
0	35%	22%	28%	11%
0.25	11%	20%	7%	18%
0.5	0%	46%	0%	44%

Percentage of individuals in the extreme 5% tail of the true polygenic score distribution that are recovered when using only private variants and shared variants in simulated European and African populations. Overall, the percentage of individuals correctly classified is low, suggesting that there will be many false negatives when using polygenic scores to identify individuals in the tails of the risk distribution. Further, as the degree of coupling between fitness effects and trait effects increases, shared variants correctly classify fewer individuals, while private variants classify more individuals correctly.

Next, we empirically validated the performance of our model to infer whether variants are private. Using data from ExAC,[42] we use our framework described above to calculate the probability that each variant is private by using the allele frequency in the non-Finnish Europeans (NFE). In Figure S1D, we plot this probability for a random subset of 10,000 variants. We see that variants above 10% frequency have a very low probability of being private and that variants below that frequency increase in probability of being private as their frequency decreases.

In addition, we classified variants in ExAC as private to "EUR" by using the simulation-based FDR of 5% ($P(\omega|i) \geq 0.23$) and checked whether those variants were present in the "AFR" subset of samples (Table S2). We see that 83% of the variants we call private are not observed in "AFR" in a sample of 10,406 chromosomes. This suggests that our empirical-based FDR is 17% and is higher than the simulation-based FDR. However, the "AFR" sample in ExAC is a mixture of African American and African samples. Importantly, African American samples are admixed between European and African populations.[42] This has the effect of introducing European variants into the "AFR" samples, making variants we expect to be private to "EUR" appear shared. Therefore, this estimate of the accuracy is most likely an underestimate.

Nonetheless, these simulations and empirical evaluations suggest that our model is able to distinguish between private and shared variants on the basis of allele frequency alone. Additionally, out of an abundance of caution, we utilize two different empirically based FDRs of 17% in downstream inferences as described below. Importantly, our determination of whether a variant is private or shared is expected to hold regardless of the sample size taken from either population (Table S2).

Inference of heritability accounted for by private variants: $h^2_{private}$

We used summary statistics for 37 different traits and diseases from the UK Biobank relating to anthropometric and blood-related traits as well as cancer-related and non-cancer related diseases (see web resources) to infer the proportion of the SNP-based heritability attributable to private variants, $h^2_{private}$. Using these data and our probabilistic method to determine whether a variant is private or not, we find that the average $h^2_{private} = 31\%$, and there is substantial variation across traits (standard deviation: 11%; Figure 4). Examining categories of diseases, we find that cancer-related diseases have $h^2_{private} = 12\%$, while non-cancer-related diseases have $h^2_{private} = 32\%$. Similarly, private variants account for ~30% of the heritability in blood-related and anthropometric traits. We observe substantial variability across different traits within a category. Two blood pressure-related traits have $h^2_{private}$ of nearly 50%, while other blood-related traits have a lower proportion.

The effect of falsely identified private variants on our inference of $h^2_{private}$

To ensure that our results from the UK Biobank data described above were not driven by shared SNPs that we mistakenly classified as private, we adjusted for an empirically based FDR. At the threshold used for classifying variants as being private ($P(\omega|i) = 0.23$), validation in the empirical data suggest the FDR is ~17% (see above). In other words, approximately 17% of SNPs that we identify as private may actually be shared. Thus, we adjusted our estimates of $h^2_{private}$ by randomly reclassifying 17% of the private SNPs as shared and re-computed $h^2_{private}$ ("17% FDR correction" in Figure 4). Despite the extremely conservative nature of this correction (because the empirical FDR is based on an admixed sample), we find that a sizeable proportion of the heritability (about 22%) still comes from private variants (Figure 4).

In addition to this conservative correction, we also performed an even more stringent correction where we sorted the SNPs we call private by their heritability and removed 17% of the SNPs that explain the most heritability. As expected, the amount of heritability from private variants goes down, but for most traits, the heritability explained by private variants is still greater than 10% ("Max FDR correction" in Figure 4). This suggests that our central claim, that private variants contribute to heritability, remains true even if our classification method is imperfect.

The effect of population stratification

Recent studies have highlighted the effects of stratification on polygenic scores.[18,19] We considered whether stratification could have an effect on our analyses. To test this, we repeated our analyses by using only those SNPs showing stronger associations with the trait. Specifically, we employed p value cutoffs, using only SNPs with a p value lower than the cutoff (Figure S3). Broadly, for quantitative traits, we observe that as the p value threshold becomes stricter, the proportion of the heritability attributable to private variants decreases. This is due to the power to detect associations for private variants. The power to detect an association will be lower for private variants than shared variants because private variants tend to have lower allele frequencies. Therefore, as the p value cutoff
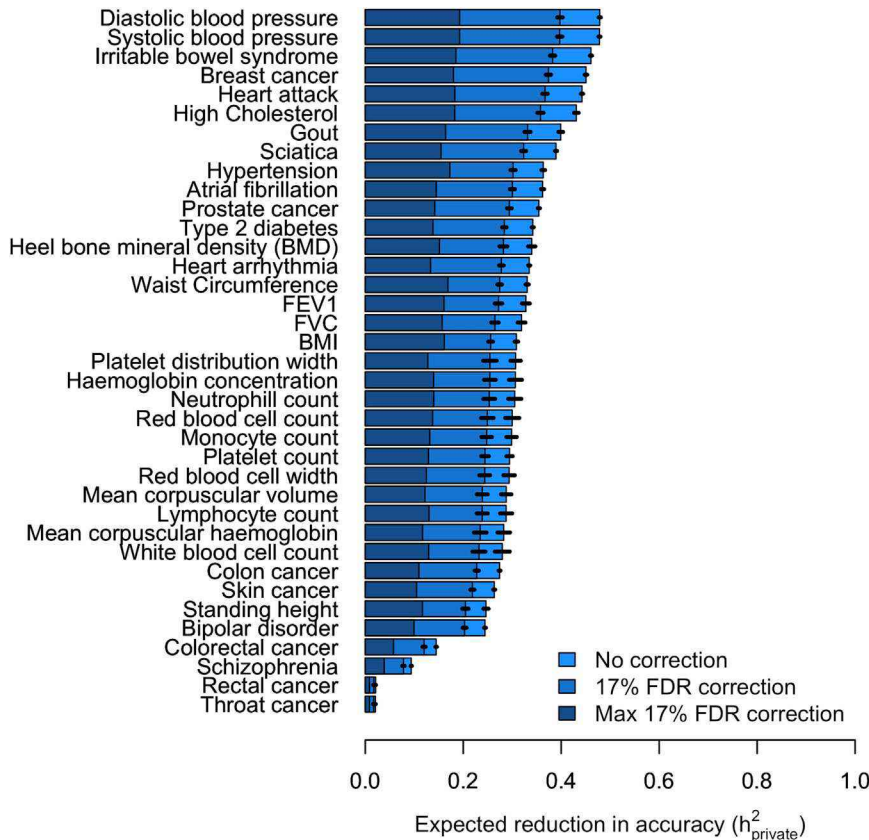
Figure 4. Estimates of the amount of heritability from private variants

The expected reduction in accuracy when transferring a polygenic score from Europe to Africa (expressed as the percentage of heritability explained by private variants) across 37 traits and diseases in the UK Biobank. We only include SNPs with an MAF $> 10^{-3}$. The mean reduction is 26.7% (SD across traits is 14.7%). "17% FDR correction" refers to randomly setting 17% of the SNPs that we call private to shared. "Max 17% FDR correction" refers to setting the 17% of the SNPs that explain the most heritability from private to shared. Lines indicate standard errors obtained via a 1 Mb block jackknife.

will be crucial to understand the full contribution of private variants to heritability.

The effect of unmodeled LD on our inferences

Our inferences of $h^2_{private}$ make the assumption that the estimated effect sizes for the GWAS SNPs were the true effect sizes of the causal variants. Further, we assumed that the variants were all independent of each other. In truth, these assumptions are violated for a variety of reasons. First, because of LD, SNPs may be correlated with one another. Second, some of the non-zero effect sizes of GWAS SNPs may be due to the fact that the GWAS SNP is tagging (in LD with) an untyped causal variant and is itself not causal. Third, even if the GWAS variants analyzed in our study are the true causal variants, their effect sizes may be misestimated by the effects at nearby SNPs in LD with them. Thus, given these challenges, we carefully considered the effect that unmodeled LD may have on our inferences (see Note S2).

First, we developed an estimator of the SNP-based heritability that downsamples the number of SNPs to be independent of each other. We checked the robustness of our results to this effect by randomly selecting a single SNP in a window and computing the proportion of heritability from private variants by using these randomly selected SNPs. We select only one SNP per window to avoid counting SNPs located nearby each other that are in LD with each other. We randomly selected SNPs to avoid biases due to the fact that more sophisticated methods for fine-mapping SNPs by using LD patterns may have different performance for different allele frequencies. We find similar results via our LD-pruned estimator compared with the full data (Note S3; Figures S5, S6, and S7). We also ensure that our estimates are sensible by estimating the proportion of additive genetic variance from variants we infer to be shared (Figure S8). If the inference procedure

decreases, we expect a lower proportion of heritability to come from private variants. We found that the total variance explained by SNPs for dichotomous traits was much lower than for quantitative traits. This effect produced a statistical artifact where the heritability from private variants tended to be very high for dichotomous traits (Figure S4).

In addition to this analysis, we were also concerned with the effect of differential population structure from rare variants.[35,44] Therefore, we checked the robustness of our results to allele frequency filters. We computed $h^2_{private}$ for atrial fibrillation, BMI, standing height, diastolic blood pressure, and type 2 diabetes with MAF cutoffs from 10^{-5}, 10^{-4}, 10^{-3}, and 10^{-2} (Figure 5). We find that although $h^2_{private}$ decreases, it still remains substantial up to a cutoff of 10^{-2}. Although this analysis removes both real and spurious signals, it suggests that private variants do indeed explain a non-negligible proportion of heritability.

Across different p value thresholds, a non-negligible proportion of heritability comes from private variants. However, this analysis does not alleviate all concerns about population stratification, as at a large enough sample size, an association due to stratification can be arbitrarily strong. Similarly, stratification could still occur when using variants at different MAF cutoffs. While these analyses provide evidence that our results are not primarily driven by stratification, they cannot completely rule it out. Further advances in controlling for stratification of rare variants
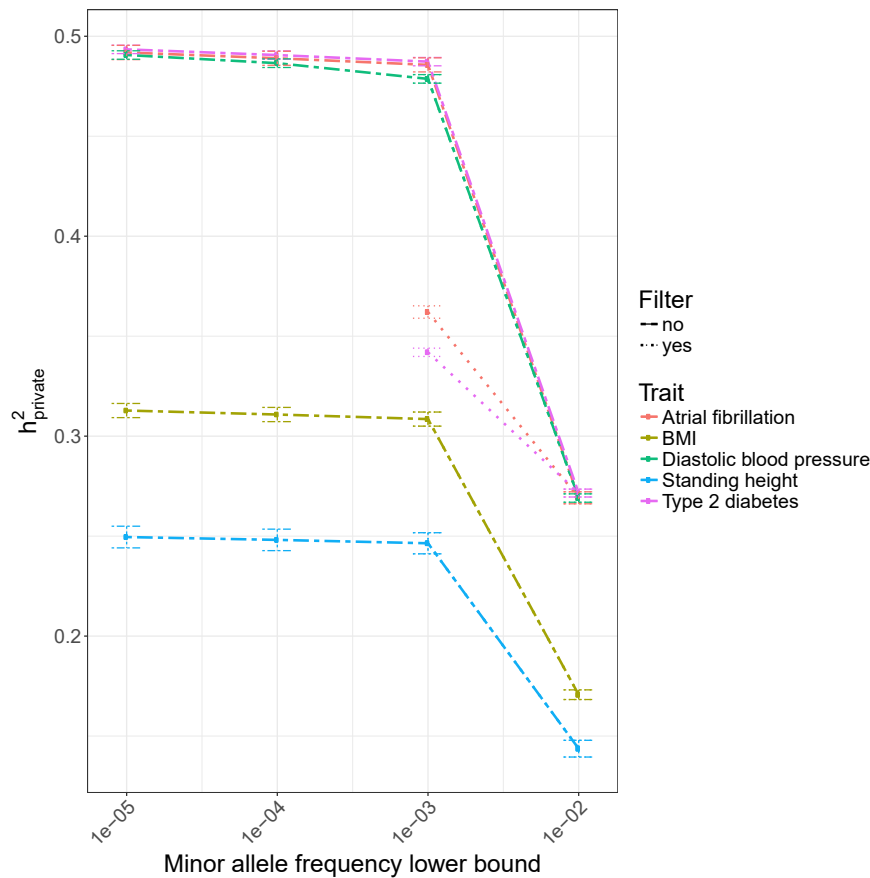
Figure 5. The effect of MAF cutoffs on heritability from private variants

"Filter" refers to a quality and allele frequency filter that removes variants with a frequency below 10^{-3}, a Hardy Weinberg equilibrium p value $< 10^{-10}$, or SNP information score < 0.8. We show results for five commonly studied traits. For BMI, diastolic blood pressure, and standing height, the "filter yes" lines are behind the "filter no" lines. This suggests that the variant filter has no effect on the estimated heritability. Lines indicate standard errors obtained via a 1 Mb block jackknife.

works correctly, this number should be $1 - \widehat{h^2_{private}}$. In Figure S8, we see that this is indeed the case.

Second, based on first principles, our estimates of $h^2{}_{private}$ most likely underestimate the true proportion due to LD between tagging and causal variants (see Note S2). Because shared variants tend to be more common, they will tend to be in LD with more (and therefore tag more) variants. Because of this effect, shared variants could have inflated marginal effect sizes compared to private variants. This would lead to overestimating the heritability from shared variants compared to private variants, making our inferences conservative. We tested for this effect in real data by testing the correlation between marginal effect sizes and recombination rate for variants we predict to be private and variants we predict to be shared (Note S4). We found that variants we predict to be private have lower correlation than variants we predict to be shared, consistent with the idea that shared variants tag more variants than private variants (we also note that this reasoning is the motivation for LD score regression[45]). In addition, coalescent simulations show that our estimator of $h^2{}_{private}$ is indeed slightly downwardly biased (Note S5; Figure S9).

Discussion

In this work, we have shown that recent population growth and negative selection create population-specific genetic architectures for phenotypes, which has the direct effect of reducing the accuracy of polygenic scores when applied between populations. The reduction in accuracy will depend on how differentiated populations are and accuracy decreases as populations become more differentiated. Another case to consider is admixed populations where some causal variants could be introduced and thus become shared variants. In these cases, we expect the utility of polygenic scores to be higher, but this will depend on how recent the admixture was and how many causal variants are transferred between populations, which can vary between individuals.

In our simulation results, we found that when there was no coupling between trait effects and fitness, approximately 30% of the heritability comes from private variants and that this proportion increases as the coupling increases. Although we expect this general pattern to hold, the specific values will depend on the distribution of fitness effect for causal alleles, the mutation target size, and the demographic history of the populations under study. We have used a distribution of fitness effects that was fit to non-synonymous variants[39] and note that the estimates of selection on causal alleles could be revised in future studies. In addition, our model with admixture fits the observed data better than a model without admixture (Figure 1C), but we may still be underestimating the number of private alleles, which would cause our estimates to be a lower bound. Nevertheless, our results suggest that a non-negligible proportion of the heritability comes from private alleles.

We find that phenotypes with a majority of heritability explained by private variants are not likely to be predicted well in non-European populations, even if effect sizes are accurately inferred. Our analysis of the UK Biobank data suggests that most traits examined here have at least 20% of the heritability explained by private variants ($h^2{}_{private} > 20\%$), indicating that cross-population polygenic scores are limited in accuracy and many

population-specific causal variants remain to be discovered. We note that our inferences on the empirical data do not make use of the Eyre-Walker τ model.[25] As such, our inferences from empirical data do not make any assumptions about the relationship between a mutation's effect on fitness and the trait.

At first glance, our result that many traits have a population-specific genetic component seems at odds with recently reported results suggesting that the genetic correlation between traits in European and East Asian populations is very high.[46,47] However, we note that both of these studies examined common variants (MAF > 5%), which are more likely to be shared. Our study explicitly considers a larger range of allele frequencies, which is more likely to include population-specific variants.

Our results have several implications for users of polygenic scores. First, we show that the transferability of polygenic scores depends on the particular trait being examined. For traits with larger values of h^2_{private} (such as diastolic and systolic blood pressure), the transferability would be lower because we find these traits derive more of their heritability from variants that are more likely to be private (h^2_{private}: $\approx 48\%$ for both). In contrast, we find that traits with lower values of h^2_{private}, such as white blood cell count, can be more easily transferred because the heritability is spread more evenly across the spectrum of MAFs (h^2_{private}: 28%). Although we include standard errors estimated via a jackknife, this procedure may not account for all the uncertainty. Therefore, specific differences across traits should be interpreted cautiously. In addition, our inferences, like those in Lam et al.[46] and Liu et al.,[47] focus on the SNP heritability rather than the total heritability of particular traits.

Several recent reviews and commentaries have pointed out the potential for misuse of polygenic scores to justify racism and white supremacy, especially when comparing polygenic scores across populations.[16,48–51] Importantly, although our study indicates that population-specific variants play a role in complex traits, it is incorrect to conclude that population-specific variants lead to differences in traits between populations. Previous simulation studies have suggested that the interplay between demography and negative selection will not lead to large differences in trait heritability between populations.[21,27] Instead, these evolutionary forces can change how the heritability is accounted for. For example, as we show here, population growth and negative selection can lead to heritability's being accounted for by lower-frequency variants that are population specific instead of common variants shared across populations. Further, non-genetic factors most likely play an important role in differences in phenotype between populations.[52]

We also highlight a crucial issue in identifying individuals in the tails of the phenotype distribution. If polygenic scores are to be used more commonly in the clinic, false-negative rates must be more closely examined across populations and phenotypes. Our work suggests that many causal variants may not be shared between populations,

indicating that variants ascertained in European populations may not be informative in other populations. This could occur because, on average, more European-specific variants have been either directly included in GWASs or imputed more often than variants specific to other non-European populations. To ensure equal predictive power of polygenic scores across populations, whole-genome sequencing-based association studies must be undertaken in non-European populations. Such studies would allow for unbiased discovery of private variants accounting for much of the heritability, resulting in improved polygenic prediction in non-European populations. Finally, large imputation panels from the relevant population of interest are necessary to include variation that is not present in Europe.

Data and code availability

The scripts required to carry out the inference of heritability from private variants can be found at https://github.com/LohmuellerLab/PRS.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.02.013.

Declaration of interests

The authors declare no competing interests.

Web resources

GWAS summary statistics, http://www.nealelab.is/uk-biobank

References

1. Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661–678.

2. Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R., Xu, C., Futema, M., Lawson, D., et al. (2015). The UK10K project identifies rare variants in health and disease. Nature *526*, 82–90.

3. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS discovery: Biology, function, and translation. Am. J. Hum. Genet. *101*, 5–22.

4. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209.

5. Vilhjálmsson, B.J., Yang, J., Finucane, H.K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P.R., Bhatia, G., Do, R., et al. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am. J. Hum. Genet. 97, 576–592.

6. Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., and Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat. Genet. 50, 1219–1224.

7. Khera, A.V., Chaffin, M., Wade, K.H., Zahid, S., Brancale, J., Xia, R., Distefano, M., Senol-Cosar, O., Haas, M.E., Bick, A., et al. (2019). Polygenic prediction of weight and obesity trajectories from birth to adulthood. Cell 177, 587–596.e9.

8. Natarajan, P., Young, R., Stitziel, N.O., Padmanabhan, S., Baber, U., Mehran, R., et al. (2017). Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. Circulation 135, 2091–2101.

9. Maas, P., Barrdahl, M., Joshi, A.D., Auer, P.L., Gaudet, M.M., Milne, R.L., Schumacher, F.R., Anderson, W.F., Check, D., Chattopadhyay, S., et al. (2016). Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. JAMA Oncol. 2, 1295–1302.

10. Scutari, M., Mackay, I., and Balding, D. (2016). Using genetic distance to infer the accuracy of genomic prediction. PLoS Genet. 12, e1006288.

11. Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., Daly, M.J., Bustamante, C.D., and Kenny, E.E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. Am. J. Hum. Genet. 100, 635–649.

12. Kim, M.S., Patel, K.P., Teng, A.K., Berens, A.J., and Lachance, J. (2018). Genetic disease risks can be misestimated across global populations. Genome Biol. 19, 179.

13. Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. Nat. Genet. 51, 584–591.

14. Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J.K., and Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. eLife 9, e48376.

15. Ragsdale, A.P., Nelson, D., Gravel, S., and Kelleher, J. (2020). Lessons learned from bugs in models of human history. Am. J. Hum. Genet. 107, 583–588.

16. Novembre, J., and Barton, N.H. (2018). Tread lightly interpreting polygenic tests of selection. Genetics 208, 1351–1355.

17. Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature 570, 514–518.

18. Berg, J.J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A.M., Mostafavi, H., Field, Y., Boyle, E.A., Zhang, X., Racimo, F., Pritchard, J.K., and Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. eLife 8, e39725.

19. Sohail, M., Maier, R.M., Ganna, A., Bloemendal, A., Martin, A.R., Turchin, M.C., Chiang, C.W., Hirschhorn, J., Daly, M.J., Patterson, N., et al. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. eLife 8, e39702.

20. Martin, A.R., Lin, M., Granka, J.M., Myrick, J.W., Liu, X., Sockell, A., Atkinson, E.G., Werely, C.J., Möller, M., Sandhu, M.S., et al. (2017). An unexpectedly complex architecture for skin pigmentation in Africans. Cell 171, 1340–1353.e14.

21. Lohmueller, K.E. (2014). The impact of population demography and selection on the genetic architecture of complex traits. PLoS Genet. 10, e1004379.

22. Keinan, A., and Clark, A.G. (2012). Recent explosive human population growth has resulted in an excess of rare genetic variants. Science 336, 740–743.

23. Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., et al. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337, 64–69.

24. Gao, F., and Keinan, A. (2014). High burden of private mutations due to explosive human population growth and purifying selection. BMC Genomics 15 (Suppl 4), S3.

25. Eyre-Walker, A. (2010). Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proc. Natl. Acad. Sci. USA 107 (Suppl 1), 1752–1756.

26. Sanjak, J.S., Long, A.D., and Thornton, K.R. (2017). A Model of compound heterozygous, loss-of-function alleles is broadly consistent with observations from complex-disease GWAS datasets. PLoS Genet. 13, e1006573.

27. Uricchio, L.H. (2020). Evolutionary perspectives on polygenic selection, missing heritability, and GWAS. Hum. Genet. 139, 5–21.

28. Hernandez, R.D., Uricchio, L.H., Hartman, K., Ye, C., Dahl, A., and Zaitlen, N. (2019). Ultrarare variants drive substantial cis heritability of human gene expression. Nat. Genet. 51, 1349–1355.

29. Gazal, S., Finucane, H.K., Furlotte, N.A., Loh, P.R., Palamara, P.F., Liu, X., Schoech, A., Bulik-Sullivan, B., Neale, B.M., Gusev, A., and Price, A.L. (2017). Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. Nat. Genet. 49, 1421–1427.

30. Gazal, S., Loh, P.R., Finucane, H.K., Ganna, A., Schoech, A., Sunyaev, S., and Price, A.L. (2018). Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. Nat. Genet. 50, 1600–1607.

31. Zeng, J., de Vlaming, R., Wu, Y., Robinson, M.R., Lloyd-Jones, L.R., Yengo, L., Yap, C.X., Xue, A., Sidorenko, J., McRae, A.F., et al. (2018). Signatures of negative selection in the genetic architecture of human complex traits. Nat. Genet. 50, 746–753.

32. Schoech, A.P., Jordan, D.M., Loh, P.R., Gazal, S., O'Connor, L.J., Balick, D.J., Palamara, P.F., Finucane, H.K., Sunyaev, S.R., and Price, A.L. (2019). Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. Nat. Commun. 10, 790.

33. Uricchio, L.H., Kitano, H.C., Gusev, A., and Zaitlen, N.A. (2019). An evolutionary compass for detecting signals of polygenic selection and mutational bias. Evol. Lett. 3, 69–79.

34. Wainschtein, P., Jain, D.P., Yengo, L., Zheng, Z., Cupples, L.A., Shadyab, A.H., McKnight, B., Shoemaker, B.M., Mitchell, B.D., Psaty, B.M., et al. (2019). Recovery of trait heritability from whole genome sequence data. bioRxiv. https://doi.org/10.1101/588020.

35. Young, A.I. (2019). Solving the missing heritability problem. PLoS Genet. *15*, e1008222.

36. de los Campos, G., Gianola, D., and Allison, D.B. (2010). Predicting genetic predisposition in humans: the promise of whole-genome markers. Nat. Rev. Genet. *11*, 880–886.

37. Haller, B.C., and Messer, P.W. (2019). SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. Mol. Biol. Evol. *36*, 632–637.

38. Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F., Gibbs, R.A., Bustamante, C.D.; and 1000 Genomes Project (2011). Demographic history and rare allele sharing among human populations. Proc. Natl. Acad. Sci. USA *108*, 11983–11988.

39. Kim, B.Y., Huber, C.D., and Lohmueller, K.E. (2017). Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. Genetics *206*, 345–361.

40. Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D., and Mountain, J.L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. Am. J. Hum. Genet. *96*, 37–53.

41. Wakeley, J., and Hey, J. (1997). Estimating ancestral population parameters. Genetics *145*, 847–855.

42. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

43. Turcot, V., Lu, Y., Highland, H.M., Schurmann, C., Justice, A.E., Fine, R.S., Bradfield, J.P., Esko, T., Giri, A., Graff, M., et al. (2018). Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. Nat. Genet. *50*, 26–41.

44. Mathieson, I., and McVean, G. (2012). Differential confounding of rare and common variants in spatially structured populations. Nat. Genet. *44*, 243–246.

45. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M.; and Schizophrenia Working Group of the Psychiatric Genomics Consortium (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.

46. Lam, M., Chen, C.Y., Li, Z., Martin, A.R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B.C., et al. (2019). Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat. Genet. *51*, 1670–1678.

47. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. *47*, 979–986.

48. Rosenberg, N.A., Edge, M.D., Pritchard, J.K., and Feldman, M.W. (2018). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. Evol. Med. Public Health *2019*, 26–34.

49. Harmon, A. (2018). Why white supremacists are chugging milk (and why geneticists are alarmed). The New York Times, October 17, 2018. https://www.nytimes.com/2018/10/17/us/white-supremacists-science-dna.html.

50. Fuentes, A., Ackermann, R.R., Athreya, S., Bolnick, D., Lasisi, T., Lee, S.H., McLean, S.A., and Nelson, R. (2019). AAPA statement on race and racism. Am. J. Phys. Anthropol. *169*, 400–402.

51. Saini, A. (2019). Superior: The Return of Race Science (Beacon Press).

52. Coop, G. (2019). Reading tea leaves? Polygenic scores and differences in traits among groups. arXiv, 1909.00892. https://arxiv.org/abs/1909.00892.

Supplemental information

Negative selection on complex traits limits

phenotype prediction accuracy between populations
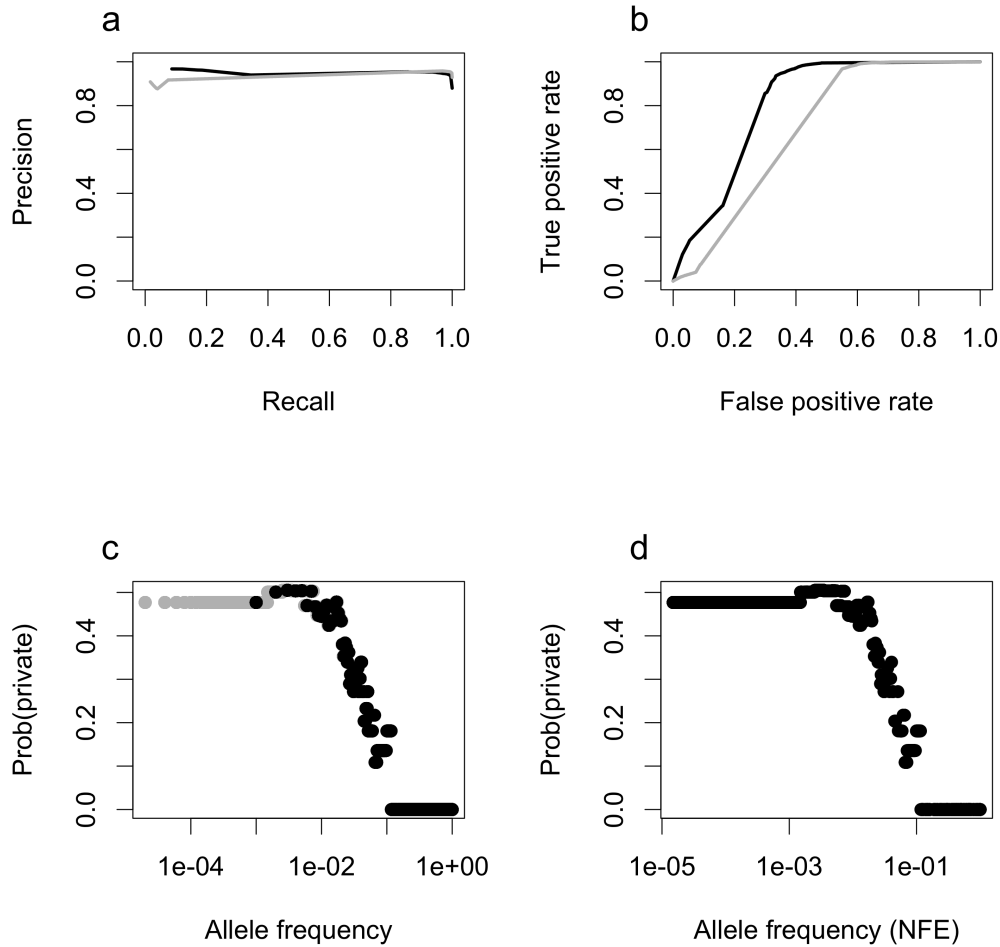
Arun Durvasula and Kirk E. Lohmueller

Figure S1: **Private variant prediction model performance.** Performance and external validation of our model for inferring whether a variant is private or shared. a) precision-recall and b) receiver operating characteristic curve. c) The probability that a variant is private versus the allele frequency (log scale) in simulated test data. We see that for variants above 10%, the probability they are private is very low. In panels a,b, and c, the grey lines and points indicate results from a simulation with 50,000 haplotypes. d) The probability that a variant is private versus the allele frequency (log scale) for 10,000 randomly sampled variants from ExAC. Variants with a frequency below 1×10^{-3} automatically set to have a probability equal to a variant at frequency 0.001 since these variants are outside the frequency range of our simulations.
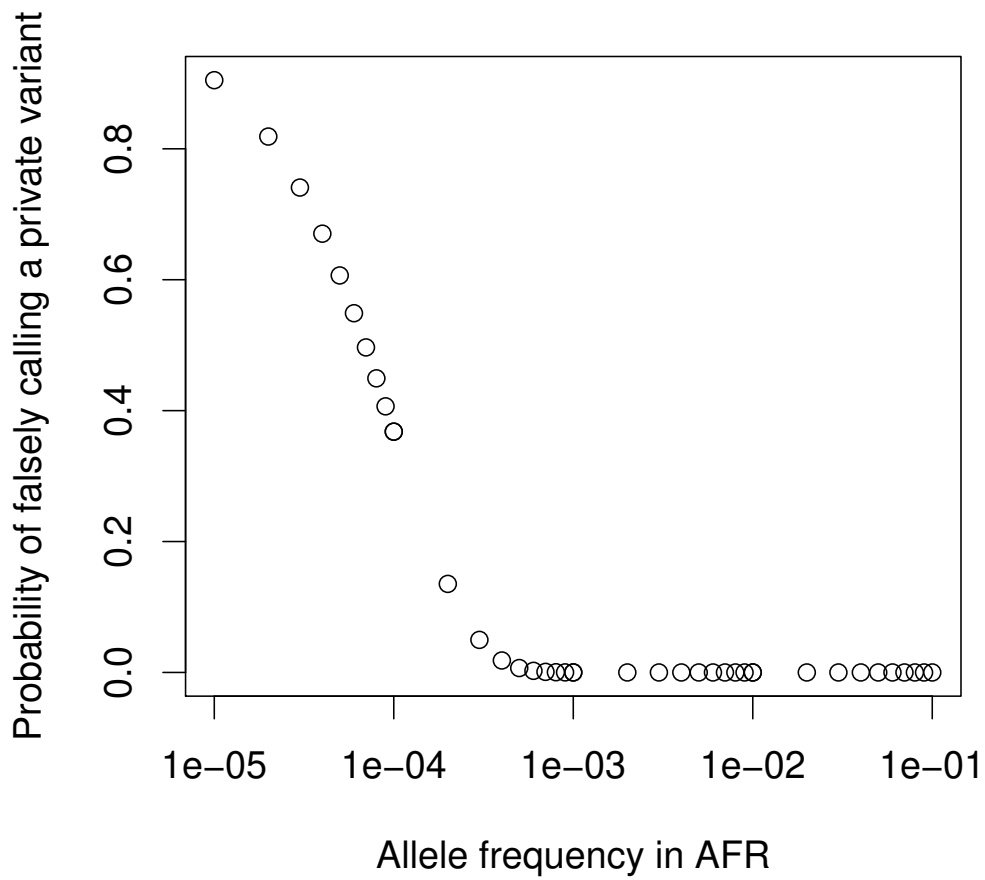
Figure S2: **Private variant misclassification probability.** The probability of misclassifying a shared variant as private to Europe as a function of the true allele frequency in a sample of 10,000 African chromosomes. For variants more common than $\approx 1 \times 10^3$, the probability is very low.
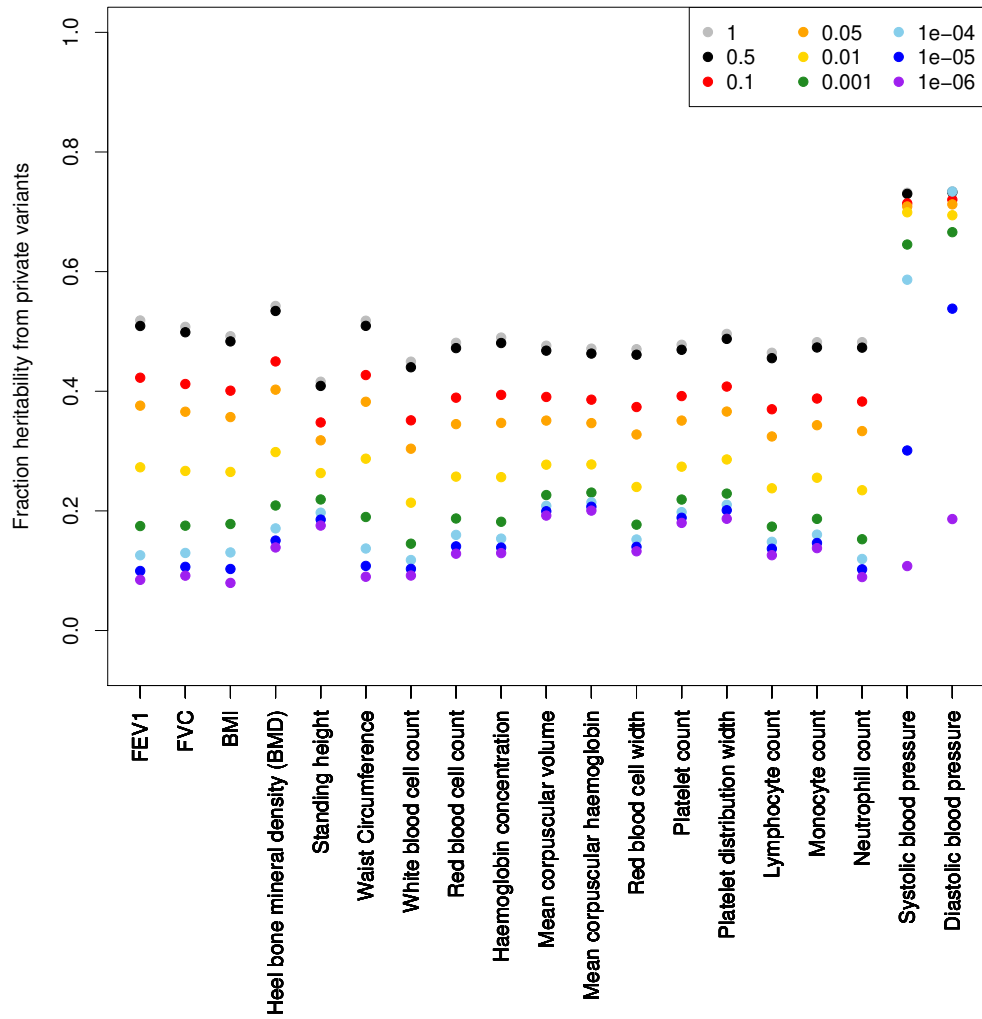
Figure S3: **Effect of *p*-value cutoffs.** The fraction of heritability explained by private variants for quantitative traits. For each trait, we calculated this fraction using a *p*-value cutoff (indicated by color). They grey dots represent no filter.
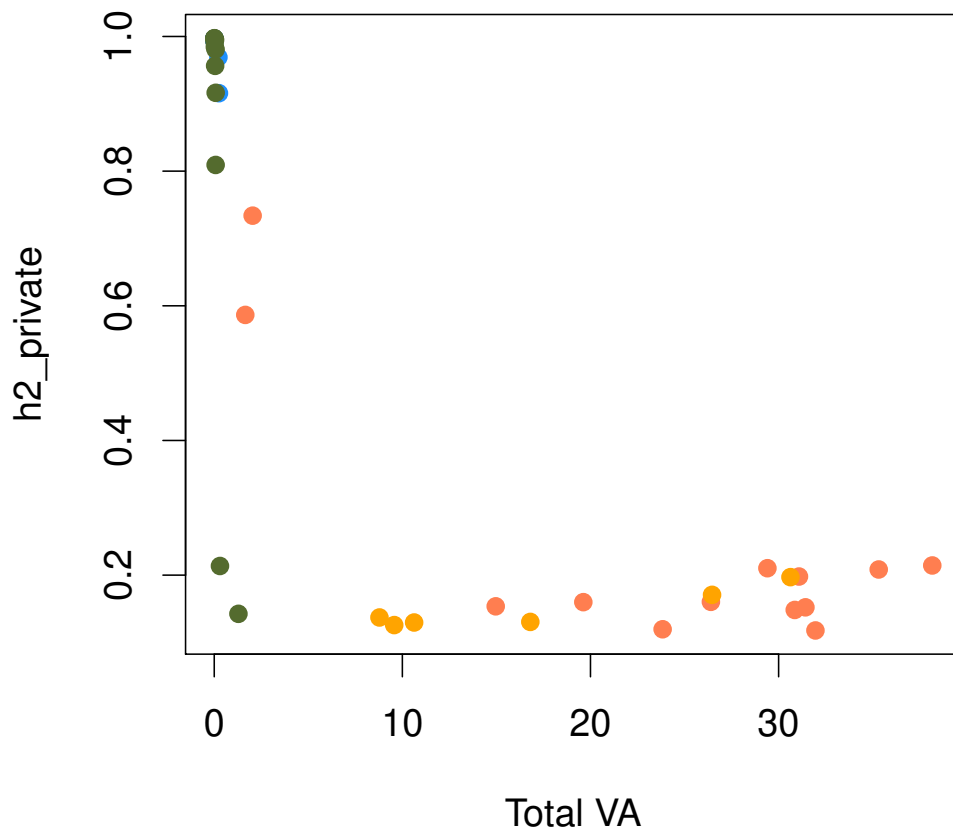
Figure S4: $h^2_{private}$ **versus total** V_A **at** $p = 1e - 4$**.** The fraction of heritability explained by private variants plotted against the total variance explained by all SNPs for each trait (each dot is a trait). For case control phenotypes (blue and green) the total V_A is very low, while it is much higher for quantitative phenotypes (orange and red)

Figure S5: **LD pruned estimator for the percentage of heritability from private variants.** We use 500 SNPs per window. Lines indicate 95% confidence intervals obtained via a jackknife over the windows.

Figure S6: **LD pruned estimator for the percentage of heritability from private variants.** We use 5,000 SNPs per window. Lines indicate 95% confidence intervals obtained via a jackknife over the windows.
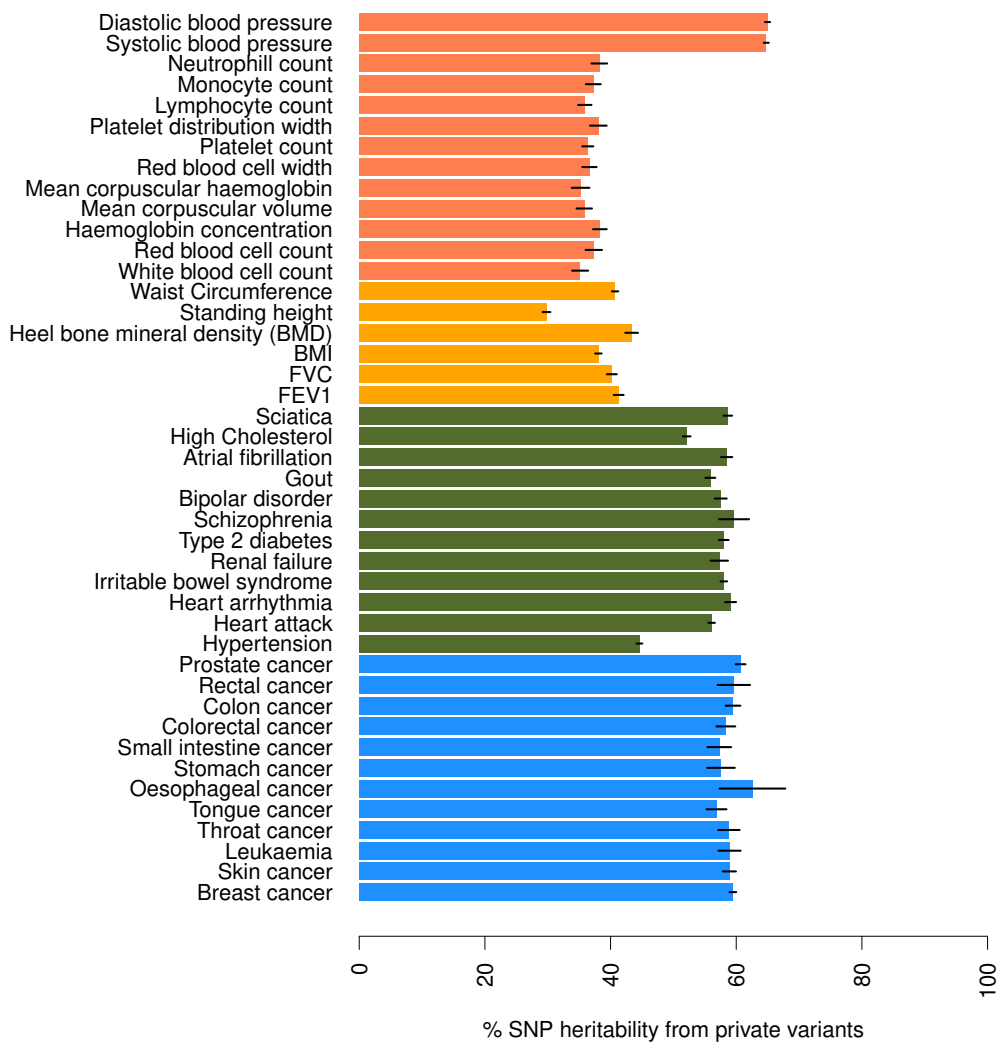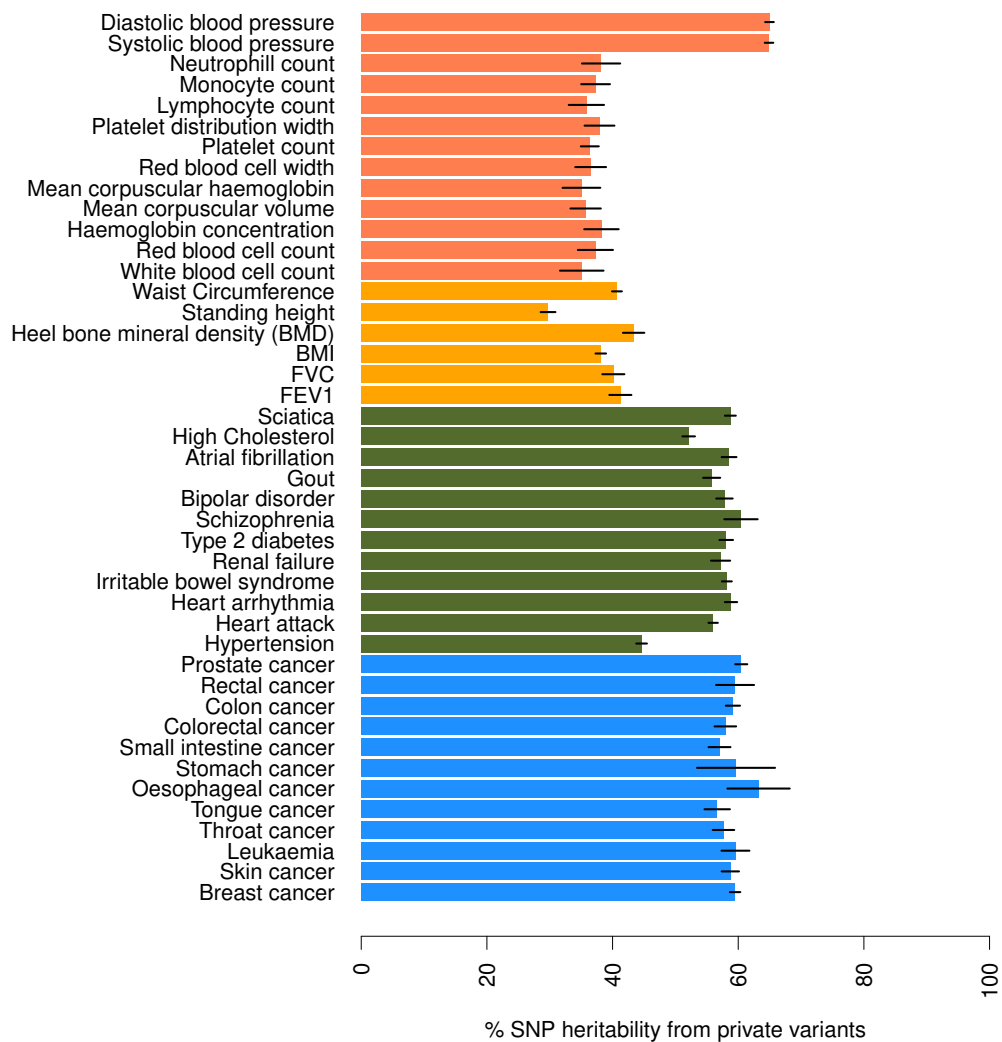
Figure S7: **LD pruned estimator for the percentage of heritability from private variants.** We use 50,000 SNPs per window. Lines indicate 95% confidence intervals obtained via a jackknife over the windows.

Figure S8: **Heritability from shared variants.** LD pruned estimator for the percentage of heritability from shared variants. We use 5,000 SNPs per window. Lines indicate 95% confidence intervals obtained via a jackknife over the windows.
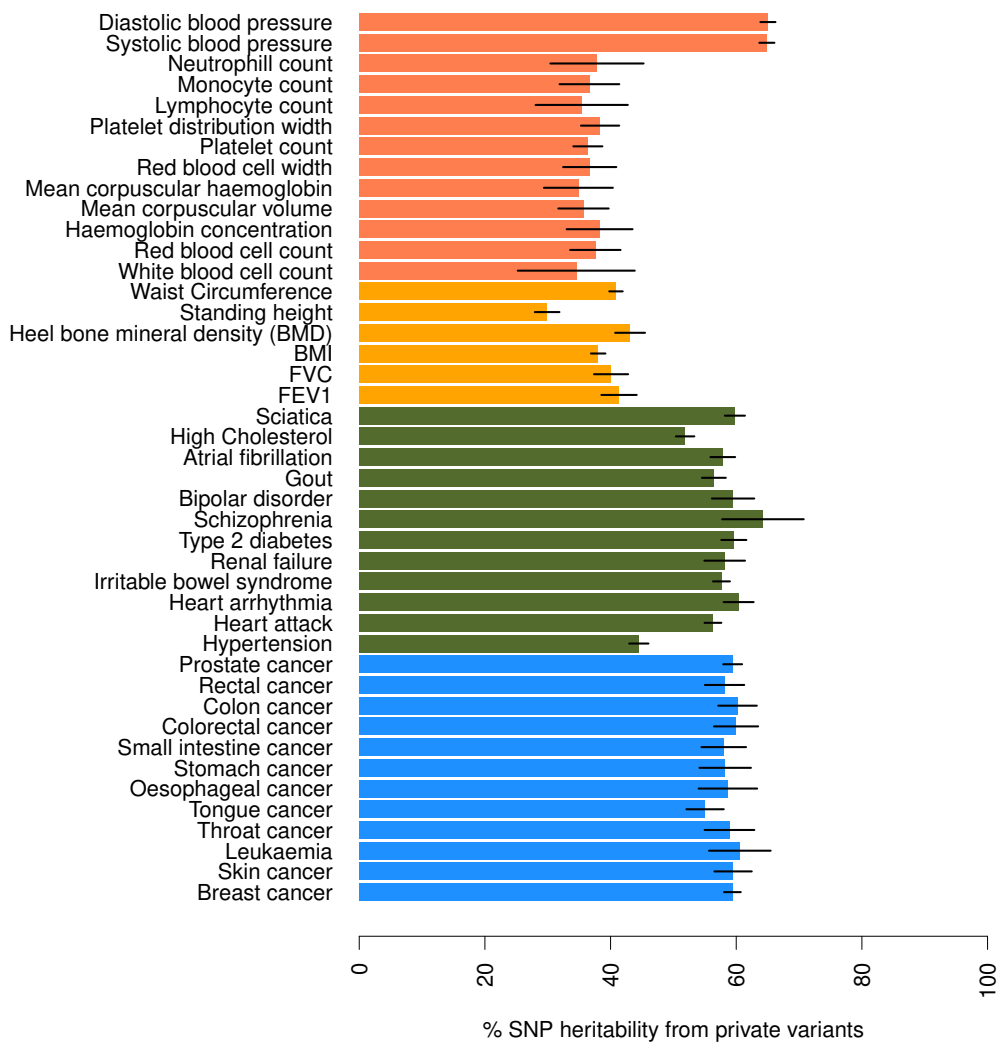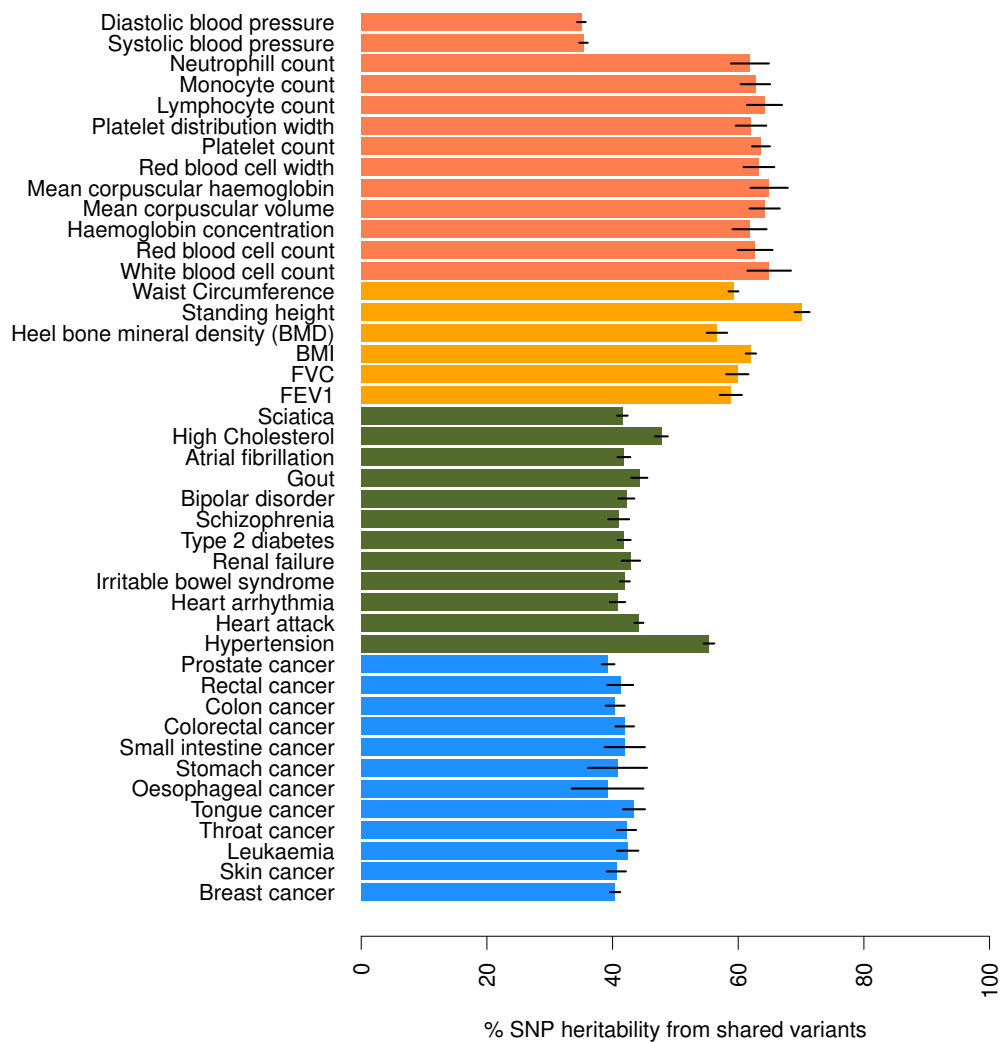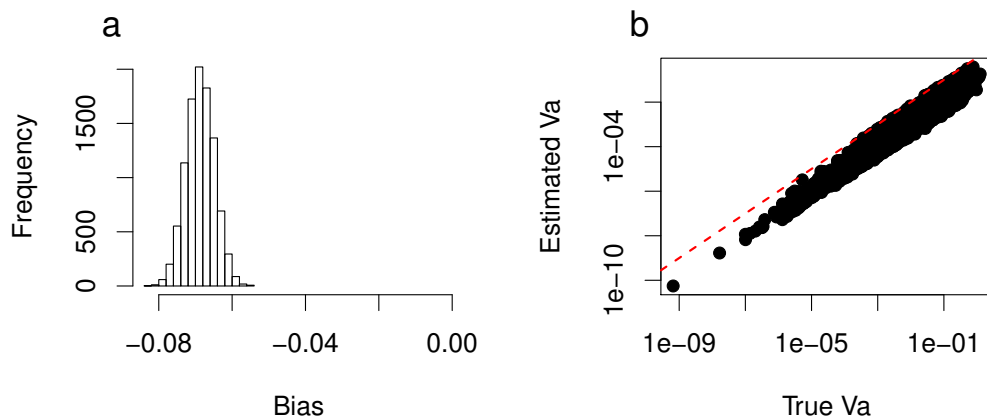
Figure S9: **Bias of the LD-pruned estimator based on coalescent simulations.** a) We plot the expected value of $\hat{h}^2_{private}$ versus the true value of $h^2_{private}$. Using 10,000 simulations, we compute $E[\hat{h}^2_{private}]$ as well as $h^2_{private}$ and plot the difference. The distribution represents bootstrap re-sampled values across simulations. We see that the estimate of the proportion of V_a from private variants is underestimated by $6 - 8\%$. b) Estimated V_a versus the true V_a from simulations for private variants. The red dashed line represents the 1:1 line. All points fall below this line, again suggesting this inference is downwardly biased.

τ	C	h^2
0	0.1	0.38
0.25	0.6	0.41
0.5	1.8	0.39

Table S1: **Scaling constant and realized heritability for simulations.** Values of C used in simulations for varying values of τ and the average resulting heritability. Values of C were chosen to obtain $h^2 \approx 0.4$.

AFR frequency cutoff	0	0.001	0.005	0.01	0.05
Predicted private, not observed in AFR	0.83	0.93	0.95	0.96	0.97
Predicted private, observed in AFR	0.15	0.06	0.04	0.03	0.01
Not predicted private, observed in AFR	0	0	0	0	0
Not predicted private, not observed in AFR	0.015	0.015	0.015	0.015	0.015

Table S2: **Empirical validation of private allele model with allele frequency filters.** We tested the ability of our model to correctly call alleles private in a sample of 66,740 chromosomes from Europe and 10,406 chromosomes from a mixture of African and African-American samples from the Exome Aggregation Consotrium (ExAC) with a posterior probability cutoff of 0.23. Because private European alleles can occur at a low frequency in the African-American samples due to admixture, we use a frequency filter in the AFR sample to remove these false shared variants.

Threshold	Sample set	FDR	FNR	FPR	TPR
0.1	10K Simulated	0.062	0.005	0.483	0.995
0.1	50K Simulated	0.044	0.001	0.692	0.999
0.1	ExAC NFE	0.144	0.001	0.908	0.999
0.23	10K Simulated	0.555	0.017	0.422	0.983
0.23	50K Simulated	0.04	0.004	0.636	0.996
0.23	ExAC NFE	0.141	0.003	0.884	0.997
0.4	10K Simulated	0.049	0.047	0.363	0.952
0.4	50K Simulated	0.039	0.011	0.608	0.989
0.4	ExAC NFE	0.138	0.005	0.865	0.995

Table S3: **Simulation and empirical validation of the private allele model across different posterior probability thresholds.** For the results presented in the main text, we use a posterior probability cutoff of 0.23, which results in an empirical FDR of $\approx 17\%$. We use the 0.1 threshold in Figures S3, S5, S6, S7, S8. "ExAC NFE" refers to an empirical test from the ExAC data set where we use the observed frequency of the allele in the African American sample to estimate the allele count in a sample of unadmixed African genomes. '10K' refers to a sample size of 10,000 chromosomes and '50K' refers to a sample size of 50,000 chromosomes. 'FDR' = false discovery rate, 'FNR' = false negative rate, 'FPR' = false positive rate, 'TPR' = true positive rate.

τ	Shared (Europe)	Private (Europe)	Shared (Africa)	Private (Africa)
0	0.91	0.38	0.96	0.34
0.25	0.76	0.62	0.83	0.57
0.5	0.56	0.83	0.58	0.79

Table S4: **Polygenic Score (PS) accuracy.** Correlation coefficient (Pearson's r^2) between the true PS and the PS using shared or private variation in European and African individuals across values of τ ($p < 2\times10^{-16}$ for all correlations). As negative selection becomes stronger, the correlation between the true PS and shared variant PS decreases, while the correlation between true PS and private variant PS increases.

S1   Application to GWAS summary statistic data

We applied our partitioning method to summary statistic GWAS data computed on individuals from the UK Biobank. We used summary statistics for 43 traits and diseases released by the Neale lab at `http://www.nealelab.is/uk-biobank/`. Details of the procedure for GWAS, including quality control can be found at the website, but we briefly summarize it here.

For sample inclusion, the authors used unrelated samples and subset to British individuals using PCA and self-reported ancestry ('white-British'/'Irish'/'White'). This resulted in 361,194 samples. SNPs were imputed using the Haplotype Reference Consortium, the UK10K project, and the 1000 Genomes data. Variants were retained if they had an INFO score > 0.8, minor allele frequency > 0.0001 (MAF $> 1 \times 10^{-6}$ if annotated as missense or protein truncating by VEP), and a Hardy-Weinberg equilibrium p-value $> 1 \times 10^{-10}$. This filter resulted in 13.7 million SNPs. The association testing was done using the first 20 principle components, sex, age, as well as interaction and non-linear terms for sex and age. We used summary statistics for GWAS performed including both sexes and for quantitative phenotypes, we used inverse rank normalized phenotypes (as opposed to the raw phenotypes).

From the summary statistic data, we use the minor allele frequency in the British cohort and the effect size to compute the additive genetic variance (Section S3) and the probability that the variant is private to Europe (See Methods in the main text). In the main text, we report results using all SNPs reported in the summary statistics. These results use a probability threshold of $P(\omega|i) = 0.23$, corresponding to a simulated FDR of $\approx 5\%$ and an empirical FDR of $\approx 17\%$. In the main text, we also present results using the FDRs to correct for the fact that some of the SNPs we infer to be private are likely false (i.e. they are truly shared).

We also show the results of our LD-pruned estimator in Supplemental figures S5, S6, and S7. For the results from the LD-pruned estimator, we vary the number of SNPs within a window between 500 SNPs ($\approx$ 100KB), 5,000 SNPs ($\approx$ 1MB), and 50,000 SNPs ($\approx$ 10MB). We also ensure that our estimates are sensible by estimating the proportion of additive genetic variance from variants we infer to be shared. If the inference procedure works correctly, this number should be $1 - \hat{h}^2_{private}$. In Supplemental figure S8, we see that this is indeed the case.

We validated our result on BMI using an external cohort from the GIANT consortium [12]. This GWAS on BMI was performed on 718,734 individuals using an exome-targeted genotyping array. Using the LD-pruned estimator and a probability cutoff of 0.1, we found the proportion of SNP heritability from private variants is 0.522 (95% confidence interval: 0.390-0.653). This value is very close to the inferred value on the full data from the UK Biobank (0.492), giving us confidence that these results are robust across association studies.

S2   The effect of linkage disequilibrium on our inferences

Our inferences of the heritability due to private variants made the assumption that the estimated effect sizes for the GWAS SNPs were the true effect sizes of the causal variants. Further, we assumed that the variants were all independent of each other. In truth, these assumptions are violated for a variety of reasons. First, due to linkage disequilibrium (LD),

SNPs may be correlated with one another. Second, some of the non-zero effect sizes of GWAS SNPs may be due to the fact that the GWAS SNP is tagging (in LD with) an untyped causal variant and is itself not causal. Third, even if the GWAS variants analyzed in our study are the true causal variants, their effect sizes may be mis-estimated by the effects at nearby SNPs in LD with them.

Thus, given these challenges, we carefully considered the effect that unmodeled LD may have on our inferences.

In principle, recent methods such as stratified LD score regression (S-LDSC; [2]), SumHer [11], or HESS [9] could provide us with the ability to examine the proportion of heritability that comes from private variants. These methods attempt to deconvolve the observed marginal association statistics, $\hat{b}_j$, by modeling linkage disequilibrium (LD) between tagging variants and causal variants. However, these methods rely on out of sample estimates of the amount of LD between variants (*i.e.* from a reference population). While estimates of LD (measured by $\hat{r}^2$) between common variants can be obtained with minimal variance in some cases, this depends greatly on allele frequency, as outlined below.

Consider two loci, A and B, with allele frequencies p_A and p_B of the minor allele. The correlation coefficient, r^2, is defined as:

$$r^2 = \frac{(p_{AB} - p_A p_B)^2}{p_A(1 - p_A)p_B(1 - p_B)}$$

An estimator of r^2 is obtained by using estimates of $\hat{p}_{AB}$, $\hat{p}_A$, and $\hat{p}_B$ [10, 5]. If the allele frequencies are assumed to be binomially distributed and if the sample size used to obtain estimates of allele frequencies is large enough, we can model these frequencies as approximately normally distributed. Then, the variance of the estimator, $Var(\hat{r}^2) \approx (1 - r^2)^2$ [3]. This relationship suggests that when the correlation is lowest, the variance will be the largest. Analysis of the mathematical properties of r^2 as a measure of LD suggests that the maximum value of r^2 is lowest when p_A is much different from p_B [13]. Taken together, this suggests that there will be much uncertainty when using LD panels to estimate LD between common and rare variants.

Here, we are interested in LD patterns between private variants, which tend to be rare, and shared variants, which tend to be common. Therefore, in lieu of using previous methods that rely on out of sample LD estimates, we consider several approaches that do not rely on knowing the patterns of LD between variants: 1) We repeated our inference only considering a sub-set of SNPs that would have a greater chance of being independent of each other, 2) We used coalescent simulations to investigate the effect of unaccounted for LD on the inference of the heritability contained in private variants, and 3) We tested whether the effect sizes of private variants or shared variants were more likely to be over-estimated due to LD. All of these analyses suggest that while unmodeled LD between SNPs may bias some of our inferences, they are biased in the direction of underestimating the heritability due to private variants. In other words, our estimates are likely to be conservative and our overall conclusion that a substantial amount of the heritability for anthropometnric and disease traits is due to private variants is robust to LD. Below we detail these results of these three analyses.

S3 Inference of $h^2_{private}$ by pruning for LD

Here, we describe an estimator for proportion of heritability that comes from private variants when using a sub-set of SNPs within a genomic window. By thinning the SNPs, we should remove some of the effects of LD between the GWAS SNPs. Our estimator has the form:

$$\hat{h}^2_{private} = \frac{\hat{V}_{A,p}}{\hat{V}_A}.$$

First, we describe a procedure for estimating $\hat{V}_{A,p}$. Consider a window w with K SNPs. We randomly select 1 SNP, k, and compute the additive genetic variance for that SNP:

$$\hat{V}_{A,p,w,k} = 2\hat{p}_k(1 - \hat{p}_k)\hat{\beta}^2_k z_k$$

where z_k is an indicator function that equals 1 when $p(\omega|\hat{p}_k) \geq t$ (see the Methods section in the main text for a description of this model) and 0 otherwise. We repeat the procedure T times and take the sum:

$$\hat{V}_{A,p,w} = \sum_k^T \hat{V}_{A,p,w,k}$$

Then, to estimate $\hat{V}_{A,p}$, we sum over the windows:

$$\hat{V}_{A,p} = \sum_w \hat{V}_{A,p,w}$$

If T is large and w contains a single causal variant, then $E[\hat{V}_{A,p,w}] \approx 2p_c(1 - p_c)(\beta_c \bar{r}^2_c)^2$, where $\bar{r}^2_c$ is the average LD of variants in the window and the causal variant[1].

A similar estimator for $\hat{V}_A$ can be obtained:

$$\hat{V}_{A,w} = 2\sum_k^T \hat{p}_k(1 - \hat{p}_k)\hat{\beta}^2_k$$

$$\hat{V}_A = \sum_w^W \hat{V}_{A,w}$$

We obtain standard errors by using a jackknife over the windows [8]. We denote the estimate of $h^2_{private}$ obtained using the full data set (*i.e.* all windows) as $\hat{h}^2_{private}$. Further, let $\hat{h}^2_{private(-w)}$ be the estimate of $h^2_{private}$ obtained by deleting window w and let $\bar{h}^2_{private} = \frac{1}{W}\sum_{w=1}^W \hat{h}^2_{private(-w)}$. We compute the standard error ($SE_{h^2_{private}}$) and bias ($B_{h^2_{private}}$) as

$$SE_{h^2_{private}} = \sqrt{\frac{w-1}{w}\sum_{w=1}^W (\hat{h}^2_{private\,w} - \bar{h}^2_{private})^2}$$

$$B_{h^2_{private}} = (w-1)(\bar{h}^2_{private} - \hat{h}^2_{private})$$

---

[1]We do not know the true allele frequency of the causal variant (p_c), so the expectation of the estimator is not exactly equal to the quantity on the right hand side. However, if r^2 is large, then $p_c \approx p_t$ where p_t is the allele frequency of the tagging variant and the expectation will hold.

We obtain 95% confidence intervals by first bias correcting the estimate ($\hat{h}^2_{privatecorr} = \hat{h}^2_{private} - B_{h^2_{private}}$). Then, the confidence interval is $\hat{h}^2_{privatecorr} \pm z(\alpha)SE_{h^2_{private}}$, where α is the confidence level, $z(\alpha) = \sqrt{2}erf^{-1}(\alpha)$ and $erf^{-1}(x)$ is the inverse error function. In all plots using this estimator, we plot the bias corrected estimate rather than the full estimate.

This approach to confidence intervals ignores correlations between windows, which could lead to double counting variants if causal variants are clustered together. We found that when we varied the window size between 500 SNPs ($\approx$ 100KB), 5,000 SNPs ($\approx$ 1MB), and 50,000 SNPs ($\approx$ 10MB), the results were consistent, suggesting that this was not a practical issue.

From the expectation above, it is clear that the estimator we propose is biased, $E[\hat{h}^2_{private}] \leq h^2_{private}$. Next, we show that this bias conservatively estimates the proportion of heritability from private variants.

When $\bar{r}_c \approx 1$, the average LD in the window between the tagging variants and causal variants is close to 1, resulting in $E[\hat{h}^2_{private}] \approx h^2_{private}$. However, such tightly linked regions will not be common in the human genome. Several papers have noted that the range of r^2 depends on the allele frequencies of the loci in question (e.g. [13]). Specifically, as the two allele frequencies diverge, the maximum r^2 value decreases. In our case, shared variants tend to have larger allele frequencies than private variants, suggesting that the maximum allele frequency difference will occur when the tagging variant is private and the causal variant is shared or when the tagging variant is shared and the causal variant is private.

In the case where there is a mismatch between the categories of the tagging and causal variants, it is possible to end up with variation contributing to the wrong additive genetic variance bin. For example, if the tagging variant is shared and the causal variant is private, V_A that should be in the private bin would be apportioned into the shared bin, leading to downward bias in the estimate of $h^2_{private}$. On the other hand, if the tagging variant is private and the causal variant is shared, V_A that should be in the shared bin will be in the private bin, leading to overestimates of $h^2_{private}$. However, in both of these cases, because of the dependency of r^2 on allele frequency, we do not expect these cases to contribute much to V_A because r^2 will be low between the tagging and causal variants. Further, we expect the second case (where causal variants are shared and the tagging variants are private) to be uncommon when a trait is under negative selection because negative selection will push large effect alleles to lower allele frequencies, increasing the probability that these alleles are private. Thus, the scenario under which our estimate of $h^2_{private}$ would be upwardly biased are likely to be unusual.

On the other hand, when both the tagging and causal variant are in the same class (*i.e.* both private or both shared), the frequencies will be more similar, leading to estimates of V_A that are reduced by a factor related to the amount of LD between the tag and the causal variant.

S4 The relationship between marginal effect size estimates and LD

We wondered whether our estimates of $h^2_{private}$ could be biased by differences in tagging causal variation between private and shared variation. Because shared variants tend to be more common, they will tend to be in LD with more (and therefore tag more) variants. If this is the case, then shared variants will have inflated marginal effect sizes compared

to private variants. This would lead to overestimating the additive genetic variance from shared variants compared to private variants, again making our inferences conservative.

To test this for this effect, we examined the summary statistics from a GWAS on BMI in the UK Biobank and correlated the marginal effect sizes with recombination rates estimated from patterns of LD in Europeans [6] for both classes of variants. We found that variants we predict to be private have a lower correlation (Spearman's $\rho = -0.04$, $P < 2 \times 10^{-16}$) between marginal effect sizes and recombination rates compared with variants we predict to be shared (Spearman's $\rho = -0.08$, $P < 2 \times 10^{-16}$). Said another way, the relationship between effect size and LD (which is one of the main ideas behind LD score regression; [1]), is stronger for shared variants than for private variants. This finding suggests that estimates of additive genetic variance from private variants are downwardly biased compared to estimates from shared variants, consistent with the intuition from Section S3 and results from our simulations (Section S5).

S5   Coalescent simulations to assess the performance of the estimator of $h^2_{private}$

We tested our intuition in Section S3 by performing coalescent simulations. We simulated genetic variation data using msprime [7] under the demography from Gutenkunst et al [4]. We simulated a single window and randomly assigned a causal variant within that window with an effect size $\beta \sim N(0, 1)$. Then, for each of the non-causal variants in the window, we simulated a GWAS by assigning marginal association statistics as $\hat{\beta} = r^2 \beta$, where r^2 is the LD between a given variant and the causal variant.

Then, to simulate our estimation procedure (Section S3), we randomly pick a SNP from the window, and compute the additive genetic variance using the 'marginal' association statistic ($\hat{V}_A = 2p(1-p)\hat{\beta}^2$). We estimate $\hat{h}^2_{private}$ using the allele frequency to infer if an allele is private. We obtain the true additive genetic variance using the true effect size (β) and determine if an allele is private based on whether it is observed in the simulated African population. We repeat this procedure 10,000 times for each window and simulate 10,000 windows.

Then, to ensure our simulation is close to our real inference procedure, we compute $E[\hat{h}^2_{private}]$ by inferring whether alleles are private or shared (rather than using the true status from the simulation). In Figure S9b, we plot the estimated additive genetic variance from each of the windows versus the true additive genetic variance from each of the windows for private variants. As expected, we see that there is a downward bias, consistent with the effects of LD on $\hat{\beta}$. This means that LD causes us to underestimate the heritability accounted for by private variants. Finally, we compare $E[\hat{h}^2_{private}]$ to $h^2_{private}$ (Figure S9a) and see that the estimator is downwardly biased as well, consistent with a greater effect of bias on private variants compared to shared variants. Therefore, we conclude that the estimator described in Section S3 is conservative with respect to the amount of heritability explained by private variants.

References

[1] Brendan K. Bulik-Sullivan, Po-Ru Loh, Hilary K. Finucane, Stephan Ripke, Jian Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick Patterson,

Mark J. Daly, Alkes L. Price, and Benjamin M. Neale. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 47(3):291–295, March 2015.

[2] Hilary K. Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, Stephan Ripke, Felix R. Day, ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, The RACI Consortium, Shaun Purcell, Eli Stahl, Sara Lindstrom, John R. B. Perry, Yukinori Okada, Soumya Raychaudhuri, Mark J. Daly, Nick Patterson, Benjamin M. Neale, and Alkes L. Price. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, 47(11):1228–1235, November 2015.

[3] R. A. Fisher. On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, 1:3–32, 1921.

[4] Ryan N. Gutenkunst, Ryan D. Hernandez, Scott H. Williamson, and Carlos D. Bustamante. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genetics*, 5(10):e1000695, October 2009.

[5] P. W. Hedrick. Gametic disequilibrium measures: proceed with caution. *Genetics*, 117(2):331–341, October 1987.

[6] International HapMap Consortium, Kelly A. Frazer, Dennis G. Ballinger, David R. Cox, David A. Hinds, Laura L. Stuve, Richard A. Gibbs, John W. Belmont, Andrew Boudreau, Paul Hardenbol, Suzanne M. Leal, Shiran Pasternak, David A. Wheeler, Thomas D. Willis, Fuli Yu, Huanming Yang, Changqing Zeng, Yang Gao, Haoran Hu, Weitao Hu, Chaohua Li, Wei Lin, Siqi Liu, Hao Pan, Xiaoli Tang, Jian Wang, Wei Wang, Jun Yu, Bo Zhang, Qingrun Zhang, Hongbin Zhao, Hui Zhao, Jun Zhou, Stacey B. Gabriel, Rachel Barry, Brendan Blumenstiel, Amy Camargo, Matthew Defelice, Maura Faggart, Mary Goyette, Supriya Gupta, Jamie Moore, Huy Nguyen, Robert C. Onofrio, Melissa Parkin, Jessica Roy, Erich Stahl, Ellen Winchester, Liuda Ziaugra, David Altshuler, Yan Shen, Zhijian Yao, Wei Huang, Xun Chu, Yungang He, Li Jin, Yangfan Liu, Yayun Shen, Weiwei Sun, Haifeng Wang, Yi Wang, Ying Wang, Xiaoyan Xiong, Liang Xu, Mary M. Y. Waye, Stephen K. W. Tsui, Hong Xue, J. Tze-Fei Wong, Luana M. Galver, Jian-Bing Fan, Kevin Gunderson, Sarah S. Murray, Arnold R. Oliphant, Mark S. Chee, Alexandre Montpetit, Fanny Chagnon, Vincent Ferretti, Martin Leboeuf, Jean-François Olivier, Michael S. Phillips, Stéphanie Roumy, Clémentine Sallée, Andrei Verner, Thomas J. Hudson, Pui-Yan Kwok, Dongmei Cai, Daniel C. Koboldt, Raymond D. Miller, Ludmila Pawlikowska, Patricia Taillon-Miller, Ming Xiao, Lap-Chee Tsui, William Mak, You Qiang Song, Paul K. H. Tam, Yusuke Nakamura, Takahisa Kawaguchi, Takuya Kitamoto, Takashi Morizono, Atsushi Nagashima, Yozo Ohnishi, Akihiro Sekine, Toshihiro Tanaka, Tatsuhiko Tsunoda, Panos Deloukas, Christine P. Bird, Marcos Delgado, Emmanouil T. Dermitzakis, Rhian Gwilliam, Sarah Hunt, Jonathan Morrison, Don Powell, Barbara E. Stranger, Pamela Whittaker, David R. Bentley, Mark J. Daly, Paul I. W. de Bakker, Jeff Barrett, Yves R. Chretien, Julian Maller, Steve McCarroll, Nick Patterson, Itsik Pe'er, Alkes Price, Shaun Purcell, Daniel J. Richter, Pardis Sabeti, Richa Saxena, Stephen F. Schaffner, Pak C. Sham, Patrick Varilly, David Altshuler, Lincoln D. Stein, Lalitha Krishnan, Albert Vernon Smith, Marcela K. Tello-Ruiz, Gudmundur A. Thorisson,

Aravinda Chakravarti, Peter E. Chen, David J. Cutler, Carl S. Kashuk, Shin Lin, Gonçalo R. Abecasis, Weihua Guan, Yun Li, Heather M. Munro, Zhaohui Steve Qin, Daryl J. Thomas, Gilean McVean, Adam Auton, Leonardo Bottolo, Niall Cardin, Susana Eyheramendy, Colin Freeman, Jonathan Marchini, Simon Myers, Chris Spencer, Matthew Stephens, Peter Donnelly, Lon R. Cardon, Geraldine Clarke, David M. Evans, Andrew P. Morris, Bruce S. Weir, Tatsuhiko Tsunoda, James C. Mullikin, Stephen T. Sherry, Michael Feolo, Andrew Skol, Houcan Zhang, Changqing Zeng, Hui Zhao, Ichiro Matsuda, Yoshimitsu Fukushima, Darryl R. Macer, Eiko Suda, Charles N. Rotimi, Clement A. Adebamowo, Ike Ajayi, Toyin Aniagwu, Patricia A. Marshall, Chibuzor Nkwodimmah, Charmaine D. M. Royal, Mark F. Leppert, Missy Dixon, Andy Peiffer, Renzong Qiu, Alastair Kent, Kazuto Kato, Norio Niikawa, Isaac F. Adewole, Bartha M. Knoppers, Morris W. Foster, Ellen Wright Clayton, Jessica Watkin, Richard A. Gibbs, John W. Belmont, Donna Muzny, Lynne Nazareth, Erica Sodergren, George M. Weinstock, David A. Wheeler, Imtaz Yakub, Stacey B. Gabriel, Robert C. Onofrio, Daniel J. Richter, Liuda Ziaugra, Bruce W. Birren, Mark J. Daly, David Altshuler, Richard K. Wilson, Lucinda L. Fulton, Jane Rogers, John Burton, Nigel P. Carter, Christopher M. Clee, Mark Griffiths, Matthew C. Jones, Kirsten McLay, Robert W. Plumb, Mark T. Ross, Sarah K. Sims, David L. Willey, Zhu Chen, Hua Han, Le Kang, Martin Godbout, John C. Wallenburg, Paul L'Archevêque, Guy Bellemare, Koji Saeki, Hongguang Wang, Daochang An, Hongbo Fu, Qing Li, Zhen Wang, Renwu Wang, Arthur L. Holden, Lisa D. Brooks, Jean E. McEwen, Mark S. Guyer, Vivian Ota Wang, Jane L. Peterson, Michael Shi, Jack Spiegel, Lawrence M. Sung, Lynn F. Zacharia, Francis S. Collins, Karen Kennedy, Ruth Jamieson, and John Stewart. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164):851–861, October 2007.

[7] Jerome Kelleher, Alison M. Etheridge, and Gilean McVean. Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLOS Computational Biology*, 12(5):e1004842, May 2016.

[8] Hans R. Kunsch. The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics*, 17(3):1217–1241, September 1989.

[9] Huwenbo Shi, Gleb Kichaev, and Bogdan Pasaniuc. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *The American Journal of Human Genetics*, 99(1):139–153, July 2016.

[10] Montgomery Slatkin. Linkage disequilibrium–understanding the evolutionary past and mapping the medical future. *Nature Reviews. Genetics*, 9(6):477–485, June 2008.

[11] Doug Speed and David J. Balding. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nature Genetics*, 51(2):277, February 2019.

[12] Valérie Turcot, Yingchang Lu, Heather M. Highland, Claudia Schurmann, Anne E. Justice, Rebecca S. Fine, Jonathan P. Bradfield, Tõnu Esko, Ayush Giri, Mariaelisa Graff, Xiuqing Guo, Audrey E. Hendricks, Tugce Karaderi, Adelheid Lempradl, Adam E. Locke, Anubha Mahajan, Eirini Marouli, Suthesh Sivapalaratnam, Kristin L. Young, Tamuno Alfred, Mary F. Feitosa, Nicholas G. D. Masca, Alisa K. Manning, Carolina Medina-Gomez, Poorva Mudgal, Maggie C. Y. Ng, Alex P. Reiner, Sailaja Vedantam, Sara M. Willems, Thomas W. Winkler, Gonçalo Abecasis, Katja K. Aben, Dewan S. Alam, Sameer E. Alharthi, Matthew Allison, Philippe Amouyel, Folkert W. Asselbergs,

Paul L. Auer, Beverley Balkau, Lia E. Bang, Inês Barroso, Lisa Bastarache, Marianne Benn, Sven Bergmann, Lawrence F. Bielak, Matthias Blüher, Michael Boehnke, Heiner Boeing, Eric Boerwinkle, Carsten A. Böger, Jette Bork-Jensen, Michiel L. Bots, Erwin P. Bottinger, Donald W. Bowden, Ivan Brandslund, Gerome Breen, Murray H. Brilliant, Linda Broer, Marco Brumat, Amber A. Burt, Adam S. Butterworth, Peter T. Campbell, Stefania Cappellani, David J. Carey, Eulalia Catamo, Mark J. Caulfield, John C. Chambers, Daniel I. Chasman, Yii-Der I. Chen, Rajiv Chowdhury, Cramer Christensen, Audrey Y. Chu, Massimiliano Cocca, Francis S. Collins, James P. Cook, Janie Corley, Jordi Corominas Galbany, Amanda J. Cox, David S. Crosslin, Gabriel Cuellar-Partida, Angela D'Eustacchio, John Danesh, Gail Davies, Paul I. W. Bakker, Mark C. H. Groot, Renée Mutsert, Ian J. Deary, George Dedoussis, Ellen W. Demerath, Martin Heijer, Anneke I. Hollander, Hester M. Ruijter, Joe G. Dennis, Josh C. Denny, Emanuele Di Angelantonio, Fotios Drenos, Mengmeng Du, Marie-Pierre Dubé, Alison M. Dunning, Douglas F. Easton, Todd L. Edwards, David Ellinghaus, Patrick T. Ellinor, Paul Elliott, Evangelos Evangelou, Aliki-Eleni Farmaki, I. Sadaf Farooqi, Jessica D. Faul, Sascha Fauser, Shuang Feng, Ele Ferrannini, Jean Ferrieres, Jose C. Florez, Ian Ford, Myriam Fornage, Oscar H. Franco, Andre Franke, Paul W. Franks, Nele Friedrich, Ruth Frikke-Schmidt, Tessel E. Galesloot, Wei Gan, Ilaria Gandin, Paolo Gasparini, Jane Gibson, Vilmantas Giedraitis, Anette P. Gjesing, Penny Gordon-Larsen, Mathias Gorski, Hans-Jörgen Grabe, Struan F. A. Grant, Niels Grarup, Helen L. Griffiths, Megan L. Grove, Vilmundur Gudnason, Stefan Gustafsson, Jeff Haessler, Hakon Hakonarson, Anke R. Hammerschlag, Torben Hansen, Kathleen Mullan Harris, Tamara B. Harris, Andrew T. Hattersley, Christian T. Have, Caroline Hayward, Liang He, Nancy L. Heard-Costa, Andrew C. Heath, Iris M. Heid, Øyvind Helgeland, Jussi Hernesniemi, Alex W. Hewitt, Oddgeir L. Holmen, G. Kees Hovingh, Joanna M. M. Howson, Yao Hu, Paul L. Huang, Jennifer E. Huffman, M. Arfan Ikram, Erik Ingelsson, Anne U. Jackson, Jan-Håkan Jansson, Gail P. Jarvik, Gorm B. Jensen, Yucheng Jia, Stefan Johansson, Marit E. Jørgensen, Torben Jørgensen, J. Wouter Jukema, Bratati Kahali, René S. Kahn, Mika Kähönen, Pia R. Kamstrup, Stavroula Kanoni, Jaakko Kaprio, Maria Karaleftheri, Sharon L. R. Kardia, Fredrik Karpe, Sekar Kathiresan, Frank Kee, Lambertus A. Kiemeney, Eric Kim, Hidetoshi Kitajima, Pirjo Komulainen, Jaspal S. Kooner, Charles Kooperberg, Tellervo Korhonen, Peter Kovacs, Helena Kuivaniemi, Zoltán Kutalik, Kari Kuulasmaa, Johanna Kuusisto, Markku Laakso, Timo A. Lakka, David Lamparter, Ethan M. Lange, Leslie A. Lange, Claudia Langenberg, Eric B. Larson, Nanette R. Lee, Terho Lehtimäki, Cora E. Lewis, Huaixing Li, Jin Li, Ruifang Li-Gao, Honghuang Lin, Keng-Hung Lin, Li-An Lin, Xu Lin, Lars Lind, Jaana Lindström, Allan Linneberg, Ching-Ti Liu, Dajiang J. Liu, Yongmei Liu, Ken S. Lo, Artitaya Lophatananon, Andrew J. Lotery, Anu Loukola, Jian'an Luan, Steven A. Lubitz, Leo-Pekka Lyytikäinen, Satu Männistö, Gaëlle Marenne, Angela L. Mazul, Mark I. McCarthy, Roberta McKean-Cowdin, Sarah E. Medland, Karina Meidtner, Lili Milani, Vanisha Mistry, Paul Mitchell, Karen L. Mohlke, Leena Moilanen, Marie Moitry, Grant W. Montgomery, Dennis O. Mook-Kanamori, Carmel Moore, Trevor A. Mori, Andrew D. Morris, Andrew P. Morris, Martina Müller-Nurasyid, Patricia B. Munroe, Mike A. Nalls, Narisu Narisu, Christopher P. Nelson, Matt Neville, Sune F. Nielsen, Kjell Nikus, Pål R. Njølstad, Børge G. Nordestgaard, Dale R. Nyholt, Jeffrey R. O'Connel, Michelle L. O'Donoghue, Loes M. Olde Loohuis, Roel A. Ophoff, Katharine R. Owen, Chris J. Packard, Sandosh Padmanabhan, Colin N. A. Palmer, Nicholette D. Palmer, Gerard Pasterkamp, Aniruddh P. Patel, Alison Pattie,

Oluf Pedersen, Peggy L. Peissig, Gina M. Peloso, Craig E. Pennell, Markus Perola, James A. Perry, John R. B. Perry, Tune H. Pers, Thomas N. Person, Annette Peters, Eva R. B. Petersen, Patricia A. Peyser, Ailith Pirie, Ozren Polasek, Tinca J. Polderman, Hannu Puolijoki, Olli T. Raitakari, Asif Rasheed, Rainer Rauramaa, Dermot F. Reilly, Frida Renström, Myriam Rheinberger, Paul M. Ridker, John D. Rioux, Manuel A. Rivas, David J. Roberts, Neil R. Robertson, Antonietta Robino, Olov Rolandsson, Igor Rudan, Katherine S. Ruth, Danish Saleheen, Veikko Salomaa, Nilesh J. Samani, Yadav Sapkota, Naveed Sattar, Robert E. Schoen, Pamela J. Schreiner, Matthias B. Schulze, Robert A. Scott, Marcelo P. Segura-Lepe, Svati H. Shah, Wayne H.-H. Sheu, Xueling Sim, Andrew J. Slater, Kerrin S. Small, Albert V. Smith, Lorraine Southam, Timothy D. Spector, Elizabeth K. Speliotes, John M. Starr, Kari Stefansson, Valgerdur Steinthorsdottir, Kathleen E. Stirrups, Konstantin Strauch, Heather M. Stringham, Michael Stumvoll, Liang Sun, Praveen Surendran, Amy J. Swift, Hayato Tada, Katherine E. Tansey, Jean-Claude Tardif, Kent D. Taylor, Alexander Teumer, Deborah J. Thompson, Gudmar Thorleifsson, Unnur Thorsteinsdottir, Betina H. Thuesen, Anke Tönjes, Gerard Tromp, Stella Trompet, Emmanouil Tsafantakis, Jaakko Tuomilehto, Anne Tybjaerg-Hansen, Jonathan P. Tyrer, Rudolf Uher, André G. Uitterlinden, Matti Uusitupa, Sander W. Laan, Cornelia M. Duijn, Nienke Leeuwen, Jessica van Setten, Mauno Vanhala, Anette Varbo, Tibor V. Varga, Rohit Varma, Digna R. Velez Edwards, Sita H. Vermeulen, Giovanni Veronesi, Henrik Vestergaard, Veronique Vitart, Thomas F. Vogt, Uwe Völker, Dragana Vuckovic, Lynne E. Wagenknecht, Mark Walker, Lars Wallentin, Feijie Wang, Carol A. Wang, Shuai Wang, Yiqin Wang, Erin B. Ware, Nicholas J. Wareham, Helen R. Warren, Dawn M. Waterworth, Jennifer Wessel, Harvey D. White, Cristen J. Willer, James G. Wilson, Daniel R. Witte, Andrew R. Wood, Ying Wu, Hanieh Yaghootkar, Jie Yao, Pang Yao, Laura M. Yerges-Armstrong, Robin Young, Eleftheria Zeggini, Xiaowei Zhan, Weihua Zhang, Jing Hua Zhao, Wei Zhao, Wei Zhao, Wei Zhou, Krina T. Zondervan, Jerome I. Rotter, John A. Pospisilik, Fernando Rivadeneira, Ingrid B. Borecki, Panos Deloukas, Timothy M. Frayling, Guillaume Lettre, Kari E. North, Cecilia M. Lindgren, Joel N. Hirschhorn, Ruth J. F. Loos, CHD Exome+ Consortium, EPIC-CVD Consortium, ExomeBP Consortium, Global Lipids Genetic Consortium, GoT2D Genes Consortium, EPIC InterAct Consortium, INTERVAL Study, ReproGen Consortium, T2D-Genes Consortium, MAGIC Investigators, and Understanding Society Scientific Group. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nature Genetics*, 50(1):26–41, 2018.

[13] Jenna M. VanLiere and Noah A. Rosenberg. Mathematical properties of the r2 measure of linkage disequilibrium. *Theoretical Population Biology*, 74(1):130–137, August 2008.