The American Journal of Human Genetics, Volume *105*

## Supplemental Data

# Extreme Polygenicity of Complex Traits

# Is Explained by Negative Selection

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

## Supplementary Figures

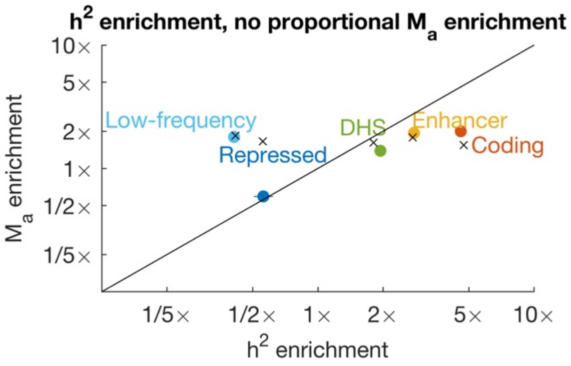


**Figure S1 Simulations with heritability enrichment and no proportional polygenicity enrichment.** Due to some sparsity being driven by differences in heritability enrichment across categories, there is lower sparsity (polygenicity enrichment) within each individual category than within their union. 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 Table S1.

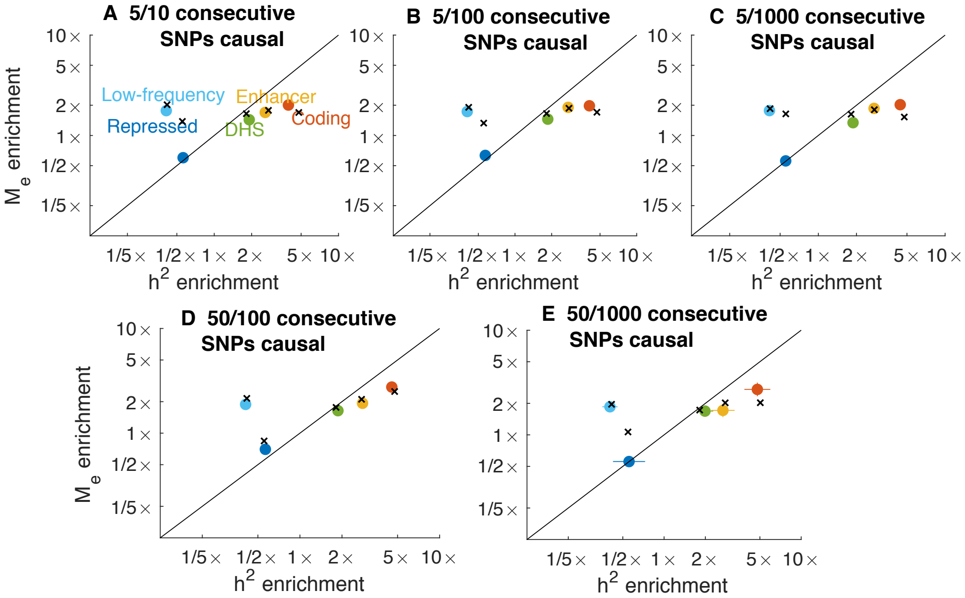


**Figure S2 Simulations with clustering of causal SNPs.** The probability for a SNP to be causal was zero for most of the genome, and nonzero for contiguous blocks of SNPs of different sizes. (A) Blocks of 10 SNPs with a 50% chance of being causal. (B) Blocks of 100 SNPs with a 5% chance of being causal. (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 Table S1.

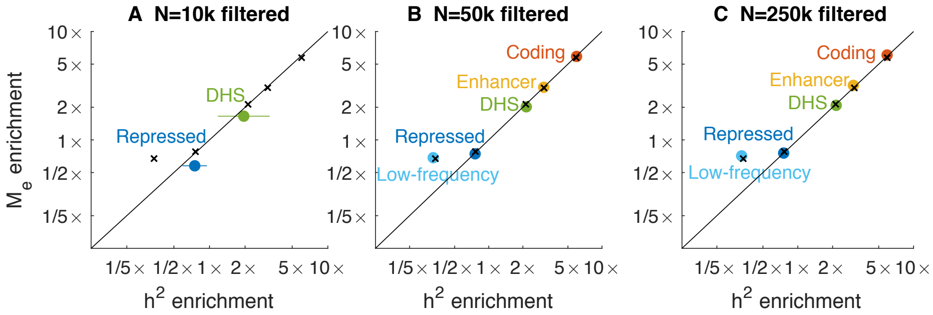


**Figure S3 Simulations with ascertainment based on power.** For each annotation, simulation runs were discarded if heritability for that annotation was not significantly different from zero ($Z > 2$) or if the standard error of the $M_e$ estimate for the annotation was larger than two times the median $M_e$ 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 Table S3 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 Table S1.

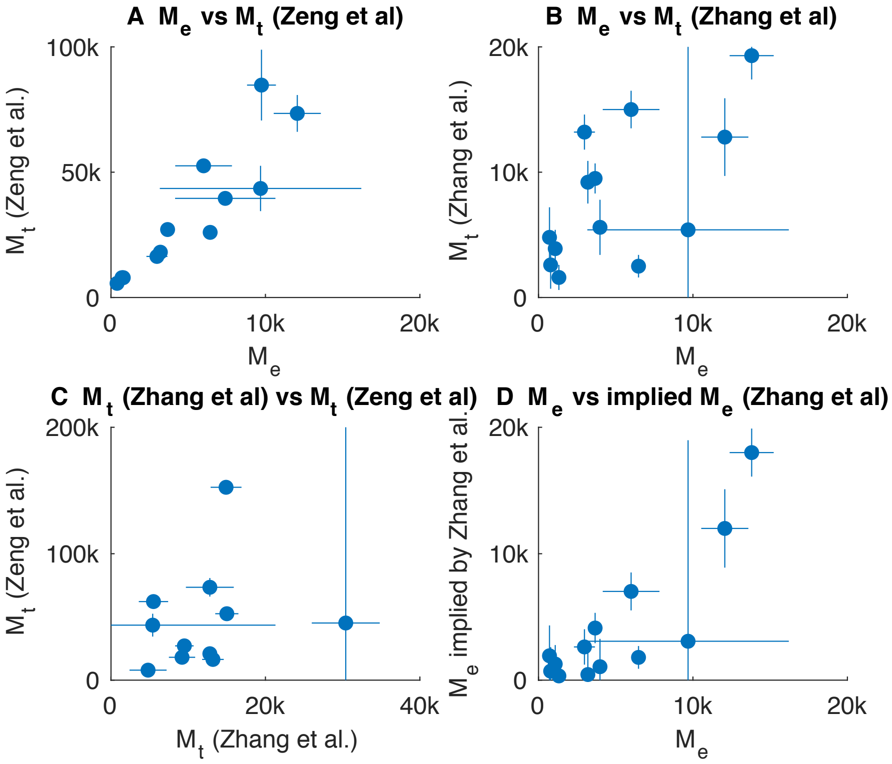
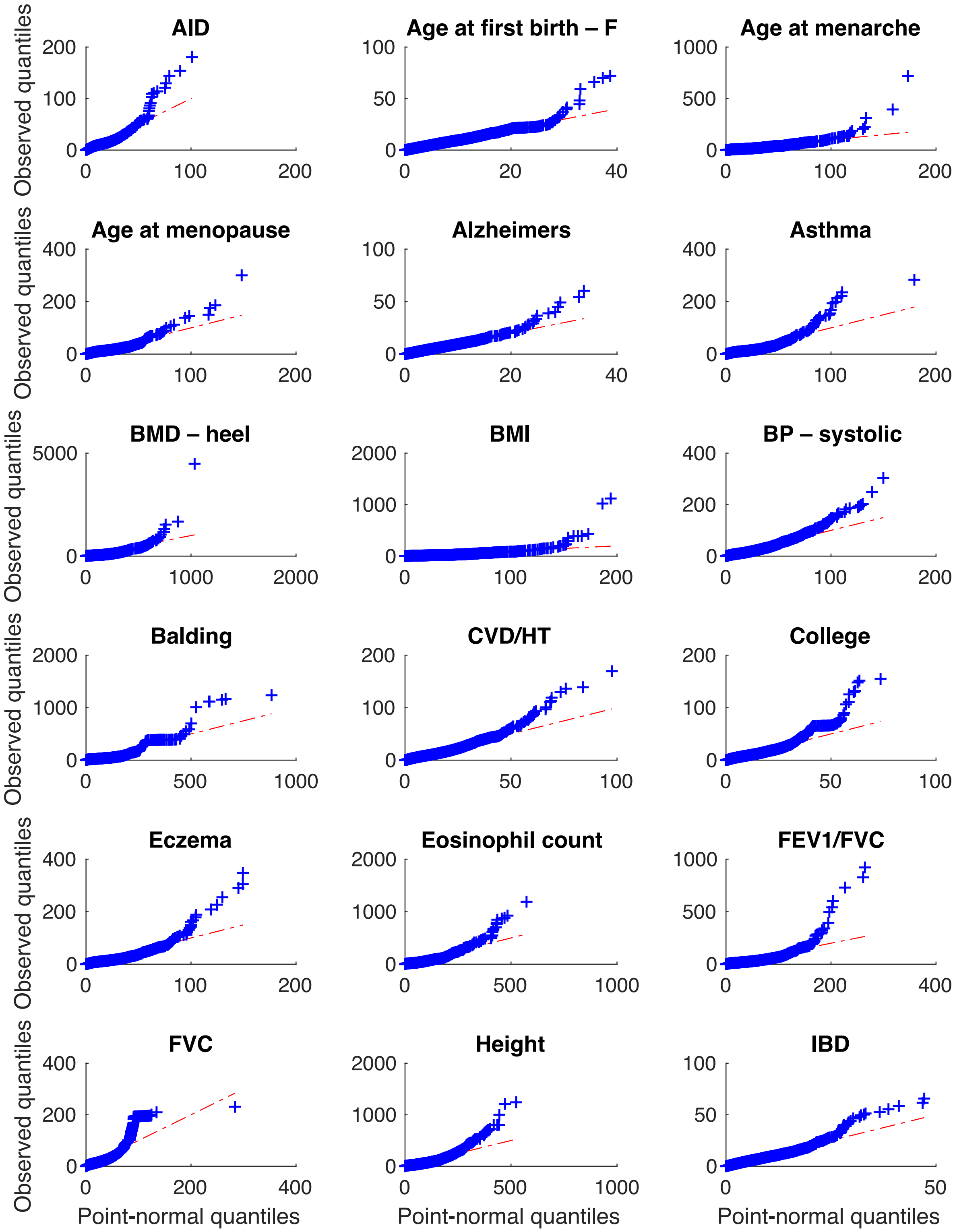


**Figure S4 Comparison of estimates of $M_e$ and $M_t$.** We compared our estimates of $M_e$, estimates of $M_t$ from Zeng et al.[11], and estimates of $M_t$ from Zhang et al.[12]. (A) Our estimates of $M_e$ were highly correlated with estimates of $M_t$ from Zeng et al. ($r = 0.90$), but $\sim 4 \times$ smaller. This difference could be due to the different definition of polygenicity, or due to LD-related upwards bias in $M_t$ estimates[11]. (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 Table S5.



AID
Age at first birth – F
Age at menarche
Age at menopause
Alzheimers
Asthma
BMD – heel
BMI
BP – systolic
Balding
CVD/HT
College
Eczema
Eosinophil count
FEV1/FVC
FVC
Height
IBD
Observed quantiles
Point-normal quantiles
0
50
100
200
300
400
500
1000
2000
5000
20
40

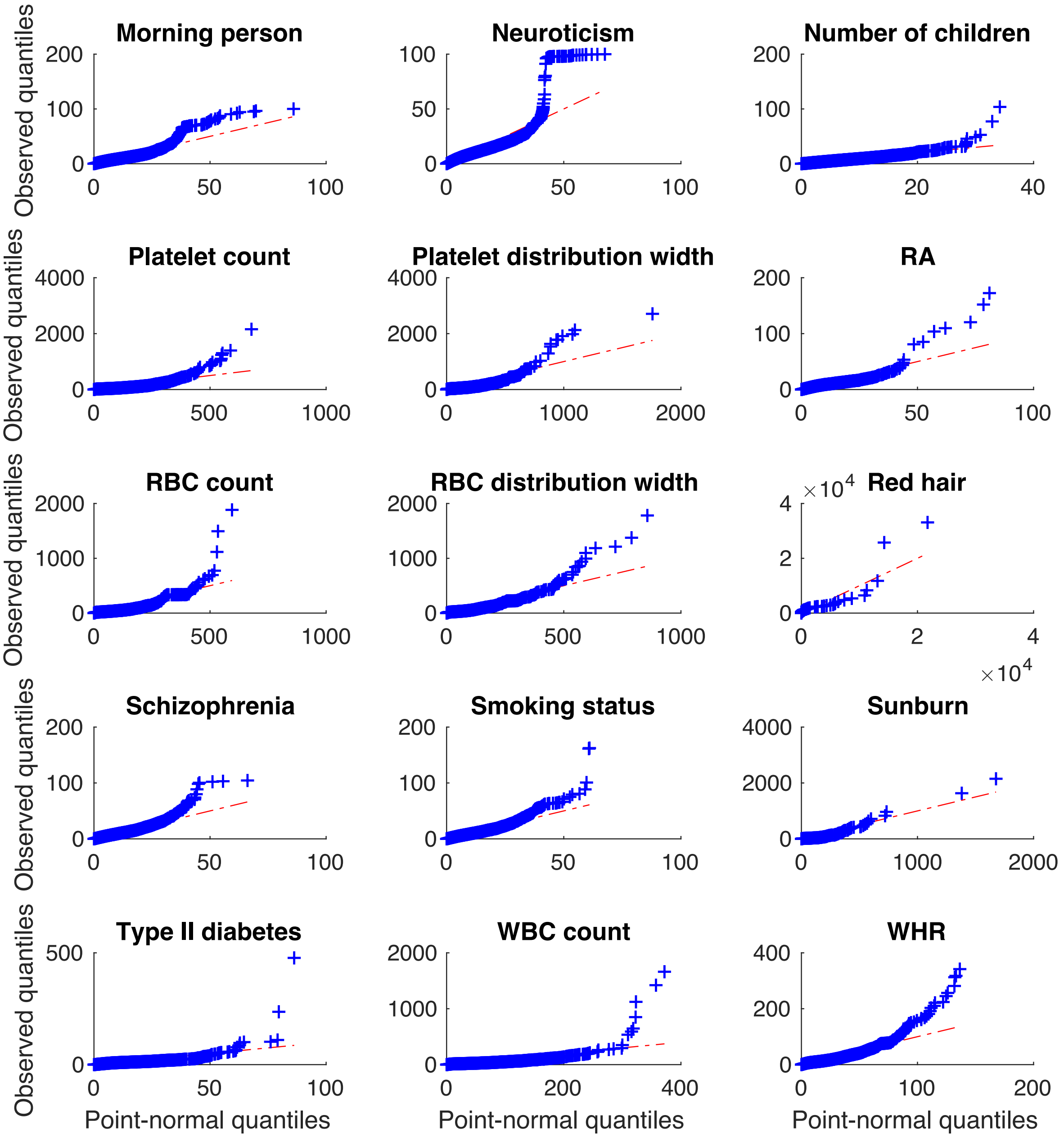


**Figure S5 QQ plots comparing the observed distribution of $\chi^2$ statistics vs. the expected distribution under a marginal point-normal model.** The point-normal model was fit by matching the mean and variance of the model to the sample mean and variance for each trait. We assume that sampling noise follows a normal distribution with variance equal to the LD score regression intercept (model 1 in Table S7). 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[12].

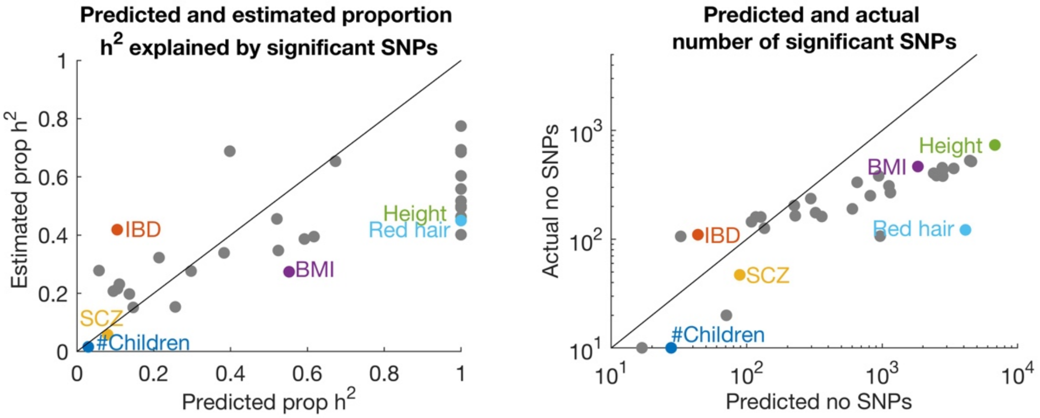


**Figure S6 Heritability explained by genome-wide significant SNPs.** $M_e$ gives a predicted upper bound on the proportion of heritability explained by significant SNPs, as well as on the number of significant SNPs (see Appendix, Properties of $M_e$). Significant SNPs $(\chi^2 > 30)$ were chosen using 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.

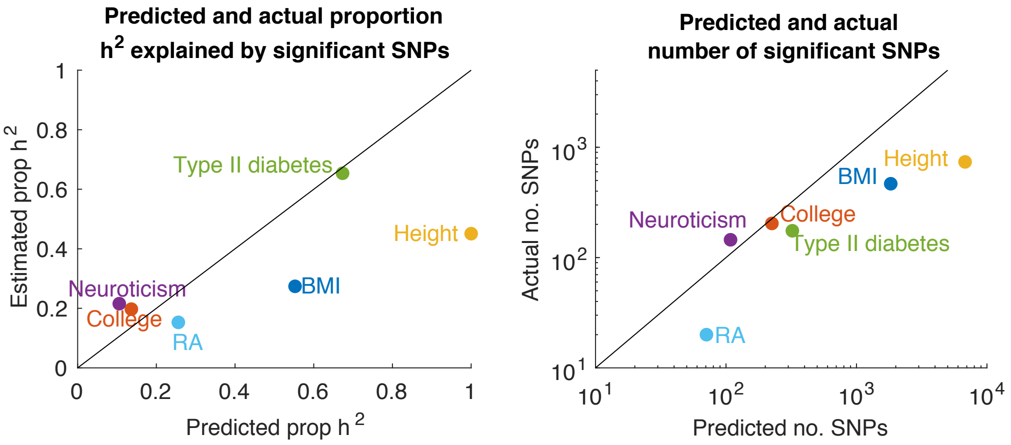


**Figure S7 Heritability explained by genome-wide significant SNPs, predicted using polygenicity estimates from independent cohorts.** $M_e$ gives a predicted upper bound on the proportion of heritability explained by significant SNPs, as well as on the number of significant SNPs (see Figure S6 caption and Appendix). For phenotypes with summary statistics from independent cohorts (see Table S8), 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).

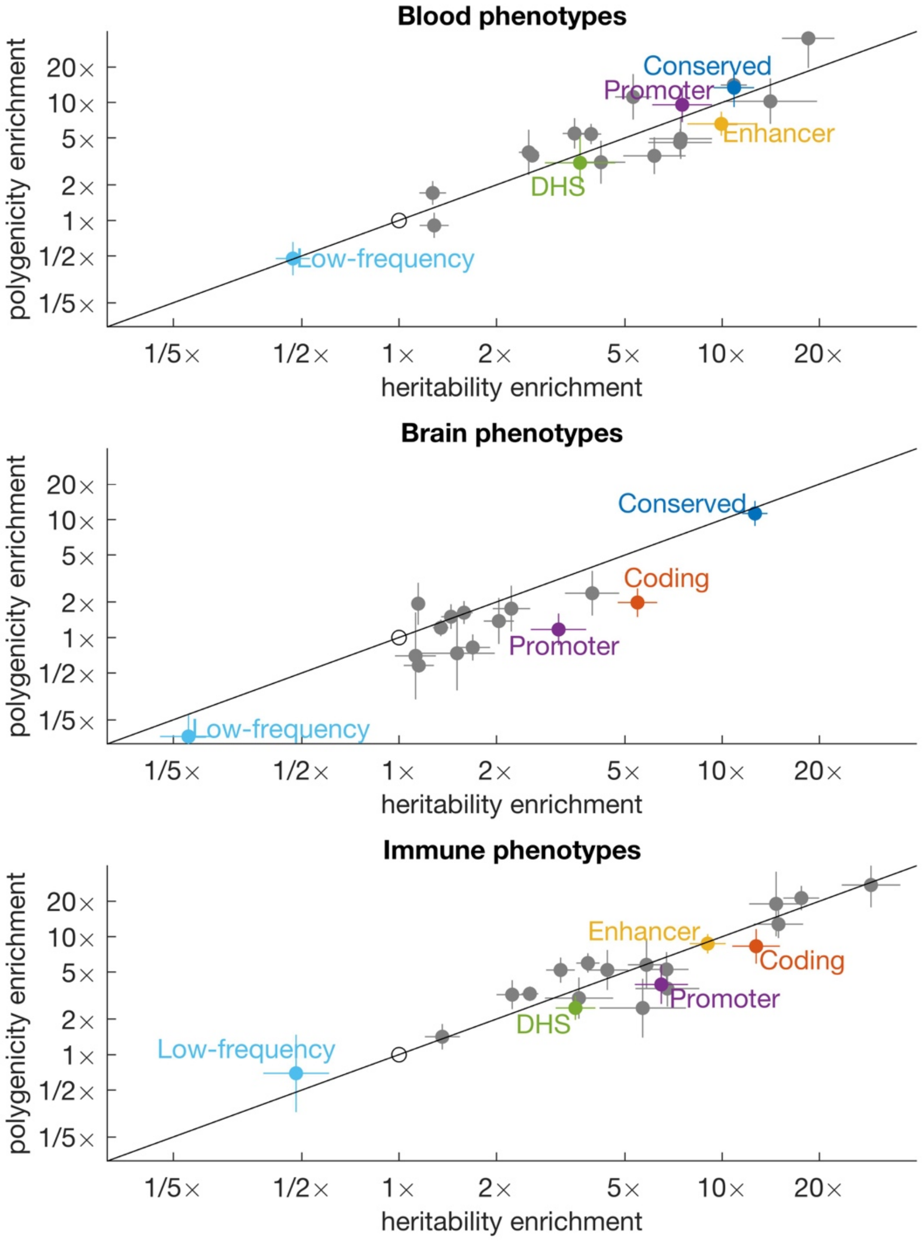


**Figure S8 Heritability and polygenicity enrichment meta-analyzed across related traits.** There were 6 blood-related phenotypes (eosinophil count, platelet count, platelet distribution width, RBC count, RBS distribution width, and WBC count); 8 brain-related phenotypes (Alzheimer's, BMI, college, morning 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.

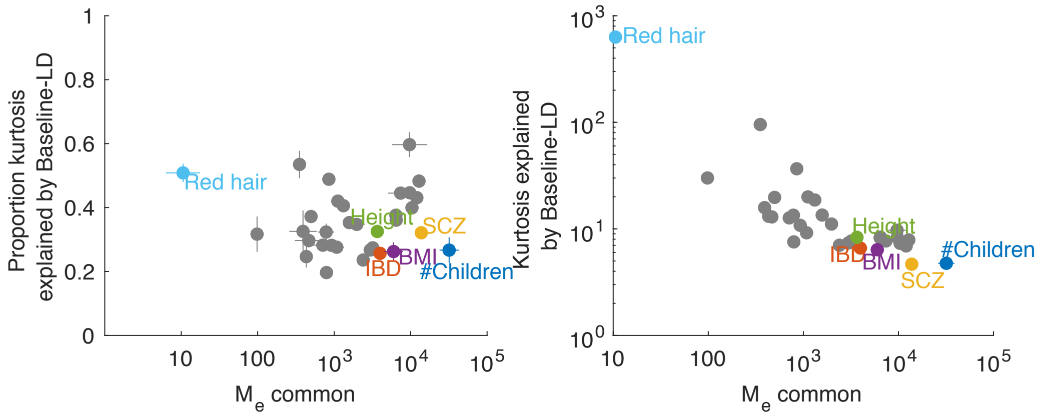


**Figure S9 Common-variant excess kurtosis explained by baseline-LD model.** Excess kurtosis is defined as $M_{eff}/M_e$; in the case of no LD, this is equal to one-third the kurtosis of the causal effect size distribution, which is one when causal effects follow a normal distribution). Some excess kurtosis 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 $\chi^2$ 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).

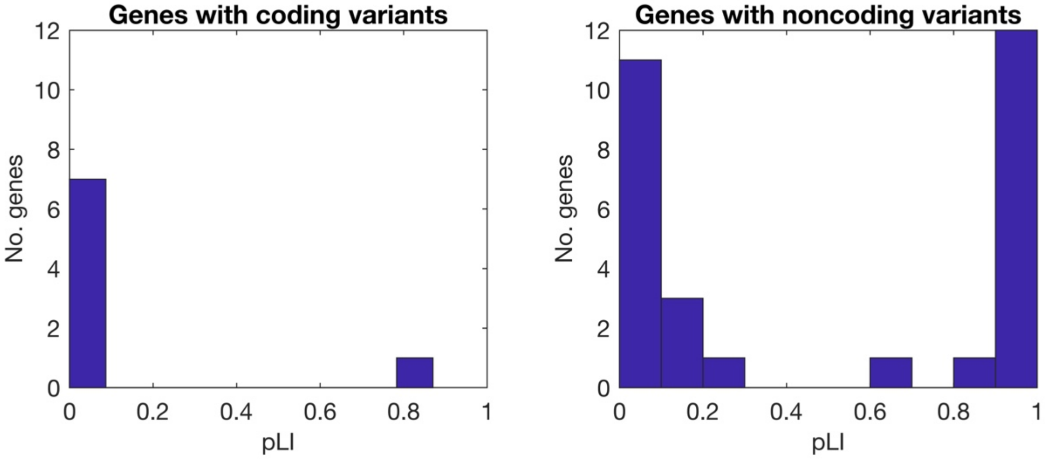


**Figure S10 Distribution of pLI values for IBD genes with fine-mapped coding and noncoding variants.** Causal variants for Crohn's disease and ulcerative colitis were fine mapped in ref. [48]. 47 fine-mapped variants with >50% posterior probability and 1-2 annotated protein-coding genes were selected. 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 Table S13.

## Supplementary Tables

**Table S1 (See Excel file) Numerical results of simulations.** Sheets (A)-(F) correspond to Figure 3A-C and Supplementary Figures 1-3, respectively.

| | **True $\log_{10} M_e$** | **Median estimated $\log_{10} M_e$** | **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 |

**Table S2 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.

| 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 |

**Table S3 Ascertainment and standard errors of polygenicity enrichment in simulations at different sample sizes.** 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 Table S2 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).

**Table S4 (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.

**Table S5 (see Excel file) Comparison of $M_e$ 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).

**Table S6 (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.[11] and estimates of $\mathrm{M_t}$ from Zhang et al.[12] We also report implied values of $\mathrm{M_e}$ based on the model (point-normal or point-normal-normal) fit by Zhang et al.

**Table S7 (see Excel file) Comparison of the observed distribution of $\boldsymbol{\chi^2}$ 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 $\chi^2$ distribution for 33 traits (Table S4). 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 Figure S5 for QQ plots); a more complex model, e.g. the point-normal-normal model of ref.[12], 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 $\chi^2$ statistics; in contrast, $M_e$ does not count large LD blocks more heavily than small LD blocks.

| 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) |

**Table S8 Comparison of common-SNP $M_e$ estimates for traits with multiple available datasets.** 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[30] (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.

**Table S9 (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).

| 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 |

**Table S10 Polygenicity and heritability enrichment for functional annotations, meta-analyzed across well-powered traits.** 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.

**Table S11 (see Excel file) Numerical results from Figure 6.**

| | 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) |

**Table S12 Polygenicity and heritability enrichment of ExAC genes.** 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.

**Table S13 (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 Figure S10A and not in panel B.