DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Genetic Architecture of Common Diseases and Complex Traits

Citation

O'Connor, Luke Jen. 2019. Genetic Architecture of Common Diseases and Complex Traits. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:42029501

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Genetic Architecture of Common Diseases and Complex Traits

A dissertation presented

by

Luke Jen O'Connor

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biomedical Informatics

Harvard University

Boston, Massachusetts

May 2019

© Luke Jen O’Connor
All rights reserved.

Dissertation advisor: Alkes L. Price Luke Jen O'Connor

Genetic architecture of common diseases and complex traits

Abstract

Over the last decade, genome-wide association studies (GWAS) have revolutionized the study of disease. For the first time, the genetic architectures of common diseases and complex traits can be observed, their associations mapped and mined; and with these advances have come new mysteries: where heritability resides, how it is mediated, and why it is as it is. In this thesis, I introduce statistical methods to probe the genetic architecture of common diseases and complex traits. In chapter I, I ask why it is that complex traits are so polygenic. I show that negative selection *flattens* the distribution of heritability across the genome relative to the underlying biology. In chapter II, I describe the landscape of pleiotropy across complex traits. I show that pleiotropy is strongest among highly polygenic traits with shared causal cell types. In chapter III, I develop method to infer causal relationships between complex traits. This method harnesses the entire polygenic signal for each trait and quantifies the asymmetry in their pleiotropic effect-size distribution, and I show that it recapitulates known biology and suggests intriguing new hypotheses.

Table of contents

1. Title page – i
2. Copyright – ii
3. Abstract – iii
4. Table of Contents – iv
5. Front matter – v
6. Body of Text
    I. Introduction – 1
    II. Distinguishing genetic correlation from causation across 52 diseases and complex traits – 6
    III. Extreme polygenicity of complex traits is explained by negative selection – 42
    IV. Defining and characterizing pleiotropic relationships across 57 diseases and complex traits – 93
    V. Conclusion – 123
7. Back matter
    I. Supplementary material for Chapter I – 124
    II. Supplementary material for Chapter II – 128

Acknowledgement

I am grateful to the many individuals whose mentorship and kindness brought me this point in my academic and personal life. These include my parents, whose wisdom, devotion, and values have always been my chief inspiration and source of strength. They include my sister, who taught me so much about love and about growth. They include Isabel, the love of my life. And they include my advisor and mentor, Alkes Price, who taught me to be a scientist.

I. Introduction

For decades, human geneticists have mapped the heritable causes of rare, mostly monogenic disorders in pedigrees using linkage[1]. The heritable basis of common diseases has been less tractable, however, because the pattern of inheritance is polygenic: heritable variation in the population and within families results from many genetic variants, none of which is sufficient on its own to cause disease[2]. This distinction - between monogenic and polygenic traits – dates to at least 1918, when R.A. Fisher showed that the continuous distribution of heritable quantitative traits could be explained by additive contributions from many alleles of small effect[3].

In the early 2000s, motivated by the hypothesis that the heritability of complex disease might be explained by numerous common genetic variants, researchers catalogued common human genetic variation[4] and undertook genome-wide association studies (GWAS) in thousands of individuals[5,6]. Unexpectedly, these early studies discovered only a few susceptibility variants, mostly with very small effect sizes. These variants explained a very small fraction of disease heritability, and the remaining heritability was termed "missing"[7]. In 2010, variance components methods were used to show that collectively, common variants do explain a large fraction of complex trait heritability; however, their effect sizes are mostly too small to be discovered individually except at extremely large sample sizes[8].

Since 2010, variance component methods have emerged as a powerful tool to analyze the genetic architecture of complex traits. They have been used to show that heritability mostly localizes to regulatory regions near genes[9,10], including cell-type-specific regulatory regions and

differentially expressed genes[11,12]. They have been used to explore the effect of negative selection on genetic architecture[13-15], to quantify the effects of population stratification[10,16], and to identify pairs of traits with shared etiology[17].

However, key questions about the genetic architecture of complex traits remain. In chapter I, I introduce a method to distinguish between genetic correlations, which are common, and causal relationships, which are relatively rare. In chapter II, I ask what evolutionary processes might explain why common diseases and complex traits are so polygenic, and I introduce a new way to define and estimate polygenicity. In chapter III, I ask what types of biological relationships and functional processes explain pleiotropy among common diseases and complex traits, and I introduce a new way to define and estimate pleiotropy. A methodological theme is the use of fourth moments of the genetic effect-size distribution, which contain information about polygenicity, pleiotropy and causality that cannot be captured using traditional variance-component methods.

References

1 Gusella, J. F. *et al.* A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234 (1983).

2 Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *science* **322**, 881-888 (2008).

3 Fisher, R. A. 009: The Correlation Between Relatives on the Supposition of Mendelian Inheritance. (1918).

4 Consortium, I. H. The international HapMap project. *Nature* **426**, 789 (2003).

5 Consortium, W. T. C. C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661 (2007).

6 International Schizophrenia Consortium *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-752, doi:10.1038/nature08185 (2009).

7 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753, doi:10.1038/nature08494 (2009).

8 Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-569, doi:10.1038/ng.608 (2010).

9 Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *The American Journal of Human Genetics* **95**, 535-552 (2014).

10 Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nature genetics* **43**, 519 (2011).

11 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* **47**, 1228 (2015).

12 Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature Genetics* **50**, 621-629, doi:10.1038/s41588-018-0081-4 (2018).

13 Zeng, J. *et al.* Signatures of negative selection in the genetic architecture of human complex traits. *Nat Genet* **50**, 746-753, doi:10.1038/s41588-018-0101-4 (2018).

14 Gazal, S. *et al.* Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nature genetics* **49**, 1421 (2017).

15 Gazal, S. *et al.* Low-frequency variant functional architectures reveal strength of negative selection across coding and non-coding annotations. *Nature Genetics* **50**, 1600-1607 (2018).

16 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291 (2015).

17 Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature genetics* **47**, 1236 (2015).

II. Distinguishing genetic correlation from causation across 52 diseases and complex traits

This chapter is published in Nature Genetics *as O'Connor, Luke J. and Alkes L. Price, "Distinguishing genetic correlation from causation across 52 diseases and complex traits" (2018). L.J.O. and A.L.P. developed the methods, analyzed the results, and wrote the manuscript. L.O. performed the analyses. We are also grateful to Ben Neale, Soumya Raychaudhuri, Chirag Patel, Sek Kathiresan, Bogdan Pasaniuc and Hilary Finucane for helpful discussions, and to Po-Ru Loh and Steven Gazal for producing BOLT-LMM summary statistics for UK Biobank traits.*

Mendelian randomization (MR), a method to infer causal relationships, is confounded by genetic correlations reflecting shared etiology. We developed a model in which a latent causal variable (LCV) mediates the genetic correlation; trait 1 is partially genetically causal for trait 2 if it is strongly genetically correlated with the LCV, quantified using the genetic causality proportion (gcp). We fit this model using mixed fourth moments $E(\alpha_1^2\alpha_1\alpha_2)$ and $E(\alpha_2^2\alpha_1\alpha_2)$ of marginal effect sizes for each trait; if trait 1 is causal for trait 2, then SNPs affecting trait 1 (large α_1^2) will have correlated effects on trait 2 (large $\alpha_1\alpha_2$), but not vice versa. In simulations, our method avoided false positives due to genetic correlations, unlike MR. Across 52 traits (average N=331k), we identified 30 causal relationships with high gcp estimates. Novel findings included a causal effect of LDL on bone mineral density, consistent with clinical trials of statins in osteoporosis.

Introduction

Mendelian Randomization (MR) is widely used to identify potential causal relationships among heritable traits, potentially leading to new disease interventions[1-12]. Genetic variants that are significantly associated with one trait, the "exposure,'' are used as genetic instruments to test for a causal effect on a second trait, the "outcome." If the exposure is causal, then variants

affecting the exposure should affect the outcome proportionally. For example, LDL[3,13] and triglycerides[4] (but not HDL[3]) causally affect coronary artery disease risk. However, pleiotropy presents a challenge for MR, especially when it produces a genetic correlation and when the exposure is highly polygenic[2,11,12,14-16]. Sometimes, this challenge can be addressed using curated genetic variants without pleiotropic effects; this approach is most appropriate for molecular traits (e.g. LDL). For other complex traits, statistical approaches have been used to reduce the likelihood of confounding, such as MR-Egger [7] and bidirectional MR [11,17,18]. However, these approaches have their own limitations.

We introduce a latent causal variable (LCV) model, under which the genetic correlation between two traits is mediated by a latent variable having a causal effect on each trait. We define trait 1 as *partially genetically causal* for trait 2 when it is strongly correlated with the causal variable, implying that part of the genetic component of trait 1 is causal for trait 2. We quantify partial causality using the *genetic causality proportion*. We showed in simulations that LCV has major advantages over MR methods, and we applied it to 52 diseases and complex traits.

Results

Overview of methods

The latent causal variable (LCV) model is based on a latent variable *L* that mediates the genetic correlation between the two traits (Figure 1a). Under the LCV model, trait 1 is *fully genetically causal* for trait 2 if it is perfectly genetically correlated with *L*; “fully'' means that the entire

genetic component of trait 1 is causal for trait 2 (Figure 1b). More generally, trait 1 is *partially genetically causal* for trait 2 if the latent variable has a stronger genetic correlation with trait 1 than with trait 2; “partially'' means that part of the genetic component of trait 1 is causal for trait 2. In order to quantify partial causality, we define the *genetic causality proportion* (gcp) of trait 1 on trait 2. The gcp ranges between 0 (no partial genetic causality) and 1 (full genetic causality). A high value of gcp (even if it is not exactly 1) implies that interventions targeting trait 1 are likely to affect trait 2. An intermediate value implies that some interventions targeting trait 1 may affect trait 2. (However, we caution that an intervention may fail to mimic genetic perturbations, e.g. due to its timing relative to disease progression.) For example, a recent study suggested either a fully causal effect of age at menarche (AAM) on height or a shared hormonal pathway affecting both traits[11]. If this shared pathway (modeled by L) has a large effect on AAM but a small effect on height, then AAM would be strictly partially genetically causal for height. Indeed, LCV produced an intermediate gcp estimate ($\widehat{\mathrm{gcp}} = 0.43(0.10)$, see below). We caution that low gcp estimates are not evidence of full genetic causality, and we refer to trait pairs with low gcp estimates as having limited partial genetic causality. LCV p-values test the null hypothesis that gcp=0, and a highly significant p-value does not imply a high gcp.

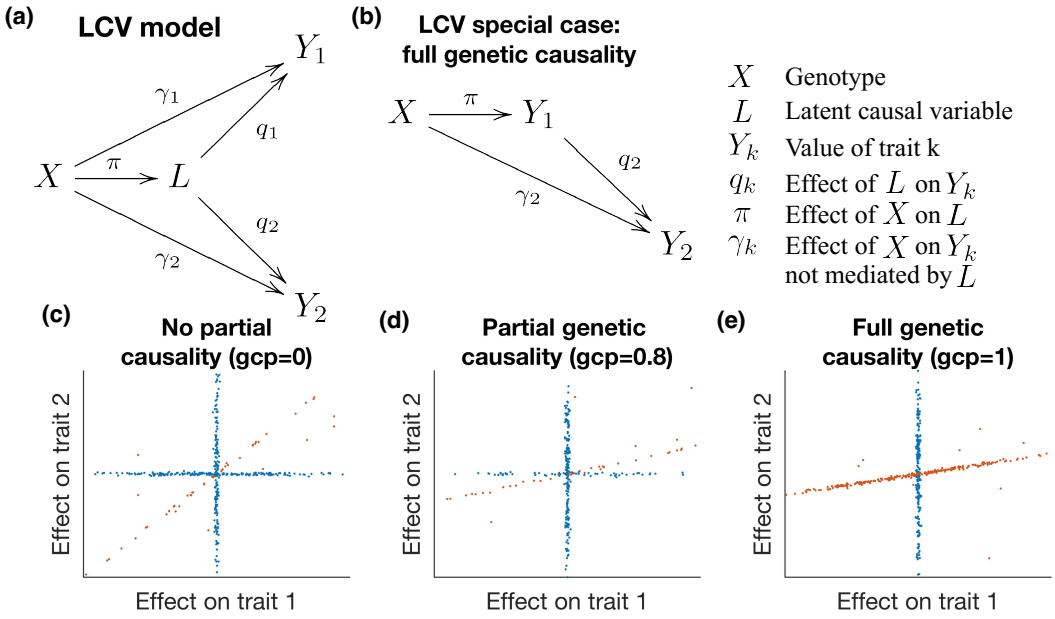


Figure 1: Illustration of the latent causal variable model. We display the relationship between genotypes X, latent causal variable L and trait values Y_1 and Y_2. (a) Full LCV model. The genetic correlation between traits Y_1 and Y_2 is mediated by L, which has normalized effects q_1 and q_2 on each trait. (b) When $q_1 = 1, Y_1$ is perfectly genetically correlated with L (so L does not need to be shown in the diagram), and we say that Y_1 is fully genetically causal for Y_2. (c) Example genetic architecture of genetically correlated traits with no genetic causality ($\text{gcp} = 0$, i.e. $q_2 = q_1 < 1$). Slight noise is added to SNP effects for illustration. Orange SNPs have correlated effects on both traits via L, while blue SNPs do not. (d) Example genetic architecture of genetically correlated traits with partial genetic causality ($\text{gcp} = 0.8$, i.e. $q_2 < q_1 < 1$). (e) Example genetic architecture of genetically correlated traits with full genetic causality ($\text{gcp} = 1$, i.e. $q_2 < q_1 = 1$).

In order to test for partial genetic causality and to estimate the gcp, we exploit the fact that if trait 1 is partially genetically causal for trait 2, then most SNPs affecting trait 1 will have proportional effects on trait 2, but not vice versa (Figure 1c-e). Instead of using thresholds to select subsets of SNPs[11], we compare the mixed fourth moments $E(\alpha_1^2\alpha_1\alpha_2)$ and $E(\alpha_2^2\alpha_1\alpha_2)$ of marginal effect sizes for each trait. The rationale for utilizing these mixed fourth moments is that if trait 1 is causal for trait 2, then SNPs with large effects on trait 1 (large α_1^2) will have proportional effects on trait 2 (large $\alpha_1\alpha_2$), so that $E(\alpha_1^2\alpha_1\alpha_2)$ will be large;

conversely, SNPs with large effects on trait 2 (large α_2^2) will generally not affect trait 1 (small $\alpha_1\alpha_2$), so that $E(\alpha_2^2\alpha_1\alpha_2)$ will be smaller. Thus, estimates of the mixed fourth moments can be used to test for partial genetic causality and to estimate the gcp. We note that LCV, unlike MR, does not distinguish between the "exposure" and the "outcome"; trait 1 and trait 2 are interchangeable labels.

LCV assumes that joint effect size distribution for two traits is a sum of two independent distributions: (1) a shared genetic component corresponding to *L*, whose values are proportional for both traits; and (2) a distribution that does not contribute to the genetic correlation (see Methods). We interpret the first distribution as "mediated'' effects (corresponding to π; Figure 1a) and the second distribution as "direct'' effects (corresponding to γ). The LCV model assumption is strictly weaker than the "exclusion restriction" assumption of MR (see Methods).

Under the LCV model, the genetic causality proportion is defined as the number *x* such that:

$$\frac{q_2^2}{q_1^2} = \left(\rho_g^2\right)^x, \qquad (1)$$

where q_k is the normalized effect of *L* on trait *k* (Figure 1a), and ρ_g is the genetic correlation[16] (note that $\rho_g = q_1 q_2$). When the gcp is equal to 1, trait 1 is fully genetically causal for trait 2; when it is positive but less than 1, trait 1 is partially genetically causal. When it

is negative, trait 2 is partially genetically causal. The gcp can be defined without making LCV (or other) model assumptions (see Methods).

In order to estimate the gcp, we utilize the following relationship between the mixed fourth moments of the marginal effect size distribution and the parameters q_1 and q_2:

$$E(\alpha_1^3\alpha_2) = \kappa_\pi\, q_1^3 q_2 + 3\rho_g, \qquad (2)$$

where π is the effect of a SNP on *L* and $\kappa_\pi = E(\pi^4) - 3$ is the excess kurtosis of π (see Methods). This equation implies that if $|E(\alpha_1^3\alpha_2)| \geq |E(\alpha_1\alpha_2^3)|$, then $q_1^2 \geq q_2^2$ and $\text{gcp} \geq 0$.

We calculate statistics S(x) for each possible value of gcp=x, using equation (2). These statistics also depend on the heritability[19], the genetic correlation[16], and the cross-trait LDSC intercept[19]. We estimate the variance of these statistics using a block jackknife and obtain an approximate likelihood function for the gcp. We compute a posterior mean gcp estimate (and a posterior standard deviation) using a uniform prior. We test the null hypothesis (that gcp=0) using S(0). Details of the method are provided in the Methods section. We have released open source software implementing the LCV method (see URLs).

Simulations

To compare the calibration and power of LCV with existing causal inference methods, we performed a wide range of null and causal simulations involving simulated summary statistics

with no LD. We compared four main methods: LCV, random-effect two-sample MR[5,9] (denoted MR), MR-Egger[7], and Bidirectional MR[11] (see Methods). We applied each method to simulated GWAS summary statistics (N=100k individuals in each of two non-overlapping cohorts; M=50k independent SNPs[20]) for two heritable traits (h^2=0.3), generated under the LCV model. LCV uses LD score regression [19]; for simulations with no LD, we use constrained-intercept LD score regression (simulations with LD are described below).

First, we performed null simulations (gcp=0) with uncorrelated pleiotropic effects (via γ; Figure 1a) and zero genetic correlation. 1% of SNPs were causal for both traits (with independent effect sizes, explaining 20% of heritability for each trait), and 4% of SNPs were causal for each trait exclusively (Figure 2a). LCV produced conservative p-values (0.0% false positive rate at $\alpha = 0.05$); our normalization of the test statistic can lead to conservative p-values when the genetic correlation is low (see Methods). All three main MR methods produced well-calibrated p-values. Even though the “exclusion restriction" assumption of MR is violated here, these results confirm that uncorrelated pleiotropic effects do not confound random-effect MR at large sample sizes[21]. (Such pleiotropy is known to cause false positives if a less conservative fixed-effect approach is used[22].) In these simulations, all methods except LCV used the set of approximately 170 SNPs (on average) that were genome-wide significant ($p < 5 \times 10^{-8}$) for trait 1 (or approximately 330 SNPs that were genome-wide significant for either trait, in the case of Bidirectional MR).

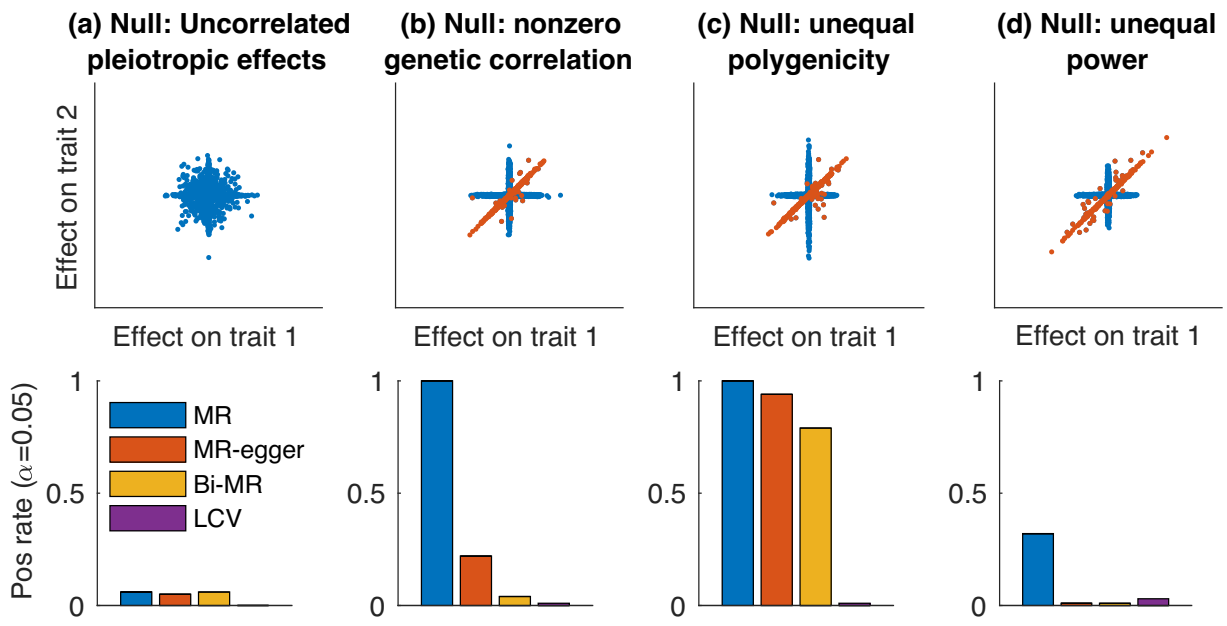


Figure 2: Null simulations with no LD to assess calibration. We compared LCV to three main MR methods (two-sample MR, MR-Egger and Bidirectional MR). We report the positive rate ($\alpha = 0.05$) for a causal (or partially causal) effect. Scatterplots illustrate the bivariate effect size distribution. (a) Null simulation (gcp=0) with uncorrelated pleiotropic effects and zero genetic correlation. (b) Null simulation with nonzero genetic correlation. (c) Null simulation with nonzero genetic correlation and differential polygenicity between the two traits. (d) Null simulation with nonzero genetic correlation and differential power for the two traits. Results for each panel are based on 4000 simulations.

Second, we performed null simulations with a nonzero genetic correlation. 1% of SNPs had causal effects on *L*, and *L* had effects $q_1 = q_2 = \sqrt{0.2}$ on each trait (so that $\rho_g = 0.2$). 4% of SNPs were causal for each trait exclusively (Figure 2b). MR and MR-Egger both produced excess false positives, while Bidirectional MR and LCV produced well-calibrated p-values. These simulations violate the MR-Egger assumption that the magnitude of pleiotropic effects on trait 2 are independent of the magnitude of effects on trait 1 (the “InSIDE" assumption) [7], as SNPs with larger effects on *L* have larger effects on both trait 1 and trait 2 on average, consistent with known limitations [22].

Third, we performed null simulations with a nonzero genetic correlation and differential polygenicity in the non-shared genetic architecture between the two traits. 1% of SNPs were causal for L with effects $q_1 = q_2 = \sqrt{0.2}$ on each trait, 2% were causal for trait 1 but not trait 2, and 8% were causal for trait 2 but not trait 1 (Figure 2c). Thus, the likelihood that a SNP would be genome-wide significant was higher for causal SNPs affecting trait 1 only than for causal SNPs affecting trait 2 only. As a result of this imbalance, Bidirectional MR (as well as other MR methods) produced excess false positives, unlike LCV.

Fourth, we performed null simulations with a nonzero genetic correlation and differential power for the two traits, reducing the sample size from 100k to 20k for trait 2. 0.5% of SNPs were causal for *L* with effects $q_1 = q_2 = \sqrt{0.5}$ on each trait, and 8% were causal for each trait exclusively (Figure 2d). Because per-SNP heritability was higher for shared causal SNPs than for non-shared causal SNPs, shared causal SNPs were more likely to reach genome-wide significance in the smaller trait 1 sample (N=20k), leading to a similar imbalance as in Figure 2c. As a result, Bidirectional MR (as well as other MR methods) produced excess false positives, while LCV produced well-calibrated p-values.

Next, we performed causal simulations (with full genetic causality) to assess whether LCV is well-powered to detect a causal effect. We caution that LCV had lower power in simulations with LD (see below). First, we chose a set of default parameters: N=25k for each trait, 5% of SNPs causal for trait 1 (the causal trait), a (fully) causal effect of size $q_2 = 0.2$ of trait 1 on trait 2, and 5% of SNPs causal for trait 2 only (Figure 3a). There were ~15 genome-wide significant

SNPs on average, explaining ~2% of h^2. LCV was well-powered to detect a causal effect at $\alpha = 0.001$, while MR had lower power and bidirectional MR and MR-Egger had low power.

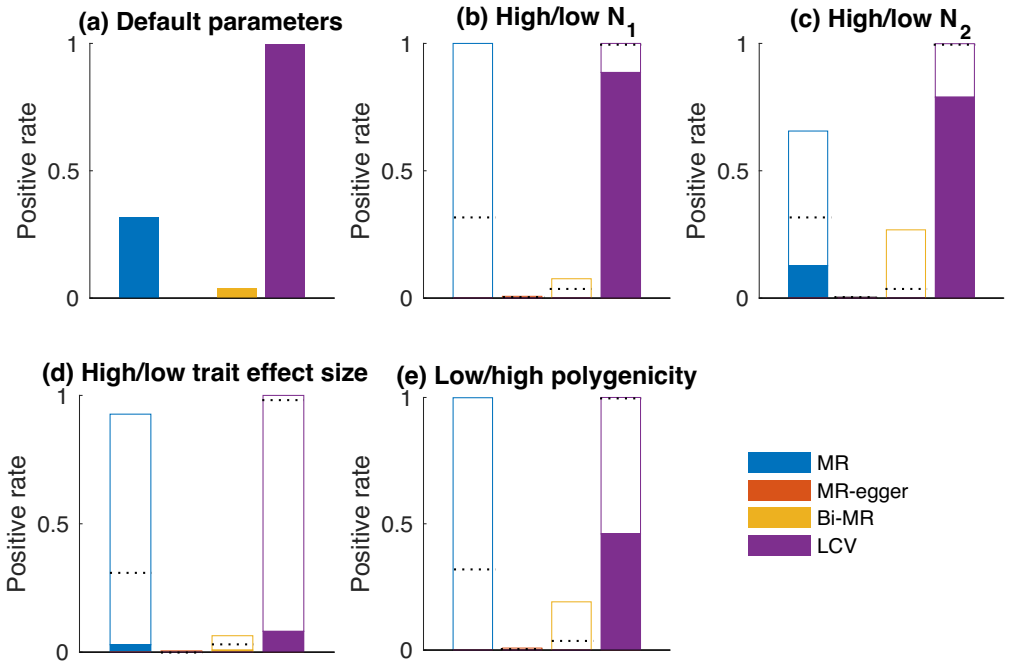


Figure 3: Causal simulations with no LD to assess power. We compared LCV to three main MR methods (two-sample MR, MR-Egger and Bidirectional MR). We report the positive rate ($\alpha = 0.001$) for a causal (or partially causal) effect. (a) Causal simulations with default parameters: $N_1 = N_2 = 25\text{k}$; $M = 50\text{k}$; $q_1 = 1, q_2 = 0.2$ (results also displayed as dashed lines in panels b-e). (b) Higher (unfilled) or lower (filled) sample size for trait 1 ($N_1 = 50\text{k}$ and $N_1 = 12.5\text{k}$ respectively). (c) Higher (unfilled) or lower (filled) sample size for trait 2 ($N_2 = 50\text{k}$ and $N_2 = 12.5\text{k}$ respectively). (d) Higher (unfilled) or lower (filled) causal effect size of trait 1 on trait 2 ($q_2 = 0.4$ and $q_2 = 0.1$ respectively). (e) Lower (unfilled) or higher (filled) polygenicity for trait 1. Results for each panel are based on 1000 simulations.

Second, we reduced the sample size for trait 1 (Figure 3b), finding that LCV had high power while the MR methods had very low power, owing to the small number of genome-wide significant SNPs. We caution that for real traits, heritability estimates can be noisy at very low sample size, potentially leading to unreliable results (see below).

Third, we reduced the sample size for trait 2 (Figure 3c). LCV had high power, while other methods had low power. The effect of trait 2 sample size on MR power was more modest than the effect of trait 1 sample size, suggesting that the number of genome-wide significant SNPs (ascertained using trait 1) is the primary limiting factor for MR power.

Fourth, we reduced the causal effect size of trait 1 on trait 2 (Figure 3d). LCV had low power, and other methods had very low power. Fifth, we increased the polygenicity of the causal trait (Figure 3e). LCV had moderate power while the MR methods had very low power, again owing to the low number of genome-wide significant SNPs.

In order to investigate potential limitations of our approach, we performed null and causal simulations under genetic architectures that violate LCV model assumptions. We simulated four types of LCV model violations: (1) null simulations with a bivariate Gaussian mixture model, where one of the mixture components generates imperfectly correlated effect sizes on the two traits; (2) null simulations with two latent causal variables; (3) causal simulations with a bivariate Gaussian mixture model; and (4) causal simulations with an additional latent confounder. LCV produced well-calibrated p-values under models of type (1) (Supplementary Figure 2.1a-c); in addition, these simulations recapitulated the limitations of existing methods (Figure 2). Models of type (2) sometimes caused LCV (and existing methods) to produce false positives (Supplementary Figure 2.1d-e); however, extreme values of the simulation parameters were required in order for LCV to produce high gcp estimates, implying that results with high gcp estimates are extremely unlikely to be false positives (Supplementary Figure 2.2). Causal

models of type (3-4) lead to reduced power for LCV (and other methods) (Supplementary Figure 2.1f-g), as well as downwardly biased gcp estimates for LCV.

Next, we performed simulations with real LD patterns to further assess the robustness of the LCV method. In null simulations with a wide range of parameter settings, LCV produced approximately well-calibrated or conservative false positive rates, except for simulations at low sample size with noisy heritability estimates. In simulations with a range of gcp values, we determined that our posterior mean gcp estimates are approximately unbiased and that our posterior standard errors are approximately well-calibrated (Supplementary Figure 2.3).

Application to real traits

We applied LCV and the MR methods to GWAS summary statistics for 52 diseases and complex traits, including summary statistics for 37 UK Biobank traits[27,28] computed using BOLT-LMM[29] (average N=429k) and 15 other traits (average N=54k) (see Methods). These traits were selected based on the significance of their heritability estimates ($Z_h > 7$), and trait pairs with very high genetic correlations ($|\rho_g| > 0.9$) were pruned. As in previous work, we excluded the MHC region from all analyses, due to its unusually large effect sizes and long-range LD patterns[19].

We applied LCV to the 429 trait pairs (32% of all trait pairs) with a nominally significant genetic correlation (p<0.05), detecting significant evidence of full or partial genetic causality for 59 trait pairs (FDR < 1%), including 30 trait pairs with $\widehat{\text{gcp}}$>0.6. We primarily focus on trait pairs with

high gcp estimates, which have greater biological interest (and are extremely unlikely to be false positives; see Simulations). Results for selected trait pairs are displayed in Figure 4, and results for the 30 trait pairs with $\widehat{gcp}$>0.6 are reported in Table 1.

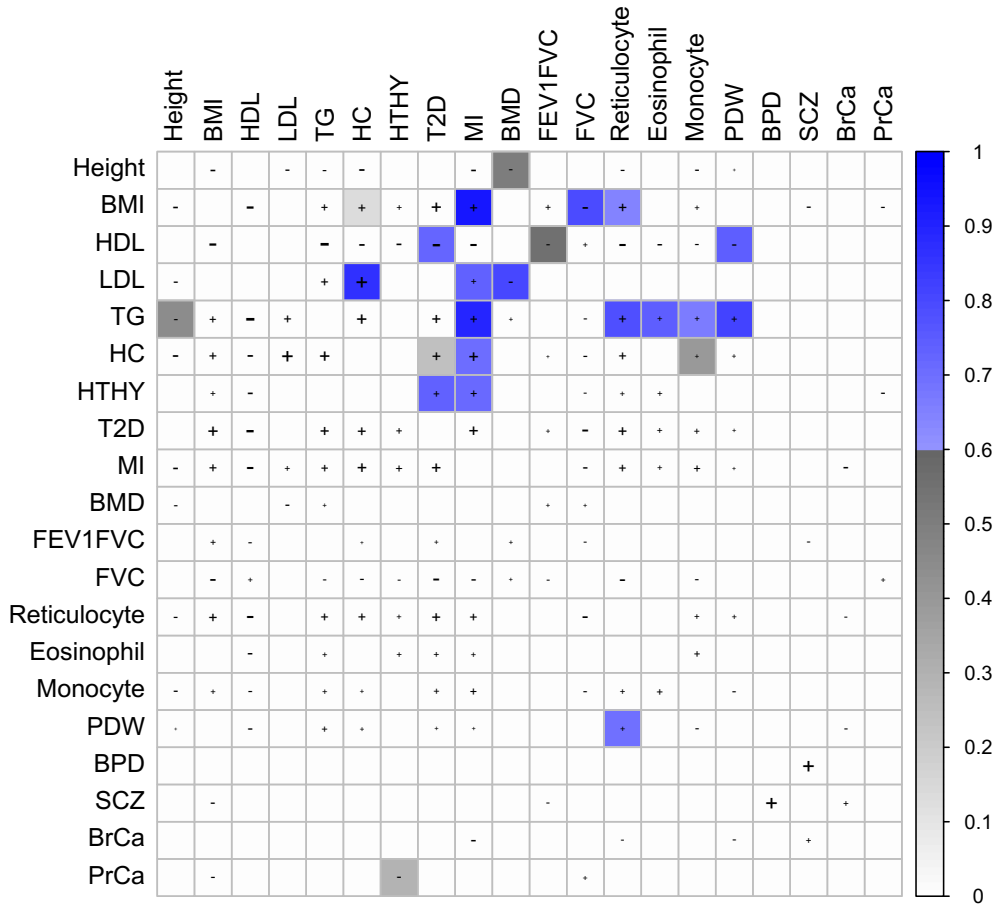


Figure 4: Partially or fully genetically causal relationships between selected complex traits. Shaded squares indicate significant evidence for a causal or partially causal effect of the row trait on the column trait (FDR <1%). Color scale indicates posterior mean $\widehat{gcp}$ for the effect of the row trait on the column trait; blue color indicates $\widehat{gcp} > 0.6$, grey color indicates $\widehat{gcp} < 0.6$. ``+'' or ``-'' signs indicate trait pairs with a nominally significant (positive or negative) genetic correlation (p<.05), and the size of the "+" or "-" size is proportional to the genetic correlation. Results for the 30 trait pairs with $\widehat{gcp}$>0.6 are reported in Table 1, and results for all 429 genetically correlated trait pairs are reported in Supplementary Table 2.1. HTHY: hypothyroidism. FG: fasting glucose. PDW: platelet distribution width. BPD: bipolar disorder. SCZ: schizophrenia. BrCa: breast cancer: PrCa: prostate cancer.

Table 1: Fully or partially genetically causal relationships between complex traits.

Trait 1	Trait 2	p_{LCV}	$\hat{\rho}_g$ (std err)	$\widehat{\text{gcp}}$ (std err)	MR ref
Triglycerides	Hypertension	1×10^{-38}	0.25 (0.04)	0.95 (0.04)	
BMI	Heart attack	5×10^{-9}	0.34 (0.09)	0.94 (0.11)	[32,38]
Triglycerides	Heart attack	2×10^{-31}	0.30 (0.06)	0.90 (0.08)	[4]
Triglycerides	BP – systolic	1×10^{-40}	0.13 (0.03)	0.89 (0.08)	
HDL	Hypertension	1×10^{-21}	-0.29 (0.06)	0.87 (0.09)	
LDL	Hi cholesterol	2×10^{-6}	0.77 (0.07)	0.86 (0.11)	
Triglycerides	Mean cell volume	2×10^{-18}	-0.20 (0.04)	0.86 (0.11)	
Triglycerides	BP – diastolic	9×10^{-39}	0.11 (0.04)	0.86 (0.10)	
Mean platelet volume	Platelet count	1×10^{-9}	-0.66 (0.03)	0.84 (0.10)	
BMI	Hypertension	3×10^{-16}	0.38 (0.03)	0.83 (0.11)	[11,38]
Triglycerides	Platelet distribution width	1×10^{-16}	0.19 (0.04)	0.81 (0.13)	
LDL	Bone mineral density – heel	7×10^{-34}	-0.12 (0.05)	0.80 (0.12)	
BMI	FVC	9×10^{-13}	-0.22 (0.03)	0.79 (0.17)	
Hi cholesterol	RBC count	0.002	0.08 (0.03)	0.79 (0.15)	
Triglycerides	Reticulocyte count	5×10^{-10}	0.33 (0.05)	0.79 (0.14)	
Type 2 Diabetes	Mean cell volume	0.004	-0.15 (0.03)	0.77 (0.20)	
HDL	RBC count	0.003	-0.13 (0.05)	0.76 (0.34)	
Triglycerides	Eosinophil count	6×10^{-17}	0.14 (0.05)	0.75 (0.16)	
Balding4	Number children – male	3×10^{-30}	-0.16 (0.05)	0.75 (0.13)	
HDL	Platelet distribution width	2×10^{-16}	-0.14 (0.04)	0.75 (0.16)	
RBC distribution width	Type 2 Diabetes	7×10^{-4}	0.11 (0.03)	0.73 (0.19)	
LDL	Heart attack	4×10^{-31}	0.17 (0.08)	0.73 (0.13)	[3,13]
Hi cholesterol	Lymphocyte count	0.004	0.18 (0.04)	0.73 (0.22)	
Platelet distribution width	Platelet count	2×10^{-7}	-0.47 (0.04)	0.73 (0.15)	
Hypothyroidism	Type 2 Diabetes	4×10^{-4}	0.22 (0.05)	0.73 (0.29)	
HDL	Type 2 Diabetes	5×10^{-7}	-0.40 (0.06)	0.72 (0.17)	
Heart attack	Breast cancer	0.01	-0.16 (0.05)	0.72 (0.24)	
Hypothyroidism	Heart attack	1×10^{-11}	0.26 (0.05)	0.72 (0.16)	
Hi cholesterol	Heart attack	5×10^{-4}	0.52 (0.12)	0.71 (0.19)	
HDL	BP – diastolic	9×10^{-17}	-0.12 (0.06)	0.70 (0.18)	

We report all significant trait pairs (1% FDR) with high gcp estimates ($\widehat{\text{gcp}}$>0.6). p_{LCV} is the p-value for the null hypothesis of no partial genetic causality; $\hat{\rho}_g$ is the estimated genetic correlation, with standard error; $\widehat{\text{gcp}}$ is the posterior mean estimated genetic causality proportion, with posterior standard error. We provide references for all MR studies supporting causal relationships between these traits that we are currently aware of. Results for all 429 genetically correlated trait pairs are reported in Supplementary Table 2.1.

Myocardial infarction (MI) had a nominally significant genetic correlation with 31 other traits, of which six had significant evidence (FDR <1%) for a fully or partially genetically causal effect on MI (Table 1); there was no evidence for a genetically causal effect of MI on any other trait. Consistent with previous studies, these traits included LDL[3,13], triglycerides[4] and BMI[30], but not HDL[3]. The effect of BMI was also consistent with prior MR studies [30-33], although these studies did not attempt to account for pleiotropic effects (also see ref. [34], which detected no effect). There was also evidence for a genetically causal effect of high cholesterol, which was unsurprising (due to the high genetic correlation with LDL) but noteworthy because of its strong genetic correlation with MI, compared with LDL and triglycerides. The result for HDL and MI did not pass our significance threshold (FDR <1%), but was nominally significant (p=0.02); we residualized HDL summary statistics on summary statistics for three established causal risk factors (LDL, BMI and triglycerides), determining that residualized HDL showed no evidence of genetic causality (p=0.8).

We also detected evidence for a fully or partially genetically causal effect of hypothyroidism on MI (Table 1). Although hypothyroidism is not as well-established a cardiovascular risk factor as high LDL, its genetic correlation with MI is comparable (Table 1), and this effect is mechanistically plausible[40,41]. While this result was robust in the conditional analysis, and there was no strong evidence for a genetically causal effect of hypothyroidism on lipid traits, it is possible that this effect is mediated by lipid traits. A recent MR study of thyroid hormone levels, at $\sim 20 \times$ lower sample size than the present study, provided evidence for a genetically causal effect on LDL but not CAD [42]. On the other hand, clinical trials have demonstrated that

treatment of subclinical hypothyroidism leads to improvement in several cardiovascular risk factors [43-47]. We also detected evidence for a fully or partially genetically causal effect of hypothyroidism on, consistent with a longitudinal association between subclinical hypothyroidism and diabetes incidence [48], as well as an effect of thyroid hormone withdrawal on glucose disposal in athyreotic patients [49].

We detected evidence for a (negative) genetically causal effect of LDL on bone mineral density (BMD; Table 1). A meta-analysis of randomized clinical trials reported that statin administration increases BMD[50]. Moreover, familial defective apolipoprotein B leads to high LDL and low BMD [51]. We performed two-sample MR using 8 SNPs that were previously curated (in ref. 3), finding modest evidence for a negative causal effect (p=0.04). Because these variants are not likely to have pleiotropic effects, this analysis provides separate evidence for a genetically causal effect.

Approximately half of significant trait pairs had low to medium gcp estimates (<0.6). Given that there is lower power to detect trait pairs with low gcp values, it is likely that partial genetic causality with gcp<0.6 is more common than full or nearly-full genetic causality with gcp>0.6. Trait pairs with low gcp estimates can suggest plausible biological hypotheses. For example, we identified a partially genetically causal effect of age at menarche (AAM) on height ($\widehat{\text{gcp}}$=0.43(0.10), Supplementary Table 2.1), suggesting that these traits are influenced by a shared hormonal pathway that is more strongly correlated with AAM than with height, as recently hypothesized [11].

A recent study reported genetic correlations between various complex traits and number of children in males and females[52]. We identified only one trait (balding in males) with a fully or partially causal effect on number of children (in males; Table 1). For college education, which has a strong negative genetic correlation with number of children ($\hat{\rho}_g = -0.31(0.07)$ and $-0.26(0.06)$ in males and females respectively), we obtained low gcp estimates with low standard errors ($\widehat{\mathrm{gcp}} = 0.00(0.09)$ and $\widehat{\mathrm{gcp}} = 0.04(0.21)$ respectively). Thus, a genetic correlation with number of children does not imply direct selection. This result does not contradict the conclusion of reference [52] that complex traits are affected by natural selection, as pleiotropic selection can also affect a trait [53].

Polygenic autism risk is positively genetically correlated with educational attainment[16] (and cognitive ability[54], a highly genetically correlated trait[57]), possibly consistent with the hypothesis that common autism risk variants persist in the population due to compensatory effects on cognitive ability [55,56]. If so, then most common variants affecting autism risk would also affect educational attainment, leading to a partially genetically causal effect of autism on educational attainment. However, we detected evidence against such an effect ($\widehat{\mathrm{gcp}} = 0.13(0.13), \hat{\rho}_g = 0.23(0.07)$).

In order to evaluate whether the limitations of MR observed in simulations (Figure 2) are also observed in analyses of real traits, we applied MR, MR-Egger and Bidirectional MR to all 429 genetically correlated trait pairs (Table 1). MR reported significant causal relationships (1%

FDR) for 271/429 trait pairs, including 155 pairs of traits for which each trait was reported to be causal for the other trait. This implausible result confirms that MR frequently produces false positives in the presence of a genetic correlation, as predicted by our simulations (Figure 2). In contrast, LCV reported a significant partially or fully genetically causal relationship for only 59 trait pairs and it never reports a causal effect in both directions. Similarly, Bidirectional MR reported a significant causal relationship for only 45 trait pairs (including 17 pairs of traits that overlapped with LCV).

Discussion

We have introduced a latent causal variable (LCV) model to identify causal relationships among genetically correlated pairs of complex traits. We applied LCV to 52 traits, finding that many trait pairs do exhibit partially or fully genetically causal relationships. Our method represents an advance for two main reasons. First, unlike existing MR methods, LCV reliably distinguishes between genetic correlation and full or partial genetic causation. Positive findings using LCV are more likely to reflect true causal effects. Second, we define and estimate the genetic causality proportion (gcp) to quantify the degree of causality. This parameter, which provides information orthogonal to the genetic correlation or the causal effect size, enables a non-dichotomous description of the causal architecture.

This study has two important limitations. First, the LCV model includes only a single intermediary and can be confounded in the presence of multiple intermediaries. However, the 30 trait pairs with $\widehat{\text{gcp}} > 0.6$ are unlikely to be false positives (see Supplementary Figure 2.2).

Second, because LCV models only two traits at a time, it cannot be used to identify conditional effects given observed confounders[4,60]. This approach was used, for example, to show that triglycerides affect coronary artery disease risk conditional on LDL[4]. However, it is less essential for LCV to model observed genetic confounders, since LCV explicitly models a latent genetic confounder.

Despite these limitations, for most pairs of complex traits we recommend using LCV instead of MR, as MR methods (including MR Egger) are easily confounded by genetic correlations. MR is more reliable when it is possible to identify variants that are likely to represent valid instruments. For example, an MR analysis identified a causal effect of vitamin D on multiple sclerosis, utilizing genetic variants near genes with well-characterized effects on vitamin D synthesis, metabolism and transport [66]. As another example, cis-eQTLs can be used as genetic instruments, as they are unlikely to be confounded by processes mediated in trans [62-64]; however, this approach has other limitations [63,65].

URLs

Open-source software implementing our method is available at https://github.com/lukejoconnor/LCV.

Methods

LCV model

The LCV random effects model assumes that the distribution of marginal effect sizes for the two traits can be written as the sum of two independent bivariate distributions (visualized in Figure 1c-e in orange and blue respectively): (1) a *shared genetic component* $(q_1\pi, q_2\pi)$ whose values are proportional for both traits; and (2) an *even genetic component* (γ_1, γ_2) whose density is mirror symmetric across both axes. Distribution (1) resembles a line through the origin, and we interpret its effects as being mediated by a latent causal variable (L) (Figure 1a); distribution (2) does not contribute to the genetic correlation, and we interpret its effects as direct effects. Informally, the LCV model assumes that any asymmetry in the shared genetic architecture arises from the action of a latent variable.

In detail, the LCV model assumes that there exist scalars q_1, q_2 and a distribution $(\pi, \gamma_1, \gamma_2)$ such that

$$(\alpha_1, \alpha_2) = (q_1\pi + q_2\pi) + (\gamma_1, \gamma_2), \quad (3)$$

$$\text{where } \pi \perp (\gamma_1, \gamma_2)$$

$$\text{and } (\gamma_1, \gamma_2) \sim (-\gamma_1, \gamma_2) \sim (\gamma_1, -\gamma_2).$$

Here α_k is the random marginal effect of a SNP of trait *k*, π is interpreted as the marginal effect of a SNP on L, and γ_k is interpreted as the non-mediated effect of a SNP on trait *k*. α and π (but not γ) are normalized to have unit variance, and all random variables have zero mean. (The symbol "$\sim$'' means "has the same distribution as.'') q_1, q_2 are the model parameters of primary interest, and we can relate them to the mixed fourth moments, which are observable (equation 2). In particular, this implies that the model is identifiable (except when the excess kurtosis $\kappa_\pi = 0$). We do not expect that κ_π will be exactly zero for any real trait, but there will be lower

power for traits with higher polygenicity. Note that we have avoided assuming any particular parametric distribution.

The LCV model assumptions are strictly weaker than the assumptions made by MR. Like LCV, a formulation of the MR assumptions is that the bivariate distribution of SNP effect sizes can be expressed in terms of two distributions. In particular, it assumes that the effect size distribution is a mixture of (1') a distribution whose values are proportional for both traits (representing all SNPs that affect the exposure Y_1) and (2') a distribution with zero values for the exposure Y1 (representing SNPs that only affect the outcome Y_2). These two distributions can be compared with distributions (1) and (2) above. Because (1') is identical to (1) and (2') is a special case of (2), the LCV model assumptions are strictly weaker than the MR assumptions (indeed, much weaker). We also note that the MR model is commonly illustrated with a non-genetic confounder affecting both traits. Our latent variable *L* is a genetic variable, and it is not analogous to the non-genetic confounder. Similar to MR, LCV is unaffected by nongenetic confounders (such a confounder may result in a phenotypic correlation that is unequal to the genetic correlation).

The genetic causality proportion (gcp) is defined as:

$$\mathrm{gcp} = \frac{\log|q_2| - \log|q_1|}{\log|q_2| + \log|q_1|} \tag{4}$$

which satisfies equation (1). gcp is positive when trait 1 is partially genetically causal for trait 2. When $\mathrm{gcp} = 1$, trait 1 is fully genetically causal for trait 2: $q_1 = 1$ and the causal effect size is $q_2 = \rho_g$ (Figure 1b,e). The LCV model is broadly related to dimension reduction techniques

such as Factor Analysis[67] and Independent Components Analysis[68], although it differs in its modeling assumptions as well as its goal (causal inference); our inference strategy (mixed fourth moments) also differs.

Under the LCV model assumptions, we derive the estimation equation (2) as follows:

$$E(\alpha_1^3\alpha_2) = E((\gamma_1 + q_1\pi)^3(\gamma_2 + q_2\pi))$$
$$= q_1^3 q_2 E(\pi^4) + 3q_1 q_2 E(\pi^2\gamma_1^2)$$
$$= q_1^3 q_2 E(\pi^4) + 3q_1 q_2 E(\pi^2)E(\gamma_1^2)$$
$$= q_1^3 q_2 E(\pi^4) + 3q_1 q_2 (1)(1 - q_1^2)$$
$$= q_1^3 q_2 (E(\pi^4) - 3) + 3q_1 q_2.$$

In the second line, we used the independence assumption to discard cross-terms of the form $\gamma_p\pi^3$ and $\gamma_1^3\pi$, and we used the symmetry assumption to discard terms of the form $\gamma_1\gamma_2^3$. In the third and fourth lines, we used the independence assumption, which implies that $E(\gamma_1^2\pi^2) = E(\gamma_1^2)E(\pi^2) = E(\gamma_1^2) = 1 - q_1^2$. The factor $E(\pi^4) - 3$ is the excess kurtosis of π, which is zero when π follows a Gaussian distribution; in order for the estimation equation to be useful, $E(\pi^4) - 3$ must be nonzero.

Estimation under the LCV model

In order to estimate the gcp and to test for partial causality, we utilize six steps. First, we use LD score regression[19] to estimate the heritability of each trait; these estimates are used to normalize the summary statistics. Second, we apply cross-trait LD score regression[16] to estimate the genetic correlation; the intercept in this regression is also used to correct for

possible sample overlap when estimating the mixed fourth moments. Third, we estimate the mixed fourth moments of the bivariate effect size distribution. Fourth, we compute test statistics for each possible value of the gcp, based on the estimated genetic correlation and on the estimated mixed fourth moments. Fifth, we jackknife on these test statistics to estimate their standard errors, similar to ref. 19, obtaining a likelihood function for the gcp. Sixth, we obtain posterior means and standard errors for the gcp using this likelihood function and a uniform prior distribution. These steps are detailed below.

First, we apply LD score regression to normalize the test statistics. Under the LCV model, the marginal effect sizes for each trait, α_1 and α_2, have unit variance. We use a slightly modified version of LD score regression[19] with LD scores computed from UK10K data [58]. In particular, we run LD score regression using a slightly different weighting scheme, matching the weighting scheme in our mixed fourth moment estimators; the weight of SNP *i* was:

$$w_i = 1/\max\left(1, \ell_{\mathrm{i}}^{\mathrm{HapMap}}\right), \quad (5)$$

where ℓ_i^{HapMap} was the estimated LD score between SNP *i* and other HapMap3 SNPs (this is approximately the set of SNPs that were used in the regression). This weighting scheme is motivated by the fact that SNPs with high LD to other regression SNPs will be over-counted in the regression (see ref. 19).

Similar to ref. 16, we improve power by excluding large-effect variants when computing the LD score intercept; for this study, we chose to exclude variants with χ^2 statistic $30 \times$ the mean, exploiting the fact that genome-wide significant SNPs are not due to population stratification

(these variants are not excluded in subsequent steps). Then, we divide the summary statistics by $s =$, where χ^2_{avg} is the weighted mean χ^2 statistic and χ^2_0 is the LD score intercept, obtaining estimates $\hat{\alpha}$ of α. (We also divide the LD score intercept by s^2.) We assess the significance of the heritability by performing a block jackknife on *s*, defining the significance Z_h as *s* divided by its estimated standard error.

Second, to estimate the genetic correlation, we apply cross-trait LD score regression [16]. Similar to above, we use a slightly modified weighting scheme (equation 5), and we exclude large-effect variants when computing the cross-trait LD score intercept. We assess the significance of the genetic correlation using a block jackknife.

Third, we estimate the mixed fourth moments $E(\alpha_1\alpha_2^3)$ using the following equation:

$$E(\hat{\alpha}_1\hat{\alpha}_2^3|\alpha_1,\alpha_2) = \alpha_1\alpha_2^3 + 3\alpha_1\alpha_2\sigma_2^2 + \alpha_2^2\sigma_{12} + 3\sigma_{12}\sigma_2^2, \quad (6)$$

where σ_2^2 is the LD score regression intercept for trait 2 (normalized by dividing by s_2^2) and σ_{12} is the cross-trait LD score regression intercept (normalized by dividing by $s_1 s_2$). For simulations with no LD, we fix $\sigma_2^2 = 1/s_2^2$ and $\sigma_{12} = 0$. Thus, we obtain an estimate $\hat{\kappa}_k$ of $\kappa_k = E(\alpha_k^2\alpha_1\alpha_2) - 3\rho_g$ by computing the weighted average of $\hat{\alpha}_1\hat{\alpha}_2^3$ over SNPs (with weights given by equation 5), and subtracting $3\hat{\alpha}_1\hat{\alpha}_2\sigma_2^2 + \alpha_2^2\sigma_{12} + 3\sigma_{12}\sigma_2^2$.

Fourth, we define a collection of statistics $S(x)$ for $x \in X = \{-1, -.01, -.02, \ldots, 1\}$ (corresponding to possible values of gcp):

$$S(x) = \frac{1}{\max\left(\frac{1}{|\rho_g|}, \sqrt{A_1(x)^2 + A_2(x)^2}\right)}$$

$$\text{where } \mathrm{A_k(x)} = \left|\rho_\mathrm{g}\right|^{-x}\hat{\kappa}_k. \tag{7}$$

The motivation for utilizing the normalization by $\sqrt{A_1(x)^2 + A_2(x)^2}$ is that the magnitude of $A_1(x)$ and $A_2(x)$ tend to be highly correlated, leading to greatly increased standard errors if we only use the numerator of S. However, the denominator tends to zero when the genetic correlation is zero, leading to instability in the test statistic and false positives. The use of the threshold leads to conservative, rather than inflated, standard errors when the genetic correlation is zero or nearly zero. We recommend only analyzing trait pairs with a significant genetic correlation, and this threshold usually has no effect on the results. Another reason not to analyze trait pairs whose genetic correlation is non-significant is that for positive LCV results, the genetic correlation provides critical information about the causal effect size and direction.

Fifth, we estimate the variance of $S(x)$ using a block jackknife with $k = 100$ blocks of contiguous SNPs, resulting in minimal non-independence between blocks. Blocks are chosen to include the same number of SNPs, and the jackknife variance is:

$$\sigma^2_{S(x)} = 101\sum_{j=1}^{100}(S_j(x) - S_{\mathrm{avg}}(x))^2 \tag{8}$$

where $S_j(x)$ is the test statistic computed on blocks $1, \ldots, j-1, j+1, \ldots 100$ and $S_{\mathrm{avg}}(x)$ is the mean of the jackknife estimates. We compute an approximate likelihood, $L(S|\mathrm{gcp} = x)$, by

assuming (1) that $L(S|\text{gcp} = x) = L(S(x)|\text{gcp} = x)$ and (2) that if $\text{gcp} = x$ then $S(x)/\sigma_{S(x)}$ follows a T distribution with 98 degrees of freedom.

Sixth, we impose a uniform prior on the gcp, enabling us to obtain a posterior mean estimate:

$$\widehat{\text{gcp}} = \frac{1}{|X|}\sum_{x \in X} xL(x) \tag{9}$$

The estimated standard error is:

$$\text{se} = \frac{1}{|X|}\sum_{x \in X} (S_j(x) - S_{\text{avg}}(x))^2 \tag{9}$$

In order to compute p-values, we apply a T-test to the statistic $S(0)$.

Outlier removal

In a secondary analysis, we applied an outlier removal procedure to determine whether our results on real traits using LCV were unduly influenced by individual loci. We computed the LCV test statistic $S(0)$ for each of the 100 jackknife blocks, discarded jackknife blocks that were >20 standard deviations from the mean, and re-ran the procedure iteratively until no outliers remained. For most trait pairs, this process results in the removal of 0 blocks; for a handful of trait pairs, it results in the removal of one or a few.

We do not recommend the broad use of this procedure, because outlier loci may contain valuable information. In particular, if any SNP affects trait 1 without affecting trait 2 proportionally, this suggests that trait 1 is not causal for trait 2. An alternative explanation is

that its effect on trait 2 is masked by an opposing pleiotropic effect, either of the same causal SNP or of a different causal SNP at the same locus. If an outlier locus is to be removed, we recommend manually examining it and determining whether its removal can be justified or whether it provides competing statistical evidence against a causal effect.

References

1. Davey Smith, George, and Shah Ebrahim. “Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease?" International journal of epidemiology 32.1 (2003): 1-22.
2. Davey Smith, George, and Gibran Hemani. “Mendelian randomization: genetic anchors for causal inference in epidemiological studies." Human molecular genetics 23.R1 (2014): R89-R98.
3. Voight, Benjamin F., et al. “Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study." The Lancet 380.9841 (2012): 572-580.
4. Do, Ron, et al. “Common variants associated with plasma triglycerides and risk for coronary artery disease." Nature genetics 45.11 (2013): 1345-1352.
5. Burgess, Stephen, Adam Butterworth, and Simon G. Thompson. “Mendelian randomization analysis with multiple genetic variants using summarized data." Genetic epidemiology 37.7 (2013): 658-635.
6. Kang, Hyunseung, et al. “Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization." Journal of the American Statistical Association 111.513 (2016): 132-144.
7. Bowden, Jack, George Davey Smith, and Stephen Burgess. “Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression." International journal of epidemiology 44.2 (2015): 512-525.

8. Bowden, Jack, et al. “Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator." Genetic epidemiology 40.4 (2016): 304-314.
9. Hemani, Gibran, et al. “MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations." BioRxiv (2016): 078972.
10. Hartwig, Fernando Pires, George Davey Smith, and Jack Bowden. “Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption." International journal of epidemiology 46.6 (2017): 1985-1998.
11. Pickrell, Joseph K., et al. “Detection and interpretation of shared genetic influences on 42 human traits." Nature genetics 48.7 (2016): 709.
12. Verbanck, Marie, et al. “Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases.'' Nature genetics 50.5 (2018): 693.
13. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. “Sequence variations in PCSK9, low LDL, and protection against coronary heart disease." New England Journal of Medicine 354 (2006): 1264-72.
14. Paaby, Annalise B., and Matthew V. Rockman. “The many faces of pleiotropy." Trends in Genetics 29.2 (2013): 63-73.
15. VanderWeele, Tyler J., et al. “Methodological challenges in Mendelian randomization." Epidemiology 25.3 (2014): 427.
16. Bulik-Sullivan, Brendan, et al. “An atlas of genetic correlations across human diseases and traits." Nature genetics 47.11 (2015): 1236-1241.

17. Welsh, Paul, et al. “Unraveling the directional link between adiposity and inflammation: a bidirectional Mendelian randomization approach." The Journal of Clinical Endocrinology \& Metabolism 95.1 (2010): 93-99.
18. Vimaleswaran, Karani S., et al. “Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts." PLoS Med 10.2 (2013): e1001383.
19. Bulik-Sullivan, Brendan K., et al. “LD Score regression distinguishes confounding from polygenicity in genome-wide association studies." Nature genetics 47.3 (2015): 291-295.
20. Yang, Jian, et al. “Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index." Nature genetics 47.10 (2015): 1114.
21. Kolesar, Michal, et al. “Identification and inference with many invalid instruments." Journal of Business \& Economic Statistics 33.4 (2015): 474-484.
22. Burgess, Stephen, and Simon G. Thompson. “Interpreting findings from Mendelian randomization using the MR-Egger method." European Journal of Epidemiology (2017): 1-13.
23. Conneely, Karen N., and Michael Boehnke. “So many correlated tests, so little time! Rapid adjustment of P values for multiple correlated tests." The American Journal of Human Genetics 81.6 (2007): 1158-1168.
24. Galinsky, Kevin J., et al. “Population structure of UK Biobank and ancient Eurasians reveals adaptation at genes influencing blood pressure." The American Journal of Human Genetics 99.5 (2016): 1130-1139.

25. Bhatia, Gaurav, et al. “Correcting subtle stratification in summary association statistics." bioRxiv (2016): 076133.
26. Goddard, Michael E., et al. “Estimating effects and making predictions from genome-wide marker data." Statistical Science 24.4 (2009): 517-529.
27. Sudlow, Cathie, et al. “UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age." PLoS medicine 12.3 (2015): e1001779.
28. Bycroft, Clare, et al. “Genome-wide genetic data on~ 500,000 UK Biobank participants." bioRxiv (2017): 163298.
29. Loh, Po-Ru, et al. “Mixed model association for biobank-scale data sets." bioRxiv (2017): 194944.
30. Holmes, Michael V., Mika Ala-Korpela, and George Davey Smith. “Mendelian randomization in cardiometabolic disease: challenges in evaluating causality." Nature Reviews Cardiology (2017): 577-590.
31. Smith, George Davey, et al. “The association between BMI and mortality using offspring BMI as an indicator of own BMI: large intergenerational mortality study." Bmj 339 (2009): b5043.
32. Nordestgaard, Brge G., et al. “The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a Mendelian randomisation approach." PLoS Med 9.5 (2012): e1001212.
33. Hgg, Sara, et al. “Adiposity as a cause of cardiovascular disease: a Mendelian randomization study." International journal of epidemiology 44.2 (2015): 578-586.

34. Holmes, Michael V., et al. “Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis." The American Journal of Human Genetics 94.2 (2014): 198-208.
35. Cole, Stephen R., et al. “Illustrating bias due to conditioning on a collider." International journal of epidemiology 39.2 (2009): 417-420.
36. Aschard, Hugues, et al. “Adjusting for heritable covariates can bias effect estimates in genome-wide association studies." The American Journal of Human Genetics 96.2 (2015): 329-339.
37. Ross, Stephanie, et al. “Mendelian randomization analysis supports the causal role of dysglycaemia and diabetes in the risk of coronary artery disease." European heart journal 36.23 (2015): 1454-1462.
38. Lyall, Donald M., et al. “Association of body mass index with cardiometabolic disease in the UK Biobank: a Mendelian randomization study." JAMA cardiology 2.8 (2017): 882-889.
39. Schunkert, Heribert, et al. “Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease." Nature genetics 43.4 (2011): 333-338.
40. Klein, Irwin, and Kaie Ojamaa. “Thyroid hormone and the cardiovascular system." New England Journal of Medicine 344.7 (2001): 501-509.
41. Grais, Ira Martin, and James R. Sowers. “Thyroid and the heart." The American journal of medicine 127.8 (2014): 691-698.
42. Zhao, Jie V., and C. Mary Schooling. “Thyroid function and ischemic heart disease: a Mendelian randomization study." Scientific reports 7:8515 (2017): 8515.

43. Monzani, F. et al. “Effect of levothyroxine on cardiac function and structure in subclinical hypothyroidism: a double blind, placebo-controlled study." J. Clin. Endocrinol. Metab. 86 (2001): 1110-1115.
44. Meier, C. et al. “TSH-controlled L-thyroxine therapy reduces cholesterol levels and clinical symptoms in subclinical hypothyroidism: a double blind, placebo-controlled trial (Basel Thyroid Study)." J. Clin. Endocrinol. Metab. 86 (2001): 4430-4863.
45. Monzani, F. et al. “Effect of levothyroxine replacement on lipid profile and intima-media thickness in subclinical hypothyroidism: a double-blind, placebo- controlled study." J. Clin. Endocrinol. Metab. 89 (2004): 2099-2106.
46. Razvi, S. et al. “The beneficial effect of L-thyroxine on cardiovascular risk factors, endothelial function, and quality of life in subclinical hypothyroidism: randomized, crossover trial." J. Clin. Endocrinol. Metab. 92 (2007): 1715-1723.
47. Nagasaki, T. et al. “Decrease of brachial-ankle pulse wave velocity in female subclinical hypothyroid patients during normalization of thyroid function: a double-blind, placebo-controlled study." Eur. J. Endocrinol. 160 (2009): 409-415.
48. Chaker, Layal, et al. “Thyroid function and risk of type 2 diabetes: a population-based prospective cohort study." BMC medicine 14.1 (2016): 150.
49. Brenta, Gabriela, et al. “Acute thyroid hormone withdrawal in athyreotic patients results in a state of insulin resistance." Thyroid 19.6 (2009): 665-669.
50. Wang, Zongze, et al. “Effects of Statins on Bone Mineral Density and Fracture Risk: A PRISMA-compliant Systematic Review and Meta-Analysis." Medicine 95.22 (2016): e3042.

51. Yerges, Laura M., et al. “Decreased bone mineral density in subjects carrying familial defective apolipoprotein B-100." The Journal of Clinical Endocrinology \& Metabolism 98.12 (2013): E1999-E2005.
52. Sanjak, Jaleal S., et al. “Evidence of directional and stabilizing selection in contemporary humans." Proceedings of the National Academy of Sciences (2017): 201707227.
53. Price, George R. “Selection and covariance." Nature 227 (1970): 520-521.
54. Clarke, T. K., et al. “Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population." Molecular psychiatry 21.3 (2016): 419-425.
55. Keller, Matthew C., and Geoffrey Miller. “Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best?" Behavioral and Brain Sciences 29.4 (2006): 385-404.
56. Mullins, Niamh, et al. “Reproductive fitness and genetic risk of psychiatric disorders in the general population." Nature communications 8 (2017): 15833.
57. Davies, Gail, et al. “Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112,151)." Molecular psychiatry 21.6 (2016): 758.
58. UK10K Consortium. “The UK10K project identifies rare variants in health and disease." Nature 526.7571 (2015): 82.
59. Ware, Jennifer J., et al. "Genome-wide meta-analysis of cotinine levels in cigarette smokers identifies locus at 4q13. 2." Scientific reports 6 (2016): 20092.

60. Burgess, Stephen, et al. “Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways." International journal of epidemiology 44.2 (2014): 484-495.
61. Schoech, Armin, et al. “Quantification of frequency-dependent genetic architectures and action of negative selection in 25 UK Biobank traits." bioRxiv (2017): 188086.
62. Gamazon, Eric R., et al. “A gene-based association method for mapping traits using reference transcriptome data." Nature genetics 47.9 (2015): 1091-1098.
63. Gusev, Alexander, et al. “Integrative approaches for large-scale transcriptome-wide association studies." Nature genetics 48 (2016): 245-252.
64. Zhu, Zhihong, et al. “Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets." Nature genetics 48 (2016):481:487.
65. The GTEx consortium, et al. “Genetic effects on gene expression across human tissues." Nature 550.7675 (2017): 204.
66. Mokry, Lauren E., et al. “Vitamin D and risk of multiple sclerosis: a Mendelian randomization study." PLoS medicine 12.8 (2015): e1001866.
67. Child, Dennis. “The essentials of factor analysis." A&C Black (2006).
68. Comon, Pierre. “Independent component analysis, a new concept?" Signal processing 36.3 (1994): 287-314.

III. Extreme Polygenicity of Complex Traits is Explained by Negative Selection

This chapter is under submission at AJHG and published on bioRxiv as O'Connor, Luke J., Armin P. Schoech, Farhad Hormozdiari, Steven Gazal, Nick Patterson and Alkes L. Price, "Polygenicity of complex traits is explained by negative selection" (2018). L.J.O., A.P.S., N.P. and A.L.P. formulated the hypothesis. L.J.O., N.P. and A.L.P. developed the methods. L.J.O., F.H. and S.G. performed the analyses. L.J.O. and A.L.P. analyzed the results and wrote the manuscript. We also thank Dr. Benjamin Neale, Dr. Hilary Finucane, Dr. Omer Weissbrod, Dr. Yakir Reshef, Dr. Xuanyao Liu and Dr. Yuval Simons for helpful comments.

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 new 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 $\sim$4x 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 (GWAS) 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 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 5a). 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 5b). However, 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 5b). 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 ref. [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 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 5a). 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 5a). In order to evaluate these predictions, we introduce a new 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.

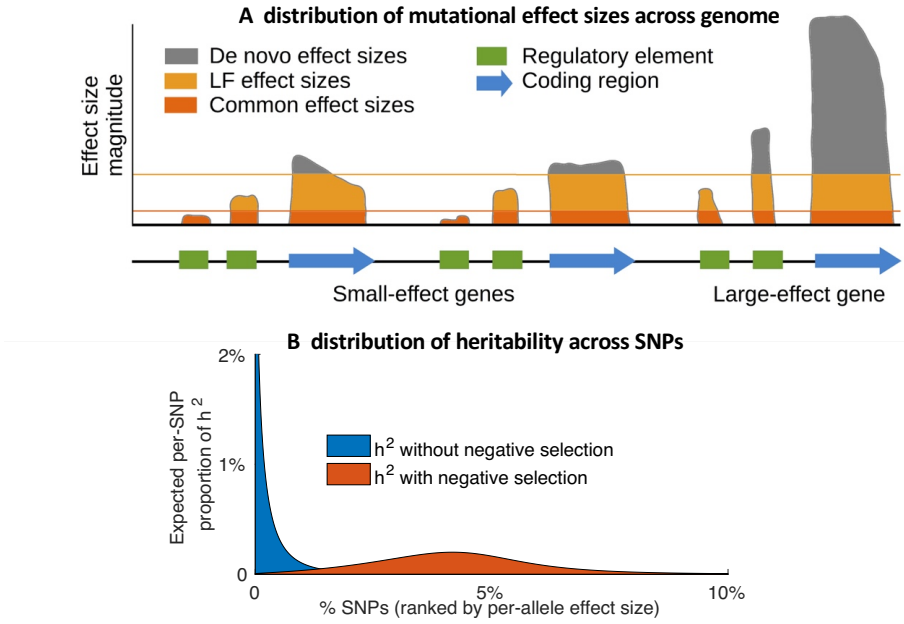


Figure 5. Illustration of flattening due to negative selection. (a) We illustrate the range of possible per-allele effect sizes for a SNP at each site for a toy example of three genes and nearby regulatory regions. Here, the distribution of de novo effects is not highly polygenic; it is dominated by 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.

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 new 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. ref. [30]), then M_e counts them as a single associated SNP, hence our use of the term "independently associated" (see Appendix II, "The effective number of independently associated SNPs").

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:

$$M_e = \frac{3M}{\kappa}, \qquad \kappa = \frac{E[\beta^4]}{E[\beta^2]^2}, \tag{1}$$

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 II, "Properties of M_e").

We define effect sizes β are 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 $\mathrm{E}_{h^2}(\alpha^2)$ (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 $\mathrm{E}_{h^2}(\alpha^2)$. This quantity can be visualized as the area under the curve in Figure 6a. For example, if 4 causal SNPs each contribute $\frac{1}{6}$ of heritability ($\frac{2}{3}$ 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 $\frac{1}{6} \times \frac{2}{3} = \frac{1}{9}$, and $M_e \approx 9$ (Figure 6a bottom). More generally, adding a large-effect SNP increases $\mathrm{E}_{h^2}(\alpha^2)$ and decreases M_e (Figure 6b). 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).

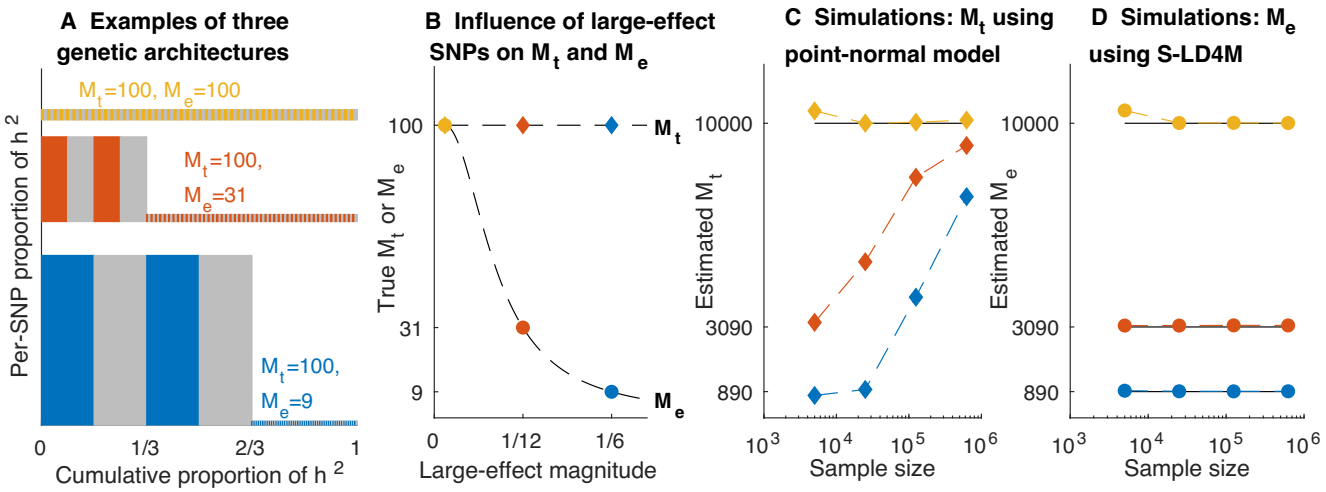


Figure 6. Comparison of the effective number of independently associated SNPs (M_e) with the total number of SNPs with nonzero effects (M_t). (a),(b) Examples of three genetic architectures with $M_t = 100$. (a) Each colored or gray block corresponds to one SNP; both height and width are proportional 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 4 large-effect SNPs. (c),(d) Simulations of the same three genetic architectures with the number of SNPs (and causal SNPs) scaled up by 100x. (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.

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 ($MAF \geq 0.05$) and low-frequency ($0.005 \leq MAF < 0.05$) 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:

$$\mathrm{E}\left(\alpha^4 | \ell^{(2)}, \ell^{(4)}\right) \approx 3\left(\ell^{(2)}\tau\right)^2 + \ell^{(4)}K, \quad (2)$$

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:

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

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, $\ell^{(2)}$, $\ell^{(4)}$, τ 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 II ("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 only report results for well-powered trait-annotation pairs, defined as having heritability Z-score greater than 2 and M_e Z-score greater than ½. Simulations suggested that log-scale estimates and estimated standard errors were approximately unbiased at these thresholds (see below). We only report meta-analyzed results 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 II ("Stratified LD fourth moments regression"). We have released open-source software implementing S-LD4M (see URLs).

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, $\widehat{\boldsymbol{\alpha}}$, as a function of the true causal effect size vector $\boldsymbol{\beta}$, was:

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

The estimated LD matrix $\hat{R}$ was equal to I in simulations with no LD. In simulations with LD, $\hat{R}$ was computed from UK Biobank data (N=460k). Sample correlations were computed for all typed and imputed SNPs (M=1.0M) within 0.1cM 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 necessary for $\hat{R}$ to be stored in memory, due to the large number of SNPs. To ensure that $\hat{R}$ 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 6a), σ_1^2 was equal to 1/10,000,

1/1,200, and 1/600 respectively. σ_2^2 was equal to $\frac{1-400\sigma_1^2}{9600}$, 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 $N = 20\text{k}$ and at $h^2 = 0.2$ and $N = 100\text{k}$.) 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 5 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 ref. [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., 10x larger effect sizes and 100x 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 $\beta^2_{SNP \to gene} s_{gene}$. 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]:

$$f(p|s_{SNP}) = \frac{1}{Z}\, p^{4N\mu-1}(1-p^{4N\mu-1}) \exp(-4N s_{SNP} p)\,, \qquad (5)$$

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 100x larger N and 100x 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 $f(p|s_{SNP})$ and approximating the integral using a sum over $\beta^2_{SNP \to gene} = \frac{1}{5000}, \frac{2}{5000}, \ldots, 8$. 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^2_{gene}.

Distinct from the 20,000 simulated genes, we analytically computed the heritability explained by a gene as a function of β^2_{gene}. In the direct selection model, s_{gene} was equal to β^2_{gene}; 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^2_{SNP} 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 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 $M = 50\text{k}$ 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 6c and Supplementary Table 3.1), consistent with recent work[13]. In general, any method that estimates M_t will only detect 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 6d). 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 ($M \approx 1.0\text{M}$ SNPs). We used $N = 50\text{k}$ 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 Methods section.

First, we assessed the ability of S-LD4M to estimate polygenicity for the set of all common and low-frequency SNPs. 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 7a and Supplementary Table 3.1). 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.

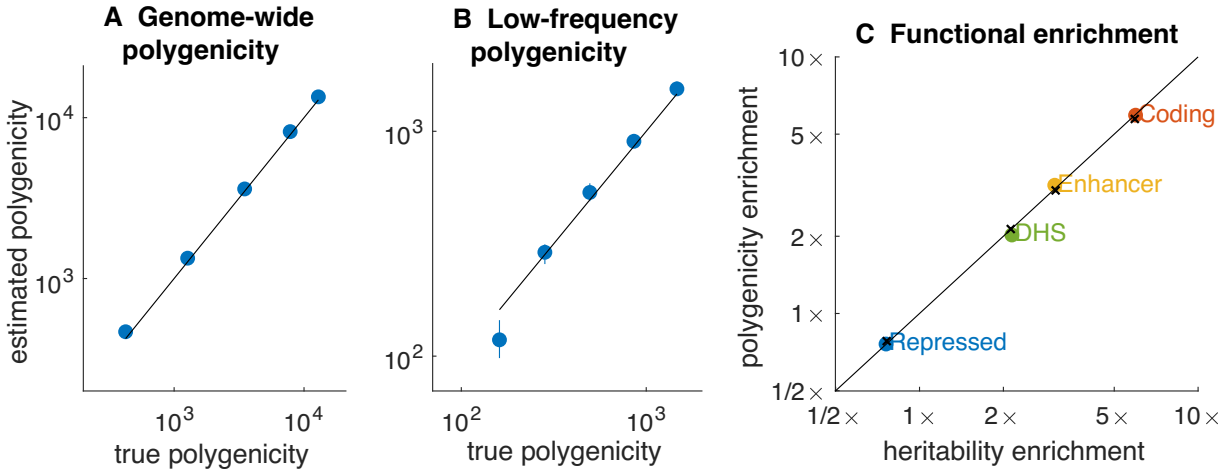


Figure 7. Accuracy of S-LD4M estimates in simulations with LD. (a) Estimates of M_e for all SNPs ($MAF = 0.5 - 50\%$). (b) Estimates of M_e for low-frequency SNPs ($MAF = 0.5 - 5\%$); common-SNP M_e is fixed at ~1,000 in these simulations. (c) Estimates of polygenicity enrichment and heritability enrichment in simulations with four functional categories. Black lines denotes y=x, and colored points denote estimates. In panel (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 Supplementary Table 3.1.

Second, we fixed the polygenicity of common SNPs ($\mathrm{MAF} > 5\%$) at $M_e \approx 1{,}000$ and progressively reduced the polygenicity of low-frequency SNPs ($\mathrm{MAF} = 0.5 - 5\%$), decreasing low-frequency heritability proportionally to fix the average effect size magnitude of causal low-frequency SNPs. Similar to Figure 7a, median estimates of low-frequency M_e were approximately unbiased for most values of M_e (Figure 7b). 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 Methods). We restrict or 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 ($\sim$8x enriched), enhancer ($\sim$5x), DNase I hypersensitivity sites (DHS, $\sim$2x), and repressed ($\sim$0.75x), similar to our results on real traits (see below). We determined that S-LD4M produces approximately unbiased estimates of polygenicity enrichment (Figure 7c). 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 (Supplementary Figure 3.1). 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, 1000 SNPs; roughly 3kb, 30kb, 300kb 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 (Supplementary Figure 3.2).

Fifth, we performed simulations at different sample sizes and included a filtering step to exclude trait-annotation pairs with inadequate power (see 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 (Supplementary Table 3.2), 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 (Supplementary Figure 3.3 and Supplementary Table 3.3). Similar results were obtained at *N*=250k (Supplementary Figure 3.3, Supplementary Table 3.2 and Supplementary Table 3.3). At *N*=10k, genome-wide polygenicity estimates remained unbiased (Supplementary Table 3.2), 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 (Supplementary Figure 3.3 and Supplementary Table 3.3). 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 (Supplementary Table 3.2) and polygenicity enrichment (Supplementary Table 3.3) were approximately well-calibrated or conservative. In particular, standard errors were moderately conservative when power was low to moderate (e.g. at $N = 10\mathrm{k}$ for genome-wide polygenicity), and well calibrated when power was high (e.g. at $N = 50\mathrm{k}$ 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 $N = 361\mathrm{k}$; Supplementary Table 3.4), including 29 UK Biobank traits[38-40] (see URLs) 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, $\mathrm{MAF} > 0.5\%$) and 75 baseline-LD model annotations[19]. As in previous work[5,19,31], 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 Methods section.

Table 2. Estimates of polygenicity for common and low-frequency SNPs across 33 complex traits.

Trait	$\log_{10}\mathrm{M_e}$ common	$\log_{10}\mathrm{M_e}$ 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 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. 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.

For most traits, M_e for common SNPs ($\mathrm{MAF} > 5\%$) ranged between 500 and 20,000 ($10^{2.7}$ and $10^{4.3}$; Table 2). 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. $\log_{10} M_{eff} = 4.69$). 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 (Supplementary Table 3.5). Estimates using 10 MAF bins were concordant with estimates using the baseline-LD model (log-scale $r = 0.95$; mean $\log_{10}$ fold difference -0.03 (s.d.=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 (s.d.=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. were based on a point-normal mixture model, and the estimates of Zhang et al. 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; Supplementary Figure 3.4a-b and Supplementary Table 3.6). The two sets of M_t estimates were only weakly correlated with each other (log-scale $r = 0.20$; Supplementary Figure 3.4c and Supplementary Table 3.6), 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. were $\sim 4\times$ larger than both our M_e estimates and the M_t estimates of Zhang et al., and the estimates of Zhang et al. 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.; 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 (s.d.=0.38)) (Supplementary Table 3.6). Finally, we compared the observed vs. expected distribution of χ^2 statistics under a point-normal model (Supplementary Figure 3.5 and Supplementary Table 3.7), 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 (Supplementary Table 3.8). (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 II, "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 $r^2 = 0.61$; Supplementary Figure 3.6a). It also provided a conservative upper bound on the number of genome-wide significant SNPs (Supplementary Figure 3.6b). Similar results were obtained when we used M_e estimates from an independent cohort (Supplementary Figure 3.7).

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 8a and Table 2). Polygenicity was 3.9x $(95\%\text{CI: } 2.9 - 5.2\text{x})$ smaller for low-frequency SNPs than for 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.

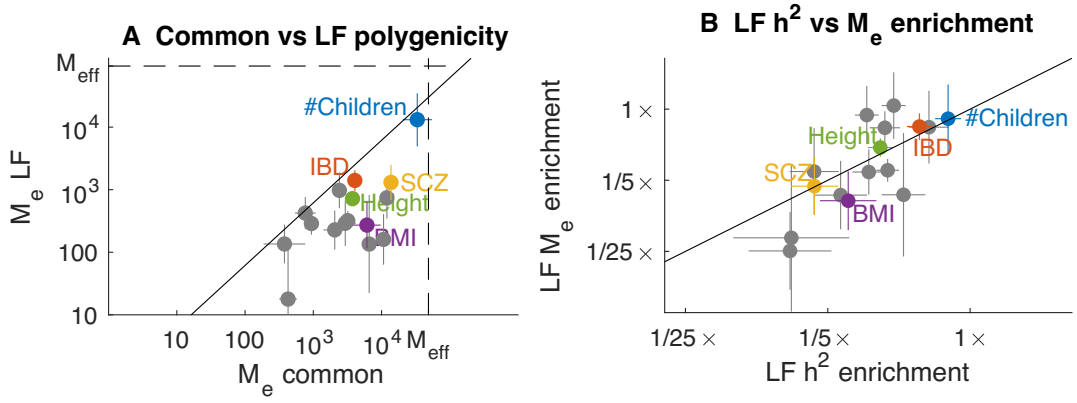


Figure 8. Comparison of common and low-frequency polygenicity across 15 complex traits. (a) Estimates of M_e for common and low-frequency SNPs. Estimates are meta-analyzed across well-powered traits. Common-variant polygenicity was ~4x greater on average than low-frequency polygenicity. Dotted lines denote the effective number of independent SNPs (M_{eff}) 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 Table 2.

The $\sim 4\times$ difference in common vs. low-frequency polygenicity is the largest that we would expect to observe, as the per-SNP *heritability* of common SNPs is also 3.9x $(95\%\text{CI: } 3.5 - 4.4\text{x})$ larger than that of low-frequency SNPs (consistent with previous estimates[12,21]). This concordance was consistent across traits (Figure 8b). The ratio h^2/M_e (i.e. the average unit of heritability; see 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 ref.[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 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 9, Supplementary Table 3.9 and Supplementary Table 3.10; 49% of trait-annotation pairs were well powered). For most annotations, polygenicity enrichment was approximately equal to heritability enrichment ($\text{regression slope} = 0.93; \text{r}^2 = 0.88$). For example, SNPs in conserved regions were 13x enriched for heritability and 14x enriched for polygenicity, and coding SNPs were 9.4x enriched for heritability and 6.6x 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 5a) 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 important functional regions are constrained by negative selection (Figure 5).

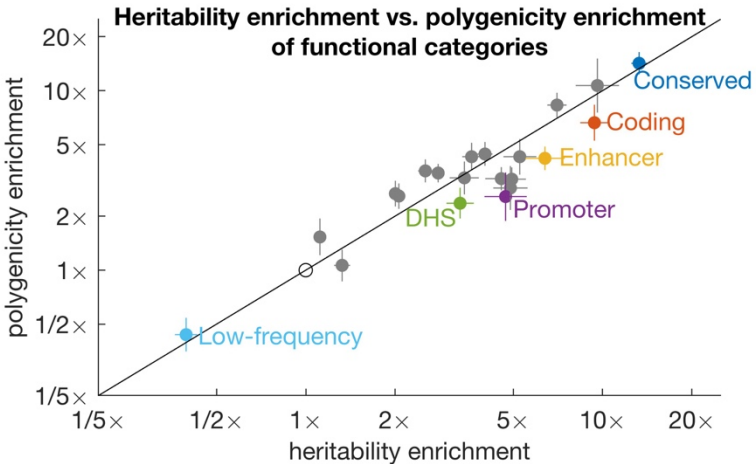


Figure 9. Estimates of polygenicity enrichment and heritability enrichment of functional categories. We report estimates for 20 functional categories plus low-frequency SNPs. Estimates are meta-analyzed across well-powered traits. Error bars denote 95% confidence intervals. Complete results for each trait are reported in Supplementary Table 3.9, and meta-analyzed results are reported in Supplementary Table 3.10.

We compared functional enrichment between groups of related traits (8 brain-related, 6 blood-related, and 5 immune-related traits; Supplementary Figure 3.8). 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 2).

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 (Supplementary Figure 3.9a). Traits with smaller M_e (larger kurtosis) had larger kurtosis explained ($r^2 = 0.78$; Supplementary Figure 3.9b).

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 5).

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 ref.[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 10a and Supplementary

Table 3.11), 5% of genes had large effect sizes (10x 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 10b and Supplementary Table 3.11), 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 Methods section.

We analytically inferred the genetic architecture that arises under each model (see Methods). Under both models, common variants have smaller per-allele effect sizes than low-frequency variants (Figure 10c-d), concordant with real traits[12,19,21]; moreover, both models produced a highly polygenic common-variant architecture (Supplementary Table 3.11). However, polygenicity differed for the two models at lower allele frequencies (Figure 10e-f). SNP-level flattening (i.e. lower polygenicity at lower allele frequencies) was only observed under the direct selection model; polygenicity was $\sim 30\text{x}$ lower for *de novo* SNPs than for common SNPs (vs. $\sim 1.1\text{x}$ under the pleiotropic selection model). Similar our results on real traits (Figure 8), low-frequency polygenicity was $\sim 4\text{x}$ smaller, and low-frequency heritability was also $\sim 4\text{x}$ smaller (Figure 10e). Thus, a $\sim 4\text{x}$ 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 5), 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 is predominantly explained by SNPs with effect sizes well below the bound, and h^2/M_e decreases (Figure 10e).

For a real trait, the difference in polygenicity between common and *de novo* SNPs could be much larger than the $\sim 30\mathrm{x}$ difference we observed under the direct-selection model. For example, when we added a few genes (0.25%) with extremely large effect sizes (100x larger than small-effect genes), the difference was ~600x, while the difference between common and low-frequency SNPs was still ~4x (Supplementary Table 3.11). The difference in polygenicity between common and *de novo* SNPs could also be slightly smaller than $\sim 30\mathrm{x}$ 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 8b). 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 10e. 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 $\sim 2\text{x}$, owing to $\sim 2.5\text{x}$ increased polygenicity for low-frequency SNPs. In contrast, the difference between common and *de novo* polygenicity increased slightly to $\sim 40\text{x}$ (Supplementary Table 3.11). 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 8).

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 10g); 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 10h). 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 (Figure 10i-j).

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].

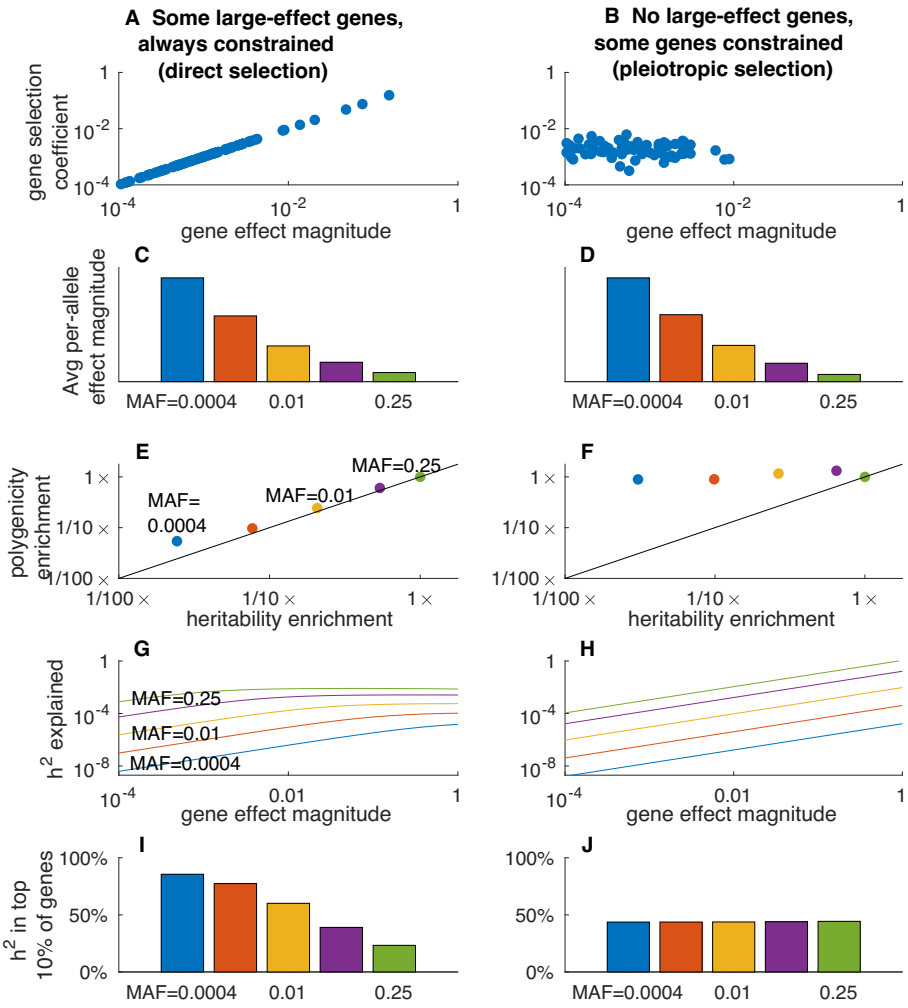


Figure 10. Gene-level flattening under an evolutionary model. In the left column (panels a,c,e,g,i), there are some large-effect genes, but direct stabilizing selection acting on the phenotype strongly constrains these genes. In the right column (panels b,d,f,h,j), there are no large-effect genes; pleiotropic stabilizing selection has varying effects on each gene, limiting common-SNP effect sizes on average. (a-b) Joint distribution of gene effect size magnitudes and selection coefficients. (c-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-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-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-j) Proportion of heritability explained by the top 10% of largest-effect genes for SNPs at different allele frequencies. Numerical results are reported in Supplementary Table 3.11.

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 50kb of 2,990 loss of function-intolerant genes from ExAC ("ExAC genic SNPs")[45,46]. These SNPs were more strongly enriched for polygenicity ($\sim 2.9\text{x}$) than for heritability ($\sim 1.7\text{x}$) (Supplementary Table 3.12). Compared with all genic SNPs (± 50kb), ExAC genic SNPs had 1.9x (95% CI: $1.7 - 2.0\text{x}$) larger polygenicity enrichment but only 1.3x (95%CI: $1.3 - 1.4\text{x}$) larger heritability enrichment, implying 0.71x (95%CI: $0.66 - 0.76\text{x}$) smaller average effect sizes (Supplementary Table 3.12). 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 8); 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 $\text{p} = 0.006$ for difference; Supplementary Figure 3.10 and Supplementary Table 3.13). 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 definition of polygenicity and a new method to estimate it, 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 $4\times$ smaller than its common-variant polygenicity, its *de novo* polygenicity may be $30-600\times$ smaller, and the two extremes would have completely different biological implications. A difference of only $30\times$ would imply that over 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 $600\times$ 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 GWAS 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 GWAS, 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 GWAS[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 (Supplementary Figure 3.1). 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 only be applied 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 only applied S-LD4M 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[19,68-70]. All of our primary analyses used the baseline-LD model,

which includes 6 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 (Supplementary Table 3.5), 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.

URLs

Open-source software implementing our method is available at: https://github.com/lukejoconnor/SLD4M. UK Biobank summary statistics are available at: https://data.broadinstitute.org/alkesgroup/UKBB/.

References

1 International Schizophrenia Consortium *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-752, doi:10.1038/nature08185 (2009).

2 Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-569, doi:10.1038/ng.608 (2010).

3 Stahl, E. A. *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* **44**, 483-489, doi:10.1038/ng.2232 (2012).

4 Loh, P. R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet* **47**, 1385-1392, doi:10.1038/ng.3431 (2015).

5 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291 (2015).

6 Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet* **47**, 1114-1120, doi:10.1038/ng.3390 (2015).

7 Moser, G. *et al.* Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. *PLoS Genet* **11**, e1004969, doi:10.1371/journal.pgen.1004969 (2015).

8 Palla, L. & Dudbridge, F. 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, doi:10.1016/j.ajhg.2015.06.005 (2015).

9 Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the genetic architecture of 30 complex traits from summary association data. *The American Journal of Human Genetics* **99**, 139-153 (2016).

10 Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **101**, 5-22, doi:10.1016/j.ajhg.2017.06.005 (2017).

11 Yang, J., Fritsche, L. G., Zhou, X., Abecasis, G. & International Age-Related Macular Degeneration Genomics, C. A Scalable Bayesian Method for Integrating Functional Information in Genome-wide Association Studies. *Am J Hum Genet* **101**, 404-416, doi:10.1016/j.ajhg.2017.08.002 (2017).

12 Zeng, J. *et al.* Signatures of negative selection in the genetic architecture of human complex traits. *Nat Genet* **50**, 746-753, doi:10.1038/s41588-018-0101-4 (2018).

13 Zhang, Y., Qi, G., Park, J.-H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits *Nature genetics* **50**, 1318, doi:https://doi.org/10.1038/s41588-018-0193-x (2018).

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

15 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753, doi:10.1038/nature08494 (2009).

16 Hirschhorn, J. N. Genomewide association studies--illuminating biologic pathways. *N Engl J Med* **360**, 1699-1701, doi:10.1056/NEJMp0808934 (2009).

17 Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186, doi:10.1016/j.cell.2017.05.038 (2017).

18 Mancuso, N. *et al.* The contribution of rare variation to prostate cancer heritability. *Nature genetics* **48**, 30 (2016).

19 Gazal, S. *et al.* Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nature genetics* **49**, 1421 (2017).

20 Gazal, S. *et al.* Low-frequency variant functional architectures reveal strength of negative selection across coding and non-coding annotations. *Nature Genetics* **50**, 1600-1607 (2018).

21 Schoech, A. P. *et al.* Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nature communications* **10**, 790 (2019).

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

23 Barton, N. H. & Keightley, P. D. Multifactorial genetics: understanding quantitative genetic variation. *Nature Reviews Genetics* **3**, 11 (2002).

24 Eyre-Walker, A. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proceedings of the National Academy of Sciences*, 200906182 (2010).

25 Agarwala, V., Flannick, J., Sunyaev, S., Altshuler, D. & Consortium, G. D. Evaluating empirical bounds on complex disease genetic architecture. *Nature genetics* **45**, 1418 (2013).

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

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

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

29 Niemi, M. E. *et al.* Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature* **562**, 268 (2018).

30 Hormozdiari, F. *et al.* Widespread allelic heterogeneity in complex traits. *The American Journal of Human Genetics* **100**, 789-802 (2017).

31 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* **47**, 1228 (2015).

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

33 Frei, O. *et al.* Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *bioRxiv*, 240275, doi:10.1101/240275 (2018).

34 O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nature Genetics* **50**, 1728-1734, doi:10.1038/s41588-018-0255-0 (2018).

35 Conneely, K. N. & Boehnke, M. So many correlated tests, so little time! Rapid adjustment of P values for multiple correlated tests. *The American Journal of Human Genetics* **81**, 1158-1168 (2007).

36 Vilhjálmsson, B. J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *The American Journal of Human Genetics* **97**, 576-592 (2015).

37 Wright, S. The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences* **23**, 307-320 (1937).

38 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).

39 Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nature genetics* **50**, 906 (2018).

40 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* **12**, e1001779 (2015).

41 UK10K consortium. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82 (2015).

42 Han, J. *et al.* A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS genetics* **4**, e1000074 (2008).

43 Sulem, P. *et al.* Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature genetics* **39**, 1443 (2007).

44 Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242 (2012).

45 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285 (2016).

46 Hormozdiari, F. *et al.* Leveraging molecular QTL to understand the genetic architecture of diseases and complex traits. *Nature genetics* **50**, 1041 (2018).

47 Hunter, D. J., Altshuler, D. & Rader, D. J. From Darwin's finches to canaries in the coal mine—mining the genome for new biology. *New England Journal of Medicine* **358**, 2760-2763 (2008).

48 Huang, H. *et al.* Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173 (2017).

49 Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337 (2015).

50 Shooshtari, P., Huang, H. & Cotsapas, C. Integrative genetic and epigenetic analysis uncovers regulatory mechanisms of autoimmune disease. *The American Journal of Human Genetics* **101**, 75-86 (2017).

51 Mahajan, A. *et al.* Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nature genetics* **50**, 559 (2018).

52 Liu, X., Li, Y. I. & Pritchard, J. K. Trans effects on gene expression can drive omnigenic inheritance. *bioRxiv*, 425108 (2018).

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

54 Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences* **109**, 1193-1198 (2012).

55 Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216 (2014).

56 Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179 (2014).

57 Howrigan, D. *et al.* Schizophrenia risk conferred by protein-coding de novo mutations. *bioRxiv*, 495036, doi:10.1101/495036 (2018).

58 Di Iulio, J. *et al.* The human noncoding genome defined by genetic diversity. *Nature genetics* **50**, 333 (2018).

59 GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204 (2017).

60 Zhou, J. *et al.* Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nature genetics* **50**, 1171 (2018).

61 Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nature genetics* **48**, 245 (2016).

62 Ernst, J. *et al.* Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nature biotechnology* **34**, 1180 (2016).

63 Zuk, O. *et al.* Searching for missing heritability: designing rare variant association studies. *Proceedings of the National Academy of Sciences* **111**, E455-E464 (2014).

64 Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185 (2014).

65 Do, R. *et al.* Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* **518**, 102 (2015).

66 Campbell, C. D. *et al.* Demonstrating stratification in a European American population. *Nature genetics* **37**, 868 (2005).

67 Boycott, K. M., Vanstone, M. R., Bulman, D. E. & MacKenzie, A. E. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature Reviews Genetics* **14**, 681 (2013).

68 Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved heritability estimation from genome-wide SNPs. *The American Journal of Human Genetics* **91**, 1011-1021 (2012).

69 Speed, D. *et al.* Reevaluation of SNP heritability in complex human traits. *Nature genetics* **49**, 986 (2017).

70 Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature genetics* **47**, 1114 (2015).

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

IV. Defining and characterizing pleiotropic relationships across 57 diseases and complex traits

The results in this chapter are yet to be published or submitted for publication. L.J.O. and A.L.P. developed the statistical methods, analyzed the results and wrote the manuscript. L.J.O. performed the analyses.

Introduction

Many common genetic variants are associated with more than one trait, a phenomenon known as *pleiotropy*[1-10]. Pleiotropy can be leveraged to increase association power[11,12], to identify shared disease mechanisms[3,4,6,9,13], and to identify modifiable risk factors[4,5,14-16].

Pleiotropy could result from at least four categories of biological explanations. First, there may be a causal relationship between the two traits[4,5,14-16], with the result that every SNP affecting one trait also affects the other. Second, there may exist one or more major shared pathways or risk factors[3,4,6,9,13,16]. Third, many of the same genes may affect both traits independently, possibly due to shared causal cell types (rather than specific shared pathways)[4,5,17]. Fourth, pleiotropic loci may harbor different causal genes for the two traits, and even different causal SNPs[4,18]. Although pleiotropy is known to be a common phenomenon, it is currently unclear which types of pleiotropy are most prevalent.

Some types of pleiotropy are expected to produce a genome-wide genetic correlation[3]. In particular, genetic correlations are usually interpreted as arising either causal relationships or shared pathways or risk factors[16,19-21] (although other explanations are possible). On the other hand, some pleiotropic mechanisms are less likely to result in a genetic correlation. First, two traits may share many of the same causal genes without sharing specific risk factors or pathways, if those genes are expressed in cell types causal for both traits. Second, if there are

multiple shared risk factors or pathways, with discordant directions of effect on each trait, their directional contributions to the genetic correlation may approximately cancel. Third, if some SNPs or haplotypes have cis-regulatory effects on multiple different genes, such that these loci can have pleiotropic effects despite the absence of pleiotropic genes, then the direction of pleiotropic effects would probably not be consistent across loci.

In order to quantify pleiotropy between complex traits and to distinguish between different types of pleiotropy, we introduce a new measure of the amount of pleiotropy between two traits, the *excess heritability overlap* (Q_e). Intuitively, this quantity measures the similarity between the Manhattan plot for two traits, whether or not the shared associations have concordant directions of effect. We characterize the landscape of correlated and uncorrelated pleiotropic relationships across 57 diseases and complex traits.

Overview of Methods

We use the term "pleiotropy" primarily to describe a relationship between two traits. Roughly, a locus is pleiotropic if it is associated with both traits, and two traits have nonzero pleiotropy if they have more pleiotropic loci than expected due to chance. However, we do not wish to define pleiotropy in a manner that relies upon a binary distinction between "associated" and "not associated," because the number of significantly associated loci inevitably increases with increasing sample size. In previous work[22], we defined polygenicity as the *effective number of associated SNPs* (Me) using fourth moments of the effect-size distribution to avoid making a binary distinction between and associated and non-associated SNPs. Here, we define the *excess*

heritability overlap (Q_e) between two traits using mixed fourth moments of their bivariate effect-size distribution.

Intuitively, Q_e quantifies the correlation between the Manhattan plots for two traits. These plots highlight sparsity in the genetic signal; loci with effects much larger than average effect sizes stand out, and if these loci are frequently shared between the two traits, then Q_e will be large.

If there is no LD between causal SNPs, the genetic correlation of a trait is proportional to the mixed second moment of its causal effect size distribution, $E[\beta_1\beta_2]$:

$$r_g = \frac{E[\beta_1\beta_2]}{\sqrt{E[\beta_1^2]E[\beta_2^2]}}.$$

The excess heritability overlap is proportional to the mixed *fourth* moment, $E[\beta_1^2\beta_2^2]$:

$$Q_e = \frac{k_{12} - s_{12}}{\sqrt{(k_{11} - s_{11})(k_{22} - s_{22})}},$$

$$k_{ij} = E(\beta_i^2\beta_j^2), \qquad s_{ij} = E\left(\beta_i^2\right)E(\beta_j^2) + 2E\left(\beta_i\beta_j\right)^2.$$

The terms s_{11}, s_{22} and s_{12} remove from the mixed fourth moments what is expected by chance under a normal distribution with the specified second moments.

In the presence of LD between causal SNPs, Q_e is defined in terms of both causal effect sizes β and marginal effect sizes α (see Methods). This has the effect that if there are many haplotypes harboring different one causal SNP for each trait, then Q_e will be positive, even if those causal SNPs are distinct from each other. However, Q_e will not be inflated merely as a consequence of

the fact that some SNPs have more LD than others, and thus more opportunity to tag causal SNPs for any trait[23].

In order to estimate Q_e, we extend LD 4th moments regression[22] to pairs of traits. We regress products of χ^2 statistics on "LD 4th moments", allowing us to bridge between the distribution of marginal effect sizes (inclusive of LD) and the distribution of causal effect sizes (without LD). We make use of the following regression equation:

$$E\left(\alpha_i^2\alpha_j^2\right) - \left(E(\alpha_i^2)E\left(\alpha_j^2\right) + 2E\left(\alpha_i\alpha_j\right)^2\right) \approx l^{(4)}(k_{ij} - s_{ij}),$$

where α_i is the *marginal* effect of a SNP on trait *i* (inclusive of LD), and $l^{(4)}$ is the LD 4th moment of that SNP, defined as the sum of r^4 values between that SNP and all SNPs (including itself).

Simulations

We performed a series of simulations to assess whether cross-trait LD4M produces robust estimates of Q_e, and to illustrate how Q_e varies under different genetic architectures. We simulated summary association statistics from the asymptotic sampling distribution for a subset of UK Biobank imputed common SNPs on chromosome 1 (M=100k SNPs). By default 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.

First, we performed simple simulations with no genetic correlation and no functional annotations. The bivariate effect-size distribution was a mixture of null SNPs, SNPs affecting trait 1 only, SNPs affecting trait 2 only, and SNPs affecting both traits with no correlation. Non-

null effect sizes were drawn from identical normal distributions for each trait. We varied the proportion of non-null SNPs that were shared, p_{12}, while fixing the total proportion of non-null SNPs for each trait, $p_1 = p_2 = 0.4\%$. Under this model, Q_e is equal to $\frac{1}{3}corr(x_1, x_2)$, where x_1 is 1 if $\beta \neq 0$ for trait 1, and zero otherwise. If effect sizes for SNPs that are non-null for both traits had identical effect-size magnitudes, instead of having effect sizes drawn from independent normal distributions, then Q_e would be equal to $corr(x_1, x_2)$. In these simulations, cross-trait LD4M produces unbiased estimates of Q_e (Figure 11).

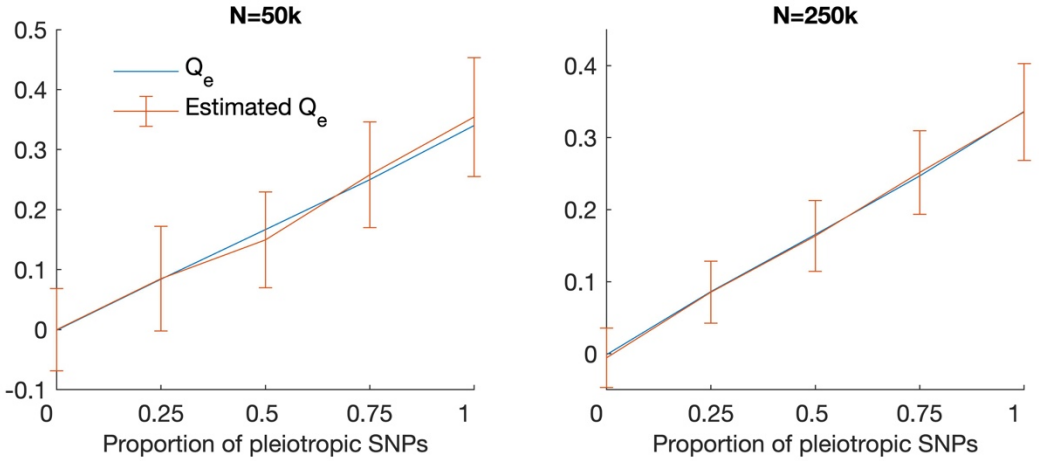


Figure 11 True and estimated values of Q_e in simulations with varying proportion of pleiotropic SNPs. For each trait, the heritability explained by pleiotropic SNPs proportional to the proportion of pleiotropic SNPs. M=100k SNPs. Points are medians and median estimates of 100 simulations; error bars are interquartile ranges.

If a SNP has a large effect on both traits, it contributes much more to Q_e than if it has a small effect on both traits. To illustrate this effect, we performed simulations where we fixed the proportion of heritability explained by pleiotropic SNPs and varied the effect size of those SNPs. In particular, we varied σ_{12}^2, the effect-size variance of these SNPs on both traits, while fixing $h_{12}^2 = p_{12}\sigma_{12}^2$, where p_{12} is the proportion of pleiotropic SNPs and h_{12}^2 is the proportion of heritability explained by pleiotropic SNPs. We also fixed σ_1^2 and σ_2^2, the effect-size variance of

non-pleiotropic SNPs on each trait (these values influence the denominator of Q_e; increasing the variance parameters increases the denominator and decreases Q_e). We found that as we increase σ_{12}^2, Q_e increases, and our estimates remain unbiased (Figure 12).

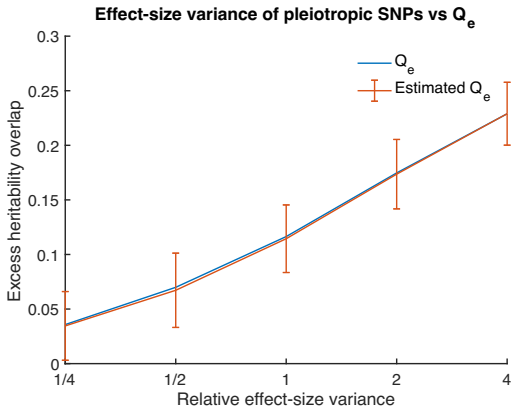


Figure 12 True and estimated Q_e in simulations varying the relative effect size of pleiotropic vs trait-specific causal SNPs. ~40% of heritability for each trait is explained by pleiotropic SNPs, with zero genetic correlation. However, this shared heritability is either spread across 2-4x as many SNPs with 2-4x smaller effect-size variance (left), or concentrated within 2-4x as many SNPs with 2-4x larger effect-size variance (right). M=100k SNPs, N=50k samples. Points are medians and median estimates of 100 simulations; error bars are interquartile ranges.

There exists a complex relationship between Q_e and r_g. On the one hand, if the genetic correlation is sparse, such that only a fraction of SNPs have large, correlated effects on both traits, then it contributes to Q_e. On the other hand, if the genetic correlation arises from an infinitesimal, normally distributed signal that is correlated across the genome, then it does not contribute to Q_e. Intuitively, an infinitesimal genetic correlation would be impossible to identify based on a Manhattan plot (which only displays unsigned χ^2 statistics), whereas a sparse genetic correlation would be readily apparent, because a small fraction of SNPs would have large effects on both traits. We illustrated these scenarios in two sets of simulations with a genetic correlation. In the first set of simulations, we specified that ½ of non-null SNPs for each

trait would affect both traits with effect sizes following a bivariate normal distribution with a specified correlation (and variance equal to that of the trait-specific non-null SNPs). At r_g=0, this is equivalent to the simulation in Figure 11. Increasing the genetic correlation increases the value of Q_e, because increasing the signed correlation of a normal distribution also increases the correlation between the effect-size magnitudes. We obtained unbiased estimates of Q_e (Figure 13a). In the second set of simulations with a genetic correlation, we modeled the effect-size distribution of each trait as having an independent sparse component (0.4% of SNPs having substantial non-null, normally distributed effects on each trait) and a correlated infinitesimal component (100% of SNPs having tiny, normally distributed effects). Each component explained half of heritability. We varied the correlation of the infinitesimal component; as expected, this had no effect on Q_e. Thus, it is possible to have a substantial genetic correlation and zero Q_e, especially for highly polygenic traits. Estimates of Q_e were unbiased (Figure 13b).

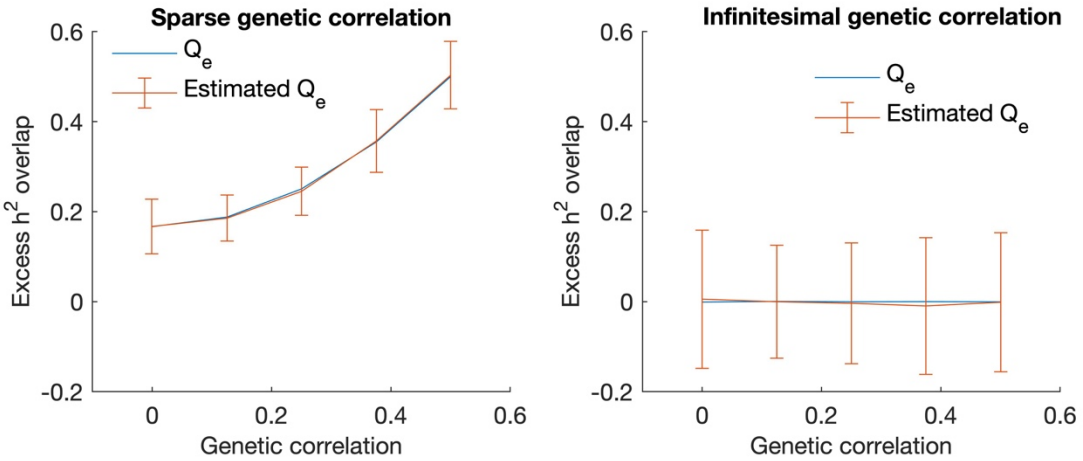


Figure 13 Simulations with a nonzero genetic correlation. In each simulation, 50% of the heritability of each trait is explained by pleiotropic SNPs. In panel a, these SNPs have a sparse effect-size distribution. In panel b, these SNPs have an infinitesimal effect-size distribution. M=100k SNPs, N=50k samples. Points are medians and median estimate of 200 simulations s; error bars are interquartile ranges.

One way that pleiotropy can arise is via shared functional enrichment: if there exists a functional annotation (e.g. coding) that is enriched for heritability for both traits, and effect sizes for the two traits are independent conditional on that functional annotation, then Q_e will be elevated. However, it should also be possible to condition on the annotation and estimate the excess heritability overlap compared with what is expected given that the annotation is enriched for both traits. We performed simulations with a single, simulated annotation spanning 10% of the genome. We specified that effect sizes for each trait would be drawn from independent point-normal distributions conditional on the annotation, but that a higher proportion of SNPs inside the annotation would be non-null. We varied the heritability enrichment of the annotation; as the shared enrichment increased, Q_e increased, but the conditional value of Q_e (Q_{annot}) did not. Estimates of Q_e and Q_{annot} were unbiased (Figure 14).

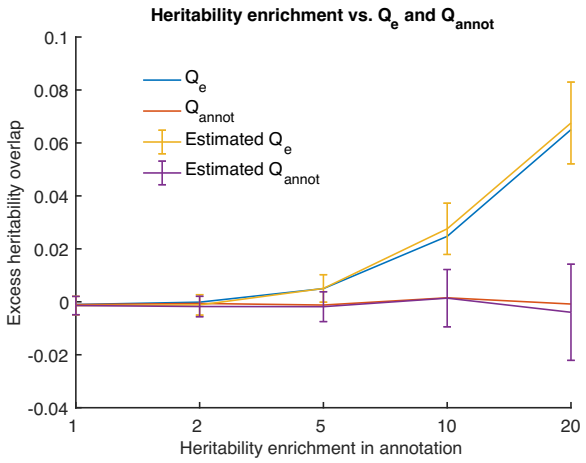


Figure 14 Simulations with shared heritability enrichment. The annotation spans 10% of the genome with contiguous blocks of 500 SNPs. Effect sizes for each trait are drawn independently conditional on the annotation, such that Q_{annot}=0. However, shared enrichment produces nonzero values of Q_e. M=100k SNPs, N=50k samples. Points are medians and median estimates of 100 simulations; error bars are interquartile ranges.

Pervasive low-level pleiotropy among 57 diseases and complex traits

We applied cross-trait LD4M and cross-trait LDSC to 57 diseases and complex traits, using summary statistics computed from UK Biobank and other publicly available data (Table 3). We excluded traits with low power and genetically redundant traits (Table 3 caption). As in previous work, we excluded the MHC region from all analyses, due to its unusually large effect sizes and long-range LD patterns.

A large percentage of trait pairs have nonzero pleiotropy: of all 1596 trait pairs, 963 (60%) had statistically significant excess heritability overlap at 5% FDR (one-tailed $p<0.03$), and 666 (42%) had statistically significant genetic correlation at 5% FDR (two-tailed $p<0.02$). 505 trait pairs (32%) passed both FDR thresholds; there were 458 trait pairs with significant Q_e but not significant r_g. Thus, these metrics provide complementary information about shared genetic architecture, and hundreds of genetically related trait pairs have statistically significant excess heritability overlap but no discernable genetic correlation.

Despite evidence of nonzero excess heritability overlap among a large percentage of traits, Q_e estimates are small for most of these traits. We estimated Q_e for 1480 well-powered trait pairs, excluding 116 trait pairs having noisy Q_e denominators (this denominator is not used to calculate statistical significance; see Methods). Although a majority (64%) of these trait pairs had significant Q_e, a majority (65%) had Q_e estimates lower than 0.05 (Figure 15). Both percentages could be underestimates: because of finite power, it is possible that all traits have true Q_e values that are strictly greater than zero, and that roughly 80% of trait pairs have Q_e less

than 0.05 (our Q_e estimate was >1.6 standard deviations greater than 0.05 for 16% of trait pairs). Thus, low-level pleiotropy is pervasive; any two traits are likely to have some pleiotropy beyond random expectation, but the amount of pleiotropy is usually small, and it may not imply meaningful shared biology.

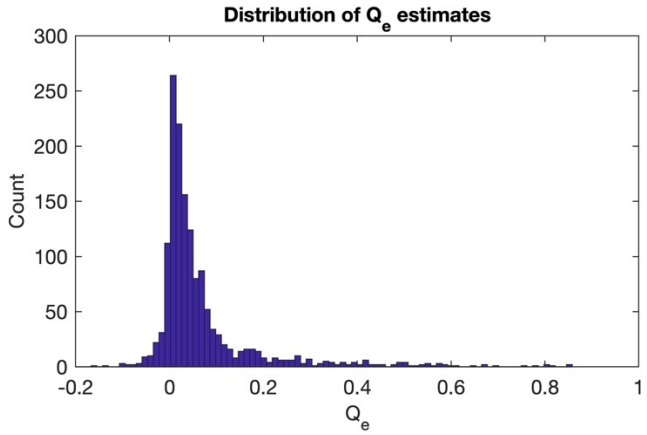


Figure 15 Distribution of Q_e estimates. We estimated Q_e for 1480 well-powered trait pairs (see Methods). The observed distribution has a mean of 0.068, a median of 0.030, and a standard error of 0.12. The root-mean-squared standard error was 0.041, and the median standard error was 0.015.

We considered the possibility that low-level pleiotropy is driven by shared heritability enrichment in known genomic annotations. Q_e measures the heritability overlap in excess of what would be expected by chance under a simple null model with frequency-dependent causal effect sizes[24,25] (see Methods). We also estimated the heritability overlap in excess of what would be expected under the richer baseline-LD model[26], which includes 97 genomic annotations related to allele frequency, LD, selection and non-cell-type-specific function. This quantity ($Q_{baselineLD}$) was barely smaller than Q_e (average difference 0.002(s.e.=0.0004)), with highly concordant point estimates (Figure 16), implying that the pleiotropy we observe is not explained by known functional and other annotations. We note that on a log scale, functional annotations explain a substantial fraction of sparsity (and pleiotropy), but the difference

between Q_e and $Q_{baselineLD}$ is on a linear scale. (For example, if 1% of SNPs explain 100% of heritability for two traits, and a functional annotation spanning 10% of SNPs contains 100% of heritability, that annotation would explain 50%=log(10)/log(100) of pleiotropy on a log scale, but only 10%=10/100 of pleiotropy on a linear scale.) Thus, pervasive low-level pleiotropy is not merely an expected consequence of heritability enrichment in known genomic annotations.

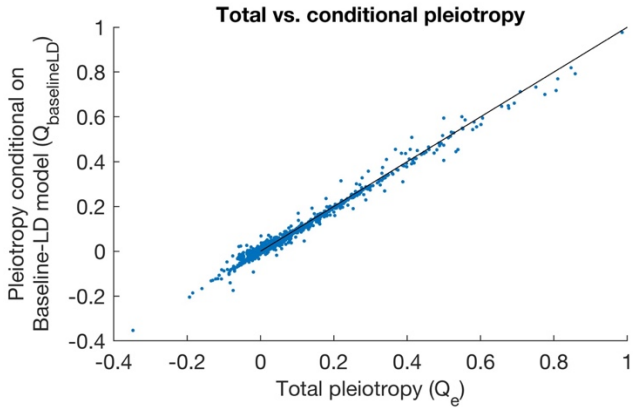


Figure 16 Comparison of total vs. conditional pleiotropy, conditional on the baseline-LD model[26]. We show all pairs of traits with estimated Q_e standard error less than 0.2. The average difference between Q_e and $Q_{baselineLD}$ was 0.0018(s.e.=0.0004). The baseline-LD 2.2 model contains 97 genomic annotations related to MAF, LD, selection, and function.

Table 3 Phenotypes used in the analysis.

Brain	Eosinophil count	Mean platelet volume
Depressive symptoms	Lymphocytes	Mean cell volume
Schizophrenia	Platelets	Monocyte count
Years of education	White blood cell count	Bone mineral density (heel)
Smoking Status	Eczema	Balding
Neuroticism	Asthma	BMI
Fluid intelligence score		Corneal Hysteresis (left)
Morning Person	**Other**	Height
Time TV	Alanine Aminotransferase	Blood pressure (systolic)
Time Walking	Albumin	Blood pressure (diastolic)
Age First Birth	Calcium	Cancer (any)
Number of Children	C Reactive Protein	Breast cancer
Number of Partners	Creatinine	Heart Attack
	Gamma Glutamyltransferase	Hypothyroidism
Immune	Glucose	Type II diabetes
Celiac	HbA1c	Basal metabolic rate
Crohn's Disease	IGF1	FEV1/FVC
Lupus	Phosphate	FVC
Primary biliary cirrhosis	Total Serum Protein	Urine sodium
Rheumatoid Arthritis	Urea	Age at Menarche
Ulcerative Colitis	Reticulocyte count	

We selected phenotypes on the basis of three criteria: first, we excluded traits with low power to detect a heritable signal, as measured by Z score for nonzero heritability (Z_h<5); second, we excluded traits with low power to detect sparsity in the heritable signal, as measured by Z score for excess kurtosis (Z_k<2); and third, we excluded traits with high pairwise genetic correlations (r_g>0.9).

Strong pleiotropy among complex traits with shared causal cell types

The distribution of Q_e estimates has a heavy tail, with a minority of trait pairs having Q_e estimates much greater than 0.05 (Figure 15). We categorized 12 trait pairs as being brain

related (e.g. schizophrenia, age at first birth, years of education) and 12 different trait pairs as being immune related (e.g. rheumatoid arthritis, asthma, lymphocyte count).

Among brain-related traits, 89% had FDR-significant Q_e (compared with 60% of all trait pairs), and Q_e estimates were much more likely to fall in the tail of the trait-pair distribution: among all 56 well-powered trait pairs, the mean Q_e estimate was 0.41(s.e.=0.03), and only 3 trait pairs (5%) had point estimates less than 0.05 (Figure 17). In contrast, for brain vs. non-brain trait pairs, the average Q_e estimate was 0.04(s.e.=0.004). Within these brain-related traits, there was no clear evidence of clustering. Consistent with previous findings[3,8], brain-related trait pairs were also more likely to have a significant genetic correlation (76% compared with 32% of all traits), and the average absolute genetic correlation was 0.25(s.e.=0.01) among brain-related traits compared with 0.10(s.e.=0.004) for brain vs. non-brain trait pairs. 7 trait pairs had high Q_e but only a modest genetic correlation (Q_e>0.4 and $|r_g|$<0.2), including schizophrenia and depressive symptoms (Q_e=0.50(0.11), r_g=0.18(0.05)). Thus, brain-related traits have much greater pleiotropy with each other than with non-brain-related traits, and this excess is partially explained by increased genetic correlations, but there also exist shared genetic associations with directions of effect that go against the genetic correlation.

Similar to brain-related traits, 86% of immune-related traits had had statistically significant Q_e estimates. However, there were three differences. First, the magnitude of these estimates was smaller on average than for brain-related traits (average Q_e=0.19(0.02) vs 0.41(0.03)). These estimates were still much higher than immune vs. non-immune trait pairs (average

Q_e=0.04(s.e.=0.005)). Second, a larger number of immune trait pairs had low Q_e estimates; 11 trait pairs had estimated Q_e<0.05, and some clustering within immune-related traits is evident (Figure 17). Third, only 35% of immune-related trait pairs had a statistically significant genetic correlation (compared with 76% of brain traits), and the average absolute genetic correlation was only modestly elevated compared with other trait pairs (0.17(0.01) for immune vs immune; 0.10(0.01) for immune vs non-immune). This is consistent with the observation that pleiotropic loci often have discordant effect directions on autoimmune diseases[6]. 4 trait pairs had high Q_e estimates and small r_g estimates (Q_e>0.2 and $|r_g|$<0.1); these included Celiac disease and primary biliary cirrhosis, which had no evidence of a genetic correlation (r_g=0.04(0.10)) but surprisingly large excess heritability overlap (Q_e=0.37(0.07)). Remarkably, these estimates would be consistent with a genetic architecture where 100% causal SNPs are shared between the two diseases and their causal effect sizes follow independent normal distributions. Thus, many but not all immune-related traits have much greater pleiotropy with each other than with non-immune-related traits, and this excess is mostly not explained by increased genetic correlations.

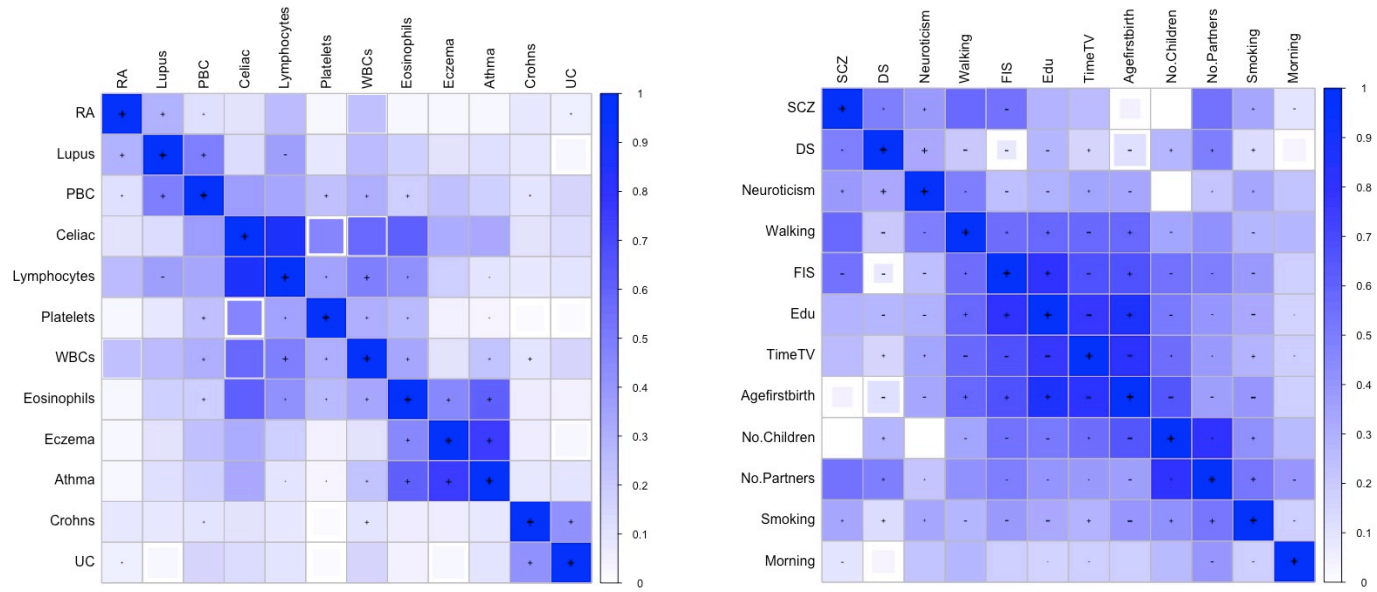


Figure 17 Pleiotropy among brain- and immune-related traits. The color of each cell indicates the point estimate of Q_e. Squares that are only partially filled are not FDR significant, with diameter proportional to significance (T statistic). For trait pairs with an FDR-significant genetic correlation, a '+' or '-' is plotted, with size proportional to the magnitude of the r_g point estimate. Traits were manually ordered to place traits with high Q_e adjacent to each other.

The large amount of pleiotropy among brain-related and immune-related traits suggests that higher resolution cell-type-specific functional enrichment may be even more informative. For each pair of traits, we computed the heritability enrichment similarity (see Methods) based on 502 cell-type-specific epigenetic annotations[27]. Enrichment similarity (S_{CTS}) was highly correlated with Q_e (r=0.61; Figure 18a-b) and modestly correlated with absolute r_g (r=0.30; Figure 18c-d). Brain- and immune-related traits had substantially greater than average enrichment similarity (average S_{CTS}=0.64 for brain, 0.55 for immune, 0.31 for all trait pairs; 0.24 between brain and immune). However, enrichment similarity was especially informative between immune-related trait pairs (r=0.76 between S_{CTS} and Q_e). It was less informative

between brain-related trait pairs (r=0.28), probably because nearly all brain-related trait pairs had very high S_{CTS}.

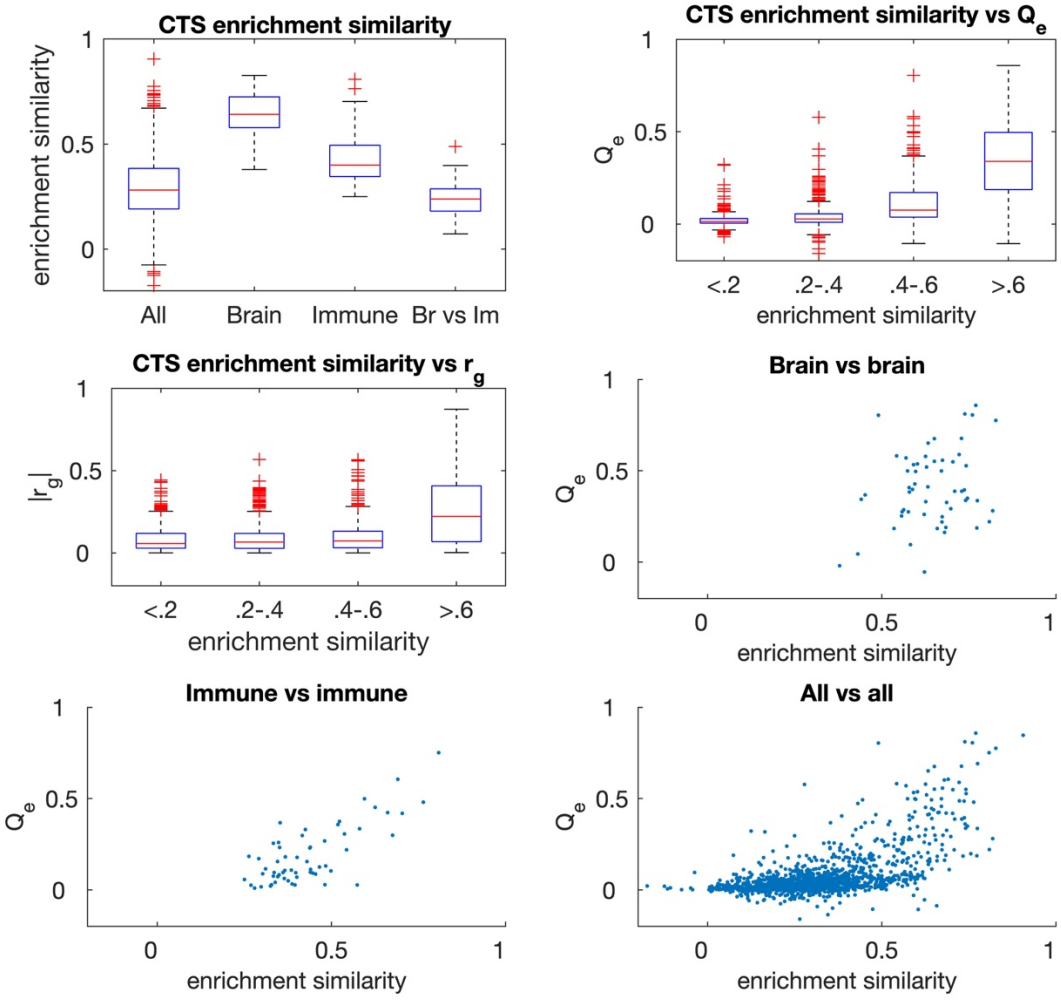


Figure 18 Comparison between cell type enrichment similarity (S_{CTS}) compared and pleiotropy. We computed the similarity between heritability enrichment profiles for each pair of traits based on 502 cell-type-specific chromatin annotations[27]. (a) Distribution of S_{CTS} values across phenotype groups. (b) S_{CTS} compared with Q_e for all traits. (c) S_{CTS} compared with r_g for all traits. (d) Scatterplot of S_{CTS} compared with Q_e for brain-related traits. (d) Scatterplot of S_{CTS} compared with Q_e for brain-related traits. (e) Scatterplot of S_{CTS} compared with Q_e for immune-related traits. (f) Scatterplot of S_{CTS} compared with Q_e for all traits.

Between brain- and immune-related traits, although there was little pleiotropy overall (average Q_e=0.03(s.e.=0.01)), an exception was ulcerative colitis (UC). UC had Q_e estimates >0.3 with three brain-related traits (and p<0.01 for each); these traits were years of education (Q_e=0.32(0.16); r_g=0.28(0.13)), time watching TV (Q_e=0.32(0.15); r_g=-0.43(0.13)), and fluid intelligence score (Q_e=0.30(0.14); r_g=0.15(0.14)) (Table 4). Besides UC, no other brain-immune trait pair had estimated Q_e>0.13, and we observed no similar results for Crohn's disease. These results indicate that a substantial fraction of UC heritability is mediated via brain-related mechanisms, consistent with evidence that psychological stress can cause increased risk of inflammatory bowel disease[28].

Table 4 Pleiotropy between brain- and immune-related traits including Ulcerative Colitis.

Trait 1	Trait 2	r_g	Q_e	Q_e p value
Ulcerative Colitis	Years of education	0.28(0.13)	0.32(0.16)	0.009174
Ulcerative Colitis	Time TV	-0.43(0.13)	0.32(0.15)	0.008287
Ulcerative Colitis	Fluid intelligence score	0.15(0.14)	0.30(0.14)	0.004022
Ulcerative Colitis	Age first birth	0.21(0.15)	0.24(0.12)	0.015337
Ulcerative Colitis	DS	0.15(0.06)	0.15(0.06)	0.001755

We display the top 5 pairs of brain- vs. immune-related traits by estimated Q_e. Many brain-related traits had substantial heritability overlap with UC, but not with other immune-related traits, including Crohn's disease.

Pleiotropy correlates strongly with polygenicity

For each trait, we computed the average amount of excess heritability overlap that it has with all other traits. These values ranged from 0 to 0.17. There was a remarkably strong positive correlation between average pleiotropy and polygenicity (Me^{22}) (r=0.77; polyfig). This correlation is partially driven by the fact that brain-related traits consistently have the greatest polygenicity and the greatest average pleiotropy, but the same trend is evident among non-brain-related traits. This observation suggests that a polygenicity and pleiotropy might result from a similar underlying phenomenon.

This trend may appear to be an unsurprising consequence of the mathematical definition of Q_e and Me, but it is not. First, one might expect that traits with greater polygenicity would be expected to have greater pleiotropy purely due to chance, but this is not the case; if the genetic architecture of two traits is independent, then Q_e will be zero, no matter the polygenicity of each. Second, one might expect Me and Q_e to covary because the fourth moment of the univariate effect-size distribution for a trait appears in the denominator of both. If this value grows and the numerator remains constant, then both Q_e and Me decrease. However, the numerator of Q_e also contains the mixed fourth moment of effect sizes for both traits, and this quantity may be expected to covary with the denominator, such that Q_e does not correlate with Me. For example, suppose that differences in polygenicity across traits are driven by differences in the number of causal genes for each trait, and for each pair of traits, a fraction *p* of causal genes are shared. Varying the number of causal genes would change the polygenicity but would not change the pleiotropy between two traits, which would be determined by *p*.

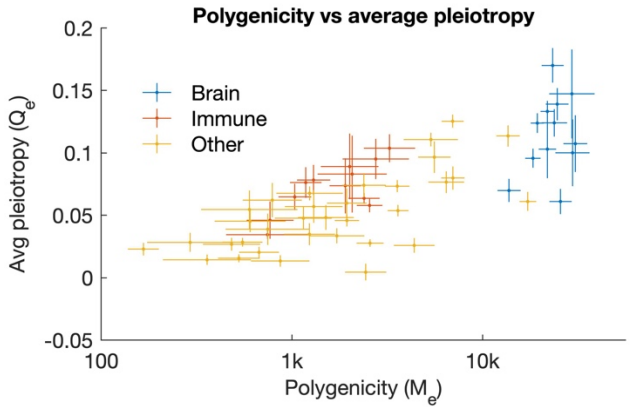


Figure 19 Comparison between polygenicity and average pleiotropy. We show the effective number of independent SNPs (M_e) for each trait, as well as the average value of Q_e between that trait and all other traits (if well-powered). Error bars represent standard errors.

Discussion

Our results suggest that pleiotropic mechanisms can be usefully sorted into three layers. The first layer comprises shared pathways, shared risk factors, and causal relationships. The second layer comprises shared causal cell types. Under the omnigenic model[17], any gene expressed in these cell types is expected to have some effect on both traits, leading to nonzero pleiotropy. The third layer comprises non-cell-type-specific regulatory (or coding) effects, leading to low-level pleiotropy. Some SNPs and regions of the genome are more functional than others; although the baseline-LD model does not explain most low-level pleiotropy, it is conceivable that there exists a set of non-cell-type-specific functional annotations that would explain these effects.

This model is consistent with several of our main observations. We observe that most trait pairs involving unrelated cell types have some low-level pleiotropy (Figure 15), consistent with third-layer mechanisms. Pleiotropy is much stronger among trait pairs with shared causal cell types,

and there are very few trait pairs with similar causal cell types but no evidence of pleiotropy, especially for brain-related traits (Figure 17 and Figure 18). This suggests that shared causal cell types are sufficient to drive substantial nonzero pleiotropy, as it is less likely that all brain-related traits have more specific shared mechanisms. However, some specific pairs of traits have much stronger pleiotropy than other trait pairs with shared causal cell types (Figure 17), suggesting specific shared pathways.

Highly polygenic traits have much greater pleiotropy than less-pleiotropic traits (Figure 19), suggesting that the same phenomenon which is responsible for differential polygenicity across traits is also responsible for differential pleiotropy. In previous work[22], we argued that *flattening* due to negative selection explains the extreme polygenicity of complex traits. Genes and regions of the genome that are directly involved in critical disease pathways are strongly constrained by negative selection, and as a result, most heritability is explained by genes and regions with smaller biological effects. The set of genes with small biological effects may include a large percentage of all genes that are expressed in causal cell types[17], and these genes may be more likely to be shared across traits than genes with large biological effects on critical disease pathways. Thus, negative selection may tilt the relative importance of first- and second-layer disease mechanisms. Traits that are under strong selection (either directly or pleiotropically) have increased polygenicity because their heritability is spread across thousands of genes expressed in relevant cell types, not concentrated in the limited number of genes that participate in critical disease pathways. Moreover, these causal cell types are more likely to be shared across traits than the specific disease pathways, causing strongly selected traits to also have greater pleiotropy.

This work has several important limitations. First, our definition of pleiotropy does not explicitly require that two traits have shared causal SNPs, and our method to estimate it does not attempt to fine-map shared associated loci. While we believe that shared causal SNPs is the most parsimonious explanation when two traits have high Q_e, it is also possible that some regions of the genome are more likely to harbor causal SNPs for both traits and that these SNPs tend to be in LD due to their proximity. Second, shared population stratification could potentially lead to inflated Q_e estimates, and it is difficult to rule out the possibility that the low-level pleiotropy we observe is partially explained by stratification. However, we have analyzed mostly UK Biobank data, which has relatively little population stratification, and the cell-type-specific patterns that we observe cannot plausibly be explained by this type of confounding. Third, we have not actually identified any of the specific mechanisms underlying the pleiotropic associations that have been described. Experimental work will be needed to validate specific instances of the types of mechanisms that are hypothesized to mediate these effects. Despite these limitations, this work represents an advance in our ability to quantify pleiotropy, and it suggests how pleiotropic relationships can be usefully categorized.

Methods

Definition of excess heritability overlap

For a pair of complex traits

$$y_1 = \boldsymbol{x\beta_1} + \epsilon_1, y_2 = \boldsymbol{x\beta_2} + \epsilon_2,$$

where bold font indicates that $\boldsymbol{x}$ and $\boldsymbol{\beta}$ are vectors, the genetic correlation can be defined as

$$r_g = corr(\boldsymbol{x\beta_1}, \boldsymbol{x\beta_2}).$$

If $\boldsymbol{\beta}$ is viewed as a random variable, this is equal to $E(\alpha_1\beta_2)/\sqrt{E(\alpha_1\beta_2)E(\alpha_2\beta_2)}$, where $\boldsymbol{\alpha} = R\boldsymbol{\beta}$ is the vector of marginal effect sizes, and non-bold α, β indicates a randomly-chosen element of the random vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$. In the absence of LD between causal SNPs, this is equal to $E(\beta_1\beta_2)/\sqrt{E(\beta_1\beta_2)E(\beta_1\beta_2)}$, which we refer to as the LD-free genetic correlation.

Analogous to the LD-free genetic correlation, we define the *LD-free excess heritability overlap* as:

$$Q_e{}' = \frac{k_{12} - s_{12}}{\sqrt{(k_{11} - s_{11})(k_{22} - s_{22})}},$$

$$k_{ij} = E(\beta_i^2\beta_j^2), \qquad s_{ij} = E\left(\beta_i^2\right)E\left(\beta_j^2\right) + 2E\left(\beta_i\beta_j\right)^2.$$

The differences $k_{ij} - s_{ij}$ are the joint fourth cumulants, and unlike moments, they behave additively: for independent random variables, the cumulant of the sum is the sum of the cumulants. The 4th joint cumulants of a bivariate normal distribution are zero; hence, this quantity can be interpreted as measuring the shared sparsity of the effect-size distribution for each trait. It is similar to the correlation between the *squared* effect sizes β_1^2, β_2^2, in that it

measures the overlap between the genetic signals for each trait without requiring that these signals have a signed correlation. The primary difference between this quantity and the correlation between the squared effect sizes is that s_{12} includes the term $2E(\beta_1\beta_2)^2$. If β_1 and β_2 follow a bivariate normal distribution with variances σ_1^2, σ_2^2 and correlation r, then the covariance between the squared effect sizes is equal to $2\sigma_1\sigma_2 r^2$, and this term is subtracted off such that the numerator of Q_e would be zero. In addition, if effect sizes for each trait follow a normal distribution, then the *denominator* is also zero, and Q_e is undefined. That is, if there is no sparsity in the genetic architecture of one of the traits, then there can be no excess heritability overlap beyond what is expected by chance, because all SNPs have the same expected heritability. Thus, Q_e can be interpreted as measuring the *shared sparsity* of the effect-size distribution for each trait.

More generally, in the presence of LD between causal SNPs, the excess heritability overlap is defined as

$$Q_e = \frac{k_{12} - s_{12}}{\sqrt{(k_{11} - s_{11})(k_{22} - s_{22})}},$$

$$k_{ij} = E\left(\beta_i^2\alpha_j^2 + 2\beta_i\beta_j\alpha_i\alpha_j - 2\beta_i^2\beta_j^2\right), \qquad s_{ij} = E(\beta_i^2)E\left(\alpha_j^2\right) + 2E\left(\beta_i\beta_j\right)E\left(\alpha_i\alpha_j\right).$$

Note that in general this quantity is not symmetric with respect to the ordering of the two traits, but it is symmetric under a mild and natural assumption.

Shared heritability enrichment in functional annotations, or in MAF- and LD-related annotations, can contribute to Q_e.

Estimation equations

Cross-trait S-LD4M relies on a fourth-moment regression equation that generalizes the S-LD4M regression equation to a pair of traits:

$$E\left(\alpha_i^2\alpha_j^2\right) - \left(E(\alpha_i^2)E\left(\alpha_j^2\right) + 2E\left(\alpha_i\alpha_j\right)^2\right) \approx l^{(4)}(k_{ij} - s_{ij}), \qquad (3)$$

where $l^{(4)}$ is the LD 4th moment of the SNP. The left-hand side is the fourth (co-) cumulant of the marginal effect size distribution. This behavior should not be surprising: for independent random variables, (co-) cumulants are additive, i.e. $k(x, y, z, w + w') = k(x, y, z, w) + k(x, y, z, w')$ if $w \perp w'$. The reason this regression equation is described as approximate is that effect sizes are not expected to be independent. In particular, we generalize the S-LD4M LD assumption as follows: for SNPs i and j,

$$cov\left(\alpha_{i1}^2, \beta_{j2}^2\right) \approx r_{ij}cov\left(\alpha_{j1}^2, \beta_{j2}^2\right).$$

This holds under various scenarios: when causal effect sizes for the two traits are independent; when causal effect sizes for the two SNPs are independent; when R has entries equal to only zero or (plus or minus) one. The derivation of the regression equation from this LD approximation is similar to the S-LD4M derivation.

In the case of multiple annotations, $l^{(4)}$ is a vector, k and s are vectors, and the term $E(\alpha_i^2)E\left(\alpha_j^2\right) + 2E\left(\alpha_i\alpha_j\right)^2$ is conditioned on the partitioned LD scores. If causal effect sizes follow a bivariate normal distribution conditional on the annotation values, then the left- and right-hand sides of (3) are both equal to zero.

Enrichment similarity

We defined the *heritability enrichment similarity* (S_{annot}) between two traits, with respect to a set of tissue- or cell-type-specific functional annotations. S_{annot} is equal to the value of Q_e that would be produced if effect size magnitudes were completely determined by the annotations and there were no genetic correlation. If two traits have identical patterns of functional enrichments, then SNPs have identical effect-size magnitudes on them, and S_{annot} is equal to 1. In equations:

$$S_{annot} = \frac{E_a\left[E[\alpha_1^2|a]E[\beta_2^2|a]\right]}{\sqrt{E_a\left[E[\alpha_1^2|a]E[\beta_1^2|a]\right]E_a\left[E[\alpha_2^2|a]E[\beta_2^2|a]\right]}}.$$

References

1 Sivakumaran, S. *et al.* Abundant pleiotropy in human complex diseases and traits. *The American Journal of Human Genetics* **89**, 607-618 (2011).

2 Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **101**, 5-22, doi:10.1016/j.ajhg.2017.06.005 (2017).

3 Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature genetics* **47**, 1236 (2015).

4 Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nature genetics* **48**, 709 (2016).

5 Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature genetics* **50**, 693 (2018).

6 Cotsapas, C. *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS genetics* **7**, e1002254 (2011).

7 Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics* **14**, 483, doi:10.1038/nrg3461 (2013).

8 Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757, doi:10.1126/science.aap8757 (2018).

9 Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet* **381**, 1371-1379 (2013).

10 Frei, O. *et al.* Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *bioRxiv*, 240275, doi:10.1101/240275 (2018).

11 Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature genetics* **50**, 229 (2018).

12 Qi, G. & Chatterjee, N. Heritability informed power optimization (HIPO) leads to enhanced detection of genetic associations across multiple traits. *PLoS genetics* **14**, e1007549 (2018).

13 Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature human behaviour*, 1 (2019).

14 Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human molecular genetics* **23**, R89-R98 (2014).

15 Smith, G. D. & Ebrahim, S. Mendelian randomization: prospects, potentials, and limitations. *International journal of epidemiology* **33**, 30-42 (2004).

16 O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *bioRxiv*, 205435 (2018).

17 Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186, doi:10.1016/j.cell.2017.05.038 (2017).

18 Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nature genetics* **48**, 245 (2016).

19 Nelson, C. P. *et al.* Genetically Determined Height and Coronary Artery Disease. *New England Journal of Medicine* **372**, 1608-1618, doi:10.1056/NEJMoa1404881 (2015).

20 Robinson, E. B. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nature Genetics* **48**, 552, doi:10.1038/ng.3529 (2016).

21 Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour*, doi:10.1038/s41562-019-0566-x (2019).

22 O'Connor, L. J. *et al.* Polygenicity of complex traits is explained by negative selection. *bioRxiv*, 420497 (2018).

23 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291 (2015).

24 Zeng, J. *et al.* Signatures of negative selection in the genetic architecture of human complex traits. *Nat Genet* **50**, 746-753, doi:10.1038/s41588-018-0101-4 (2018).

25 Schoech, A. P. *et al.* Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nature communications* **10**, 790 (2019).

26 Gazal, S. *et al.* Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nature genetics* **49**, 1421 (2017).

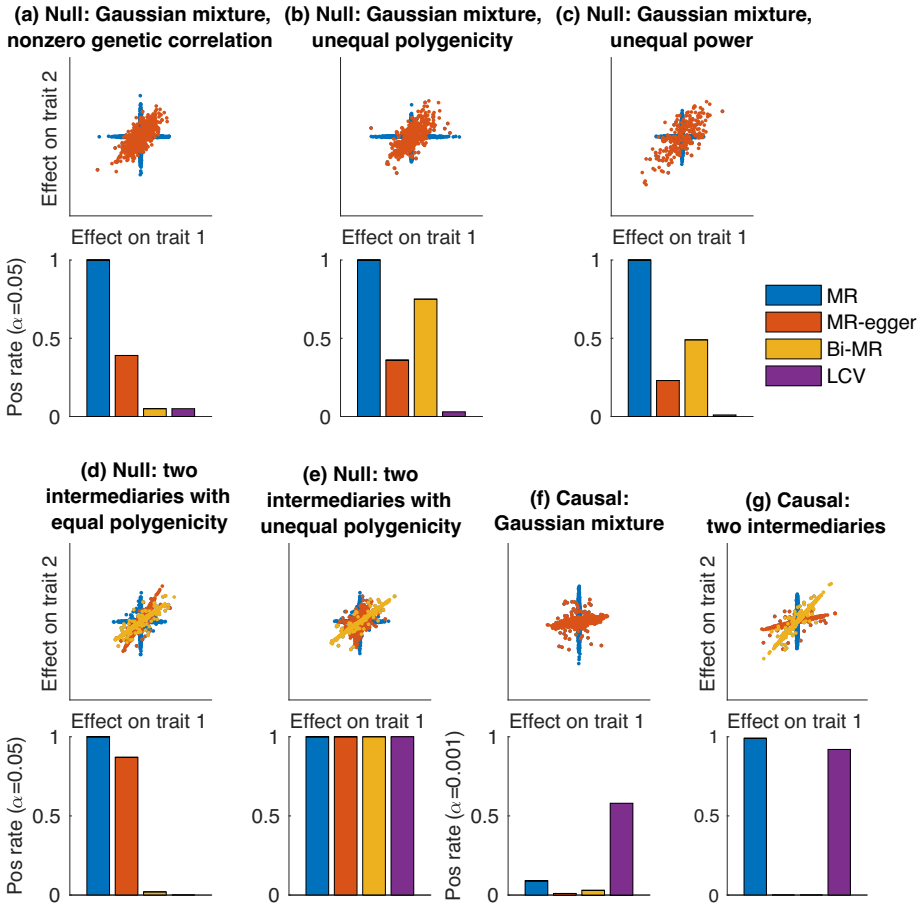
27 Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature Genetics* **50**, 621-629, doi:10.1038/s41588-018-0081-4 (2018).

28 Elsenbruch, S. Abdominal pain in Irritable Bowel Syndrome: a review of putative psychological, neural and neuro-immune mechanisms. *Brain, behavior, and immunity* **25**, 386-394 (2011).

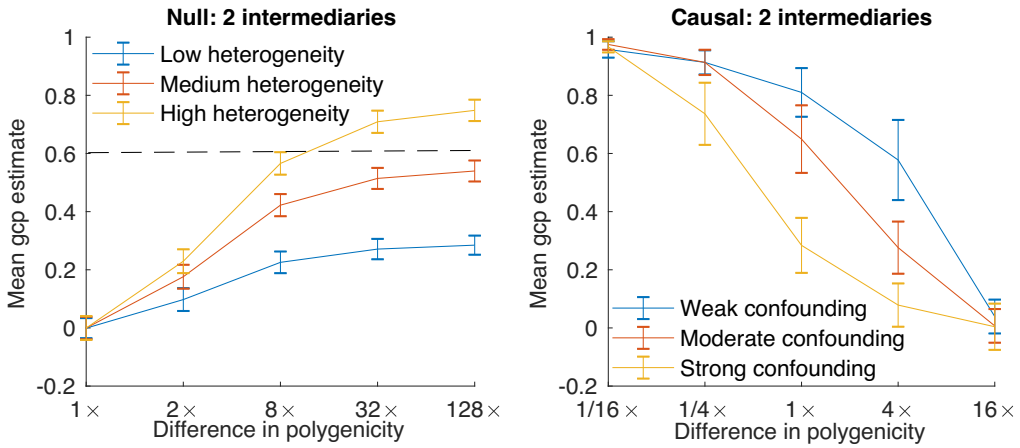
Conclusion

In Chapter I, I introduced a statistical method to quantify causal relationships between complex traits. In Chapter II, I introduced the concept of *flattening*, as well as a mathematical definition of polygenicity and a method to estimate it. In Chapter III, I categorized pleiotropic relationships and introduced a statistical method to quantify pleiotropy. Taken together, this research advances our understanding of the genetic architecture of complex traits.

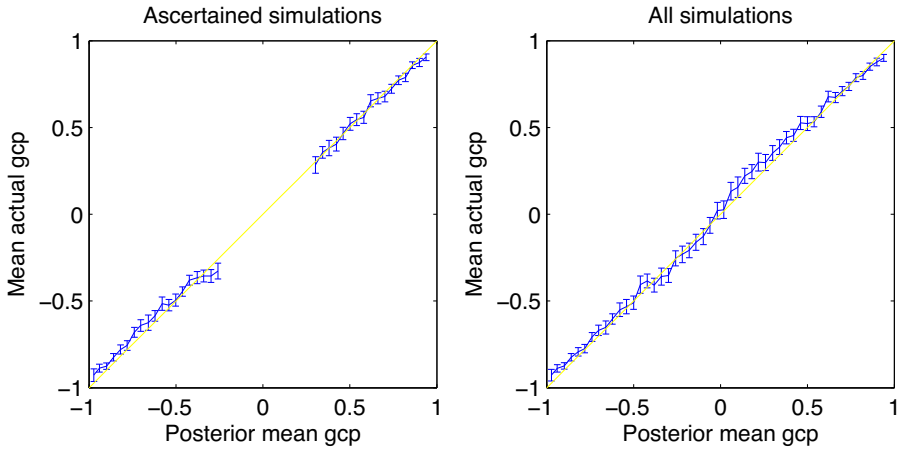
Appendix I. Supplementary material for Chapter II



Supplementary Figure 2.1 Null and causal simulations with no LD and LCV model violations. We report the positive rate ($\alpha = 0.05$ for null simulations, $\alpha = 0.001$ for causal simulations) for two-sample MR, MR-Egger, Bidirectional MR and LCV. Panels (a)-(c) correspond to Gaussian mixture model extensions of the models in Figure 2b-d. Panels (f) and (g) correspond to causal analogues of the models in panels (a) and (d), respectively. We also display scatterplots illustrating the bivariate distribution of true SNP effect sizes on the two traits. (a) Null simulation with nonzero SNP effects drawn from a mixture of Gaussian distributions; one mixture component has correlated effects on each trait. (b) Null simulation with SNP effects drawn from a mixture of Gaussian distributions, and differential polygenicity between the two traits. (c) Null simulation with SNP effects drawn from a mixture of Gaussian distributions, and unequal power between the two traits. (d) Null simulation with two intermediaries having different effects on each trait. (e) Null simulation with two intermediaries having different effects on each trait and unequal polygenicity for the two intermediaries. (f) Causal simulation with SNP effects drawn from a mixture of Gaussian distributions; all SNPs affecting trait 1 also affect trait 2, but the relative effect sizes were noisy. (g) Causal simulation with an additional genetic confounder (i.e. a second intermediary) mediating part of the genetic correlation. Results for each panel are based on 1000 simulations.



Supplementary Figure 2.2 Mean gcp estimates in simulations with LCV model violations. Error bars show standard deviations based on 1000 simulations. (a) Null simulation with two intermediaries having possibly unequal polygenicity. The two intermediaries had either a slightly, moderately, or highly heterogenous effect on the two traits; that is, when heterogeneity was high, intermediary 1 had a much larger effect on trait 1 while intermediary 2 had a much larger effect on trait 2. Then, we specified a certain difference in polygenicity between the two traits (measured by the proportion of causal SNPs). (b) Casual simulation with an additional latent confounder. The latent confounder explained either a low, medium or high proportion of the genetic correlation. We varied the polygenicity of the confounder and of the causal trait, such that a 16x difference in polygenicity indicates that 16x more SNPs were causal for the causal trait than for the genetic confounder.



Supplementary Figure 2.3 Unbiasedness of posterior mean gcp estimates in simulations with LD and random true gcp values. Estimated values of gcp were binned and averaged, and mean true values of gcp are plotted for each bin, with standard errors. Points above the line indicate that gcp estimates were downwardly biased (toward -1). (a) Ascertained simulations (43%) with significant genetic correlation ($p<0.05$) and evidence for partial causality ($p<0.001$). Only bins with count at least 10 are plotted. (b) All 10k simulations.

Supplementary Table 2.1 (see excel file) Results of LCV and MR on 429 trait pairs with a nominally significant genetic correlation.

Appendix II. Supplementary material for Chapter III

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 Methods section). Third, we provide a concise definition that provides less intuition.

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

$$M_e = \frac{3M}{\kappa_e}, \quad \kappa_e = \frac{3\mathrm{E}(\alpha^2\beta^2) - 2\mathrm{E}(\beta^4)}{\mathrm{E}(\alpha\beta)^2}. \tag{6}$$

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

$$M_e{}' = \frac{3M}{\kappa_e{}'}, \qquad \kappa_e{}' = \frac{\mathrm{E}(\beta^4)}{\mathrm{E}(\beta^2)^2}. \tag{7}$$

The difference between M_e and $M_e{}'$ is analogous to the difference between heritability $\mathrm{E}(\boldsymbol{\beta}^T R\boldsymbol{\beta})$ and an alternative definition of heritability that does not account for LD between variants, $\mathrm{E}(\boldsymbol{\beta}^T\boldsymbol{\beta})$. In the presence of perfect LD, $M_e{}'$ and $E(\boldsymbol{\beta}^T\boldsymbol{\beta})$ 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:

$$\boldsymbol{\beta} \sim \mathrm{N}(0, \Sigma) \tag{8}$$

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 $\kappa_e{}'$ 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:

$$M_e = \frac{h^2}{\mathrm{E}_{h^2}(\alpha^2)}. \tag{9}$$

$M_e{}'$ can be defined in a similar manner:

$$M_e{}' = \frac{h^2}{\mathrm{E}_{h^2}(\beta^2)}\ . \tag{10}$$

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 $\mathrm{E}(\boldsymbol{\beta}^T R \boldsymbol{\beta}) = \mathrm{E}(\boldsymbol{\beta}^T \boldsymbol{\beta})$[8,11].) Let $S =$

$\Sigma^{\frac{1}{2}} R \Sigma^{\frac{1}{2}}$. 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:

$$M_e = \frac{\mathrm{Tr}(S)^2}{\mathrm{Tr}(S^2)}. \quad (11)$$

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,

$$\mathrm{P}_{h^2}(\alpha^2 > T) \leq \frac{\mathrm{E}_{h^2}(\alpha^2)}{T}. \quad (12)$$

This bound is relatively tight when $T \gg \mathrm{E}_{h^2}(\alpha^2)$. When T is a significance threshold (for example, $T = \frac{30}{N}$ 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:

$$\mathrm{E}(r^2) = h^2 \mathrm{E}_{h^2}\left(\frac{\alpha^2}{\alpha^2 + \frac{1}{N}}\right). \quad (13)$$

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:

$$\mathrm{E}(r^2) \approx N h^2 \mathrm{E}_{h^2}(\alpha^2). \quad (14)$$

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[41] (M_{eff}). M_{eff} is defined as the number of SNPs divided by the average LD score, or in notation similar to equation (11), as

$$M_{eff} = \frac{\mathrm{Tr}(R)^2}{\mathrm{Tr}(R^2)}. \quad (15)$$

We note that M_e might be close to M_{eff} 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_{eff} even if only 1 SNP per LD block is causal. We also note that the value of M_e is strongly dependent on the range of allele frequencies that is specified, as rare SNPs (which have less LD)

contribute strongly to M_{eff}. 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, $\mathrm{E}(\boldsymbol{\beta}^T R \boldsymbol{\beta}) = \mathrm{Tr}(\mathrm{R}\Sigma)$. It is not shared by $M_e{}'$ or by the LD-independent definition of heritability, $E(\boldsymbol{\beta}^T \boldsymbol{\beta}) = \mathrm{Tr}(\Sigma)$.

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[16]) and LD fourth moment, respectively, for a randomly chosen SNP:

$$\ell_i^{(p)} = \sum_j r_{ij}^p \tag{16}$$

The regression equation is:

$$\mathrm{E}\left(\alpha^4 | \ell^{(2)}, \ell^{(4)}\right) \approx 3\mathrm{E}\left(\alpha^2 | \ell^{(2)}, \ell^{(4)}\right)^2 + \ell^{(4)} K. \tag{17}$$

In the first term, $\mathrm{E}\left(\alpha^2 | \ell^{(2)}, \ell^{(4)}\right) = \ell^{(2)}\tau$, where the coefficient τ is the variance of causal effect sizes[11,16]. In the second term, K is related to M_e:

$$K = 3\mathrm{E}(\beta^2)[\mathrm{E}_{h^2}(\alpha^2) - \mathrm{E}(\alpha^2)]. \tag{18}$$

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

In the case that there are P functional annotations, $\ell_{\mathrm{i}}^{(4)}$ and K are vectors of size $1 \times \mathrm{P}$ and $\mathrm{P} \times 1$ 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*:

$$\mathrm{cov}\left(\alpha_i^2,\beta_j^2\right) \approx r_{ij}^2 \mathrm{var}\left(\beta_j^2\right). \quad (19)$$

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 hand side but not the right hand side. We weaken (19) to obtain our LD approximation:

$$\mathrm{cov}\left(\alpha_i^2, \beta_j^2\right) \approx r_{ij}^2 \mathrm{cov}\left(\alpha_j^2, \beta_j^2\right). \quad (20)$$

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 (19) implies equation (20) by applying equation (19) to the case $i = j$. Equation (20) 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 (20) 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 (Supplementary Figure 3.2).

Equations (19-20) have second-moment analogues that can be used to justify LD score regression[11,16]. LDSC has been justified by assuming that correlated SNPs have uncorrelated effect sizes; this assumption can be stated as:

$$\mathrm{E}\left(\alpha_i \beta_j\right) = r_{ij} \mathrm{E}\left(\beta_j^2\right), \quad (21)$$

which is analogous to (19). 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 (20):

$$\mathrm{E}\left(\alpha_i \beta_j\right) \approx r_{ij} \mathrm{E}\left(\alpha_j \beta_j\right). \quad (22)$$

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

$$\mathrm{E}(\alpha_i^2) = \sum_j r_{ij}\mathrm{E}\big(\alpha_i\beta_j\big) \approx \sum_j r_{ij}^2\mathrm{E}\big(\alpha_j\beta_j\big) = \ell_i^{(2)}E(\alpha\beta). \quad (23)$$

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

$$E(\hat{\alpha}^2|\alpha) = \alpha^2 + \frac{1}{N}, E(\hat{\alpha}^4|\alpha) = \alpha^4 + 6\,\alpha^2/N + 3/N^2\,, \quad (24)$$

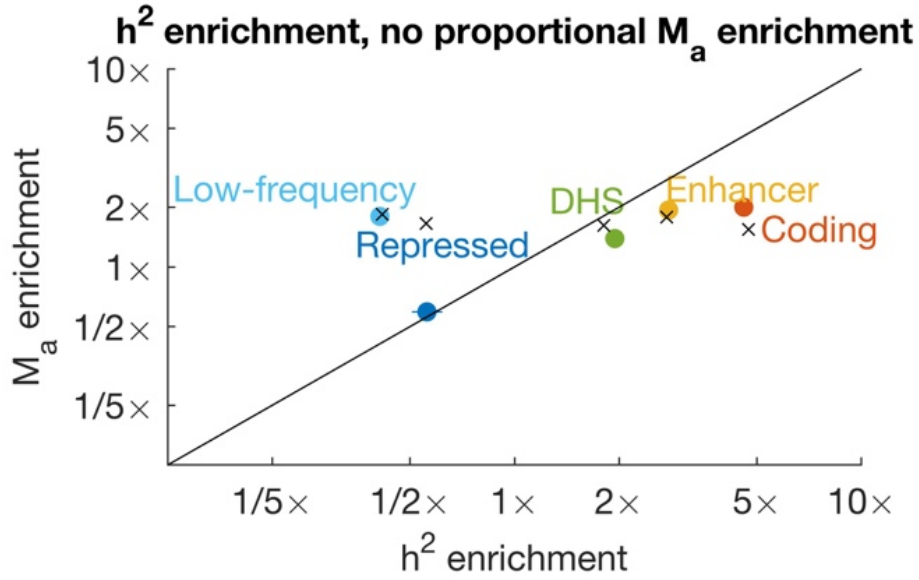
where in practice we use the LD score regression intercept divided by N, denoted $1/\widehat{N}$, instead of $1/N$. We caution that this equation assumes that sampling error follows a normal distribution, which may be false for rare SNPs. In the presence of population stratification, it may approximately 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[11] (see below) to estimate $E(\alpha_i^2)$ and (β_i^2) for each SNP, conditional on their respective LD scores and annotation values. Second, we regress the $M \times 1$ vector

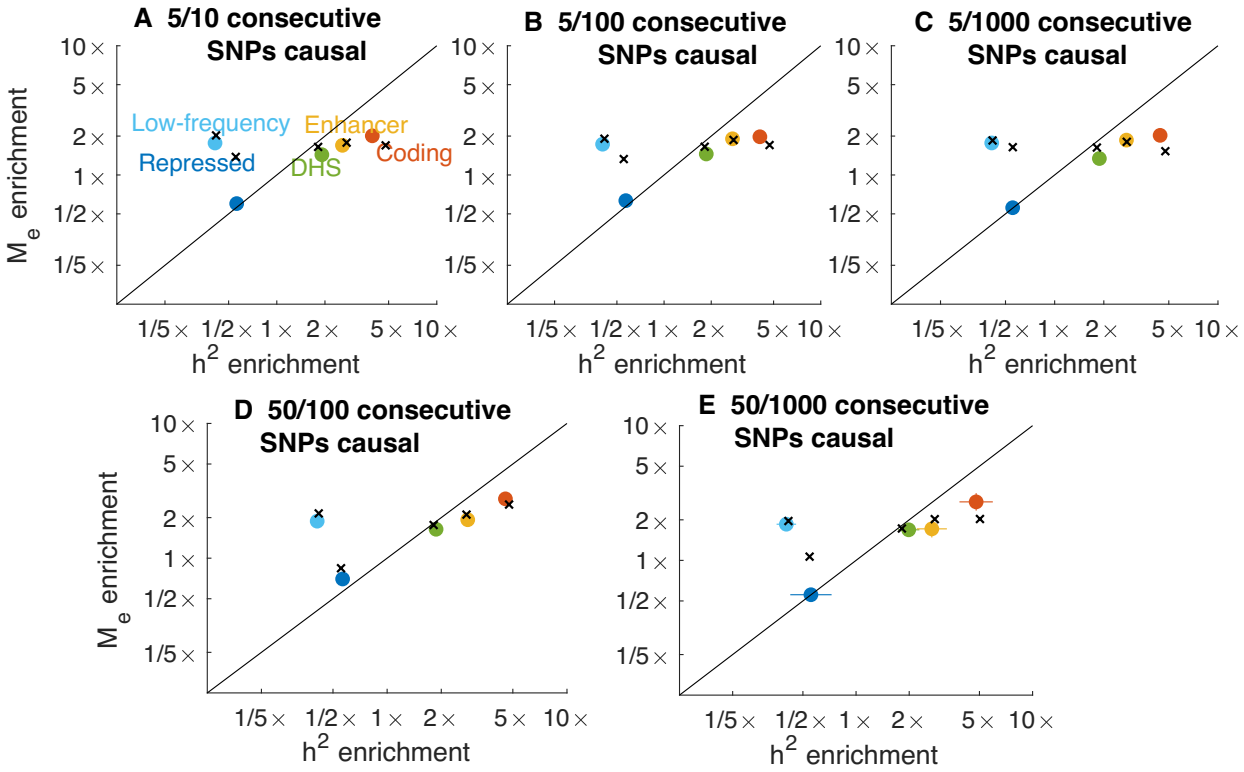
$$\widehat{\boldsymbol{\alpha}}^4 - \frac{6\widehat{\boldsymbol{\alpha}}^2}{N} + \frac{3}{N^2} - 3\big(\boldsymbol{\ell}^{(2)}\tau\big)^2, \quad (25)$$

whose expectation is $\boldsymbol{\alpha}^4 - 3E\left(\boldsymbol{\alpha}^2 \middle| \boldsymbol{\ell}^{(2)}\right)^2$, 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. For each annotation A_p, we add $3E\left(E(\alpha^2)E(\beta^2)|A_p\right)$ to K_p and estimate M_e for each category.

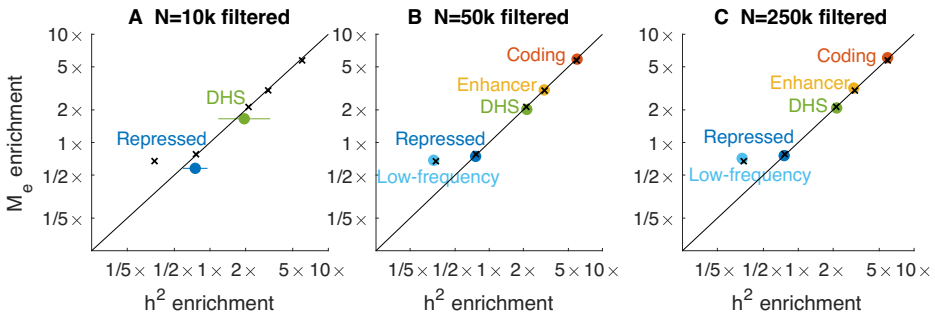
Our modified version of S-LDSC uses a slightly modified weighting scheme and does not exclude large-effect SNPs. The regression weight of each SNP is 1 divided by the LD score for that SNP to all common and low-frequency SNPs; this choice is similar to the original version of S-LDSC[11], but slightly 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.)



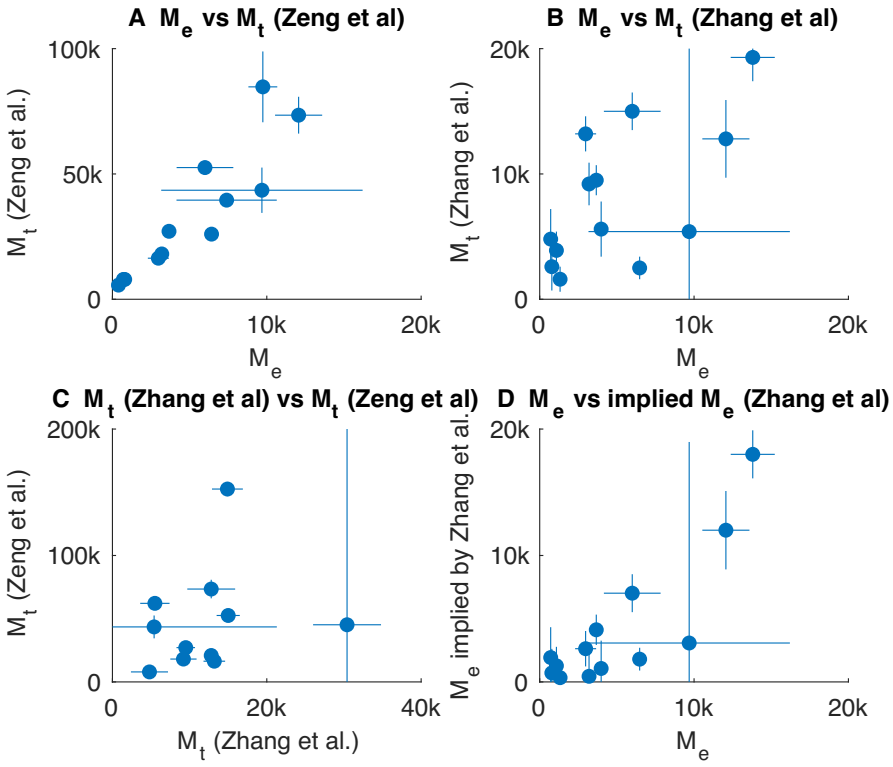
Supplementary Figure 3.1 Simulations with heritability enrichment and no proportional polygenicity enrichment. Due to some sparsity being driven by differences in heritability enrichment across categories, there is lower sparsity (polygenicity enrichment) within each individual category than within their union. We observed strong downward bias for the repressed category, which is depleted for heritability; we hypothesize that this bias is the result of imperfect resolution to distinguish LD to this category from LD to nearby SNPs, which leads to inflated estimated fourth moments because the nearby SNPs have larger causal effect sizes. This bias has little effect on our estimates for categories depleted for heritability because the nearby SNPs have smaller causal effect sizes than the SNPs in the category, and therefore very little effect on fourth moments; it also has little effect on our estimates for low-frequency SNPs because these SNPs are never in strong LD with common SNPs. In analyses of real traits, we do not report polygenicity enrichment estimates for categories that are depleted for heritability. Based on 1,000 simulations. Error bars indicate 95% confidence intervals. Numerical results are reported in Supplementary Table 3.1.



Supplementary Figure 3.2 Simulations with clustering of causal SNPs. The probability for a SNP to be causal was zero for most of the genome, and nonzero for contiguous blocks of SNPs of different sizes. (a) Blocks of 10 SNPs with a 50% chance of being causal. (b) Blocks of 100 SNPs with a 5% chance of being causal. (c) Blocks of 1,000 SNPs with a 0.5% chance of being causal. (d) Blocks of 100 SNPs with a 25% chance of being causal. Based on 1,000 simulations. Error bars indicate 95% confidence intervals. Numerical results are reported in Supplementary Table 3.1.

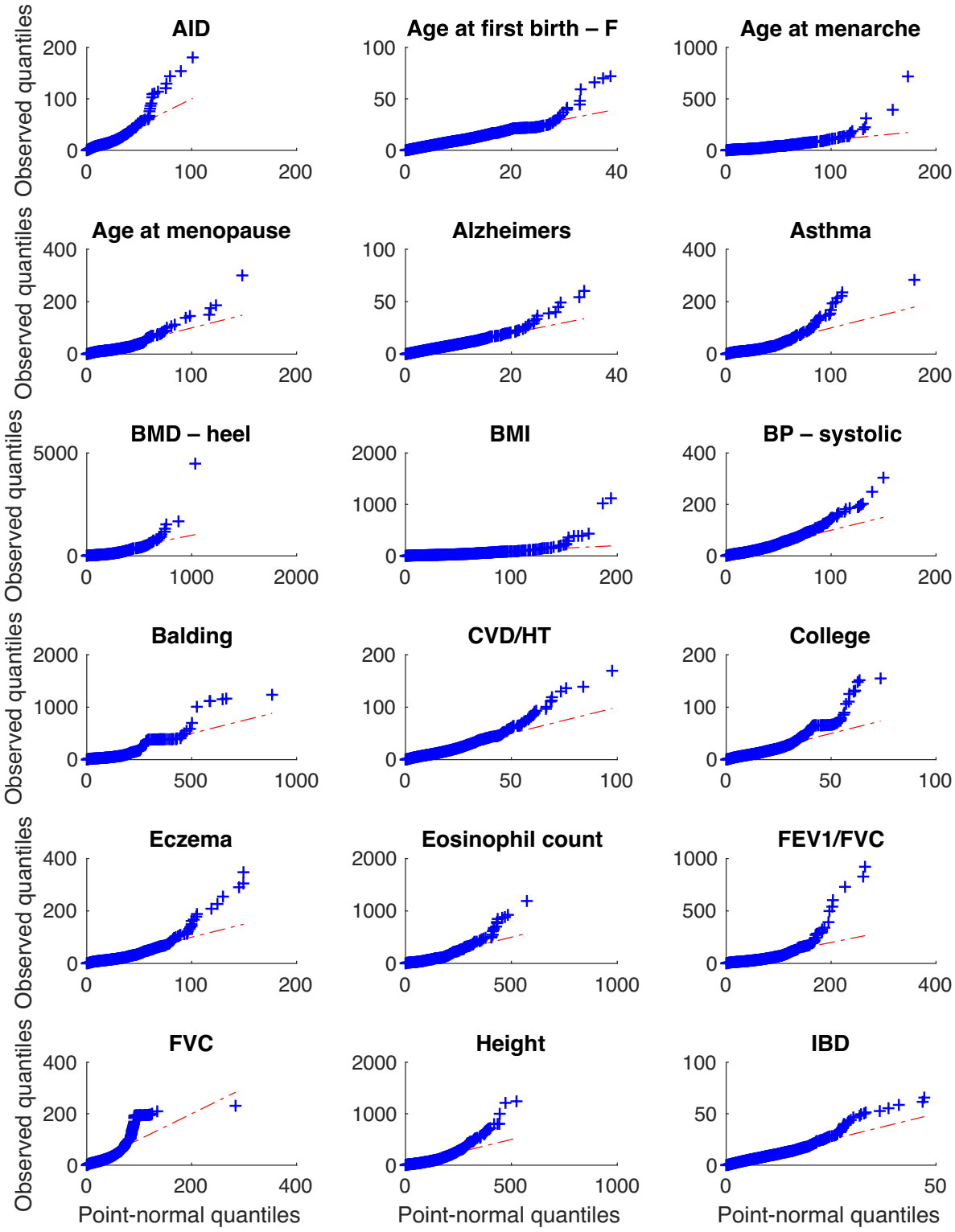


Supplementary Figure 3.3 Simulations with ascertainment based on power. For each annotation, simulation runs were discarded if heritability for that annotation was not significantly different from zero ($Z > 2$) or if the standard error of the M_e estimate for the annotation was larger than two times the median M_e point estimate. Results are not shown for annotations where the fraction of retained simulations was less than 1%, and confidence intervals are large for annotations where the fraction of retained simulations is low. See Supplementary Table 3.3 for the fraction of simulations that were retained in each case. Based on 1,000 simulations (before filtering). Error bars indicate 95% confidence intervals. Numerical results are reported in Supplementary Table 3.1.

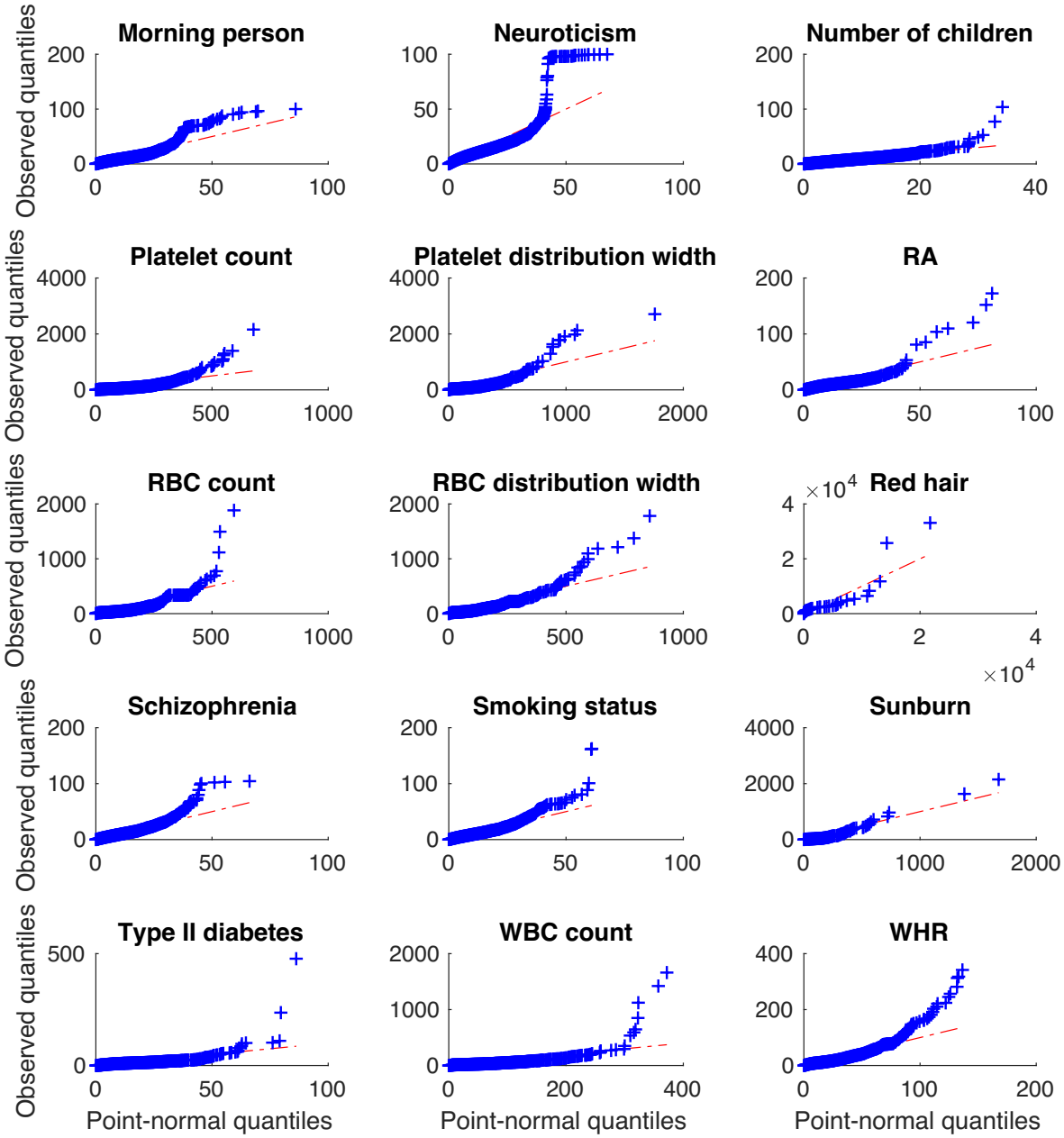


Supplementary Figure 3.4 Comparison of estimates of M_e and M_t. We compared our estimates of M_e, estimates of M_t from Zeng et al.[13], and estimates of M_t from Zhang et al.[26]. (a) Our estimates of M_e were highly correlated with estimates of M_t from Zeng et al. ($r = 0.90$), but $\sim 4\times$ smaller. This difference could be due to the different definition of polygenicity, or due to LD-related upwards bias in M_t estimates[13]. (b) Our estimates of M_e were correlated with estimates of M_t from Zhang et al. ($r = 0.62$). These estimates were derived under either a point-normal or a point-normal-model model (depending on a model selection step). (c) Estimates of M_t were only modestly correlated between Zeng et al. and Zhang et al. ($r = 0.20$). The Zeng et al. estimates were about $4\times$ larger, and the Zhang et al. estimates had much larger standard errors. These differences may result from the different models used by the two studies (point-normal vs. point-normal-normal for most traits, respectively), from different sample sizes, or from the different statistical heuristics used to account for LD. (d) We computed the M_e values that would be implied under the estimated models of Zhang et al. (assuming no LD between causal SNPs), and we compared these implied values with our M_e estimates. These estimates were mostly concordant in magnitude and highly correlated ($r = 0.85$). We note that the M_e values that would be implied by the estimates of Zeng et al. are similar to their M_t estimates (because $M_e = M_t$ under a point-normal model), except for the effect of their allele frequency dependent variance parameter; we did not attempt to calculate this because it depends on the site frequency spectrum of the set of SNPs that was used in their study. Numerical results are reported in Supplementary Table 3.5.

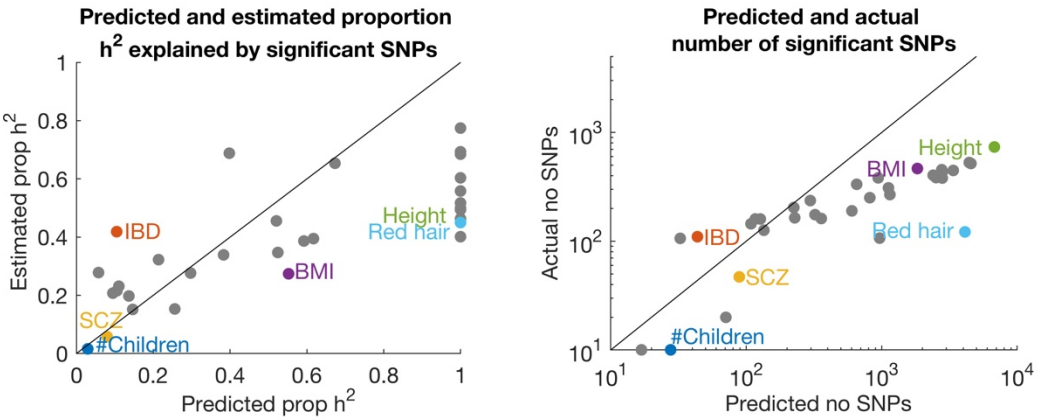
Supplementary Figure 3.5



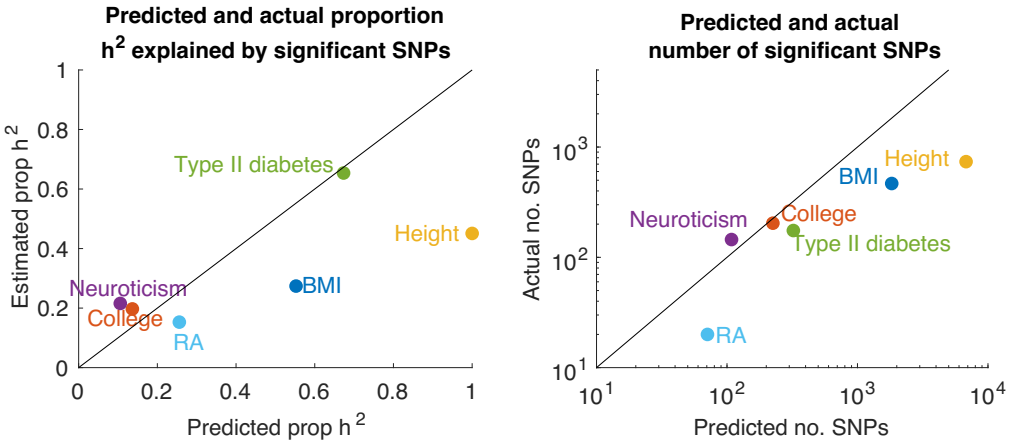
Supplementary Figure 3.5 (continued)



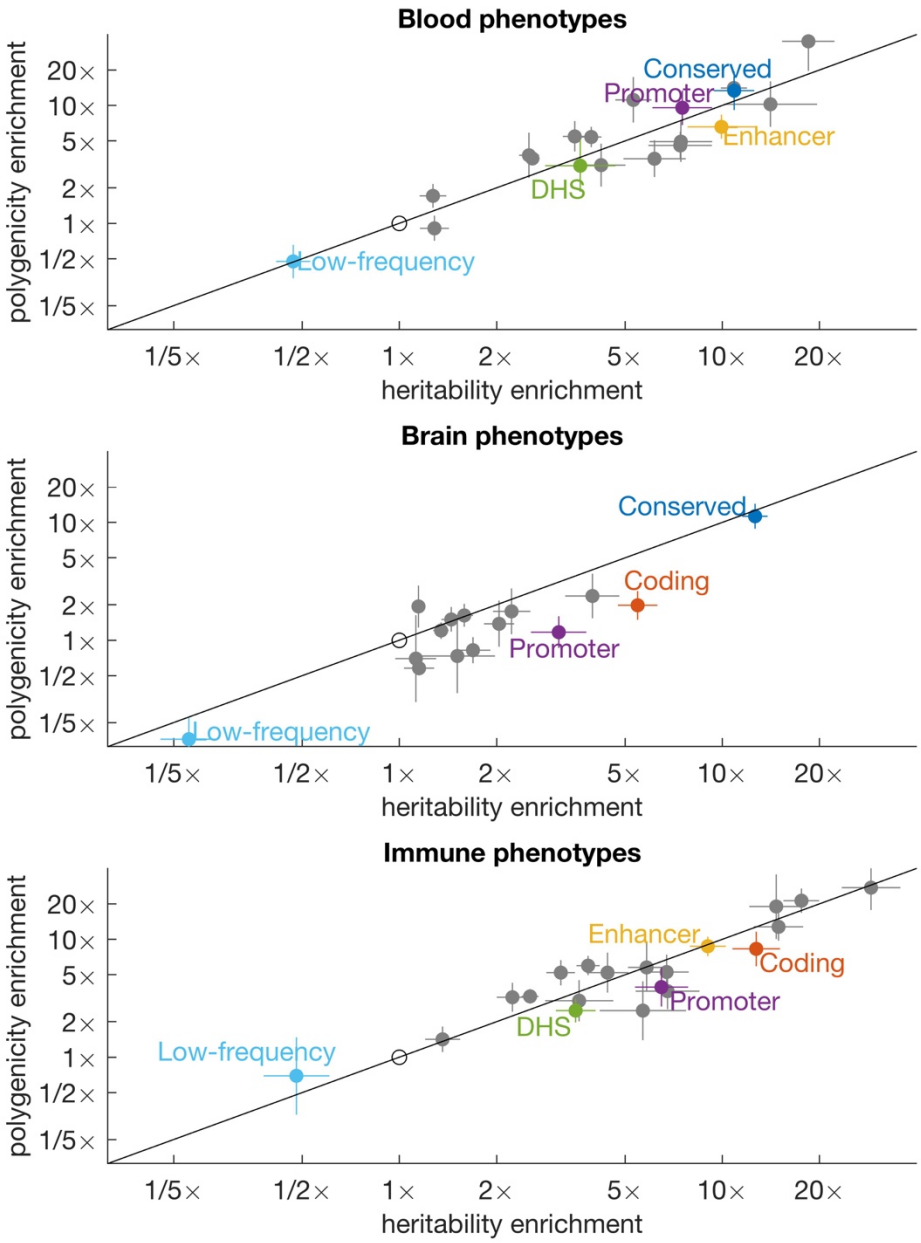
QQ plots comparing the observed distribution of χ^2 statistics vs. the expected distribution under a marginal point-normal model. The point-normal model was fit by matching the mean and variance of the model to the sample mean and variance for each trait. We assume that sampling noise follows a normal distribution with variance equal to the LD score regression intercept (model 1 in Supplementary Table 3.7). We did not perform any LD pruning or weighting. For most traits, the largest-effect SNPs consistently have larger effects than expected under the model, consistent with the observation that a point-normal-normal model usually fits better[26].



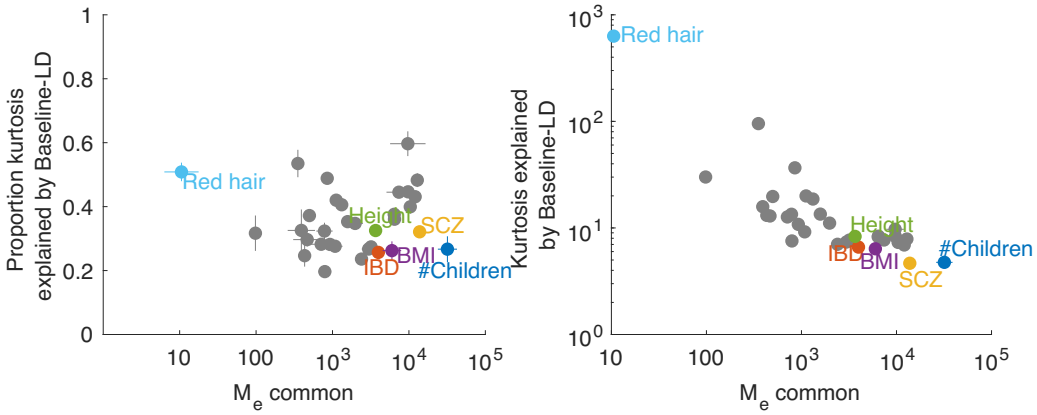
Supplementary Figure 3.6 Heritability explained by genome-wide significant SNPs. M_e gives a predicted upper bound on the proportion of heritability explained by significant SNPs, as well as on the number of significant SNPs (see Appendix, Properties of M_e). Significant SNPs $(\chi^2 > 30)$ were chosen using a greedy pruning procedure, iteratively selecting the most significant SNP that is at least 0.5cM from any previously-selected SNP. We estimated the proportion of heritability explained by these SNPs as the sum of their estimated marginal effect size magnitudes. We caution that this estimate may be upwardly biased due to winner's curse and due to subtle LD between selected SNPs.



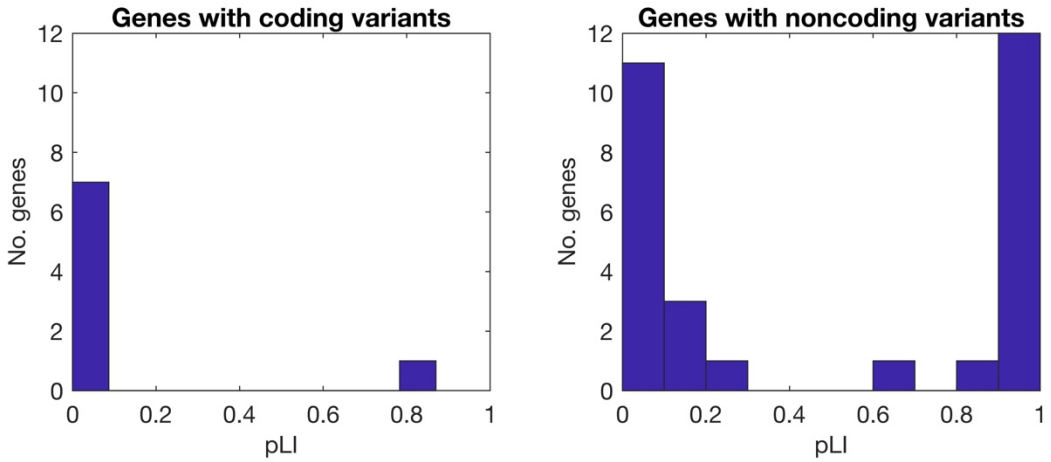
Supplementary Figure 3.7 Heritability explained by genome-wide significant SNPs, predicted using polygenicity estimates from independent cohorts. M_e gives a predicted upper bound on the proportion of heritability explained by significant SNPs, as well as on the number of significant SNPs (see Supplementary Figure 3.6 caption and Appendix). For phenotypes with summary statistics from independent cohorts (see Supplementary Table 3.8), we computed this bound based on the polygenicity estimated from the older study, the observed-scale heritability estimated from the newer study, and the sample size of the newer study. (We note that if the heritability of the older study were used instead, then unequal observed-scale heritability between the studies would lead to poor estimates).



Supplementary Figure 3.8 Heritability and polygenicity enrichment meta-analyzed across related traits. There were 6 blood-related phenotypes (eosinophil count, platelet count, platelet distribution width, RBC count, RBS distribution width, and WBC count); 8 brain-related phenotypes (Alzheimer's, BMI, college, morning person, neuroticism, smoking status, number of children and schizophrenia); and 5 immune-related phenotypes (all autoimmune, asthma, eczema, IBD, and RA). Results for each annotation are meta-analyzed across well-powered traits within each group, and each annotation is plotted if at least three traits had a well-powered polygenicity estimates for that annotation.



Supplementary Figure 3.9 Common-variant excess kurtosis explained by baseline-LD model. Excess kurtosis is defined as M_{eff}/M_e; in the case of no LD, this is equal to one-third the kurtosis of the causal effect size distribution, which is one when causal effects follow a normal distribution). Some excess kurtosis is expected due to differences in per-SNP heritability across categories; S-LDSC was used to estimate the causal effect size of each regression SNP, and these estimates were multiplied by the observed χ^2 statistics to estimate the sparsity explained by the model. (a) The proportion of excess kurtosis explained was defined as the ratio of logs (log-excess kurtosis explained divided by log-total excess kurtosis). It ranged between 20% and 50% for most traits, and there was no clear tendency for traits with greater polygenicity to have greater or smaller percent sparsity explained. (b) The amount of excess kurtosis explained was strongly negatively correlated with polygenicity (and positively correlated with total excess kurtosis).



Supplementary Figure 3.10 Distribution of pLI values for IBD genes with fine-mapped coding and noncoding variants. Causal variants for Crohn's disease and ulcerative colitis were fine mapped in ref. [57]. 47 fine-mapped variants with >50% posterior probability and 1-2 annotated protein-coding genes were selected. We caution that pLI values are computed based on allele frequencies of loss of function variants, so high-pLI genes will never have common LoF variants (although they may have common missense variants). There were 7 SNPs (all noncoding) with 2 annotated genes; pLI values were averaged for these loci. There were 3 genes with multiple causal variants (ranging from 2-6), including both coding and noncoding variants; these were counted only once, as coding genes. pLI values were significantly smaller for genes with coding variants than for genes with noncoding variants (single-tailed rank-sum test $p = 0.006$). Results for each SNP are reported in Supplementary Table 3.13.

Supplementary Table 3.1 (See Excel file) Numerical results of simulations. Sheets (a)-(f) correspond to Figure 3a-c and Supplementary Figures 1-3, respectively.

Supplementary Table 3.3 Ascertainment and standard errors of polygenicity enrichment in simulations at different sample sizes.

	$\mathbf{True\ log_{10} M_e}$	**Median estimated** $\mathbf{log_{10} M_e}$	**Normalized errors**
N=10k	3.22	3.30	0.53
N=50k	3.22	3.26	0.92
N=250k	3.22	3.26	1.01

Supplementary Table 3.2 Polygenicity estimates and jackknife standard errors in simulations at different sample sizes. We report true and estimated polygenicity for all common SNPs, as well as the standard deviation of the normalized errors. The normalized errors are defined as $log\, M_e\ - log\, \widehat{M_e}$ divided by the jackknife standard error. Their standard deviation is equal to one if the jackknife standard errors are perfectly calibrated, and less than 1 when the jackknife standard errors are conservative.

Supplementary Table 3.3 Ascertainment and standard errors of polygenicity enrichment in simulations at different sample sizes.

Category	Fraction retained			Normalized errors		
	N=10k	N=50k	N=250k	N=10k	N=50k	N=250k
Coding	0.007	0.39	0.5	NA	0.66	0.66
Enhancer	0.02	0.36	0.57	0.5	0.5	0.68
DHS	0.11	0.72	0.77	0.36	0.59	0.65
Repressed	0.24	0.69	0.81	0.43	0.61	0.64
Low-frequency	0.01	0.88	0.97	NA	0.7	0.81

For each annotation, simulation runs were discarded if heritability for that annotation was not significantly different from zero ($Z < 2$) or if the standard error of the M_e estimate for the annotation was larger than two times the median M_e point estimate. For each annotation, the fraction of simulation runs that were ascertained is reported, as well as the standard deviation of the normalized errors for the ascertained simulations (see Supplementary Table 3.2 caption). Jackknife standard errors for polygenicity enrichment are nearly proportional to standard errors for M_e, since there is relatively little noise in the denominator (the estimate of M_e for all common and LF SNPs).

Supplementary Table 3.4 (see Excel file) Datasets analyzed. 29 UK Biobank traits were selected to have low pairwise genetic correlations and high power, as measured by the significance of the S-LDSC heritability estimate; 28 of these were analyzed in ref.[20], and red hair pigmentation was added. Four additional diseases were selected on the basis of availability of summary statistics for low-frequency SNPs. Seven additional datasets were used for replication.

Supplementary Table 3.5 (see Excel file) Comparison of $\mathbf{M_e}$ estimates with and without the baseline-LD model. We used either the full baseline-LD model (as in Table 1), only the 10 common MAF bins, or no annotations (except the base annotation containing all common SNPs).

Supplementary Table 3.6 (see Excel file) Comparison of estimates of $\mathbf{M_e}$ and $\mathbf{M_t}$. We report estimates of $\mathrm{M_e}$, estimates of $\mathrm{M_t}$ from Zeng et al.[13] and estimates of $\mathrm{M_t}$ from Zhang et al.[26] We also report implied values of $\mathrm{M_e}$ based on the model (point-normal or point-normal-normal) fit by Zhang et al.

Supplementary Table 3.7 (see Excel file) Comparison of the observed distribution of χ^2 statistics with the expected distribution under a marginal point-normal model. We report the observed vs. expected normalized variance (variance divided by mean squared) and skewness of the χ^2 distribution for 33 traits (Supplementary Table 3.4). Two point-normal models are considered: first, we matched the mean and variance of the model to the sample mean and variance for each trait; second, we set the proportion of non-null SNPs equal to M_e/M_{eff}. Under both models, we assume that sampling noise follows a normal distribution with variance equal to the LD score regression intercept. We did not perform any LD pruning or weighting. Under the first model, the variance matches perfectly, but the skewness of the observed distribution is greater than expected under the model for most traits. This suggests that the largest-effect SNPs consistently have larger effects than expected under a marginal point normal model (see Supplementary Figure 3.5 for QQ plots); a more complex model, e.g. the point-normal-normal model of ref.[26], may fit better. The trait with the greatest difference between observed and expected skewness was BMI (49 observed vs 6.7 expected). This may explain why initial BMI GWAS found more associations than initial height GWAS (20 observed vs 8.7 expected for height), despite the fact that their overall polygenicity is similar (Table 1). Under the second model, the variance does not match perfectly; it was slightly smaller than expected for most traits (mean $\log_{10}$ fold difference -0.21). This difference is expected: a large LD block is more likely to be associated and represents a larger fraction of all SNPs than a small LD block, increasing the apparent polygenicity of the distribution of χ^2 statistics; in contrast, M_e does not count large LD blocks more heavily than small LD blocks.

Supplementary Table 3.8 Comparison of common-SNP $\mathbf{M_e}$ estimates for traits with multiple available datasets.

Trait	Reference	N_{tot}	$M_{\text{regression}}$	$\log_{10} M_{\text{e}}$	$M_{common} E[\beta^2]$
Height	UKBB	458k	10M	3.56(0.02)	0.38(0.01)
Height	Lango Allen et al. 2010	131k	1.0M	3.49(0.11)	0.27(0.01)
BMI	UKBB	458k	10M	3.78(0.12)	0.23(0.01)
BMI	Speloites et al. 2010	122k	1.0M	3.51(0.33)	0.17(0.01)
College	UKBB	455k	10M	4.08(0.05)	0.026(0.001)
Years of education	Rietveld et al. 2013	127k	1.0M	4.26(0.18)	0.12(0.01)
Years of education	Okbay et al. 2016	329k	1.0M	4.23(0.09)	0.16(0.01)
Type II Diabetes	UKBB	459k	10M	2.85(0.14)	0.0014(0.0001)
Type II Diabetes	Morris et al. 2012	61k	1.0M	2.85(0.38)	0.065(0.013)
Neuroticism	UKBB	372k	10M	3.99(0.23)	0.97(0.03)
Neuroticism	Okbay et al. 2016	171k	1.0M	4.26(0.10)	0.13(0.01)
Rheumatoid arthritis	Okada et al. 2014	104k	4.6M	3.03(0.10)	0.10(0.01)
Rheumatoid arthritis	UKBB	459k	10M	3.08(0.20)	0.16(0.03)

Standard errors are also reported. No M_e estimates were significantly different ($p < 0.05$ assuming independent errors) for any pair of datasets. N_{tot}: total number of samples. $M_{regression}$: number of regression SNPs. We also report the number of common SNPs times the estimated effect-size variance estimated by S-LDSC[11] (slightly modified; see Appendix, Stratified LD fourth moments regression). While this quantity is equivalent to heritability under some scenarios, it is subject to biases that do not affect our estimates of M_e (or heritability enrichment or polygenicity enrichment), such as differences in prevalence or trait definition among data sets (particularly between the binary college attendance trait and the continuous years of education trait) and possible genomic control correction.

Supplementary Table 3.9 (See Excel file) Complete results of S-LD4M and S-LDSC on 33 traits. Results are reported for well-powered trait-annotation pairs (see Material and Methods).

Supplementary Table 3.10 Polygenicity and heritability enrichment for functional annotations, meta-analyzed across well-powered traits.

Category	Heritability enrichment	Polygenicity enrichment	Number of traits	Proportion of SNPs
Conserved (LindbladToh)	13.29(0.51)	14.23(1.28)	21	0.03
TSS (Hoffman)	9.61(1.00)	10.70(2.33)	10	0.02
Coding (UCSC)	9.38(0.64)	6.63(0.97)	16	0.02
Weak Enhancer (Hoffman)	8.56(1.21)	4.86(0.62)	9	0.02
Super Enhancer (Vahedi)	7.03(0.33)	8.32(0.83)	18	0.02
Typical Enhancer (Vahedi)	6.40(0.60)	4.20(0.40)	13	0.02
Enhancer (Andersson)	5.56(2.06)	1.39(1.12)	2	0
Promoter Flanking (Hoffman)	5.46(1.50)	0.98(0.54)	2	0.01
DGF (ENCODE)	5.27(0.42)	4.28(0.61)	18	0.15
Enhancer (Hoffman)	4.94(0.34)	3.22(0.30)	19	0.04
UTR 3 (UCSC)	4.89(0.41)	2.87(0.51)	11	0.01
Promoter (UCSC)	4.71(0.48)	2.57(0.50)	12	0.05
CTCF (Hoffman)	4.59(1.25)	0.61(0.27)	1	0.02
TFBS (ENCODE)	4.56(0.35)	3.23(0.30)	18	0.14
H3K9ac (Trynka)	4.02(0.15)	4.44(0.42)	21	0.14
H3K4me3 (Trynka)	3.62(0.16)	4.29(0.47)	23	0.14
Fetal DHS (Trynka)	3.42(0.24)	3.27(0.44)	14	0.09
DHS (Trynka)	3.31(0.22)	2.36(0.29)	15	0.18
Super Enhancer (Hnisz)	2.79(0.08)	3.47(0.26)	27	0.17
H3K27ac (PGC2)	2.53(0.09)	3.57(0.33)	26	0.28
H3K27ac (Hnisz)	2.06(0.04)	2.58(0.26)	28	0.4
Transcribed (Hoffman)	1.33(0.05)	1.06(0.14)	23	0.36
Intron (UCSC)	1.12(0.03)	1.53(0.22)	23	0.4
Low-frequency (UK10K)	0.40(0.02)	0.44(0.06)	15	0.3

The number of traits used in the meta-analysis is indicated; traits were excluded if the heritability estimate for a trait-annotation pair was not significantly different from zero, or if the standard error on the M_e estimate was greater than 4 times the median point estimate for that annotation across traits. Annotations were excluded if the number of remaining traits was less than 10 or if the meta-analyzed heritability enrichment estimate was less than 1 (except for the low-frequency category). Standard errors are also reported.

Supplementary Table 3.11 (see Excel file) Numerical results from Figure 6.

Supplementary Table 3.12 Polygenicity and heritability enrichment of ExAC genes.

	Enrichment	$\log_{10}$Enrichment
Polygenicity	1.82	0.27 (0.02)
Heritability	1.35	0.13(0.007)
Average unit of heritability	0.71	-0.15 (0.02)

Enrichments are reported for SNPs in and near ExAC LoF-intolerant genes[32] compared with SNPs near any gene (in and near is defined as the gene body plus or minus 50kb). The average unit of heritability is equal to the heritability divided by the polygenicity (Figure 2; see Material and Methods). Standard errors are also reported.

Supplementary Table 3.13 (see Excel file) Fine-mapped IBD genes from ref.[35] harboring coding and noncoding variants. Coding and noncoding SNPs with $> 50\%$ posterior probability and 1-2 nearby genes are listed. pLI values are averaged for SNPs with 2 genes. Genes with both coding and noncoding variants are included in Supplementary Figure 3.10a and not in panel b.