ARTICLE

Extreme Polygenicity of Complex Traits Is Explained by Negative Selection

Luke J. O'Connor,[1,2,*] Armin P. Schoech,[1] Farhad Hormozdiari,[1] Steven Gazal,[1] Nick Patterson,[3] and Alkes L. Price[1,3,*]

Complex traits and common diseases are extremely polygenic, their heritability spread across thousands of loci. One possible explanation is that thousands of genes and loci have similarly important biological effects when mutated. However, we hypothesize that for most complex traits, relatively few genes and loci are critical, and negative selection—purging large-effect mutations in these regions—leaves behind common-variant associations in thousands of less critical regions instead. We refer to this phenomenon as *flattening*. To quantify its effects, we introduce a mathematical definition of polygenicity, the effective number of independently associated SNPs (M_e), which describes how evenly the heritability of a trait is spread across the genome. We developed a method, stratified LD fourth moments regression (S-LD4M), to estimate M_e, validating that it produces robust estimates in simulations. Analyzing 33 complex traits (average N = 361k), we determined that heritability is spread ~4× more evenly among common SNPs than among low-frequency SNPs. This difference, together with evolutionary modeling of new mutations, suggests that complex traits would be orders of magnitude less polygenic if not for the influence of negative selection. We also determined that heritability is spread more evenly within functionally important regions in proportion to their heritability enrichment; functionally important regions do not harbor common SNPs with greatly increased causal effect sizes, due to selective constraint. Our results suggest that for most complex traits, the genes and loci with the most critical biological effects often differ from those with the strongest common-variant associations.

Introduction

Genome-wide association studies (GWASs) have revealed that common diseases and complex traits are heritable and highly polygenic.[1–14] There are usually no large-effect common SNPs, and heritability is evenly spread across thousands of small-effect SNPs. The polygenic distribution of heritability presents a challenge, as small-effect SNPs are difficult to detect (leading to "missing heritability"[15]) and are difficult to interpret.[16,17]

One factor contributing to the polygenic distribution of heritability is the complexity of the underlying biology: many genes and regions of the genome have a nonzero phenotypic effect if mutated. A plausible explanation for the large mutational target is that cellular networks are densely connected, such that nearly every gene expressed in a relevant cell type has a small phenotypic effect (the "omnigenic model").[17] A striking implication of this model is that disease genes with direct phenotypic effects would explain a minority of disease heritability.

Although biological complexity clearly contributes to the polygenicity of complex traits, negative selection may also be a critical factor. Biological complexity determines the effect-size distribution of new mutations. We hypothesized that this distribution—in contrast to that of heritability—is dominated by a relatively small number of large-effect genes and loci (Figure 1A). In the absence of negative selection, the resulting distribution of heritability would be highly concentrated in these large-effect regions, and hence only moderately polygenic (Figure 1B). How-

ever, in the presence of negative selection, the heritability explained by any single SNP is limited, and mutations in these large-effect regions would not become common. As a result, heritability would be spread much more evenly across large- and small-effect regions alike (Figure 1B). We refer to this phenomenon—negative selection causing the distribution of heritability to be extremely polygenic—as *flattening*.

Negative selection has previously been shown to influence complex-trait genetic architectures. It limits the average per-allele effect sizes of common variants,[12,18–21] especially in coding regions and brain regulatory elements.[20] It also causes average effect sizes to vary with linkage disequilibrium (LD) and allele age.[19] Various models have been proposed for the mathematical relationship between phenotypic effect sizes and selection coefficients.[22–28] However, it is currently unclear whether negative selection merely limits common-variant heritability on average without affecting how evenly it is spread across the genome—as in the model of Zeng et al.[12] and in one of the evolutionary models that we explore below—or whether it actually reshapes the genome-wide distribution of heritability.

Here, we investigate the hypothesis that negative selection flattens the distribution of heritability across the genome, explaining the extreme polygenicity of complex traits. We evaluate two specific predictions. First, heritability should be spread more evenly across common SNPs than across low-frequency SNPs. Ideally, we would compare the polygenicity of common SNPs and new

A Distribution of mutational effect sizes across genome

De novo effect sizes | Regulatory element
LF effect sizes | Coding region
Common effect sizes

Effect size magnitude

Small-effect genes | Large-effect gene

B Distribution of heritability across SNPs

Expected per-SNP proportion of h^2

h^2 without negative selection
h^2 with negative selection

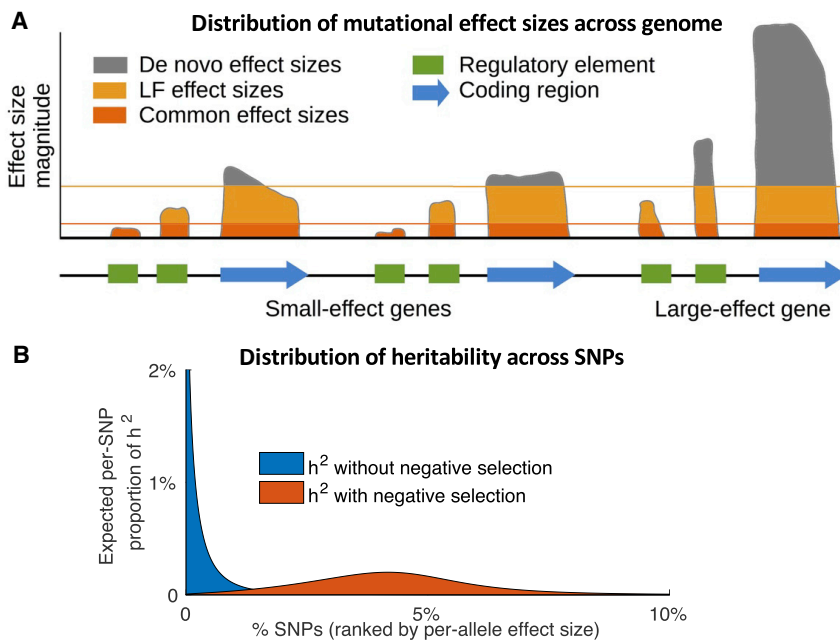% SNPs (ranked by per-allele effect size)

Figure 1. Illustration of Flattening due to Negative Selection

(A) We illustrate the range of possible per-allele effect sizes for a SNP at each site for a toy example of three genes and nearby regulatory regions. Here, the distribution of *de novo* effects is not highly polygenic; it is dominated by coding mutations in a single large-effect gene (although other genes also harbor small effects). Negative selection imposes an upper effect size bound (possibly soft) on common variants (and, to a lesser extent, low-frequency variants), resulting in increased polygenicity. Within functionally important regions (e.g., coding), a larger proportion of variants have effect sizes near the bound, leading to especially large polygenicity. In practice, this bound may vary across the genome, but we hypothesize that it is much more even than the effect-size distribution of *de novo* variants.

(B) We illustrate the expected per-SNP proportion of heritability for SNPs ranked by per-allele effect size, for a hypothetical trait whose *de novo* effect-size distribution has a mixture of small- and large-effect mutations. In the absence of negative selection (blue), heritability is concentrated among a limited number of large-effect SNPs. In the presence of negative selection (orange), large-effect SNPs are prevented from becoming common, and thus explain little heritability; instead, heritability is spread across a large number of SNPs with small effects.

mutations, but this comparison is not possible using GWAS data; instead, we compare common and low-frequency SNPs, reasoning that the polygenicity of low-frequency SNPs will lie in between the polygenicity of common SNPs and the polygenicity of new mutations (Figure 1A). Second, functionally important regions of the genome should not harbor common SNPs with greatly increased causal effect sizes, as the maximum effect size is determined by the strength of selection. Instead, their heritability enrichment should be predominantly explained by increased polygenicity: a larger proportion of SNPs in functionally important regions should have effect sizes close to the upper bound imposed by selection (Figure 1A). In order to evaluate these predictions, we introduce a mathematical definition of polygenicity, the effective number of independently associated SNPs (M_e). We develop a method, stratified LD fourth moments regression (S-LD4M), to estimate M_e from summary association statistics. We validate S-LD4M in simulations and apply it to summary statistics for 33 diseases and complex traits.

Material and Methods

Defining Polygenicity: The Effective Number of Independently Associated SNPs

In this manuscript, we use the term "polygenicity" to describe the phenomenon that heritability is spread evenly across many genes and loci, such that the strongest genetic associations explain a limited fraction of heritability (leading to "missing heritability"[1,2,15]). According to this definition, schizophrenia, which is predominantly driven by thousands of small-effect common

variants, is extremely polygenic.[1,4,13] In contrast, severe neurodevelopmental disorders, which are predominantly driven by deleterious mutations in one or a few genes, are not highly polygenic—despite the fact that they are also influenced by many common variants (cumulatively explaining <10% of variance).[29] Previous studies defined polygenicity as the total number of SNPs with nonzero effects (i.e., the total number of independently associated SNPs) (M_t),[1,3,6–8,12–14] but this definition fails to describe how heritability is spread across causal SNPs; M_t could be similar for schizophrenia and a predominantly monogenic developmental disorder, despite their qualitatively different genetic architectures. Moreover, in practice, estimates of M_t implicitly rely on a detection threshold, causing them to vary with sample size and with the parametric model that is specified[13] (see Simulations). Variability in sample size and power could bias comparisons among traits or between common and low-frequency SNPs.

We introduce a mathematical definition of polygenicity, the effective number of independently associated SNPs (M_e), which quantifies how evenly heritability is spread across loci. Roughly, if all associated SNPs have similar effect sizes, then M_e is equal to M_t. However, if a small number of SNPs explain a large proportion of heritability, then the number of independently associated SNPs can be literally large but *effectively* small—in terms of missing heritability, risk prediction, and biological implications—and M_e can be much smaller than M_t. Despite the widespread use of M_t to define polygenicity,[1,3,6–8,12–14] we use the term "polygenicity" to refer to M_e, as it is the even distribution of heritability—and not the number of SNPs with nonzero effects—that makes the genetic architecture of complex traits different from that of predominantly monogenic disorders.[29] We note that if several causal SNPs are in very strong linkage disequilibrium (LD) (see e.g., Hormozdiari et al.[30]), then M_e counts them as a single associated SNP, hence our use of the term "independently associated" (see Appendix A).
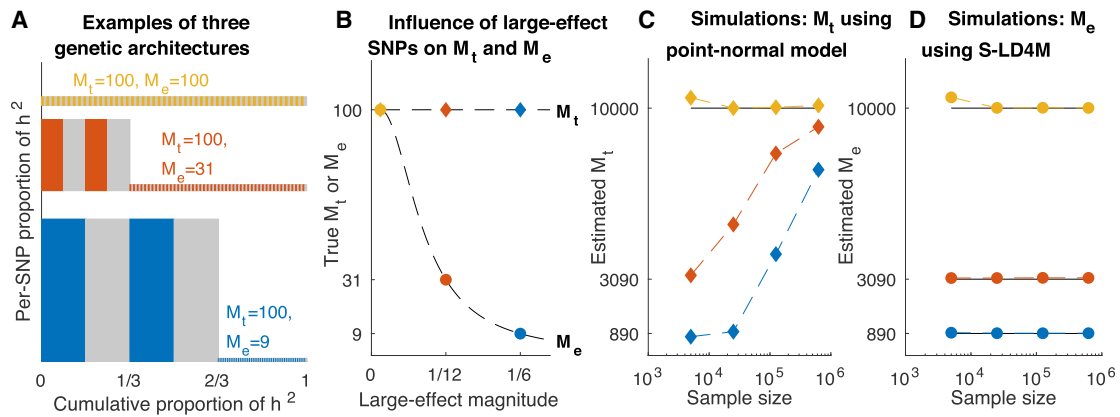
Figure 2. Comparison of the Effective Number of Independently Associated SNPs (M_e) with the Total Number of SNPs with Nonzero Effects (M_t)

(A and B) Examples of three genetic architectures with $M_t = 100$.

(A) Each colored or gray block corresponds to one SNP; both height and width are proportional to the expected proportion of heritability explained by that SNP. The average unit of heritability, denoted $E_{h^2}(\alpha^2)$, is the average height (equal to the total area) of the colored and gray regions. M_e is equal to $h^2/E_{h^2}(\alpha^2)$.

(B) M_t and M_e as a function of the effect size magnitude of the four large-effect SNPs.

(C and D) Simulations of the same three genetic architectures with the number of SNPs (and causal SNPs) scaled up by 100×.

(C) Estimates of M_t under a point-normal model, at different sample sizes.

(D) Estimates of M_e using S-LD4M, at different sample sizes. Error bars denote 95% confidence intervals (based on 1,000 simulations) but are smaller than the data points.

If there is no LD between causal SNPs, the heritability of a trait is proportional to the second moment of its effect size distribution, $E[\beta^2]$, while M_e is inversely proportional to the *fourth* moment, $E[\beta^4]$. The fourth moment heavily weights SNPs with the largest effects, such that if m large-effect SNPs explain most heritability, then M_e will be close to m, irrespective of how many other SNPs may have small nonzero effects. We define:

$$M_e = \frac{3M}{\kappa}, \quad \kappa = \frac{E[\beta^4]}{E[\beta^2]^2}, \quad \text{(Equation 1)}$$

where M is the number of SNPs, β is the causal effect size of a SNP in standardized units (see below), and κ is the kurtosis (normalized fourth moment) of β. We note that M_e is closely related to missing heritability and to polygenic prediction accuracy (see Appendix A, Properties of M_e).

We define effect sizes β in standardized units, i.e., the number of standard deviations increase in phenotype per 1 standard deviation increase in genotype. In these units, the squared effect size of a SNP is equal to the heritability it explains, and the kurtosis describes how evenly heritability is spread across SNPs. Thus, a rare SNP explaining little heritability will contribute little to M_e, even if its per-allele effect size is large.

We refer to h^2 divided by M_e as the average unit of heritability, denoted $E_{h^2}(\alpha^2)$ (where α is the standardized marginal effect size of a SNP, inclusive of LD; see below); heritability is spread across loci in units of average size $E_{h^2}(\alpha^2)$. This quantity can be visualized as the area under the curve in Figure 2A. For example, if 4 causal SNPs each contribute 1/6 of heritability (2/3 of heritability in total) and 96 other causal SNPs have much smaller effects (standardized causal effects drawn from a mixture of two Gaussian distributions[13]), then the average unit of heritability is approximately ($1/6 \times 2/3 = 1/9$), and $M_e \approx 9$ (Figure 2A bottom). More generally, adding a large-effect SNP increases $E_{h^2}(\alpha^2)$ and decreases M_e (Figure 2B). In the case of a neurodevelopmental disorder predominantly caused by nearly penetrant mutations in one or a

few genes, with thousands of common variants also having a small cumulative effect,[29] M_e will be small (e.g., ∼10), while M_t will be large (e.g., ∼10,000).

M_e can be defined for categories of SNPs, such as low-frequency SNPs or coding SNPs. To compare categories of different size, we divide M_e by the number of SNPs in each category. We refer to differences in per-SNP M_e simply as differences in polygenicity. We define polygenicity enrichment as the per-SNP M_e of all SNPs in a category divided by the per-SNP M_e of all SNPs; analogously, we define heritability enrichment as the per-SNP heritability of all SNPs in a category divided by the per-SNP heritability of all SNPs (similar to previous work[31]). "All SNPs" refers to common (MAF ≥ 0.05) and low-frequency (0.005 ≤ MAF < 0.05) SNPs. Enrichment can be either >1 or <1 (i.e., depletion).

Estimating Polygenicity: Stratified LD Fourth Moments Regression

We developed a method, stratified LD fourth moments regression (S-LD4M), to estimate M_e from summary association statistics and LD from a reference panel. S-LD4M regresses squared χ^2 statistics (i.e., fourth powers of signed Z scores) on LD fourth moments, defined as sums of r^4 values to each category of SNPs. Intuitively, S-LD4M provides a bridge between the fourth moment of the marginal effect size distribution (i.e., squared χ^2 statistics) and the fourth moment of the causal effect size distribution (which is related to M_e). S-LD4M relies on the following regression equation:

$$E(\alpha^4 \mid \ell^{(2)}, \ell^{(4)}) \approx 3(\ell^{(2)}\tau)^2 + \ell^{(4)}K, \quad \text{(Equation 2)}$$

where α is the marginal effect size of a SNP in per-normalized-genotype units; $\ell^{(2)}$ is the LD score (LD second moment) of a SNP; $\ell^{(4)}$ is the LD fourth moment, defined as the sum of r^4 values over SNPs; τ is the S-LDSC regression coefficient;[31] and K is the S-LD4M regression coefficient. K is proportional to the

excess kurtosis of the effect-size distribution, compared with an infinitesimal model, and it is related to M_e:

$$K = 3\mathrm{E}(\beta^2)\left[\mathrm{E}_{h^2}(\alpha^2) - E(\alpha^2)\right] = \frac{3h^2}{M}\left[\frac{h^2}{M_e} - \frac{h^2}{M_{indep}}\right], \quad \text{(Equation 3)}$$

where h^2 is the heritability tagged by reference-panel SNPs, M is the number of SNPs, and M_{indep} is the effective number of independent SNPs.[32] Under an infinitesimal genetic architecture, $M_e = M_{indep}$ and $K = 0$. $E(\beta^2)$ and $E(\alpha^2)$ are estimated using a slightly modified version of stratified LD score regression (S-LDSC).[31] Thus, S-LD4M uses Equation 2 to estimate K and Equation 3 to estimate M_e from K. In the case of multiple annotations, $\ell^{(2)}, \ell^{(4)}, \tau$, and K are all vectors with one entry per annotation. Standard errors are computed by jackknifing on contiguous blocks of SNPs, similar to previous work.[5] The runtime of S-LD4M scales linearly with the number of regression SNPs and quadratically with the number of annotations. Details are provided in Appendix A (Stratified LD Fourth Moments Regression). We note that previous studies have used fourth moments of the bivariate effect-size distribution (across two traits) to quantify pleiotropy[33] and causality.[34]

We report results only for well-powered trait-annotation pairs, defined as having heritability Z-score greater than 2 and M_e Z-score greater than $^1/_2$. Simulations suggested that log-scale estimates and estimated standard errors were approximately unbiased at these thresholds (see below). We report meta-analyzed results only for functional annotations for which at least 10 out of the 33 traits analyzed had well-powered estimates for that functional annotation.

In order to aggregate enrichment estimates across traits, we performed meta-analyses on a logarithmic scale, for both heritability enrichment and polygenicity enrichment. The meta-analyzed estimate was computed as the unweighted average of the log-enrichment estimates for well-powered traits. Standard errors were computed by jackknifing on the mean (i.e., without assuming that each trait has independent standard errors).

Our moment-based approach has some similarities with S-LDSC, which estimates heritability by regressing χ^2 statistics on LD scores (LD second moments).[31] We use a slightly modified version of S-LDSC to estimate heritability enrichment. We run S-LD4M and S-LDSC using the baseline-LD model,[19] which includes 75 coding, conserved, regulatory, MAF-related, and LD-related annotations. When estimating the polygenicity and polygenicity enrichment of a category of SNPs, we restrict to well-powered traits. Further details are provided in Appendix A (Stratified LD Fourth Moments Regression). We have released open-source software implementing S-LD4M (see Web Resources).

Simulations

We simulated summary statistics directly using the asymptotic sampling distribution,[35] rather than by simulating individual-level data. Specifically, the distribution of the summary statistic vector, $\widehat{\alpha}$, as a function of the true causal effect size vector β, was:

$$\widehat{\alpha} \sim N\left(\widehat{R}\beta, \frac{1}{N}\widehat{R}\right). \quad \text{(Equation 4)}$$

The estimated LD matrix $\widehat{R}$ was equal to I in simulations with no LD. In simulations with LD, $\widehat{R}$ was computed from UK Biobank data (N = 460k). Sample correlations were computed for all typed and imputed SNPs ($M = 1.0$M) within 0.1 M of each other on chromosome 1. Blocks of 5,000 SNPs were used, and R was set to zero for pairs of SNPs in different blocks. These choices were neces-

sary for $\widehat{R}$ to be stored in memory, due to the large number of SNPs. To ensure that $\widehat{R}$ was positive semidefinite, negative eigenvalues were discarded, as previously described.[34]

Simulations with No LD

We simulated 50,000 total SNPs, 9,600 small-effect SNPs, and 400 large-effect SNPs. For small- and large-effect SNPs, effect sizes were drawn from a normal distribution with mean zero and variance σ_1^2 and σ_2^2, respectively; in the three simulations (corresponding to the three example architectures in Figure 2A), σ_1^2 was equal to 1/10,000, 1/1,200, and 1/600, respectively. σ_2^2 was equal to $(1 - 400\sigma_1^2/9600)$, resulting in a heritability of 1. (We note that in all simulations, only Nh^2 affects the results; for example, identical results are obtained at $h^2 = 1$ and N = 20k and at $h^2 = 0.2$ and N = 100k.) We performed simulations at N = 5,000, 25,000, 125,000, and 625,000.

Simulations with LD

We performed two sets of simulations with real LD patterns (see above). First, we performed simulations with no functional annotations. We specified that a certain percentage of common SNPs (MAF > 0.05) and a certain percentage of low-frequency SNPs (0.05 > MAF > 0.005) were causal. Conditional on being causal, effect sizes were drawn from a mixture of two normal distributions, with probability $p_1 = 0.1$ and $p_2 = 0.9$, and variance $\sigma_1^2 = 4\sigma_2^2$. The variance parameter was frequency dependent; it was proportional to $[p(1-p)]^{0.25}$, where p is the allele frequency. We note that in simulations where we varied the polygenicity of low-frequency SNPs, we did not change the effect-size variance in order to fix the average per-SNP heritability. Thus, as we decreased the low-frequency polygenicity, low-frequency per-SNP heritability was decreased proportionally.

Second, we performed simulations with five real functional annotations (coding, enhancer, promoter, DHS, repressed). We specified a different probability of being causal for SNPs in each category: 1/2, 1/4, 3/20, 3/80, and 1/20, for SNPs in each annotation and SNPs in no annotation at all, respectively. When a SNP was in multiple annotations, the probability of being causal was the maximum of the respective values. Conditional on being causal, SNPs had independent and identically distributed effect sizes; their effects were drawn from a mixture of two normal distributions, with probability $p_1 = 0.1$ and $p_2 = 0.9$, and variance $\sigma_1^2 = 4\sigma_2^2 \propto [p(1-p)]^{0.25}$. The variances were scaled so that the total (expected) heritability was 0.2.

Point-Normal Model Estimator

We implemented a maximum-likelihood estimator of the total number of SNPs with nonzero effects (M_t) under a point-normal model with known heritability and no LD. Because there was no LD, the likelihood was equal to the product over SNPs of the per-SNP likelihoods, as described in Vilhjálmsson et al.[36] We performed a grid search to estimate the proportion of causal SNPs, and the estimate of M_t was equal to the estimated proportion of causal SNPs times the total number of SNPs. We note that in the case of LD, the likelihood does not factor, and it is computationally difficult to compute the exact likelihood, motivating the use of sophisticated estimators and heuristics.[12,13] In the much easier case of no LD, these approaches are expected to have similar performance as the optimal maximum-likelihood estimator.

Evolutionary Modeling

We used evolutionary modeling to investigate the impact of flattening on the distribution of heritability across genes. In our evolutionary modeling, each SNP affected one gene and each

gene affected both the trait and fitness. We specified a distribution of effect sizes for SNPs on genes (identical for every gene) and a joint distribution for the effect size and selection coefficient of each gene. The effect of a SNP on the trait was its effect on its gene times the gene effect size; its selection coefficient of a SNP was its effect on its gene times the gene selection coefficient. Using an analytical formula for the distribution of allele frequencies conditional on selection coefficients, we obtained the joint distribution of trait effects and allele frequencies for each gene. We chose to perform analytical calculations based on the stationary distribution of allele frequencies rather than forward simulations because none of the qualitative phenomena we sought to illustrate are dependent on linkage disequilibrium or background selection.

We considered two models. Under the first model, 5% of genes had large effect sizes (β_{gene}), and these genes had large selection coefficients (s_{gene}). We refer to this model as the "direct selection model" because it could arise from direct selection acting on the trait. Under the second model, there were no large-effect genes, and different genes had different selection coefficients. We refer to this model as the "pleiotropic selection" model because the amount of selection acting on a gene is independent of its phenotypic effect.

Under the direct selection model, s_{gene} was equal to β_{gene}^2, and β_{gene} followed a mixture of normal distributions; 95% of genes had variance 10^{-3}, and 5% of genes had variance 10^{-1} (i.e., 10× larger effect sizes and 100× larger effect size magnitudes). As a result of this choice, the *de novo* effect size distribution is dominated by large-effect genes. In one secondary analysis, we also added a third mixture component, where 0.5% of genes had variance 10. In another secondary analysis, we increased the strength of selection; s_{gene} was equal to $5\beta_{gene}^2$.

Under the pleiotropic selection model, s_{gene} was independent of β_{gene}; β_{gene} followed a normal distribution with variance 10^{-3} (so that there were no large-effect genes), and s_{gene} followed a gamma distribution with parameters $k = 5/2$ and $\theta = 1/1,250$. As a result of this choice, the *de novo* effect-size distribution was polygenic.

Under each model, we specified a distribution of effect sizes for SNPs on each gene, denoted $\beta_{SNP \to gene}$. Each SNP affected one gene, and the distribution was identical for every gene. For the direct selection model, we specified an inverse-gamma distribution with parameters $k = 100$ and $\theta = 1$. This is a heavy-tailed distribution; it causes the expected heritability explained by a gene to plateau as a function of the gene effect size, rather than decreasing after attaining a maximum. For the pleiotropic selection model, we specified a mixture of normal distributions; 75% of SNPs had variance 0.2, and 25% of SNPs had variance 0.02. This choice causes the effect size distribution to be highly polygenic, despite the lack of flattening in this model; it also leads to an appropriate relationship between allele frequency and per-SNP heritability.

The effect size of a SNP on the trait, denoted β_{SNP}, was equal to $\beta_{SNP \to gene}\beta_{gene}$. The selection coefficient of a SNP, denoted s_{SNP}, was equal to $\beta_{SNP \to gene}^2 s_{gene}$. For each gene, we computed the joint distribution of β_{SNP} and allele frequency p, using the formula for the probability density of p conditional on s_{SNP}:[37]

$$f(p \mid s_{SNP}) = \frac{1}{Z} p^{4N\mu - 1}(1 - p^{4N\mu - 1}) \exp(-4N s_{SNP} p), \quad \text{(Equation 5)}$$

where Z is a constant, $N = 100$, and $\mu = 1/200$ (so that the exponent $4N\mu - 1$ is equal to 1). We note that the model is overparameterized; for example, identical results would be obtained with 100× larger N and 100× smaller s_{gene} and μ.

We simulated 20,000 genes. For each gene, we computed the second and fourth moments of β_{SNP} for SNPs at allele frequencies $p = 0.25, 0.05, 0.01, 0.002, 0.004$, by multiplying the probability density function of β_{SNP} by $f(p \mid s_{SNP})$ and approximating the integral using a sum over $\beta_{SNP \to gene}^2 = (1/5000), (2/5000), \ldots, 8$. Averaging over genes, we computed the variance in per-allele effect sizes, the per-SNP heritability (variance times heterozygosity), and the relative polygenicity (inverse of kurtosis). We also computed the proportion of heritability explained by the 10% of genes with largest β_{gene}^2.

Distinct from the 20,000 simulated genes, we analytically computed the heritability explained by a gene as a function of β_{gene}^2. In the direct selection model, s_{gene} was equal to β_{gene}^2; in the pleiotropic selection model, for the purpose of this specific analysis, s_{gene} was fixed at 2×10^{-4}. The heritability explained by a gene at a given allele frequency was defined as the variance of β_{SNP}^2 at that allele frequency, times the heterozygosity.

Results

Simulations

We performed simple simulations with no LD to evaluate the S-LD4M estimator of M_e (the effective number of independently associated SNPs) and a maximum-likelihood estimator of M_t (the total number of SNPs with nonzero effects) under a point-normal model (PN) (see Material and Methods). This estimator is expected to produce similar results as previous estimators of M_t under the same point-normal model.[3,8,11–14] We simulated 400 large-effect SNPs and 9,600 small-effect SNPs, with $M = 50k$ total SNPs. The mixture of different causal effect sizes, violating the point-normal model, is consistent with evidence for real traits.[12,13] When there was a large difference between the causal effect sizes of large- and small-effect SNPs, PN produced biased and sample size-dependent estimates of M_t (Figure 2C and Table S1), consistent with recent work.[13] In general, any method that estimates M_t will detect only SNPs with effect sizes greater than some threshold, where that threshold depends on modeling assumptions and power. In contrast, S-LD4M does not depend on parametric modeling assumptions, and it produces unbiased estimates of M_e (Figure 2D). Power-dependent bias would be especially problematic for comparing common vs. low-frequency polygenicity, due to lower power for low-frequency SNPs.

Next, we performed a series of simulations using real LD patterns to determine whether S-LD4M produces reliable estimates of polygenicity in realistic settings. We simulated summary association statistics from the asymptotic sampling distribution[35] for UK Biobank imputed SNPs on chromosome 1 ($M \approx 1.0M$ SNPs). We used $N = 50k$ samples and $h^2 = 0.2$, to approximately match Nh^2/M (and hence the expected χ^2 statistic) for UK Biobank traits. We included MAF-dependent genetic architectures[12,21] and causal effect size heterogeneity. Details of each simulation are provided in the Material and Methods.

First, we assessed the ability of S-LD4M to estimate polygenicity for the set of all common and low-frequency SNPs.

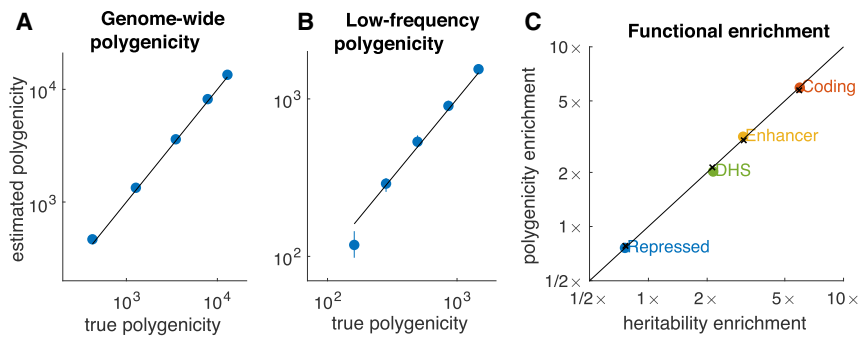**Figure 3. Accuracy of S-LD4M Estimates in Simulations with LD**

(A) Estimates of M_e for all SNPs (MAF = 0.5%–50%).

(B) Estimates of M_e for low-frequency SNPs (MAF = 0.5%–5%); common-SNP M_e is fixed at ~1,000 in these simulations.

(C) Estimates of polygenicity enrichment and heritability enrichment in simulations with four functional categories. Black lines denotes y = x, and colored points denote estimates. In (C), × denotes true values. Error bars denote 95% confidence intervals (based on 1,000 simulations) but are smaller than the data points in most cases. Numerical results are reported in Table S1.

We determined that median estimates of M_e were approximately unbiased across a wide range of true values, although slight bias was observed for very low values of M_e (Figure 3A and Table S1). We report medians instead of means due to noise in the denominator of the estimator, which leads to instability in the mean. In general, S-LD4M is expected to produce approximately unbiased median estimates except when power is very low.

Second, we fixed the polygenicity of common SNPs (MAF > 5%) at $M_e \approx 1,000$ and progressively reduced the polygenicity of low-frequency SNPs (MAF = 0.5%–5%), decreasing low-frequency heritability proportionally to fix the average effect size magnitude of causal low-frequency SNPs. Similar to Figure 3A, median estimates of low-frequency M_e were approximately unbiased for most values of M_e (Figure 3B). Some bias was observed at very low values of low-frequency M_e, owing to very low power (due to very low per-SNP heritability; see Material and Methods). We restrict our analyses of real traits to well-powered traits (see below). These estimates indicate that S-LD4M can be used to compare the polygenicity of common and low-frequency SNPs.

Third, we performed simulations with four real functional annotations, simulating equal heritability enrichment and polygenicity enrichment: coding (~8× enriched), enhancer (~5×), DNase I hypersensitivity sites (DHS, ~2×), and repressed (~0.75×), similar to our results on real traits (see below). We determined that S-LD4M produces approximately unbiased estimates of polygenicity enrichment (Figure 3C). We also considered an alternative model of heritability enrichment, under which polygenicity was approximately constant across functional categories. Estimates were approximately unbiased for enriched functional categories but biased for the repressed annotation (Figure S1). Despite the fact that this genetic architecture is less realistic (see below), we avoid reporting estimates of polygenicity for depleted functional annotations on real traits.

Fourth, we performed simulations to assess whether genetic architectures with non-random clustering of causal SNPs would bias our estimates of polygenicity enrichment for SNPs in functional categories. Such clustering is expected in real data (e.g., due to biologically important

genes), and it could potentially lead to bias, as S-LD4M uses an LD approximation whose accuracy depends on LD between causal SNPs (however, we do not assume that linked SNPs have independent true causal effect sizes). We simulated clusters of either 5 or 50 causal SNPs (on average) across three genomic length scales (10, 100, 1,000 SNPs; roughly 3 kb, 30 kb, 300 kb on average). We determined that median enrichment estimates were approximately unbiased in each case, indicating that our LD approximation is robust to non-random clustering of causal SNPs (Figure S2).

Fifth, we performed simulations at different sample sizes and included a filtering step to exclude trait-annotation pairs with inadequate power (see Material and Methods). We performed simulations at $N = 10k$, $N = 50k$, and $N = 250k$, estimating both the genome-wide polygenicity and the polygenicity enrichment of five categories (four functional categories and low-frequency SNPs). At the default setting of $N = 50k$, genome-wide polygenicity estimates were unbiased (Table S2), and there was adequate power to estimate polygenicity enrichment: 61% of simulated traits were retained on average across categories, and polygenicity enrichment estimates were approximately unbiased (Figure S3 and Table S3). Similar results were obtained at $N = 250k$ (Figure S3, Tables S2 and S3). At $N = 10k$, genome-wide polygenicity estimates remained unbiased (Table S2), but there was limited power to estimate polygenicity enrichment: only 8% of simulated traits were retained on average (including ~0% of simulated traits for coding SNPs and for low-frequency SNPs), and polygenicity enrichment estimates were downward biased (Figure S3 and Table S3). Our analyses of real traits more closely correspond to $N = 50k$ both in terms of average χ^2 statistic and in terms of the proportion of traits retained (49% on average across categories). Nonetheless, we avoid reporting estimates of polygenicity enrichment for annotations having well-powered enrichment estimates for fewer than 10 out of the 33 traits that we analyzed.

Finally, we tested the calibration of our jackknife-based standard errors, once again considering different sample sizes and including a filtering step to exclude trait-annotation pairs with inadequate power. We assessed calibration using the normalized mean squared error, i.e., the actual

squared error divided by the estimated squared error. We determined that our jackknife standard errors for both genome-wide polygenicity (Table S2) and polygenicity enrichment (Table S3) were approximately well calibrated or conservative. In particular, standard errors were moderately conservative when power was low to moderate (e.g., at $N = 10k$ for genome-wide polygenicity), and well calibrated when power was high (e.g., at $N = 50k$ for genome-wide polygenicity).

Polygenicity of Common SNPs across 33 Complex Traits

We applied S-LD4M to publicly available summary association statistics for 33 diseases and complex traits (average $N = 361k$; Table S4), including 29 UK Biobank traits[38–40] (see Web Resources) and 4 additional common diseases. LD scores and LD fourth moments were computed using LD estimated from UK10K[41] ($M = 8.5$ million SNPs after QC, MAF > 0.5%) and 75 baseline-LD model annotations.[19] As in previous work,[18,19,23,31] we excluded the major histocompatibility complex, which has unusually large effect sizes and long-range LD due to balancing selection. (This choice increases M_e estimates for immune-related traits.) Details of our analyses are provided in the Material and Methods.

For most traits, M_e for common SNPs (MAF > 5%) ranged between 500 and 20,000 ($10^{2.7}$ and $10^{4.3}$; Table 1). Fecundity- and brain-related traits were remarkably polygenic, with M_e estimates greater than 10,000. For number of children, the most polygenic trait, M_e was almost as large as effective number of independent SNPs, corresponding to an infinitesimal trait ($\log_{10} M_e = 4.52$ (0.13) vs. $\log_{10} M_{indep} = 4.69$). This estimate implies that most common SNPs are associated with this trait (though a much smaller fraction may be causal). Schizophrenia was also extremely polygenic ($\log_{10} M_e = 4.14$ (0.04)), consistent with previous work.[1,4,13] Particularly high polygenicity for fecundity- and brain-related traits could result from particularly strong negative selection (see below), if these traits are strongly impacted by direct or pleiotropic selection; however, it is also possible that these traits have greater biological complexity. Red hair pigmentation and sunburn were the least polygenic traits, consistent with known large-effect common SNPs for pigmentation traits[42,43] (M_t may be much larger than M_e for these traits). We also estimated M_e without including annotations from the baseline-LD model, either using the 10 common MAF bins only or using no annotations (Table S5). Estimates using ten MAF bins were concordant with estimates using the baseline-LD model (log-scale $r = 0.95$; mean $\log_{10}$ fold difference -0.03 (SD = 0.21)), and estimates using no annotations were approximately concordant but slightly smaller on average (log-scale $r = 0.95$; mean $\log_{10}$ fold difference -0.13 (SD = 0.22)). This suggests that S-LD4M estimates are not contingent on accurately modeling LD-dependent architectures, which is critical in other contexts.

We compared our estimates of M_e with estimates of the total number of SNPs with nonzero effects (M_t) from Zeng et al.[12] and from Zhang et al.[13] The estimates of Zeng et al.[12] were based on a point-normal mixture model, and the estimates of Zhang et al.[13] were based on either a point-normal model or a 3-component (point-normal-normal) model, as determined via a model selection step. Our M_e estimates were correlated with both sets of M_t estimates (log-scale $r = 0.90$ and $r = 0.63$, respectively; Figures S4A and S4B and Table S6). The two sets of M_t estimates were only weakly correlated with each other (log-scale $r = 0.20$; Figure S4C and Table S6), largely due to discordant M_t estimates for intelligence and depression (neither of which was analyzed in this study). The M_t estimates of Zeng et al.[12] were ~4× larger than both our M_e estimates and the M_t estimates of Zhang et al.,[13] and the estimates of Zhang et al.[13] had much larger standard errors. Consistent with our simulations and with the stated limitations of these studies, this discordance confirms that estimates of M_t will vary from study to study based on the parametric model that is assumed and the sample size of the study. We also compared our estimates of M_e to the values of M_e that would be implied under the model fit by Zhang et al.;[13] these estimates were approximately concordant, both in relative value (log-scale $r = 0.85$ across 14 traits) and on average (mean $\log_{10}$ fold difference 0.19 (SD = 0.38)) (Table S6). Finally, we compared the observed vs. expected distribution of χ^2 statistics under a point-normal model (Figure S5 and Table S7), and we determined that a point-normal model does not fit the observed distribution of χ^2 statistics.

We performed three secondary analyses. First, for 6 of the 33 traits, summary association statistics from independent cohorts were available. S-LD4M produced concordant M_e estimates for common SNPs on these data sets, despite the smaller sample size and much smaller number of regression SNPs (Table S8). (We did not estimate low-frequency M_e for these data sets, because summary statistics were not available for low-frequency SNPs.) Second, across all 33 traits, our estimates of M_e were not significantly correlated with effective sample size (Spearman $r = -0.23$, $p = 0.19$). Third, M_e provides an upper bound on the proportion of heritability explained by genome-wide significant loci for a trait (see Appendix, Properties of M_e). We compared this bound with a direct estimate of this proportion (which may be upwardly biased by winner's curse), and determined that the predicted bound corresponded fairly well with the estimate (Spearman $r^2 = 0.61$; Figure S6A). It also provided a conservative upper bound on the number of genome-wide significant SNPs (Figure S6B). Similar results were obtained when we used M_e estimates from an independent cohort (Figure S7).

Unequal Polygenicity of Common and Low-Frequency SNPs across 33 Complex Traits

We compared per-SNP M_e estimates for common and low-frequency SNPs across 15 well-powered traits (Figure 4A and Table 1). Polygenicity was 3.9× (95% CI: 2.9–5.2×), smaller for low-frequency SNPs than for

Table 1. Estimates of Polygenicity for Common and Low-Frequency SNPs across 33 Complex Traits

Trait	$\log_{10}M_e$ Common	$\log_{10}M_e$ LF	Sample Size (N or N_{eff})
Infinitesimal trait	4.69	4.98	
Number of children	4.50 (0.13)	3.95 (0.23)	457k
Schizophrenia	4.14 (0.04)	3.10 (0.16)	111k
Smoking status	4.11 (0.02)	NA	458k
College	4.08 (0.05)	2.86 (0.20)	389k
Morning person	4.02 (0.02)	2.16 (0.25)	411k
Age at first birth – F	3.99 (0.04)	NA	169k
Neuroticism	3.99 (0.23)	NA	372k
FVC	3.87 (0.16)	NA	372k
CVD including HT	3.81 (0.03)	NA	400k
BP – systolic	3.81 (0.02)	2.31 (0.42)	423k
BMI	3.78 (0.12)	2.53 (0.21)	458k
IBD	3.60 (0.04)	3.19 (0.10)	86k
Height	3.56 (0.02)	2.91 (0.05)	458k
WHR	3.50 (0.04)	2.54 (0.09)	458k
Age at menarche	3.47 (0.09)	2.47 (0.26)	242k
FEV1/FVC	3.38 (0.05)	3.01 (0.16)	372k
WBC count	3.30 (0.10)	2.42 (0.18)	445k
Eczema	3.20 (0.05)	NA	324k
Asthma	3.12 (0.03)	NA	187k
Eosinophil count	3.05 (0.04)	NA	440k
RA	3.03 (0.10)	NA	85k
Platelet count	2.97 (0.04)	2.51 (0.10)	444k
AID	2.93 (0.05)	NA	197k
BMD – heel	2.90 (0.08)	NA	446k
Alzheimer's	2.90 (0.11)	NA	47k
Type II diabetes	2.85 (0.14)	NA	74k
RBC distribution width	2.70 (0.05)	NA	443k
RBC count	2.67 (0.18)	NA	445k
Platelet distribution width	2.64 (0.08)	1.36 (0.54)	445k
Balding	2.59 (0.18)	2.16 (0.20)	180k
Age at menopause	2.55 (0.06)	NA	143k
Sunburn	1.99 (0.08)	NA	343k
Red hair	1.02 (0.22)	NA	81k

We report common variant estimates for all traits, and low-frequency estimates for well-powered traits (see Material and Methods). We also report the sample size; for binary traits we report the effective sample size, defined as $N_{eff} = 4/(1/N_{case} + 1/N_{control})$. The first row reports the effective number of independent SNPs.[32] M_e is close to this value when marginal effect sizes approximately follow a normal distribution, which does not imply that every SNP is causal. Abbreviations: FVC, forced vital capacity; CVD including HT, cardiovascular diseases including hypertension (most cases have HT); BP, blood pressure; BMI, body mass index; IBD, inflammatory bowel disease; WHR, waist-hip ratio; FEV1, forced expiratory volume; WBC, white blood cell count; RA, rheumatoid arthritis; AID, autoimmune and inflammatory diseases; BMD, bone mineral density; RBC, red blood cell.

common SNPs on average, with substantial variation across traits. This difference represents a lower bound on the difference in polygenicity between common SNPs and *de novo* mutations. We are not currently aware of any possible explanation for this difference except for the influence of negative selection.

The ~4× difference in common vs. low-frequency polygenicity is the largest that we would expect to observe, as
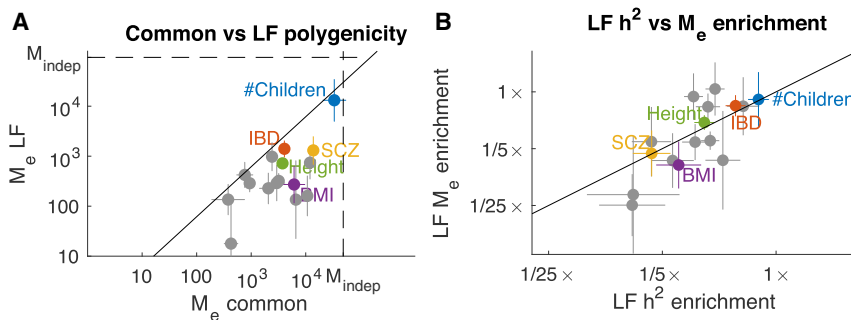
Figure 4. Comparison of Common and Low-Frequency Polygenicity across 15 Complex Traits

(A) Estimates of M_e for common and low-frequency SNPs. Estimates are meta-analyzed across well-powered traits. Common-variant polygenicity was $\sim 4\times$ greater on average than low-frequency polygenicity. Dotted lines denote the effective number of independent SNPs (M_{indep}) for common and low-frequency SNPs, respectively, corresponding to an infinitesimal (Gaussian) architecture. The solid line denotes equal per-SNP M_e.

(B) Estimates of polygenicity enrichment and heritability enrichment for low-frequency SNPs (compared to all common and low-frequency SNPs). The solid line denotes equal enrichment. Error bars denote 95% confidence intervals. Numerical results are reported in Tables 1 and S9.

the per-SNP heritability of common SNPs is also $3.9\times$ (95% CI: 3.5–4.4×) larger than that of low-frequency SNPs (consistent with previous estimates[12,21]). This concordance was consistent across traits (Figure 4B). The ratio h^2/M_e (i.e., the average unit of heritability; see Material and Methods) is expected to either decrease or remain constant at decreasing allele frequencies; individual low-frequency SNPs are not expected to explain more heritability than individual common SNPs. Our estimates indicate that this ratio remains constant at allele frequencies above ~ 0.005. This observation suggests three conclusions.

First, negative selection imposes a bound on per-SNP heritability that is approximately the same for common and low-frequency SNPs. (In units of per-allele effect size, the bound is higher for low-frequency SNPs.) This would occur under a model where selection coefficients scale with squared per-allele effect sizes, e.g., under the model of Simons et al.[28] and under the evolutionary models we consider below. We caution that the bound may vary across the genome, although it does not appear to vary across functional annotations (see below).

Second, the heritability of both common and low-frequency SNPs is predominantly explained by SNPs with standardized effect sizes near the effect-size bound, resulting in similar values of h^2/M_e. This suggests that the *de novo* effect-size distribution is dominated by mutations with effect sizes at least as large as per-allele effect size bound for low-frequency SNPs. If mutations with effect sizes approaching the common-SNP effect-size bound explained an equal proportion of variance in the *de novo* effect-size distribution as mutations with effect sizes approaching the low-frequency effect size bound, then they would also explain an equal proportion of low-frequency heritability. However, these SNPs would contribute little to $E(\beta^4)$, and h^2/M_e would be $\sim 2\times$ smaller for low-frequency SNPs than for common SNPs (note that $h^2/M_e = E(\beta^4)/3E(\beta^2)$, ignoring LD; see Material and Methods).

Third, the difference in polygenicity between common SNPs and new mutations is likely to be much greater than $4\times$ on average. The only way for this difference to be only $\sim 4\times$ would be for the *de novo* effect-size distribution to be abruptly truncated near the per-allele effect-size bound for low-frequency SNPs; the variance of this distribution would need to be driven by mutations with effect sizes confined to a narrow range. Instead, it is more likely that this distribution is not abruptly truncated, and mutations with increasingly large effects explain increasing proportions of variance in the *de novo* effect-size distribution, up to some point well beyond the low-frequency effect-size bound. If so, then the *de novo* effect-size distribution would be much more sparse than the effect-size distribution of low-frequency SNPs, and the polygenicity of common SNPs would be $\gg 4\times$ the polygenicity of the *de novo* effect-size distribution. Indeed, we observed a $\geq 30\times$ difference in our evolutionary models (see below).

Polygenicity of Functional Categories across 33 Complex Traits

We compared estimates of polygenicity enrichment with estimates of heritability enrichment across 25 main functional categories from the baseline-LD model[19] and meta-analyzed results across well-powered traits for 21 categories with at least 10 well-powered traits (Figure 5, Tables S9 and S10; 49% of trait-annotation pairs were well powered). For most annotations, polygenicity enrichment was approximately equal to heritability enrichment (regression slope = 0.93; $r^2 = 0.88$). For example, SNPs in conserved regions were $13\times$ enriched for heritability and $14\times$ enriched for polygenicity, and coding SNPs were $9.4\times$ enriched for heritability and $6.6\times$ enriched for polygenicity. (These functional enrichments for the union of common and low-frequency SNPs are larger than the corresponding enrichments for common SNPs, due to larger functional enrichment for low-frequency SNPs.[20]) Thus, heritability enrichment in functional categories is predominantly driven by differences in polygenicity, rather than differences in effect-size magnitude. In particular, h^2/M_e (i.e., the average unit of heritability) is constant across functional annotations, and the upper effect-size bound imposed by negative selection (Figure 1A) is approximately constant across annotations. In contrast, *de novo* mutations are expected to have much larger effect sizes in functionally important regions.[44] Thus, genetic signals of
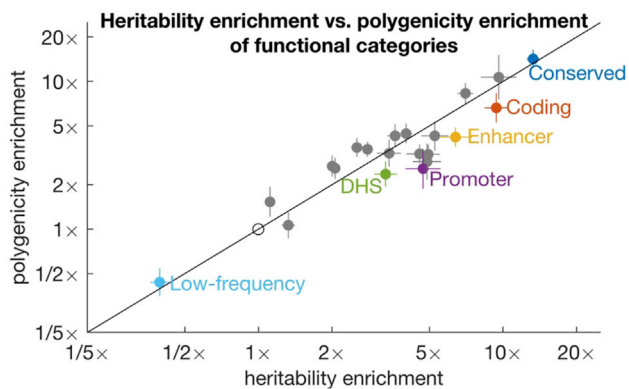
Figure 5. Estimates of Polygenicity Enrichment and Heritability Enrichment of Functional Categories
We report estimates for 20 functional categories plus low-frequency SNPs. Estimates are meta-analyzed across well-powered traits. Error bars denote 95% confidence intervals. Complete results for each trait are reported in Table S9 and meta-analyzed results are reported in Table S10.

important functional regions are constrained by negative selection (Figure 1).

We compared functional enrichment between groups of related traits (8 brain-related, 6 blood-related, and 5 immune-related traits; Figure S8). Brain-related traits had smaller functional enrichments both for heritability and for polygenicity, consistent with previous findings.[19,20,31] Smaller functional enrichment could be explained by stronger negative selection for these traits, which may strongly limit the enrichment of any functional category.[20] Stronger negative selection would also be consistent with greater genome-wide polygenicity for these traits (Table 1).

To investigate the relationship between functional enrichment and genome-wide polygenicity, we quantified the kurtosis explained by functional annotations (where kurtosis κ is inversely proportional to M_e; Equation 1). The proportion of kurtosis explained by functional annotations from the baseline-LD model ranged between 7% and 42% on the logarithmic scale (Figure S9A). Traits with smaller M_e (larger kurtosis) had larger kurtosis explained ($r^2 = 0.78$; Figure S9B).

GWAS Signals of Biologically Important Genes Are Constrained by Negative Selection

Flattening may lead to increased polygenicity not only at the level of SNPs, but also at the level of genes: as a consequence of negative selection, heritability is spread more evenly across genes, in comparison with the effect-size distribution of *de novo* mutations. GWAS effect sizes may be similar for SNPs near genes with small effects and for SNPs near genes with critical effects, and top GWAS SNPs may often implicate small-effect genes (Figure 1).

In order to investigate the impact of flattening on the distribution of heritability across genes, we explored two evolutionary fitness models: a model with both gene-level and SNP-level flattening, and a model with neither gene-

level nor SNP-level flattening that is potentially realistic in other respects. Although these specific models depend on many unknown parameters, they provide examples of qualitative phenomena: the first model illustrates that flattening at the level of SNPs may result from flattening at the level of genes, and the second model illustrates that it is possible for negative selection to cause common variants and rare variants to have different average effect sizes but the same level of polygenicity (as in the model of Zeng et al.[12]). Under both models, SNPs affect genes and genes affect the trait. Each gene has a trait effect size and a selection coefficient. In the first model (Figure 6A and Table S11), 5% of genes had large effect sizes (10× larger than small-effect genes), and these genes are always strongly constrained. We refer to this model as the "direct selection" model because it could arise from direct selecting acting on the trait (but see below). In the second model (Figure 6B and Table S11), there are no large-effect genes and different genes have different levels of constraint. We refer to this model as the "pleiotropic selection" model. We did not consider a pleiotropic model with unconstrained large-effect genes, which would lead to an unrealistic non-polygenic architecture. Further details of the two models are provided in the Material and Methods.

We analytically inferred the genetic architecture that arises under each model (see Material and Methods). Under both models, common variants have smaller per-allele effect sizes than low-frequency variants (Figures 6C and 6D), concordant with real traits;[12,19,21] moreover, both models produced a highly polygenic common-variant architecture (Table S11). However, polygenicity differed for the two models at lower allele frequencies (Figures 6E and 6F). SNP-level flattening (i.e., lower polygenicity at lower allele frequencies) was only observed under the direct selection model; polygenicity was ~30× lower for *de novo* SNPs than for common SNPs (vs. ~1.1× under the pleiotropic selection model). Similar our results on real traits (Figure 4), low-frequency polygenicity was ~4× smaller, and low-frequency heritability was also ~4× smaller (Figure 6E). Thus, a ~4× difference between common and low-frequency polygenicity is consistent with a much greater difference between common and *de novo* polygenicity (moreover, see below).

The concordance between low-frequency polygenicity enrichment and heritability enrichment is expected: the selection coefficient of a SNP scales with its squared per-allele effect size, and as a result, the bound on per-SNP heritability is approximately a constant function of allele frequency. Most heritability is explained by SNPs with effect sizes near the bound, so h^2/M_e is similar for common and low-frequency SNPs. However, in units of per-allele effect size, the bound is higher for low-frequency SNPs (Figure 1), so fewer variants approach the bound, leading to lower heritability and proportionally lower polygenicity. At allele frequencies smaller than 0.002, very few SNPs have effect sizes near the bound, so the heritability
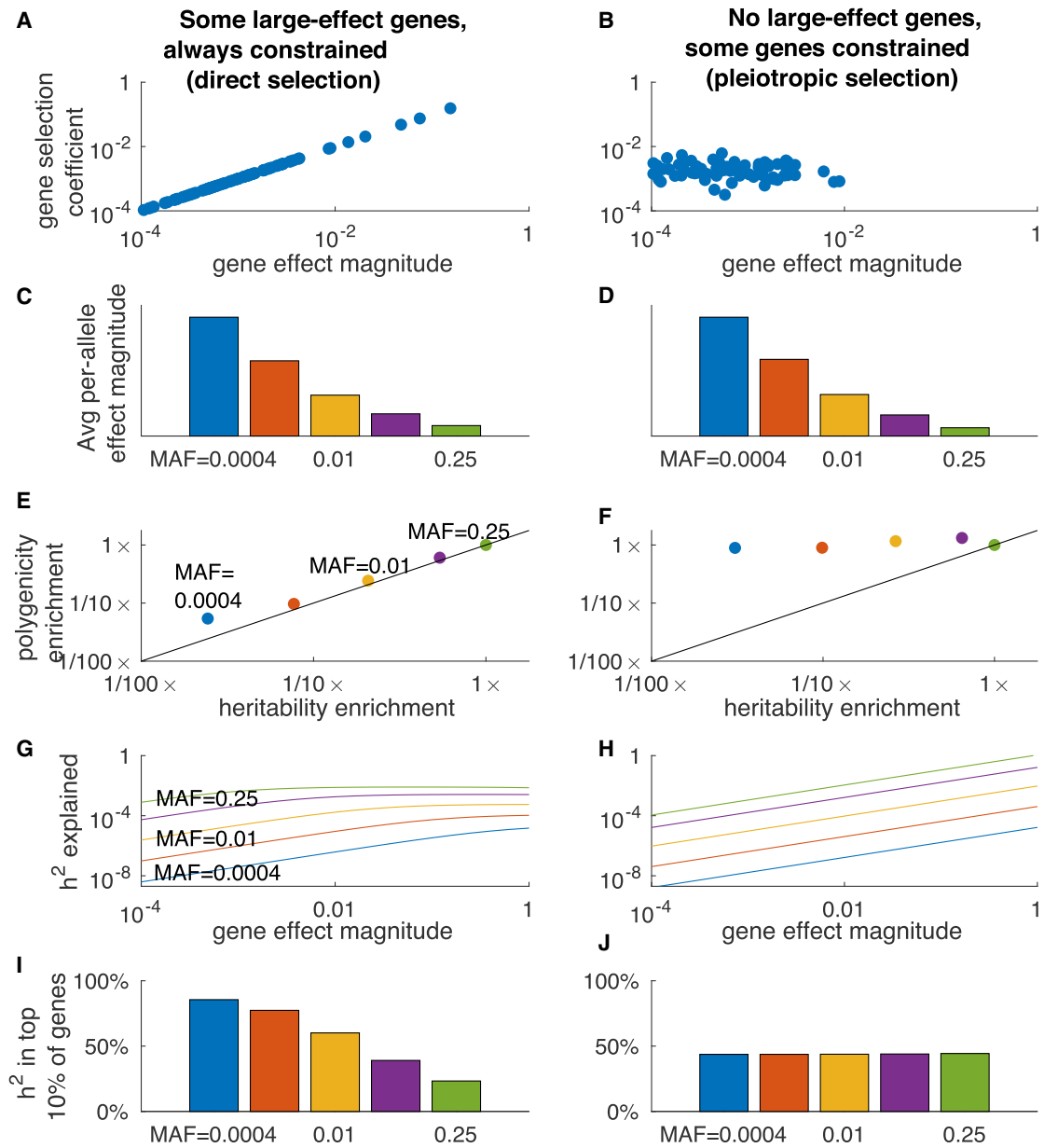
Figure 6. Gene-Level Flattening under an Evolutionary Model
In the left column (A, C, E, G, I), there are some large-effect genes, but direct stabilizing selection acting on the phenotype strongly constrains these genes. In the right column (B, D, F, H, J), there are no large-effect genes; pleiotropic stabilizing selection has varying effects on each gene, limiting common-SNP effect sizes on average.
(A and B) Joint distribution of gene effect size magnitudes and selection coefficients.
(C and D) Average squared per-allele effect sizes at different allele frequencies. The strength of selection was chosen to produce similar common-variant effect sizes in both columns.
(E and F) Heritability and polygenicity enrichment at different allele frequencies (relative to MAF = 0.25). Polygenicity at MAF = 0.25 is approximately equal for the two columns, due to the different distributions of gene effect sizes.
(G and H) Expected heritability explained by a single gene as a function of its effect size, for SNPs at different frequencies. In (G), the selection coefficient is proportional to the effect size. In (H), the selection coefficient is held constant.
(I and J) Proportion of heritability explained by the top 10% of largest-effect genes for SNPs at different allele frequencies. Numerical results are reported in Table S11.

is predominantly explained by SNPs with effect sizes well below the bound, and h^2/M_e decreases (Figure 6E).

For a real trait, the difference in polygenicity between common and *de novo* SNPs could be much larger than the ~30× difference we observed under the direct-selec-tion model. For example, when we added a few genes (0.25%) with extremely large effect sizes (100× larger than small-effect genes), the difference was ~600×, while the difference between common and low-frequency SNPs was still ~4× (Table S11). The difference in polygenicity

between common and *de novo* SNPs could also be slightly smaller than ~30× if the distribution of gene effect sizes is abruptly truncated, i.e., if there are no genes with effect sizes larger than the genes that explain most of the heritability of low-frequency SNPs. However, we observed that low-frequency SNPs have roughly equal heritability and polygenicity enrichment (Figure 4B). This concordance implies that at decreasing minor allele frequency, per-SNP heritability and polygenicity follow a linear relationship at allele frequencies greater than ~0.005, similar to Figure 6E. While this trend is expected to plateau eventually, it is unlikely to plateau extremely abruptly, suggesting that the polygenicity of new mutations is even smaller than the polygenicity of variants at allele frequency ~0.005, and much smaller than the polygenicity of low-frequency SNPs overall. Thus, we expect that the difference between the polygenicity of common and *de novo* SNPs is greater than one order of magnitude for most complex traits; complex traits would be far less polygenic if not for the influence of negative selection.

We caution that the difference between common vs. low-frequency polygenicity (and heritability) may be a poor proxy for the difference between common vs. *de novo* polygenicity when selection is sufficiently strong. When we increased the strength of selection in the direct selection model, we observed that the effects of flattening saturated at lower allele frequencies: despite stronger selection, the difference between common and low-frequency polygenicity decreased to ~2×, owing to ~2.5× increased polygenicity for low-frequency SNPs. In contrast, the difference between common and *de novo* polygenicity increased slightly to ~40× (Table S11). Thus, the strength of selection affects both the total amount of flattening and the range of allele frequencies over which differential polygenicity is observed. This effect may explain why number of children, though extremely polygenic (perhaps due to particularly strong selection), has a small difference in polygenicity between common and low-frequency SNPs (Figure 4).

We analytically computed the heritability explained by a gene (gene-heritability) as a function of its effect size, for SNPs at different allele frequencies. Under the direct selection model, common variant gene-heritability was approximately constant as a function of gene effect size (except for genes with near-zero effect), illustrating that flattening can act at the level of genes (Figure 6G); the effect was weaker for rare SNPs. Under the pleiotropic selection model (which has no SNP-level flattening), there was also no gene-level flattening (Figure 6H). We computed the proportion of heritability explained by the top 10% of genes (ranked by effect size), at different MAF strata. This proportion was strongly frequency dependent under the direct selection model, but not under the pleiotropic selection model (Figures 6I–6J).

These results suggest that large-effect disease genes are always constrained, and that one way that this constraint could arise is direct selection acting on the disease itself.

However, some forms of pleiotropic selection may produce similar effects as direct selection; for example, in the case of schizophrenia, pleiotropic selection on neurodevelopment broadly may mimic the effects of direct selection on schizophrenia specifically. Therefore, our results do not imply that direct selection is more important than pleiotropic selection and do not contradict models of selection that are primarily pleiotropic.[24,28]

If flattening occurs at the level of genes, then polygenicity should be increased near strongly constrained genes. We estimated the heritability and polygenicity of SNPs within 50 kb of 2,990 loss of function-intolerant genes from ExAC ("ExAC genic SNPs").[45,46] These SNPs were more strongly enriched for polygenicity (~2.9×) than for heritability (~1.7×) (Table S12). Compared with all genic SNPs (±50 kb), ExAC genic SNPs had 1.9× (95% CI: 1.7–2.0×) larger polygenicity enrichment but only 1.3× (95% CI: 1.3–1.4×) larger heritability enrichment, implying 0.71× (95% CI: 0.66–0.76×) smaller average effect sizes (Table S12). These estimates suggest that ExAC genes are more likely to be causal but also are more strongly constrained relative to their effect sizes when mutated. Moreover, they confirm that negative selection at the level of genes affects polygenicity at the level of SNPs.

If the GWAS signal of critical disease genes is constrained, then top GWAS loci should include a mixture of weak perturbations to critical, strongly constrained disease genes (like a "canary in a coal mine"[47]) and strong perturbations to less critical, weakly constrained genes. In particular, common coding variants—usually representing strong perturbations—are more likely to harbor top associations for any gene, leading to increased polygenicity among coding variants (Figure 4); therefore, they may be less likely to implicate critical, strongly constrained disease genes. We tested this prediction for 37 fine-mapped IBD GWAS loci,[48] comparing the probability of loss-of-function intolerance (pLI)[45] between genes harboring coding or noncoding causal risk SNPs. Indeed, 0/8 candidate genes containing fine-mapped coding variants had high pLI (≥0.9), compared to 12/29 candidate genes near fine-mapped noncoding variants (rank-sum test p = 0.006 for difference; Figure S10 and Table S13). Although pLI is different from gene effect size, this difference suggests that critical disease genes can be detected by prioritizing SNPs with subtle, context-dependent regulatory effects on their target gene[49,50] over SNPs with overt coding effects.[51]

Discussion

Our flattening hypothesis makes directly testable predictions about complex trait polygenicity. Using a new mathematical definition of polygenicity, we compared the polygenicity of 33 complex traits across allele frequencies and functional categories. We determined that low-frequency variants have lower polygenicity than

common variants and that biologically important functional categories have higher polygenicity in proportion to their higher heritability, consistent with the flattening hypothesis. Our results demonstrate that negative selection not only constrains common-variant effect sizes on average but flattens their distribution across the genome, explaining the extreme polygenicity of complex traits. The effect-size distribution of new mutations (not yet affected by negative selection) is far less polygenic, probably by orders of magnitude. Thus, the genes and loci with the most critical effects when mutated may not harbor the strongest common-variant associations.

Recent studies[17,52] proposed an "omnigenic model" of complex traits: a limited number of "core genes" have direct effects on a trait, but due to densely connected cellular networks, thousands of other "peripheral genes"—perhaps including every gene expressed in a relevant cell type—also contribute to heritability. Our results support the distinction between core and peripheral genes, suggesting that only a limited number of genes have critical phenotypic effects if mutated, and that these genes explain little heritability. (An alternative would be that many genes have direct phenotypic effects,[53] e.g., if a common disease is actually the union of multiple phenotypically similar but mechanistically distinct diseases or subtypes.[54]) One might expect that core genes, even if they explain a minority of heritability due to the larger number of peripheral genes, would usually harbor the strongest common-variant associations; however, our results suggest that they may not, due to being strongly constrained by selection.

A key question is how many genes and loci have critical effects, and how this number varies among traits. Our evolutionary modeling indicated that for a trait whose low-frequency polygenicity is 4× smaller than its common-variant polygenicity, its *de novo* polygenicity may be 30–600× smaller, and the two extremes would have completely different biological implications. A difference of only 30× would imply that more than 100 genes (depending on common-variant polygenicity) have critical effects, possibly spanning many pathways, cell types, and stages of disease progression. On the other hand, a difference of 600× would suggest a much smaller number of critical genes and pathways, with many other genes and pathways having auxiliary effects. In autism, a study of parent-child trios[55] estimated that a few hundred genes harbor penetrant coding mutations, with considerable uncertainty. Other traits may have completely different *de novo* effect size distributions; for example, schizophrenia has much weaker total *de novo* enrichment, and *de novo* point mutations have unclear penetrance and polygenicity.[56,57] This question could be addressed by estimating polygenicity for increasingly rare variants. We only analyzed SNPs at allele frequencies greater than 0.005, due to poor imputation accuracy at lower frequencies. Given well-powered exome-sequencing data, it may be possible to apply S-LD4M to rare coding SNPs, probably aggregating association signals across genes.

Polygenicity has not precluded GWASs from producing biological insights,[10] and our results suggest guidelines to accelerate their progress. First, GWAS follow-up studies should prioritize genes with evidence of constraint, as most critical disease genes are constrained. Loss-of-function variants provide a useful metric of constraint;[45] for GWASs, a maximally informative constraint metric might incorporate noncoding variation,[58] gene expression data,[59] and functional predictions.[60] Second, counter-intuitively, follow-up studies should prioritize associations that do not map to coding regions or large-effect regulatory elements—instead having weak or context-dependent regulatory effects—since SNPs with strongly deleterious effects on their target gene are less likely to implicate strongly pathogenic genes. As our ability to interrogate the gene-regulatory effects of GWAS SNPs improves,[60–62] a potential pitfall would be to prioritize GWAS SNPs with the largest regulatory effects. Third, rare-variant based evidence from exome-sequencing studies, even if underpowered, can be used to prioritize GWAS genes. Indeed, exome-sequencing studies[63–65] have been viewed as an attractive complement to GWASs,[17] and our results support this perspective. However, moderately rare coding variants may suffer the same limitations as common regulatory variants.

Finally, this study has several limitations. First, the nonparametric approach that we used to define and estimate polygenicity may not be optimal for every application. A recent study[13] fit a parametric model involving a mixture of normal distributions; this approach may provide more accurate estimates of missing heritability as a function of sample size, and it may more accurately predict the performance of risk prediction methods that make similar parametric assumptions.[36] Second, S-LD4M can produce biased estimates for depleted functional annotations, albeit in unrealistic settings (Figure S1). However, this bias does not affect enriched annotations or low-frequency variants, which are not in strong LD with common variants, and we have avoided reporting polygenicity estimates for depleted functional annotations. Third, S-LD4M produces noisy estimates for some annotations and traits, making it necessary to perform meta-analyses across well-powered traits, which may not be representative of all traits. Comparisons of heritability enrichment and polygenicity enrichment are not biased by this filtering process, as the same traits are used to estimate both types of enrichment. Fourth, S-LD4M can potentially be biased due to population stratification, and it should be applied only to data sets where stratification is well controlled. In particular, S-LD4M assumes that any stratification leads to approximately uniform inflation of χ^2 statistics; this assumption could be violated at loci under positive selection, such as at the LCT locus for height.[66] However, we have applied S-LD4M only to datasets that were corrected for population stratification (including UK Biobank, a relatively homogenous study). Fifth, although our evolutionary modeling supports the hypothesis that flattening

affects the distribution of heritability across genes, an alternative is that SNP-level flattening results from increased allelic heterogeneity for common variants near large-effect genes. However, this explanation would require that there be many independent associations per gene. Despite evidence of allelic heterogeneity, it is not a common phenomenon that multiple independent, similarly strong associations implicate the same gene.[30] Sixth, we have not analyzed any rare diseases, which may differ from complex traits both in their biological complexity and in their relationship with fitness; our findings do not have implications for their genetic architectures.[67] Seventh, inferences about components of heritability can potentially be biased by failure to account for LD-dependent architectures.[6,19,68,69] All of our primary analyses used the baseline-LD model, which includes six LD-related annotations.[19] The baseline-LD model is supported by formal model comparisons using likelihood and polygenic prediction methods, as well as analyses using a combined model incorporating alternative approaches.[71] There can be no guarantee that the baseline-LD model perfectly captures LD-dependent architectures; however, our estimates of M_e were similar with or without including annotations from the baseline-LD model (Table S5), suggesting that our estimates are unlikely to be affected by imperfect modeling of LD-dependent architectures. Despite these limitations, this study advances our understanding of genetic architecture and the evolutionary processes that shape it.

Appendix A

The Effective Number of Independently Associated SNPs

Here, we provide three definitions of the effective number of independent causal SNPs (M_e). First, we define M_e in terms of the mixed fourth moments of the distribution of causal and marginal effect sizes. This generalizes the definition in Equation 1, which corresponds to the special case of no LD between causal SNPs. Second, we provide an alternative definition (under a random effects model) that formalizes the illustration in Figure 2A, involving the average unit of heritability explained by a causal SNP (see Material and Methods). Third, we provide a concise definition that provides less intuition. We show that these definitions are equivalent below (see Equivalence of M_e Definitions).

First, let β denote the random vector of causal effect sizes for common and low-frequency SNPs. Let R be the fixed LD matrix, and let $\alpha = R\beta$ denote the random vector of marginal effect sizes. We use the notation α, β to denote randomly chosen entries of α, β. The heritability is $h^2 = \mathrm{E}(\beta^T R \beta) = M\mathrm{E}(\alpha\beta)$. M_e is defined as:

$$M_e = \frac{3M}{\kappa_e}, \quad \kappa_e = \frac{3\mathrm{E}(\alpha^2\beta^2) - 2\mathrm{E}(\beta^4)}{\mathrm{E}(\alpha\beta)^2}. \quad \text{(Equation A1)}$$

Another possible definition of polygenicity is the effective number of causal SNPs, denoted M_e':

$$M_e' = \frac{3M}{\kappa_e'}, \quad \kappa_e' = \frac{\mathrm{E}(\beta^4)}{\mathrm{E}(\beta^2)^2}. \quad \text{(Equation A2)}$$

The difference between M_e and M_e' is analogous to the difference between the heritability $\mathrm{E}(\beta^T R \beta)$ and an alternative definition of heritability that does not account for LD between variants, $\mathrm{E}(\beta^T \beta)$. In the presence of perfect LD, M_e' and $E(\beta^T\beta)$ are unidentifiable without making a strong assumption (roughly, that linked SNPs have independent causal effect sizes; see below). The two definitions of polygenicity are equivalent in the special case of no LD between causal SNPs (because $\alpha = \beta$ whenever $\beta \neq 0$), as in Equation 1. M_e' is different from the total number of causal SNPs, except when causal effect sizes follow a point-normal distribution. Because it is unlikely that causal effect sizes follow a normal distribution, we view the distinction between M_e and M_t as having greater importance than the distinction between M_e and M_e'.

Second, we consider a non-i.i.d. normal model:

$$\beta \sim N(0, \Sigma) \quad \text{(Equation A3)}$$

where Σ is a diagonal matrix with diagonal entries $\sigma_1^2, \ldots, \sigma_M^2$. This flexible model generalizes the point-normal model. h^2 is equal to $\mathrm{Tr}(\Sigma)$. Note that if $\Sigma \propto I$, then κ_e' is equal to 3 and M_e' is equal to M. Above, we used the notation $\mathrm{E}(.)$ to denote a uniform average across SNPs. We use the notation $\mathrm{E}_{h^2}(.)$ to denote an average across components of heritability, i.e., a weighted average where the probability of choosing SNP i is equal to σ_i^2/h^2. Using this notation, we refer to $\mathrm{E}_{h^2}(\alpha^2)$ as the average unit of heritability, and M_e is equal to:

$$M_e = \frac{h^2}{\mathrm{E}_{h^2}(\alpha^2)}. \quad \text{(Equation A4)}$$

M_e' can be defined in a similar manner:

$$M_e' = \frac{h^2}{\mathrm{E}_{h^2}(\beta^2)}. \quad \text{(Equation A5)}$$

Third, M_e can also be defined without specifying that Σ is a diagonal matrix. (The assumption that Σ is diagonal is similar to the common assumption that $\mathrm{E}(\beta^T R \beta) = \mathrm{E}(\beta^T\beta)$.[2,31]) Let $S = \Sigma^{1/2} R \Sigma^{1/2}$. In the diagonal case, S is a weighted LD matrix, where the rows and columns are weighted by the expected effect size of each SNP. The heritability is equal to $\mathrm{Tr}(S)$ (in the diagonal case, $\mathrm{Tr}(S) = \mathrm{Tr}(\Sigma)$). M_e is equal to:

$$M_e = \frac{\mathrm{Tr}(S)^2}{\mathrm{Tr}(S^2)}. \quad \text{(Equation A6)}$$

This definition, though concise and natural, provides little intuition.

Properties of M_e

M_e has notable mathematical properties.

- *Identifiability.* If there are two SNPs in perfect LD, there is no way to tell based on GWAS data whether one or both of them are causal. If a definition of polygenicity differs depending on whether one or both are causal, then it is unidentifiable. Indeed, M_e' is larger if both SNPs are causal than if only one SNP is causal. However, because both SNPs have identical values of $E(\alpha^2)$, the value of $E_{h^2}(\alpha^2)$ (and therefore M_e) is unaffected.

- *Missing Heritability.* M_e gives an upper bound on the proportion of heritability explained by SNPs whose effect sizes exceed a specified threshold T, such as the genome-wide significance threshold. This proportion can be denoted $P_{h^2}(\alpha^2 > T)$. Because α^2 is nonnegative,

$$P_{h^2}(\alpha^2 > T) \leq \frac{E_{h^2}(\alpha^2)}{T}. \qquad \text{(Equation A7)}$$

This bound is relatively tight when $T \gg E_{h^2}(\alpha^2)$. When T is a significance threshold (for example, $T = (30/N)$ corresponds to genome-wide significance), this bound is most relevant at low sample size.

- *Polygenic Prediction Accuracy.* Intuitively, increased polygenicity makes prediction more difficult. If Σ is given, there is a simple expression for the optimal risk prediction accuracy; this expression provides an upper bound on prediction accuracy in the case that Σ is not given:

$$E(r^2) = h^2 E_{h^2} \left(\frac{\alpha^2}{\alpha^2 + \frac{1}{N}} \right). \qquad \text{(Equation A8)}$$

(see Polygenic Prediction Accuracy below for derivation). At large N, prediction accuracy converges to h^2. At small N, it is approximately a linear function of sample size with slope inversely proportional to M_e:

$$E(r^2) \approx N h^2 E_{h^2}(\alpha^2) = N h^4 / M_e. \qquad \text{(Equation A9)}$$

In practice, polygenic prediction is usually performed using large datasets for which this approximation is not appropriate.

- *Effective Number of Independent SNPs.* Under an infinitesimal model, where every SNP is causal with a normally distributed causal effect size, M_e is equal to the effective number of independent SNPs[32] (M_{indep}). M_{indep} is defined as the number of SNPs divided by the average LD score, or in notation similar to Equation A6, as

$$M_{indep} = \frac{\text{Tr}(R)^2}{\text{Tr}(R^2)}. \qquad \text{(Equation A10)}$$

We note that M_e might be close to M_{indep} even when M_e' is much smaller than M. For example, if the genome comprises perfect LD blocks of 100 SNPs each, then M_e can

be equal to M_{indep} even if only 1 SNP per LD block is causal. We also note that the value of M_{indep} is strongly dependent on the range of allele frequencies that is specified, as rare SNPs (which have less LD) contribute strongly to M_{indep}. In contrast, M_e will not diverge if many rare SNPs are included, as these SNPs explain little heritability.

- *Symmetry.* M_e is symmetric with respect to the two fixed parameters in the random effects model, R and Σ: $M_e(R, \Sigma) = M_e(\Sigma, R)$. This property is shared by the LD-dependent definition of heritability, $E(\boldsymbol{\beta}^T R \boldsymbol{\beta}) = \text{Tr}(R\Sigma) = \text{Tr}(\Sigma R)$. It is not shared by M_e' or by the LD-independent definition of heritability, $E(\boldsymbol{\beta}^T \boldsymbol{\beta}) = \text{Tr}(\Sigma)$.

Stratified LD Fourth Moments Regression

S-LD4M is justified by an approximate regression equation, which states that the expected value of α^4 for SNP i is approximately proportional to the LD fourth moment of SNP i, with a proportionality constant that can be used to estimate M_e.

Let $\ell^{(2)}, \ell^{(4)}$ denote the LD second moment (LD score[5]) and LD fourth moment, respectively, for a randomly chosen SNP:

$$\ell_i^{(p)} = \sum_j r_{ij}^p \qquad \text{(Equation A11)}$$

The regression equation is:

$$E(\alpha^4 \mid \ell^{(2)}, \ell^{(4)}) \approx 3 E(\alpha^2 \mid \ell^{(2)}, \ell^{(4)})^2 + \ell^{(4)} K. \qquad \text{(Equation A12)}$$

In the first term, $E(\alpha^2 \mid \ell^{(2)}, \ell^{(4)}) = \ell^{(2)} \tau$, where the coefficient τ is the variance of causal effect sizes.[5,31] In the second term, K is related to M_e:

$$K = 3 E(\beta^2) \left[E_{h^2}(\alpha^2) - E(\alpha^2) \right]. \qquad \text{(Equation A13)}$$

Note that there are three kinds of expectations in Equations A12 and A13. First, $E(\alpha^2 \mid \ell^{(2)}, \ell^{(4)})$ and $E(\alpha^4 \mid \ell^{(2)}, \ell^{(4)})$ are conditioned on LD. Second, $E(\alpha^2)$ and $E(\beta^2)$ are not conditioned on LD; rather, they represent unweighted averages over all reference SNPs. Third, $E_{h^2}(a^2)$ is the average across components of heritability (not uniformly across SNPs).

In the case that there are P functional annotations, $\ell_i^{(4)}$ and K are vectors of size $1 \times P$ and $P \times 1$, respectively (see below). Similar to S-LDSC, we make an additivity assumption for the fourth moments of SNPs in the intersection of annotations. Our simulations violate this assumption, but it does not appear to result in bias.

This regression equation relies on an LD approximation. Roughly, the LD approximation states that if SNP i is in LD with causal SNP j, then the expected marginal effect size of SNP i is proportional to the *marginal* effect size of SNP j. This approximation is exact in important special cases, suggesting that it will be robust in practice. First, we

consider the stronger assumption that would be needed to estimate M_e', which is roughly that linked SNPs have independent causal effect sizes. More precisely, we would need to assume that the contribution of causal SNP j to the marginal effect size of SNP i, not conditional on other SNPs, is proportional to the causal effect size of SNP j:

$$cov\left(\alpha_i^2, \beta_j^2\right) \approx r_{ij}^2 var\left(\beta_j^2\right). \qquad \text{(Equation A14)}$$

Note that the expectations are not conditioned on other SNPs; if causal SNPs tend to cluster together, then the approximation will be poor, because the presence of causal SNP j would suggest additional nearby causal SNPs, inflating the left side but not the right side. We weaken Equation A14 to obtain our LD approximation:

$$cov\left(\alpha_i^2, \beta_j^2\right) \approx r_{ij}^2 cov\left(\alpha_j^2, \beta_j^2\right). \qquad \text{(Equation A15)}$$

Now, we are assuming that the contribution of causal SNP j to the marginal effect size of SNP i is proportional to the *marginal* effect size of SNP j. Note that Equation A14 implies Equation A15 by applying Equation A14 to the case $i = j$. We show that Equation A15 implies Equation A12 below (Derivation of Regression Equation). Equation A15 is not violated when other causal SNPs k are in strong LD with SNP j ($r_{jk}^2 \approx 1$), since they would contribute to α_i^2 in proportion to their contribution to α_j^2 (because $r_{ik}^2 \approx r_{ij}^2 r_{jk}^2$). However, it could be violated if there were other causal SNPs in weak LD with SNPs i and j (violating $r_{ik}^2 \approx r_{ij}^2 r_{jk}^2$). Thus, Equation A15 is exact both in the case that there is no clustering of causal SNPs and in the case that there is very tight clustering; this type of approximation is expected to be robust in intermediate cases as well, and our simulations support this intuition (Figure S2).

Equations A14 and A15 have second-moment analogues that can be used to justify LD score regression.[5,31] LDSC has been justified by assuming that correlated SNPs have uncorrelated effect sizes; this assumption can be stated as:

$$E\left(\alpha_i \beta_j\right) = r_{ij} E\left(\beta_j^2\right), \qquad \text{(Equation A16)}$$

which is analogous to Equation A14. This assumption allows LDSC to estimate a non-LD-dependent definition of heritability, $ME(\beta^2)$. However, a weaker assumption is also possible, analogous to Equation A15:

$$E\left(\alpha_i \beta_j\right) \approx r_{ij} E\left(\alpha_j \beta_j\right). \qquad \text{(Equation A17)}$$

This weaker assumption still holds when positively correlated SNPs tend to have positively (or negatively) correlated causal effect sizes. It allows LDSC to estimate $E(\alpha\beta)$, which is proportional to $\beta^T R \beta$, as follows:

$$E\left(\alpha_i^2\right) = \sum_j r_{ij} E\left(\alpha_i \beta_j\right) \approx \sum_j r_{ij}^2 E\left(\alpha_j \beta_j\right) = \ell_i^{(2)} E(\alpha\beta).$$

$$\text{(Equation A18)}$$

Thus, although LDSC has previously been justified using the assumption that correlated SNPs have uncorrelated effect sizes, only this weaker assumption is strictly necessary in order to estimate the LD-dependent definition of heritability.

Equation A12 relates $E(\alpha^4)$ with M_e. However, in practice, we do not observe α, but rather a noisy estimate $\widehat{\alpha}$. We can correct for sampling noise using:

$$E\left(\widehat{\alpha}^2 \mid \alpha\right) = \alpha^2 + \frac{1}{N}, \quad E\left(\widehat{\alpha}^4 \mid \alpha\right) = \alpha^4 + 6\alpha^2/N + 3/N^2,$$

$$\text{(Equation A19)}$$

where in practice we use the LD score regression intercept divided by N, denoted $1/\widehat{N}$, instead of $1/N$. We caution that this equation assumes that sampling error follows a normal distribution, which may be false for rare SNPs. In the presence of population stratification, it may approximately hold if stratification is relatively even across the genome, but it would be violated if some regions of the genome have been under population-specific positive selection.

We perform a two-step inference procedure. First, we use a slightly modified version of S-LDSC[31] (see below) to estimate $E(\alpha_i^2)$ and $E(\beta_i^2)$ for each SNP, conditional on their respective LD scores and annotation values. Second, we regress the $M \times 1$ vector

$$\widehat{\boldsymbol{\alpha}}^4 - \frac{6\widehat{\alpha}^2}{N} + \frac{3}{N^2} - 3\left(\ell^{(2)}\tau\right)^2, \qquad \text{(Equation A20)}$$

whose expectation is $\boldsymbol{\alpha}^4 - 3E(\alpha^2 \mid \ell^{(2)})^2$, on the $M \times P$ matrix $\boldsymbol{l}^{(4)}$ to obtain an estimate of the $1 \times P$ vector K. ($\boldsymbol{x}^k$ denotes element-wise exponentiation.) The regression is weighted: the weight of SNP i is 1 divided by the LD fourth moment of SNP i (to all common and low-frequency SNPs). This choice prevents over-counting of high-LD regions (see below). For each annotation A_p, we add $3E(E(\alpha^2)E(\beta^2) \mid A_p)$ to K_p and estimate M_e for each category.

Our modified version of S-LDSC uses a slightly modified weighting scheme and does not exclude large-effect SNPs. The regression weight of each SNP is 1 divided by the LD score for that SNP to all common and low-frequency SNPs; this choice is similar to the original version of S-LDSC,[31] but slightly modified for consistency with the weights used in S-LD4M. We do not exclude large-effect SNPs because these SNPs are important for estimating fourth moments; their exclusion would lead to upwardly biased estimates of M_e.

Equivalence of M_e Definitions

M_e can be defined in three equivalent ways. First, it can be defined in terms of fourth moments of the effect size distribution:

$$M_e = \frac{3M}{\kappa_e}, \quad \kappa_e = \frac{3E(\alpha^2\beta^2) - 2E(\beta^4)}{E(\alpha\beta)^2}. \qquad \text{(Equation A21)}$$

Second, it can be defined in terms of the average unit of heritability. Suppose that $\beta \sim N(0, \Sigma)$, where Σ is a diagonal matrix with entries $\sigma_1^2, ..., \sigma_M^2$. The average unit of heritability is defined as:

$$E_{h^2}(\alpha^2) = \frac{1}{h^2} \sum_i \sigma_i^2 E(\alpha_i^2 \mid R, \Sigma). \qquad \text{(Equation A22)}$$

$E_{h^2}(\alpha^2)$ is proportional to κ_e :

$$3E(\alpha^2 \beta^2) = \frac{3}{M} \sum_i E(\beta_i^2 \alpha_i^2 \mid \Sigma)$$

$$= \frac{3}{M} \sum_{i,j} r_{ij}^2 E(\beta_i^2 \beta_j^2 \mid \Sigma)$$

$$= \frac{3}{M} \sum_i \left[E(\beta_i^4 \mid \Sigma) + \sum_{j \neq i} r_{ij}^2 E(\beta_i^2 \Sigma) E(\beta_j^2 \mid \Sigma) \right]$$

$$= \frac{3}{M} \sum_i \left[2\sigma_i^4 + \sum_{i,j} r_{ij}^2 \sigma_i^2 \sigma_j^2 \right]$$

$$= 2E(\beta^4) + \frac{3}{M} \sum_i \sigma_i^2 E(\alpha_i^2 \mid \acute{O}), \qquad \text{(Equation A23)}$$

where we have used the fact that $E(\beta_i^4 \mid \Sigma) = 3\sigma_i^4$. Rearranging,

$$\kappa_e = \frac{3E(\alpha^2 \beta^2) - 2E(\beta^4)}{E(\alpha\beta)^2}$$

$$= \frac{3}{M} \sum_i \sigma_i^2 E(\alpha_i^2 \mid \Sigma) E(\alpha\beta)^2 \qquad \text{(Equation A24)}$$

$$= \frac{3}{E(\alpha\beta)} E_{h^2}(\alpha^2).$$

Substituting $h^2 = ME(\alpha\beta)$, we have a second definition of M_e:

$$M_e = h^2 E_{h^2}(\alpha^2). \qquad \text{(Equation A25)}$$

Third, define $S = \Sigma^{1/2} R \Sigma^{1/2}$. Then:

$$Tr(S^2) = \sum_{i,j} r_{ij}^2 \sigma_i^2 \sigma_j^2 = \sum_i \sigma_i^2 \sum_j r_{ij}^2 \sigma_j^2$$

$$= h^2 E_{h^2}(\alpha^2). \qquad \text{(Equation A26)}$$

Thus, we obtain another equivalent definition:

$$M_e = \frac{Tr(S)^2}{Tr(S^2)}, \qquad \text{(Equation A27)}$$

where $Tr(S) = h^2$. This definition is slightly more general than Equation A25, since it does not require that Σ is a diagonal matrix. (Note that in the diagonal case, $Tr(S) = Tr(\Sigma)$; more generally, these definitions are different, corresponding to the difference between $E(\beta^2)$ and $E(\alpha\beta)$.)

Polygenic Prediction Accuracy

If Σ is given, then it is clear what the optimal risk prediction scheme is. Given an estimate $\widehat{\alpha}$ of the marginal effect-size vector α, the expected phenotypic value of an individual with genotype X is:

$$E(X\beta \mid \widehat{\alpha}, \Sigma, X) = XE(\beta \mid \widehat{\alpha}, \Sigma). \qquad \text{(Equation A28)}$$

The expected prediction r^2 is:

$$r^2 = E((X\beta)(XE(\beta \mid \widehat{\alpha}, \Sigma)) \mid \Sigma)$$

$$= E(\beta^T R E(\beta \mid \widehat{\alpha}, \Sigma))$$

$$= E(\beta^T E(\alpha \mid \widehat{\alpha}, \Sigma)). \qquad \text{(Equation A29)}$$

Conveniently, we have eliminated $E(\beta \mid \cdot)$ from the expression, so the optimal risk prediction accuracy does not depend on R^{-1} (although the optimal risk predictor may). Now, $\widehat{\alpha} \sim N(\alpha, (1/N)R)$, so:

$$E(\alpha \mid \widehat{\alpha}, \Sigma) = \frac{\widehat{\alpha} E(\alpha^2 \mid \Sigma)}{\frac{1}{N} + E(\alpha^2 \mid \Sigma)}. \qquad \text{(Equation A30)}$$

Taking an expectation over SNPs,

$$r^2 = ME\left(\alpha\beta \frac{E(\alpha^2 \mid \Sigma)}{\frac{1}{N} + E(\alpha^2 \mid \Sigma)} \mid \Sigma \right)$$

$$= h^2 E_{h^2}\left(\;doublehyphen; 10pt \frac{\alpha^2}{\frac{1}{N} + \alpha^2} \right) \; doublehyphen; 10pt.$$

$$\text{(Equation 31)}$$

When N is large, r^2 converges to h^2; when N is small, r^2 is approximately $Nh^2 E_{h^2}(\alpha^2) = Nh^4/M_e$.

Derivation of Regression Equation

We assume that:

$$cov\left(\alpha_i^2, \beta_j^2 \mid \ell_i^{(2)}, \ell_i^{(4)} \right) \approx r_{ij}^2 cov\left(\alpha_j^2, \beta_j^2 \mid \ell_i^{(2)}, \ell_i^{(4)} \right)$$

$$\text{(Equation A32)}$$

We use this approximation as follows. First, we split up $E(\alpha_i^4)$:

$$E\left(\alpha_i^4 \mid \ell_i^{(2)}, \ell_i^{(4)} \right) = E\left(\alpha_i^2 \left[\sum_j r_{ij}^2 \beta_j^2 + \sum_{k \neq j} r_{ik} r_{ij} \beta_k \beta_j \right] \mid \ell_i^{(2)}, \ell_i^{(4)} \right).$$

Next, we use the fact that

$$E\left(\alpha_i^2 \left[\sum_j r_{ij}^2 \beta_j^2 \right] \right) = E\left(\left[\sum_j r_{ij}^2 \beta_j^2 \right]^2 \right)$$

and that

$$E\left(\alpha_i^2 \left[\sum_{k \neq j} r_{ik} r_{ij} \beta_k \beta_j \right] \right) = E\left(2 \sum_{k \neq j} r_{ik}^2 r_{ij}^2 \beta_k^2 \beta_j^2 \right)$$

to obtain:

$$E\left(\alpha_i^4 \mid \ell_i^{(2)}, \ell_i^{(4)}\right) = E\left(\left[\sum_j r_{ij}^2 \beta_j^2\right]^2 + 2\sum_{j \neq k} r_{ij}^2 r_{ik}^2 \beta_j^2 \beta_k^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)$$

$$= 3\sum_j r_{ij}^2 \left[E\left(\alpha_i^2 \beta_j^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right) - \frac{2}{3} r_{ij}^2 E\left(\beta_j^4 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)\right].$$

(Equation A33)

Now, we are ready to use Equation A32 to break down $E(\alpha_i^2 \beta_j^2) = cov(\alpha_i^2, \beta_j^2) + E(\alpha_i^2)E(\beta_j^2)$:

$$E\left(\alpha_i^4 \mid \ell_i^{(2)}, \ell_i^{(4)}\right) = 3\sum_j r_{ij}^2 \left[cov\left(\alpha_i^2, \beta_j^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)\right.$$
$$+ E\left(\alpha_i^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right) E\left(\beta_j^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)$$
$$\left. - \frac{2}{3} r_{ij}^2 E\left(\beta_j^4 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)\right]$$

$$\approx 3\sum_j r_{ij}^2 \left[r_{ij}^2 \left(E\left(\alpha_j^2 \beta_j^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right) - E\left(\alpha_j^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)\right.\right.$$
$$\left.\times E\left(\beta_j^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)\right) + E\left(\alpha_i^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right) E\left(\beta_i^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)$$
$$\left. - \frac{2}{3} r_{ij}^2 E\left(\beta_j^4 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)\right]$$

$$= 3E\left(\alpha_i^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)^2 + \sum_j r_{ij}^4 \left[3E\left(\alpha_j^2 \beta_j^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)\right.$$
$$\left. - 3E\left(\alpha_j^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right) E\left(\beta_j^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right) - 2E\left(\beta_j^4 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)\right].$$

(Equation 34)

Similar to LD score regression, we assume that SNPs in LD with regression SNPs (i.e., SNPs j which are in LD with SNP i) are representative of a larger population of SNPs (e.g., all common SNPs), allowing us to replace $E(\bullet_j \mid \ell_i^{(2)}, \ell_i^{(4)})$ with $E(\bullet_j)$:

$$E\left(\alpha_i^4 \mid \ell_i^{(2)}, \ell_i^{(4)}\right) = 3E\left(\alpha_i^2 \mid \ell_i^{(2)}, \ell_i^{(4)}\right)^2 + \left(3E(\alpha^2 \beta^2)\right.$$
$$\left. - 2E(\beta^4) - 3E(\alpha^2)E(\beta^2)\right)\ell_i^{(4)}.$$

(Equation A35)

We restate this equation for a randomly chosen SNP (rather than for a particular SNP i):

$$E\left(\alpha^4 \mid \ell^{(2)}, \ell^{(4)}\right) = 3E\left(\alpha^2 \mid \ell^{(2)}\right)^2 + \ell^{(4)} K, \quad \text{(Equation A36)}$$

where

$$K = 3E\left(\beta^2\right)\left[E_{h^2}\left(\alpha^2\right) - E\left(\alpha^2\right)\right]. \qquad \text{(Equation A37)}$$

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2019.07.003.

Web Resources

SLD4M open-source software, https://github.com/lukejoconnor/SLD4M
UK Biobank summary statistics, https://data.broadinstitute.org/alkesgroup/UKBB/

References

1. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., Sklar, P.; and International Schizophrenia Consortium (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748–752.
2. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42, 565–569.
3. Stahl, E.A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B.F., Kraft, P., Chen, R., Kallberg, H.J., Kurreeman, F.A., et al.; Diabetes Genetics Replication and Meta-analysis Consortium; and Myocardial Infarction Genetics Consortium (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat. Genet. 44, 483–489.
4. Loh, P.R., Bhatia, G., Gusev, A., Finucane, H.K., Bulik-Sullivan, B.K., Pollack, S.J., de Candia, T.R., Lee, S.H., Wray, N.R., Kendler, K.S., et al.; Schizophrenia Working Group of Psychiatric Genomics Consortium (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nat. Genet. 47, 1385–1392.
5. 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.
6. Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A., Lee, S.H., Robinson, M.R., Perry, J.R., Nolte, I.M., van Vliet-Ostaptchouk, J.V., et al.; LifeLines Cohort Study (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. 47, 1114–1120.
7. Moser, G., Lee, S.H., Hayes, B.J., Goddard, M.E., Wray, N.R., and Visscher, P.M. (2015). Simultaneous discovery, estimation

and prediction analysis of complex traits using a bayesian mixture model. PLoS Genet. *11*, e1004969.

8. Palla, L., and Dudbridge, F. (2015). A Fast Method that Uses Polygenic Scores to Estimate the Variance Explained by Genome-wide Marker Panels and the Proportion of Variants Affecting a Trait. Am. J. Hum. Genet. *97*, 250–259.

9. Shi, H., Kichaev, G., and Pasaniuc, B. (2016). Contrasting the genetic architecture of 30 complex traits from summary association data. Am. J. Hum. Genet. *99*, 139–153.

10. 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.

11. Yang, J., Fritsche, L.G., Zhou, X., Abecasis, G.; and International Age-Related Macular Degeneration Genomics Consortium (2017). A Scalable Bayesian Method for Integrating Functional Information in Genome-wide Association Studies. Am. J. Hum. Genet. *101*, 404–416.

12. 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.

13. Zhang, Y., Qi, G., Park, J.-H., and Chatterjee, N. (2018). Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat. Genet. *50*, 1318–1326.

14. Zhu, X., and Stephens, M. (2018). Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. Nat. Commun. *9*, 4361.

15. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.

16. Hirschhorn, J.N. (2009). Genomewide association studies—illuminating biologic pathways. N. Engl. J. Med. *360*, 1699–1701.

17. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell *169*, 1177–1186.

18. Mancuso, N., Rohland, N., Rand, K.A., Tandon, A., Allen, A., Quinque, D., Mallick, S., Li, H., Stram, A., Sheng, X., et al.; PRACTICAL consortium (2016). The contribution of rare variation to prostate cancer heritability. Nat. Genet. *48*, 30–35.

19. 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.

20. Gazal, S., Loh, P.-R., Finucane, H., Ganna, A., Schoech, A., Sunyaev, S., and Price, A. (2018). Low-frequency variant functional architectures reveal strength of negative selection across coding and non-coding annotations. Nat. Genet. *50*, 1600–1607.

21. 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.

22. Keightley, P.D., and Hill, W.G. (1988). Quantitative genetic variability maintained by mutation-stabilizing selection balance in finite populations. Genet. Res. *52*, 33–43.

23. Barton, N.H., and Keightley, P.D. (2002). Understanding quantitative genetic variation. Nat. Rev. Genet. *3*, 11–21.

24. 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.

25. Agarwala, V., Flannick, J., Sunyaev, S., Altshuler, D.; and GoT2D Consortium (2013). Evaluating empirical bounds on complex disease genetic architecture. Nat. Genet. *45*, 1418–1427.

26. Thornton, K.R., Foran, A.J., and Long, A.D. (2013). Properties and modeling of GWAS when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect. PLoS Genet. *9*, e1003258.

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

28. Simons, Y.B., Bullaughey, K., Hudson, R.R., and Sella, G. (2018). A population genetic interpretation of GWAS findings for human quantitative traits. PLoS Biol. *16*, e2002985.

29. Niemi, M.E.K., Martin, H.C., Rice, D.L., Gallone, G., Gordon, S., Kelemen, M., McAloney, K., McRae, J., Radford, E.J., Yu, S., et al. (2018). Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. Nature *562*, 268–271.

30. Hormozdiari, F., Zhu, A., Kichaev, G., Ju, C.J.-T., Segrè, A.V., Joo, J.W.J., Won, H., Sankararaman, S., Pasaniuc, B., Shifman, S., and Eskin, E. (2017). Widespread allelic heterogeneity in complex traits. Am. J. Hum. Genet. *100*, 789–802.

31. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al.; ReproGen Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; and RACI Consortium (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.

32. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.

33. Frei, O., Holland, D., Smeland, O.B., Shadrin, A.A., Fan, C.C., Maeland, S., O'Connell, K.S., Wang, Y., Djurovic, S., Thompson, W.K., et al. (2019). Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. Nat. Commun. *10*, 2417.

34. O'Connor, L.J., and Price, A.L. (2018). Distinguishing genetic correlation from causation across 52 diseases and complex traits. Nat. Genet. *50*, 1728–1734.

35. Conneely, K.N., and Boehnke, M. (2007). So many correlated tests, so little time! Rapid adjustment of P values for multiple correlated tests. Am. J. Hum. Genet. *81*, 1158–1168.

36. 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.; Schizophrenia Working Group of the Psychiatric Genomics Consortium, Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) study (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am. J. Hum. Genet. *97*, 576–592.

37. Wright, S. (1937). The distribution of gene frequencies in populations. Proc. Natl. Acad. Sci. USA *23*, 307–320.

38. 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.

39. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A.P., and Price, A.L. (2018). Mixed-model association for biobank-scale datasets. Nat. Genet. *50*, 906–908.

40. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. *12*, e1001779.

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

42. Han, J., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., Hankinson, S.E., Hu, F.B., Duffy, D.L., Zhao, Z.Z., et al. (2008). A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet. *4*, e1000074.

43. Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Magnusson, K.P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G., et al. (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat. Genet. *39*, 1443–1452.

44. Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V., et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature *485*, 242–245.

45. 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.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

46. Hormozdiari, F., Gazal, S., van de Geijn, B., Finucane, H.K., Ju, C.J.-T., Loh, P.-R., Schoech, A., Reshef, Y., Liu, X., O'Connor, L.J., et al. (2018). Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. Nat. Genet. *50*, 1041–1047.

47. Hunter, D.J., Altshuler, D., and Rader, D.J. (2008). From Darwin's finches to canaries in the coal mine–mining the genome for new biology. N. Engl. J. Med. *358*, 2760–2763.

48. Huang, H., Fang, M., Jostins, L., Umićević Mirkov, M., Boucher, G., Anderson, C.A., Andersen, V., Cleynen, I., Cortes, A., Crins, F., et al.; International Inflammatory Bowel Disease Genetics Consortium (2017). Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature *547*, 173–178.

49. Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343.

50. Shooshtari, P., Huang, H., and Cotsapas, C. (2017). Integrative genetic and epigenetic analysis uncovers regulatory mechanisms of autoimmune disease. Am. J. Hum. Genet. *101*, 75–86.

51. Mahajan, A., Wessel, J., Willems, S.M., Zhao, W., Robertson, N.R., Chu, A.Y., Gan, W., Kitajima, H., Taliun, D., Rayner, N.W., et al.; ExomeBP Consortium; MAGIC Consortium; and GIANT Consortium (2018). Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nat. Genet. *50*, 559–571.

52. Liu, X., Li, Y.I., and Pritchard, J.K. (2019). Trans effects on gene expression can drive omnigenic inheritance. Cell *177*, 1022–1034.

53. Wray, N.R., Wijmenga, C., Sullivan, P.F., Yang, J., and Visscher, P.M. (2018). Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. Cell *173*, 1573–1580.

54. Zuk, O., Hechter, E., Sunyaev, S.R., and Lander, E.S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. USA *109*, 1193–1198.

55. Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216–221.

56. Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M., et al. (2014). De novo mutations in schizophrenia implicate synaptic networks. Nature *506*, 179–184.

57. Howrigan, D., Rose, S.A., Samocha, K.E., Fromer, M., Cerrato, F., Chen, W.J., Churchhouse, C., Chambert, K., Chandler, S.D., Daly, M.J., et al. (2018). Schizophrenia risk conferred by protein-coding de novo mutations. bioRxiv. https://doi.org/10.1101/495036.

58. di Iulio, J., Bartha, I., Wong, E.H.M., Yu, H.-C., Lavrenko, V., Yang, D., Jung, I., Hicks, M.A., Shah, N., Kirkness, E.F., et al. (2018). The human noncoding genome defined by genetic diversity. Nat. Genet. *50*, 333–337.

59. Battle, A., Brown, C.D., Engelhardt, B.E., Montgomery, S.B.; GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; and eQTL manuscript working group (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

60. Zhou, J., Theesfeld, C.L., Yao, K., Chen, K.M., Wong, A.K., and Troyanskaya, O.G. (2018). Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. Nat. Genet. *50*, 1171–1179.

61. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., de Geus, E.J., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.

62. Ernst, J., Melnikov, A., Zhang, X., Wang, L., Rogov, P., Mikkelsen, T.S., and Kellis, M. (2016). Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. Nat. Biotechnol. *34*, 1180–1190.

63. Zuk, O., Schaffner, S.F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M.J., Neale, B.M., Sunyaev, S.R., and Lander, E.S. (2014). Searching for missing heritability: designing rare variant association studies. Proc. Natl. Acad. Sci. USA *111*, E455–E464.

64. Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E.,

Kähler, A., et al. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. Nature *506*, 185–190.

65. Do, R., Stitziel, N.O., Won, H.-H., Jørgensen, A.B., Duga, S., Angelica Merlini, P., Kiezun, A., Farrall, M., Goel, A., Zuk, O., et al.; NHLBI Exome Sequencing Project (2015). Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. Nature *518*, 102–106.

66. Campbell, C.D., Ogburn, E.L., Lunetta, K.L., Lyon, H.N., Freedman, M.L., Groop, L.C., Altshuler, D., Ardlie, K.G., and Hirschhorn, J.N. (2005). Demonstrating stratification in a European American population. Nat. Genet. *37*, 868–872.

67. Boycott, K.M., Vanstone, M.R., Bulman, D.E., and MacKenzie, A.E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat. Rev. Genet. *14*, 681–691.

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

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

71. Gazal, S., Marquez-Luna, C., Finucane, H.K., and Price, A.L. (2018). Reconciling S-LDSC and LDAK models and functional enrichment estimates. bioRxiv. https://doi.org/10.1101/256412.