



Published in final edited form as:

*Nature*. 2023 February ; 614(7948): 492–499. doi:10.1038/s41586-022-05684-z.

## Polygenic architecture of rare coding variation across 394,783 exomes

**Daniel J. Weiner**<sup>†,\*1,2,3</sup>, **Ajay Nadig**<sup>†,\*1,2,3</sup>, **Karthik A. Jagadeesh**<sup>1,4</sup>, **Kushal K. Dey**<sup>1,4</sup>, **Benjamin M. Neale**<sup>1,2,3</sup>, **Elise B. Robinson**<sup>2,3,5</sup>, **Konrad J. Karczewski**<sup>1,2,3</sup>, **Luke J. O'Connor**<sup>†,1</sup>

<sup>1</sup>Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

<sup>2</sup>Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

<sup>3</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

<sup>4</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

<sup>5</sup>Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

### Summary

Both common and rare genetic variants influence complex traits and common diseases. Genome-wide association studies have discovered thousands of common-variant associations, and more recently, large-scale exome sequencing studies have identified rare-variant associations in hundreds of genes<sup>1-3</sup>. However, rare-variant genetic architecture is not well characterized, and the relationship between common- and rare-variant architecture is unclear<sup>4</sup>. Here, we quantify the heritability explained by gene-wise burden of rare coding variants across 22 common traits and diseases in 394,783 UK Biobank exomes<sup>5</sup>. Rare coding variants ( $AF < 1e-3$ ) explain 1.3% (SE = 0.03%) of phenotypic variance on average – much less than common variants – and most burden heritability is explained by ultra-rare loss-of-function variants ( $AF < 1e-5$ ). Common and rare variants implicate the same cell types, with similar enrichments, and they have pleiotropic effects on the same pairs of traits, with similar genetic correlations. They partially colocalize at individual genes and loci, but not to the same extent: burden heritability is strongly concentrated in significant genes, while common-variant heritability is more polygenic, and burden heritability is also more strongly concentrated in constrained genes. Finally, we find that burden heritability for schizophrenia and bipolar disorder<sup>6,7</sup> is approximately 2%. Our results indicate that rare coding variants will implicate a tractable number of large-effect genes, that common and rare associations

<sup>†</sup>Correspondence: [dweiner@broadinstitute.org](mailto:dweiner@broadinstitute.org) (D.J.W.), [anadig@broadinstitute.org](mailto:anadig@broadinstitute.org) (A.N.), [loconnor@broadinstitute.org](mailto:loconnor@broadinstitute.org) (L.J.O).

\*Equal contributions, order decided randomly

Author contributions

D.J.W., A.N. and L.J.O. conceived and designed experiments. K.J. and K.K.D. suggested analyses. B.M.N., E.B.R., K.K. and L.J.O. supervised the project. D.J.W., A.N. and L.J.O. performed analyses. D.J.W., A.N. and L.J.O. wrote the manuscript.

Competing interests

KJK is a consultant for Vor Biopharma and AlloDx. BMN is a member of the scientific advisory board at Deep Genomics and Neumora, consultant of the scientific advisory board for Camp4 Therapeutics and consultant for Merck. The remaining authors have no competing interests.

are mechanistically convergent, and that rare coding variants will contribute only modestly to missing heritability and population risk stratification.

---

## Introduction

Genome-wide association studies have discovered thousands of common variants that are associated with common diseases and traits. Common variants have small effect sizes individually, but they combine to explain a large fraction of common disease heritability<sup>8,9</sup>. More recently, sequencing studies have identified hundreds of genes harboring rare coding variants, and these variants can have much larger effect sizes<sup>1-3,5</sup>. However, it is unclear how much heritability rare variants explain in aggregate, or more generally how common- and rare-variant architecture compare: whether they are equally polygenic; whether they implicate the same genes, cell types and genetically correlated risk factors; whether rare variants will contribute meaningfully to population risk stratification.

To characterize common-variant architecture, a productive approach has been to quantify components of heritability by aggregating subtle associations across the genome. This approach has been used to address the problem of “missing heritability”<sup>9-11</sup>, to quantify the shared genetic basis of related diseases and traits<sup>12-14</sup>, to prioritize disease-relevant cell types and regulatory elements<sup>15-18</sup>, and to quantify the effect of negative selection on common-variant architecture<sup>19-22</sup>.

For rare variants, however, heritability estimation is more challenging<sup>23</sup>. Most rare alleles are observed only once or twice, leading to low statistical power, and confounding due to uncorrected population stratification and cryptic relatedness is a major concern. Wainshtein et al.<sup>24</sup> estimated that common and rare variants combine to explain most of the twin-heritability of height and BMI, but their estimates for the rare-variant contribution specifically had very wide confidence intervals.

To characterize rare variant genetic architecture, we estimated the heritability explained by gene-wise burden of rare and ultra-rare coding alleles, while avoiding confounding due to population stratification. Analyzing association statistics from 394,783 UK Biobank exomes<sup>5</sup> together with common-variant association data from the same individuals<sup>25</sup>, we find that the burden heritability due to rare coding variants is modest (1.3%  $\pm$  0.03%), and we systematically compare the architecture of common and rare variants.

## Results

### Estimation of burden heritability

In sequencing studies, most rare variants are observed in only one or a few individuals, motivating the use of burden tests that aggregate minor alleles within genes<sup>26</sup>. We define *burden heritability* as the fraction of phenotypic variance explained by minor allele burden in each gene under a random effects model (Figure 1A; see Methods). It is a component of the total coding-variant heritability, and it is statistically tractable even for singletons. The alleles comprising the burden are stratified by their predicted functional impact, and we focus primarily on predicted loss-of-function (pLoF) variants, whose gene-wise burden

is expected to explain the majority of their total heritability (due to their similar functional consequences) (Figure 1A). For missense variants the burden heritability is an unknown fraction of their total; for example, if 50% of alleles are deleterious and 50% are null, the burden heritability is one half the total heritability.

We developed *burden heritability regression* (BHR) to estimate burden heritability and to partition it across genes and alleles (see Methods). BHR inputs variant-level association summary statistics and allele frequencies. It regresses burden test statistics on “burden scores,” which are related to the combined allele frequency, and it estimates burden heritability from the regression slope (Figure 1B, Supplementary Tables 1-2). Similar to LD score regression<sup>10</sup>, this approach distinguishes heritable signal, which affects the slope of the regression, from confounding due to population stratification and relatedness, which affect its intercept. BHR relies on the assumption that genes with larger or smaller burden scores do not have larger or smaller per-allele effect sizes, which might be violated due to selection-related effects; we use two approaches to avoid selection-related bias (Methods).

We evaluated the performance of BHR in two sets of simulations. First, we analyzed simulated summary statistics at whole-genome scale, sampled directly without any individual-level data (Methods). These simulations included negative selection and population stratification. They did not include linkage disequilibrium (LD); this choice is approximately realistic for ultra-rare variants, which have extremely little LD (Supplementary Table 3). BHR produced unbiased estimates of the burden heritability, and in non-null simulations, it was well powered to detect a burden heritability of 0.5% (Figure 1C). In additional simulations (Extended Data Figure 1, Supplementary Table 4), BHR produced approximately unbiased estimates with different amounts of selection, different amounts of population stratification (including minor-allele biased stratification), different ranges of allele frequencies, and different sample sizes.

Second, we simulated individual-level genotype and phenotype data in small-scale null simulations (10,000 individuals, 1,000 genes), and we compared BHR with GCTA-LDMS<sup>11</sup>, which has been used to estimate heritability in whole genome sequencing studies. The genotypes were sampled from forward simulations with independent sites (no LD) and migration among demes with different mean phenotypes (see Methods). We considered three models, including two with potentially problematic minor allele-biased stratification; as expected, this led to bias, but it was possible to detect and correct for it by subtracting the exome-wide mean minor-allele effect size (Extended Data Figure 2a,c). No bias was observed in the simulation with non-minor-allele-biased stratification. Results were similar in simulations with and without selection, indicating that synonymous variants can be used as a negative control for pLoF and missense variants. GCTA also produced upwardly biased estimates in these simulations, but conditioning on deme label corrected for the bias (Extended Data Figure 2b,d). This exact approach cannot be used in real-data analyses, but principal components may approximate it, to an extent that is difficult to predict in simulations.

## Burden heritability of 22 complex traits

We analyzed publicly available UK Biobank exome sequencing association statistics from Genebase<sup>5</sup> for 22 complex traits and up to 394,783 individuals of European ancestry, including 18 continuous traits and 4 common diseases (see Data Availability and Supplementary Table 5). We analyzed 6.9 million coding variants in 17,318 protein coding genes (see Methods). Within each gene, variants were stratified into three allele frequency bins ( $MAF < 1e-5$ ,  $1e-5 - 1e-4$ ,  $1e-4 - 1e-3$ ); we refer to  $MAF < 1e-5$  variants as ultra-rare, and to  $MAF = 1e-5 - 1e-3$  variants as rare. Variants were also stratified into four functional categories (pLoF, missense damaging, missense benign, and synonymous) (Figure 2A, Supplementary Table 6); missense functional predictions were obtained using PolyPhen2.<sup>27</sup> Heritability estimates are reported on an observed scale, and liability-scale estimates are reported in Supplementary Table 7.

We estimate that on average across traits, gene-wise burden of rare and ultra-rare pLoF and damaging missense variants explain 1.3% ( $SE = 0.03\%$ ) of phenotypic variance (Figure 2B). All 22 traits had nonzero burden heritability at a nominal significance level (Supplementary Tables 7-8). Burden heritability concentrates among variants with the most severe predicted functional consequences: pLoF variants explain the majority of burden heritability, followed by damaging missense variants, while benign missense variants and synonymous variants explain little or no heritability (Figure 2B). Rare variants explained less burden heritability than ultra-rare variants. These estimates are corrected for within-gene LD, which causes inflation in the burden test statistics in proportion to the number of alleles per gene (Methods). With common-variant summary statistics for the same traits in UK Biobank, we estimated common variant SNP-heritability using LD score regression (Methods). A much larger fraction of phenotypic variance is explained by common variants (median 13%), and common variant and burden heritability are highly correlated (Figure 2C, Supplementary Table 10).

Inflation in exome association test statistics due to uncorrected population stratification is a major concern, especially when estimating heritability. The BHR intercept quantifies the inflation in burden test statistics due to sampling variation and most forms of confounding (analogous to the LD Score Regression intercept<sup>10</sup>), as well as overdispersion. (Fixing the intercept at  $1/n$  resulted in inflated burden heritability estimates; Supplementary Figure 1). We evaluated the robustness of this approach in three analyses. First, we quantified minor allele-biased population stratification, which could produce upward bias in BHR, by calculating the mean minor-allele effect size of synonymous variants (see Methods). This effect was nonzero but very small, and we quantified the resulting bias in our heritability estimates (0.005% on average; Extended Data Figure 3). Second, we computed null pLoF burden statistics by randomly permuting the major and minor alleles; as expected, BHR produced heritability estimates not significantly different from zero (Extended Data Figure 4). Third, we computed the correlation between pLoF burden scores and synonymous burden statistics, and they were uncorrelated, as expected (Supplementary Figure 2).

Accordingly, we used the BHR intercept to quantify the amount of residual population stratification in the burden test statistics. For ultra-rare pLoF variants, on average across

traits, confounding and overdispersion explained 4% of variance in the test statistics, sampling variation explained 85%, and genuine burden heritability explained the remaining 10% (Figure 2D, Supplementary Table 6). For ultra-rare synonymous variants, burden heritability explained 0% of variance; confounding and overdispersion explained 4% of variance, and sampling variation explained 94% (Supplementary Table 6). The estimated amount of inflation due to population stratification (~4%) implies a family-wise error rate of more than 0.05 but less than 0.1 for most traits (Supplementary Table 6).

We performed three additional sensitivity analyses. First, we considered frequency-dependent burden weights, motivated by the known dependence of common-variant effect sizes on allele frequency<sup>28</sup>; estimates were nearly identical (Supplementary Figure 3). Second, we performed a joint regression with a shared intercept for different frequency bins and functional categories; again, results were nearly identical (Supplementary Figure 4). Third, we varied the number of gene constraint bins, and no change was observed with more than five bins (the number we use) (Supplementary Figure 5).

Two recent papers reported that rare variants from whole-genome sequencing data are an important source of heritability for complex traits. Wainschtein et al.<sup>24</sup> reported the heritability explained by rare and low-frequency variants (MAF=1e-4 - 0.01) is 0.3 (SE 0.1) for height and 0.29 (SE 0.25) for BMI; Jang et al.<sup>29</sup> similarly reported that rare variants explain a large fraction of heritability for smoking phenotypes, with large standard errors. Unlike our burden estimates, these estimates include noncoding SNPs and SNPs at intermediate allele frequencies (0.001-0.01), and they do not aggregate variants by gene. Because of these differences, our rare variant heritability estimates are smaller but much better powered: 0.037 (SE=0.001) for height, 0.012 (SE=0.001) for BMI, and 0.006 (SE=0.001) for smoking status, respectively (Supplementary Tables 8 and 11).

### Concentration within significant genes

In GWAS, a consistent observation has been that common traits are highly polygenic, with numerous loci of small effect<sup>31,32</sup>. In contrast, most rare genetic diseases are caused by large-effect mutations in a much smaller number of genes, and it is unclear whether the rare-variant genetic architecture of common diseases is highly polygenic like common variants or more oligogenic like rare diseases. We quantified the proportion of burden heritability that is explained by exome-wide significant genes (Methods), and we compared the extent to which common- and rare-variant heritability is concentrated in large-effect genes and regions of the genome.

17 of 22 traits had at least one significantly associated gene in Genebase<sup>5</sup> (Methods), and they had a median of 6 significant genes per trait (Supplementary Tables 12-13). These genes explained a substantial proportion of the burden heritability (median: 19%; Figure 3A), after partially correcting for winner's curse<sup>33</sup>; see Methods and Supplementary Figure 6. For LDL cholesterol levels, *APOB* alone explained 39% (SE = 4%) of burden heritability, and for diabetes, *GCK* explained nearly 15% (SE = 4%).

In contrast, individual common-variant associations are dramatically smaller as a fraction of common-variant heritability (Figure 3B, Supplementary Table 14). Even aggregating

common-variant heritability across large LD blocks (most > 1Mb), top rare-variant associated genes (out of 17,318) explain a much larger fraction of heritability than top LD blocks (out of 1,651) (Extended Data Figure 5, Supplementary Table 15). The difference in common- vs. rare-variant polygenicity can be explained by “flattening” due to negative selection, as we previously hypothesized<sup>19</sup> (see Discussion).

We sought to reconcile the difference in polygenicity with the observation that rare-variant associations are strongly enriched near GWAS loci<sup>3</sup>. For traits with at least 5 significant genes, we quantified the fraction of common variant heritability mediated by those genes using the Abstract Mediation Model (AMM), which fully accounts for uncertainty in which SNPs regulate which genes<sup>34</sup> (Supplementary Table 16). We confirm that rare-variant associated genes are enriched for common variant heritability; for example, the 81 exome-wide significant genes for height explain 9.5% of its common variant heritability (SE = 2.5%). However, these same genes explain 32.1% of burden heritability (SE = 2.0%) and other traits exhibit similar patterns (Figure 3C). The same is observed for individual genes (Extended Data Figure 6).

For the cancer phenotype, which is a composite of multiple cancer types, the seven exome-wide significant genes (*MSH2*, *BRCA1*, *BRCA2*, *APC*, *ATM*, *PALB2*, *CHEK2*) explain 33% (SE = 4%) of its burden heritability. Noting that all of these genes are well-known tumor suppressors, we analyzed known tumor suppressors and oncogenes from the Cancer Gene Census<sup>35</sup> (CGC). Indeed, the 172 CGC tumor suppressor genes explain nearly half of the burden heritability (48%, SE = 10%) (Figure 3D, Supplementary Table 17). In contrast, the 101 oncogenes do not explain any burden heritability (1%, SE = 2%). These results are concordant with the known biology of tumor suppressors and oncogenes. They contrast with common-variant architecture: tumor suppressor genes only mediate 5% (SE = 4%) of common-variant heritability, and the seven exome-wide significant genes mediate 0% (SE = 2%) (Figure 3D).

### Enrichment of constrained genes

We investigated the contribution of different gene sets to the burden heritability, defining the burden heritability enrichment of a gene set as its fraction of burden heritability divided by its fraction of burden variance (approximately the fraction of minor alleles, not of genes) (see Methods). We estimated common variant gene-mediated enrichments for the same gene sets using AMM.

First, we analyzed sets of genes that are differentially expressed in trait-matched cell and tissue types (see Methods, Supplementary Table 13). For these gene sets, burden heritability enrichments and common-variant enrichments were approximately equal (Figure 4A). For example, in a set of 3,396 red blood cell-expressed genes, the enrichment was 2.1x (SE = 0.4x) for common variants and 2.2x (SE = 0.3x) for rare variants. Even though common-variant heritability is spread across more genes compared with burden heritability, it is equally concentrated in specific cell types, consistent with the cell-type-centric omnigenic model<sup>36</sup> (see Discussion).

Next, we compared common- vs. rare-variant enrichments across the spectrum of selective constraint<sup>37</sup>. Rare variant enrichments were larger than common variant enrichments in constrained genes for 21/22 traits (Figure 4B). Fluid intelligence score had a rare variant enrichment of 8.1x (SE = 1.0x), compared with a common variant enrichment of 2.2x (SE = 0.3x). From the 1st to the 5th quintiles of constraint, rare variant enrichments decayed from a median of 4.5x to 0.3x, while common variant enrichments had a lower maximum (2.1x) and a similar minimum (0.5x) (Figure 4C). These observations are consistent with the expected effect of negative selection, which prevents both coding and regulatory variants affecting highly constrained genes from becoming common in the population<sup>19,22,38,39</sup> (see Discussion).

For phenotypes that directly affect fitness, loss-of-function alleles are expected to be deleterious almost exclusively, since if gene loss were protective, the gene would be lost. Indeed, pLoF variants in constrained genes are associated with childlessness in UK Biobank<sup>40</sup>. Moreover, a standard approach in severe psychiatric and neurodevelopmental disorders is to aggregate pLoFs across a set of candidate genes<sup>41-43</sup> (this approach cannot be used to estimate burden heritability, as not all candidate genes are causal). We calculated the genome-wide mean minor allele effect of ultra-rare pLoFs on each trait (Supplementary Table 7). These values were uncorrelated with the corresponding synonymous effects, indicating that they are not driven by minor-allele biased population stratification (Extended Data Figure 3) (see Methods). Traits with large mean minor-allele effect sizes tended to have a strong burden heritability enrichment in constrained genes (Extended Data Figure 7), consistent with the hypothesis that these traits are directly under selection (but not providing evidence against the importance of pleiotropic selection<sup>44</sup>).

### Burden genetic correlations

Exome-sequencing studies often aggregate pLoF and damaging missense variants to maximize power<sup>6,45</sup>, raising the question of whether damaging missense variants generally act via loss of function. We used BHR to compute burden genetic correlations between pLoF and damaging missense variants (Figure 5A, see Methods, Supplementary Table 18). We observed a mean burden genetic correlation of 0.64 (SE = 0.10), implying that pLoF and missense variants in the same genes often have divergent phenotypic effects. One explanation is that deleterious missense variants frequently act via mechanisms other than partial loss of function. Alternatively, PolyPhen2 predicted damaging variants approximate pLoFs in some genes but not others.

Common-variant effect sizes are often correlated across traits, providing evidence of shared biological mechanisms<sup>12</sup>. We estimated pairwise burden genetic correlations from ultra-rare pLoF variants among an extended group of 37 traits (Supplementary Table 5). 197 correlations passed a nominal threshold for statistical significance, and 55 passed a Bonferroni threshold (Supplementary Table 19). For the same group of UKB traits, we also computed common variant genetic correlations using LDSC<sup>12</sup> (Methods, Supplementary Table 19). Both common and rare variants had correlated effects within clusters of closely related traits (e.g. LDL/Triglycerides/High Cholesterol, Calcium/Albumin, Neuroticism/

Depression) and also within less obvious trait pairs (FVC/BMI, Osteoarthritis/Depression) (Figure 5B; for all 37 traits, see Extended Data Figure 8).

More generally, rare-variant genetic correlations were concordant with those from common variants, but they were stronger by 1.6x on average (Figure 5C). A potential explanation is that pleiotropic genes are more strongly constrained<sup>44</sup>, which would dampen common-variant genetic correlations. A different possibility is that coding effects are less cell-type specific and therefore more pleiotropic, but we did not observe stronger genetic correlations among common coding variants (Extended Data Figure 9; see Methods). We note that rare-variant genetic correlations, similar to common-variant correlations, can be an artifact of cross-trait assortative mating<sup>46</sup> (Supplementary Figure 7).

### Schizophrenia and bipolar disorder

Damaging variants in constrained genes are strongly associated with neuropsychiatric disorders<sup>45,47,48</sup>. We applied BHR to summary statistics from recent exome-sequencing studies of schizophrenia (SCHEMA study<sup>6</sup>: 24,248 cases, 97,322 controls) and bipolar disorder (BipEx study<sup>7</sup>: 14,210 cases, 14,422 controls) (Methods). Following the original reports, we analyzed ultra-rare variants with minor allele count less than 5 (MAF < 2e-5 for SCZ, MAF < 9e-5 for BPD).

We estimate that schizophrenia and bipolar disorder have a pLoF burden heritability of 1.7% (SE = 0.3%) and 1.8% (SE = 0.3%), respectively (on a liability scale) (Figure 6A, Supplementary Table 20). These estimates were larger than those of the UK Biobank traits except for height, consistent with their high common variant heritability. The pLoF burden genetic correlation between bipolar disorder and the two main schizophrenia cohorts was 0.39 (SE = 0.22) and 0.51 (SE = 0.28), roughly consistent with estimates of their common-variant genetic correlation of 0.72<sup>49</sup> (Supplementary Table 21). The burden heritability due to ultra-rare damaging missense variants (MPC > 2)<sup>50</sup> was 0.35%, SE = 0.12% for schizophrenia and 0.14%, SE = 0.12% for bipolar disorder. There was no evidence of nonzero burden heritability for synonymous variants (Figure 6A).

The SCHEMA study<sup>6</sup> identified 9 autosomal genome-wide significant genes associated with schizophrenia, and we estimate that they explain 7% (SE = 1.5%) of the burden heritability. Larger studies will discover many additional significant genes, and the same will probably occur for bipolar disorder, which has a high burden heritability but no exome-wide significant genes in the BipEx sample.

A consistent observation in exome-based analyses of neuropsychiatric disorders is an enrichment of significant associations in constrained genes<sup>6,7</sup>. Indeed, in the top quintile of constrained genes, burden heritability is 9.6x (SE = 1.2x) enriched for schizophrenia and 4.6x (SE = 1.1x) enriched for bipolar disorder (Figure 6B). For schizophrenia, we estimate that constrained genes explain 70% (SE = 9%) of its burden heritability (Supplementary Table 20).



## Discussion

Rare protein-coding genetic variation is a rich source of biological insight. Rare diseases are often caused by mutations in one or a handful of genes, and the discovery of those genes has led to effective therapies<sup>51,52</sup>. For common diseases and complex traits, the role of rare variation has been debated<sup>24,53</sup>. In this study, we found that rare loss-of-function variants comprise ~1% of phenotypic variance for most traits, that burden heritability concentrates among ultra-rare variants in highly constrained, large-effect disease genes, and that these genes are modestly enriched for common-variant heritability. Our findings make us highly optimistic about the potential for rare coding associations to inform our understanding of common disease biology, for two reasons.

First, for rare, syndromic forms of common diseases (e.g., *MC4R*-driven obesity), a critical question is whether their causal genes are relevant to common variant liability as well. If common and rare variants converge on the same disease-causing processes, therapeutics targeting rare-variant associated genes have the potential to benefit a large number of patients, not only the few who carry specific mutations. Reassuringly, we find that common- and rare-variant associations are mechanistically convergent: rare-variant associated genes are enriched for common-variant heritability (Figure 3C), common and rare variants implicate the same cell types and tissues (Figure 4A), and they have pleiotropic effects on the same pairs of traits (Figure 5C). These findings provide quantitative, genome-wide confirmation of previous reports that common and rare variants implicate overlapping genes<sup>3,6,34</sup> and pathways<sup>43,54</sup>.

Second, rare-variant architecture is much less polygenic. Already, exome-wide significant genes explain a substantial proportion of the total burden heritability for well-powered traits (Figure 3A), suggesting that large-effect mechanisms involve a tractable number of genes. Many of these significant genes are drug targets, and accordingly, drug target gene sets<sup>55</sup> explain a large fraction of burden heritability for some traits (Extended Data Figure 10); in contrast, common-variant polygenicity has been a challenge for translational efforts<sup>4</sup>.

The differences that we observe between common- and rare-variant architecture support the flattening hypothesis, which we previously proposed as an explanation for extreme common-variant polygenicity<sup>19</sup>. Under the flattening hypothesis, a small fraction of genes and regions of the genome have large effect sizes when mutated, and these loci dominate the rare-variant heritability. However, these genes are constrained, limiting their common-variant associations, and common-variant heritability is spread across a much larger number of loci with much smaller effects. This hypothesis provides an explanation for the differences we observe between common vs. rare-variant polygenicity (Figure 3A-B) and between their enrichments in constrained genes (Figures 4C, 6B).

Our results also support aspects of the omnigenic model<sup>36</sup>, which posits that there are a limited number of “core genes” with biologically interpretable effects and a much larger number of “peripheral genes” whose effects are mediated by highly connected intracellular networks. The limited polygenicity of rare variants, with relatively few large-effect genes, is consistent with the limited number of hypothesized core genes. Core genes

and exome-wide significant genes may differ, for example because some peripheral genes may have large effects<sup>56</sup>, but they should largely overlap. In addition, we find that cell-type-specific heritability enrichments are similar between common and rare variants (Figure 4a), consistent with the hypothesis that core genes and peripheral genes are expressed in the same cell types.

Just as negative selection affects the distribution of heritability among genes, it also affects the fraction of heritability in protein-coding versus regulatory regions. Gazal et al.<sup>22</sup> found that coding variants explain a much larger fraction of heritability for low-frequency variants (~26%) than for common variants (~8%), consistent with the expected effect of negative selection. If this trend continues at even lower frequencies, then rare and ultra-rare noncoding variants would explain little heritability and would not explain “missing heritability.”

Polygenic risk scores derived from common variants may stratify individuals into clinically meaningful groups<sup>57-59</sup>. The growing accessibility of whole exome and genome sequencing raises the question of whether these genetic profiles should expand to include both common and rare variants. On the one hand, our estimates suggest that rare coding variation will have modest predictive power on average in the population, since most patients do not carry a large-effect risk allele<sup>60</sup>. On the other, these variants are highly relevant to the individuals who do carry them, and screening for these variants can be especially valuable to individuals who have been ascertained by phenotype or family history<sup>61</sup>. Moreover, rare variant-associated genes may implicate disease processes that are relevant to patients without those specific mutations.

Our analysis has a number of limitations. First, it is limited to coding variants, and we do not quantify the contribution of rare noncoding variants. Second, for missense variants in particular, burden heritability might represent a fraction of the total rare coding heritability, due to overdispersion effects (Figure 1A). We stratified missense variants by their PolyPhen2 predicted effect, but with a more sophisticated approach, it would be possible to capture a larger fraction of the total missense heritability. Third, BHR is a method-of-moments estimator operating on burden association statistics, and this approach is not expected to have highest possible statistical power; a method based on individual-level data, especially if it used maximum likelihood rather than the method of moments, might produce more precise estimates of the burden heritability. Fourth, our analysis is limited to European-ancestry participants in the UK Biobank, which reflects a well-documented bias in human genetics research<sup>62</sup>. Fourth, the UK Biobank is a relatively healthy population cohort<sup>63</sup>, which limits our power to analyze diseases. For the same reason, the UK Biobank sample might be depleted of deleterious genetic variation<sup>40</sup>, potentially causing decreased burden heritability in this population.

In GWAS, widespread sharing of summary statistics has been catalytic. We have released open-source software implementing the full suite of BHR analyses (Code Availability), and we advocate for sequencing studies to share variant-level association statistics, including variant frequencies, functional annotations, and per-allele effect sizes, which are sufficient for analysis using BHR.

## Methods

### Definition of burden heritability

Let  $X_g$  be the mean-centered genotype matrix for gene  $g$ , and let  $Z_g$  be the standardized genotype matrix, whose columns have zero mean and unit variance. We define the *burden* for gene  $g$  as the mean-centered minor allele count for each individual:

$$b_g := X_g \mathbf{1}_{m_g \times 1} = Z_g w_g \quad \#(1)$$

where  $\mathbf{1}_{m_g \times 1}$  is the all-ones vector,  $m_g$  is the number of variants in gene  $g$ , and  $w_g$  is the vector of burden weights. The entries of  $w_g$  are the standard deviations of the corresponding columns of  $X_g$ ; under Hardy-Weinberg equilibrium, they are equal to  $\sqrt{2p(1-p)}$  (where  $p$  is the allele frequency).

Let  $y$  be the  $n \times 1$  standardized phenotype vector, and let  $\tilde{\beta}_g$  be the  $m_g \times 1$  vector of per-allele effect sizes:

$$\tilde{\beta}_g := E(X_g^T X_g)^{-1} E(X_g^T y). \quad \#(2)$$

Let  $\beta_g$  be the vector of per-normalized genotype effect sizes, or correlations:

$$\beta_g := E(Z_g^T Z_g)^{-1} E(Z_g^T y) = w_g \circ \tilde{\beta}_g \quad \#(3)$$

where  $\circ$  denotes the element-wise product. The *burden effect size*  $\gamma_g$  is the correlation between the burden  $b_g$  and the phenotype  $y$ :

$$\begin{aligned} \gamma_g &:= \frac{E(b_g^T y)}{\sqrt{E(y^T y) E(b_g^T b_g)}} \\ &= \frac{w_g^T E(Z_g^T y)}{\sqrt{n w_g^T E(Z_g^T Z_g) w_g}} \\ &= \frac{w_g^T \beta_g}{\sqrt{w_g^T w_g}}. \end{aligned} \quad \#(4)$$

We assumed that there is no LD, such that  $E(Z_g^T Z_g) = nI$ , in the third line (see below).

Burden heritability is defined under a random effects model for the burden effect sizes  $\gamma$ . Suppose that the vector of per-allele effect sizes  $\tilde{\beta}_g$  has mean  $\mu_g \mathbf{1}$  and zero covariance. Then the per-normalized genotype effect size vector  $\beta$  has mean  $\mu_g w_g$ , and the burden effect  $\gamma_g$  has mean

$$E(\gamma_g \mid \mu_g) = \frac{w_g^T E(\beta_g \mid \mu_g)}{\sqrt{w_g^T w_g}} = \frac{\mu_g w_g^T w_g}{\sqrt{w_g^T w_g}} = \mu_g \mid \mid w_g \mid \mid. \quad \#(5)$$

We define the burden heritability of gene  $g$  as:

$$h_{\text{burden}}^2(g) := E(\gamma_g | \mu_g)^2 = \mu_g^2 | | w_g | | ^2. \quad \#(6)$$

The total burden heritability across a set of genes  $A$  is:

$$h_{\text{burden}}^2(A) = \sum_{g \in A} E(h_{\text{burden}}^2(g)) = \sum_{g \in A} E(\mu_g^2 | | w_g | | ^2). \quad \#(7)$$

The burden heritability is a component of the total heritability. For gene  $g$ , its total heritability (without LD) is:

$$h_{\text{total}}^2(g) = E(\text{var}(Z_g \beta_g | \beta_g)) = E(\beta_g^T \beta_g) \geq h_{\text{burden}}^2(g). \quad \#(8)$$

### Burden heritability regression

Burden test statistics, which are commonly used to identify associated genes, are essentially burden effect estimates. The burden effect estimate,  $\gamma_g$ , is the sample correlation between  $b_g$  and  $y$ :

$$\hat{\gamma}_g := \frac{b_g^T y}{\sqrt{(b_g^T b_g)(y^T y)}}. \quad \#(9)$$

(It is related to the burden  $\chi^2$  statistic:  $n\hat{\gamma}_g^2 \sim \chi_1^2 | H_0$ ). Without LD, and without correlated stratification effects (see below),  $\hat{\gamma}_g$  has mean  $\gamma_g$  and variance:

$$\text{var}(\hat{\gamma}_g | \mu_g) \approx n^{-1} + a + \text{var}(\gamma_g | \mu_g). \quad \#(10)$$

There are three terms.  $n^{-1}$  is the ordinary sampling variation, which is the approximate term; the approximation is accurate when the burden effect is small.  $a$  quantifies inflation due to population stratification and cryptic relatedness, and we assume that it is not gene specific (see below). The third term quantifies overdispersion-related sampling variation in the true value of  $\gamma_g$ . If variants in the same gene have uncorrelated overdispersion effects with a constant effect-size variance  $d_g$  in per-standard deviation units, then the overdispersion term is:

$$\text{var}(\gamma_g | \mu_g, \sigma_g^2) = \frac{w_g^T (E(\beta_g \beta_g^T) - E(\beta_g) E(\beta_g)^T) w_g}{w_g^T w_g} = \frac{w_g^T (\sigma_g^2 I) w_g}{w_g^T w_g} = d_g. \quad \#(11)$$

Combining equations 5, 10 and 11:

$$E(\hat{\gamma}_g^2 | \mu_g) = \mu_g^2 | | w_g | | ^2 + d_g + n^{-1} + a. \quad \#(12)$$

The BHR regression equation is obtained by taking an average value of  $\mu_g^2$  and  $d_g$  across genes. Let  $\tau = E(\mu_g^2)$  and  $d = E(d_g)$ . The BHR regression equation is:

$$E(\hat{\gamma}^2) = \tau \|w\|^2 + d + n^{-1} + a. \quad \#(13)$$

The first term is used to estimate the burden heritability, and the other terms are the regression intercept.

### Minor-allele biased population stratification

A potential source of bias for BHR is minor-allele biased population stratification. Specifically, let  $\alpha_g$  be the random vector of normalized stratification effects for minor alleles in gene  $g$ ; we generally assume that

$$E(\alpha_g \alpha_g^T) = aI \quad \#(14)$$

where  $a$  is the non-gene-specific inflation parameter. Under this assumption, the contribution of stratification to the BHR equation is:

$$E\left(\frac{(w_g^T \alpha_g)^2}{w_g^T w_g}\right) = a. \quad \#(15)$$

However, minor alleles may have nonzero mean effect sizes due to stratification, such that

$$E(\alpha_g \alpha_g^T) = aI + m_g m_g^T \quad \#(16)$$

where  $m_g = m w_g$  is the mean effect due to stratification. This type of stratification could plausibly arise when a small fraction of individuals in the study come from a certain subpopulation, such that variants specific to that subpopulation are observed at low frequencies. It could also occur when one subpopulation is bottlenecked, causing its frequency spectrum to shift. In this scenario, the contribution of stratification effects is:

$$E\left(\frac{(w_g^T \alpha_g)^2}{w_g^T w_g}\right) = a + m^2 \frac{\|w_g\|^2}{\|w_g\|^2} \quad \#(17)$$

and the BHR slope will be inflated by  $m^2$ .

Minor-allele biased stratification would cause the mean minor-allele effect size to be nonzero genome wide, possibly motivating a genome-wide mean centering approach. For pLoF variants, however, it is biologically plausible that their causal effect sizes have nonzero mean (especially for traits such as autism and schizophrenia<sup>41-43</sup>). To distinguish between these possibilities, we calculate the genome-wide mean minor allele effect size for pLoF and synonymous variants separately. Let  $w_{genome}(syn)$  be the concatenated vector of synonymous burden weights across the genome; we compute the genome-wide mean synonymous minor allele effect,  $\hat{m}_{genome}(syn)$ , as:

$$\widehat{m}_{genome}(\text{syn}) = | | w_{genome}(\text{syn}) | |^{-2} w_{genome}(\text{syn})^T \widehat{\beta}. \quad \#(18)$$

We estimate the contribution of minor allele biased stratification to the burden heritability as:

$$h_{\text{bias}}^2 = \widehat{m}_{genome}^2(\text{syn}) | | w_{genome} | |^2 \quad \#(19)$$

because  $\widehat{m}_{genome}^2$  is the estimate of the upward bias in the BHR regression slope, by equation 17. For pLoF variants, we estimate the bias in their heritability estimates as

$$\widehat{m}_{genome}^2(\text{syn}) | | w_{genome}(\text{pLoF}) | |^2.$$

### Independence assumption and selection-related bias

BHR assumes that  $E(\mu_g^2)$  is not correlated with  $| | w_g | |^2$ . In general in a regression analysis, if the slope depends on the independent variable, it leads to bias. Here, the most plausible reason for non-independence is that genes under selective constraint have smaller burden scores and larger mean effect sizes; this would produce downward bias in the heritability estimates.

We use two approaches to mitigate this potential bias. First, we bin genes by their observed vs. expected number of pLoF variants in gnomAD (a measure of selective constraint)<sup>37</sup>. With this approach, we only require the weaker assumption that  $E(\mu_g^2)$  is uncorrelated with  $| | w_g | |^2$  within bins. We use five bins of approximately equal size. This approach is analogous to the use of LD-related annotations by Gazal et al. to address bias due to LD-dependent architecture in stratified LD score regression (S-LDSC)<sup>21</sup>.

Second, we incorporate null burden statistics that effectively fix the BHR intercept and ameliorate bias in its slope. (Even in the absence of bias, this approach is useful to increase power). We define random null burden weights vectors  $v_g$ , whose burden weights are randomly sign flipped compared with  $w_g$  (but identical in magnitude). Burden statistics computed using null burden weights are equally affected by noise, confounding, and overdispersion effects, but they contain very little burden signal.

In detail, let  $v_g = w_g \circ [\pm 1, \dots, \pm 1]$  be the null burden weights for gene  $g$ . The null burden effect size is:

$$\delta_g = \frac{v_g^T \beta_g}{\sqrt{v_g^T v_g}}. \quad \#(20)$$

If the mean minor-allele effect is  $\mu_g$ , the mean of  $\delta_g$  is:

$$E(\delta_g) = \frac{\mu_g v_g^T w_g}{\sqrt{v_g^T v_g}}. \quad \#(21)$$

The regression equation for the null burden statistics is

$$E(\hat{\delta}_g^2) = E(\mu_g^2) \frac{(v_g^T w_g)^2}{v_g^T v_g} + d + n^{-1} + a. \quad \#(22)$$

The null burden scores,  $(v_g^T w_g)^2 / v_g^T v_g$ , are much smaller than the original burden scores, as the random sign flipping causes  $v_g^T w_g$  to be small; therefore, the intercept of the regression is effectively constrained to be approximately equal to the mean null burden statistic.

Any number of these null burden statistics can be incorporated into the regression. We use five null burden statistics per gene, which is enough that including a larger number has little effect (Supplementary Figure 8).

### Large-effect genes as fixed effects

Large-effect genes introduce noise in BHR. We identified genes with a significant association at a Bonferroni-significant exome-wide significance threshold (i.e., 0.05 / number of genes by a  $\chi^2$  test). We excluded these genes from the regression and instead included them as fixed effects, adding their squared burden effect size estimates to the heritability directly. This approach is appropriate because the effect size estimates of significant genes are less likely to reflect confounding, and it greatly reduces the standard error of the regression estimator. The estimated heritability explained by each gene was  $h^2(g) = \hat{\gamma}_g^2 - (n^{-1} + \hat{a})$ , where  $(n^{-1} + \hat{a})$  is the BHR intercept.

### Standard errors calculation

We estimated standard errors in the regression using a block jackknife, as previously described<sup>10</sup>. We used 100 contiguous blocks of genes with around 170 genes per block. Significant genes are excluded from the block jackknife procedure, and uncertainty in their effect size estimates is incorporated using the delta method. The delta method is also used to calculate the standard error for the fraction of burden heritability mediated by significant genes, the enrichment of burden heritability in particular annotations, and the genetic correlation.

In detail, let  $\theta$  be a vector of parameters with covariance matrix  $\Sigma$ . For a function  $g(\theta)$ , the sampling variance is approximately:

$$\text{var}(g(\theta)) \approx \nabla g \Sigma \nabla g^T. \quad \#(23)$$

We apply this formula as follows:

- *Standard error of the fraction of burden heritability in a particular gene set.* Let  $h_{total}^2$  be the total burden heritability estimated by BHR, and  $h_k^2$  be the burden heritability in annotation  $k$  estimated by BHR. The fraction of burden heritability in annotation  $k$  is:

$$g(\theta) = g \begin{pmatrix} h_k^2 \\ h_{total}^2 \end{pmatrix} = \frac{h_k^2}{h_{total}^2}. \quad \#(24)$$

The covariance matrix of  $\theta$ ,  $\Sigma$ , is computed via block jackknife.

- *Standard error of the fraction of burden heritability in a particular gene annotation under the mixed model.* In the mixed effects model, genes with exome-wide significant associations are modelled as fixed effects. Let  $h_{random}^2$  be the total burden heritability estimated by the BHR random effects model excluding significant genes, and let  $h_k^2$  be the burden heritability in annotation  $k$  estimated by BHR, excluding significant genes. Let  $\hat{\gamma}_{EWS}$  denote the vector of burden effect sizes for exome-wide significant genes. Let  $M(k)$  be the diagonal matrix with dimension equal to the number of significant genes whose diagonal entries are 1 for genes in annotation  $k$ , and 0 otherwise. The fraction of burden heritability in annotation  $c$  is:

$$g(\theta) = g \begin{pmatrix} \hat{\gamma}_{EWS} \\ h_k^2 \\ h_{random}^2 \end{pmatrix} = \frac{h_k^2 + \hat{\gamma}_{EWS} M(k) \hat{\gamma}_{EWS}^T}{h_{random}^2 + \hat{\gamma}_{EWS} \hat{\gamma}_{EWS}^T}. \quad \#(25)$$

The variances and covariance of  $h_k^2$  and  $h_{total}^2$  are computed via a block jackknife. The variance of  $\hat{\gamma}_{EWS}$  is estimated as the intercept of the BHR random effects model, and their covariance is assumed to be zero with each other and with  $h_k^2$  and  $h_{total}^2$ .

- *Standard error of the burden genetic correlation under the mixed model.* Let  $h_{random\ 1}^2$  and  $h_{random\ 2}^2$  be the random-effects burden heritability of traits 1 and 2 respectively, excluding significant genes. Let  $\rho$  be the burden genetic covariance between trait 1 and trait 2 excluding significant genes, computed with the cross-trait BHR model. Let  $\hat{\gamma}_{EWS\ 1}$  denote the vector of burden effect sizes for significant genes (in per-s.d. units) for trait 1. Let  $\hat{\gamma}_{EWS\ 2}$  denote the vector of burden effect sizes for significant genes for trait 2. The burden genetic correlation under the mixed model is:

$$g \begin{pmatrix} h_{random\ 1}^2 \\ h_{random\ 2}^2 \\ \rho_{random} \\ \hat{\gamma}_{EWS\ 1} \\ \hat{\gamma}_{EWS\ 2} \end{pmatrix} = \frac{\rho_{random} + \hat{\gamma}_{EWS\ 1}^T \hat{\gamma}_{EWS\ 2}}{\sqrt{(h_{random\ 1}^2 + \hat{\gamma}_{EWS\ 1}^T \hat{\gamma}_{EWS\ 1})(h_{random\ 2}^2 + \hat{\gamma}_{EWS\ 2}^T \hat{\gamma}_{EWS\ 2})}} \quad \#(26)$$

The variances and covariances of  $h_{total\ 1}^2$ ,  $h_{total\ 2}^2$ , and  $\rho$  are computed via a block jackknife. The estimated variances of  $\hat{\gamma}_{EWS\ 1}$  and  $\hat{\gamma}_{EWS\ 2}$  are the BHR intercepts for traits 1 and 2 respectively, and the covariance between  $\hat{\gamma}_{EWS\ 1}$  and  $\hat{\gamma}_{EWS\ 2}$  is the BHR cross-trait intercept. The exome-wide significant effects are assumed to have no



covariance with  $h_c^2$ ,  $h_{total}^2$ , or  $\rho$ . The covariance between  $u$  and  $v$  is the intercept from the cross-trait BHR model.

### Stratified regression equation and heritability enrichment

BHR can be used to model any number of gene-level annotations. Let  $A_g$  be the row vector of annotation values for gene  $g$ . Similar to S-LDSC<sup>16</sup>, we model the effect-size variance of gene  $g$  as a linear function of  $A_g$ :

$$E(\mu_g^2) = A_g \tau \quad \#(27)$$

where  $\tau$  is the regression slope. This choice is necessary in order for  $\tau$  to be estimated using linear regression (other choices give rise to least-squares estimators without closed-form solutions). The gene-stratified regression equation becomes

$$E(\hat{\gamma}_g^2) = | | w_g | | ^2 A_g \tau + d + n^{-1} + a \quad \#(28)$$

where we assume that overdispersion and confounding effects do not vary across gene sets.

We define the burden heritability enrichment of a gene set as the fraction of heritability divided by the fraction of burden scores. Let the cumulative burden score for gene set  $k$  be

$$w(k) = \sum_{\text{genes } g} A_{gk} | | w_g | | ^2 \quad \#(29)$$

and let its estimated burden heritability be

$$h^2(k) = \sum_{\text{genes } g} A_{gk} | | w_g | | ^2 A_g \tau. \quad \#(30)$$

Letting  $w(0)$ ,  $h^2(0)$  denote the cumulative burden score and the burden heritability across all genes, the burden heritability enrichment of gene set  $k$  is:

$$e(k) = \frac{h^2(k)w(0)}{h^2(0)w(k)}. \quad \#(31)$$

This definition differs from the fraction of heritability divided by the fraction of genes; for example, constrained genes have smaller burden scores on average, so their burden heritability enrichment is greater than their fraction of heritability divided by their fraction of genes.

### Whole-exome scale simulations

We simulated gene burden statistics under approximately realistic genetic architectures without LD. We simulated 18,000 genes with between 1 and 1,000 possible variants per gene (drawn from a uniform distribution). We chose the mean effect size for each gene,  $\mu_g$ , from a sparse mixture of normal distributions. In simulations with overdispersion, we also

included nonzero gene-specific effect-size variance parameters,  $d_g$ . Then, we drew per-allele effect sizes for variants within each gene from gene-specific normal distributions:

$$\tilde{\beta}_g \sim N(\mu_g \mathbf{1}, d_g \mathbf{I}). \quad \#(32)$$

To model negative selection, we simulated effect sizes on 100 independent traits, and we defined a selection coefficient for each variant in proportion to its sum of squared effect sizes across traits. This choice follows the stabilizing pleiotropic selection model of Simons et al.<sup>44</sup> The selection coefficients were scaled to a desired mean selection coefficient.

We sampled allele frequencies from the neutral spectrum, such that the probability of observing an allele at allele count  $n_j$  was proportional to  $1/n_j$ , where  $1 \leq n_j \leq n$ . We approximated the effect of selection of the allele frequency spectrum by discarding variants whose sampled allele frequency was  $\frac{n_j}{n} > 4Ns$ , where  $s$  was the selection coefficient and  $N$  was  $1e4$ . This approach allows millions of variants to be sampled efficiently.

After sampling the allele frequencies  $p$ , we set the per-normalized-genotype effect sizes to  $\beta \propto \tilde{\beta}(2p(1-p))^{1/2}$ , normalizing them so that the burden heritability of variants in the allele frequency bin under consideration matched the desired value.

We computed the observed over expected number of variants in each gene by dividing the number of variants with frequency greater than zero by the number of variants (between 1 and 1000), computing o/e bins from these values.

We sampled effect-size estimates for each variant from a normal distribution, which is appropriate for a continuous trait:

$$\hat{\beta}_j \sim N(\beta_j, n^{-1} + a). \quad \#(33)$$

Simulation parameters for each simulation are provided in Extended Data Figure 1 legend. For the “realistic” simulations, the fraction of causal genes with large, medium and small effects was  $4e-4$ ,  $2e-3$ ,  $1e-2$  respectively, and their per-allele effect size variance before normalization was 5, 1,  $1/5$ . The mean selection coefficient was  $N\bar{s} = 10$ . The sample size was  $5e5$ , the number of genes was  $1.8e4$ , and the true burden heritability was either 0 or 0.005. The variance of the stratification effects was  $1e-7$ , the mean minor-allele-biased stratification effect was  $1e-5$ , and there was no overdispersion.

### Small-scale simulations with individual level data

We simulated individual-level genotype and phenotype data in small-scale null simulations (10,000 individuals, 1,000 genes, 1-100 variants per gene,  $h^2 = 0$ ). We sampled genotypes and a continuous phenotype (“height”) in forward simulations under three steady-state demographic models involving 3-4 demes:

1. “North-south” stratification with a northern, central, and southern deme, and migration between adjacent demes (rate: 0.01 per generation). We simulated a

large environmental effect, with taller simulated height (+0.5 SD) in the northern deme and shorter height (−0.5 SD) in the southern deme. Population size was  $N=1e4$  in each deme.

2. “North-bottlenecked” stratification with the same migration patterns and the same mean phenotypes in each deme, but different population sizes:  $N=5e3$ ,  $1e4$ , and  $2e4$  in the northern, central and southern demes respectively.
3. “Local” stratification, with three demes of equal size ( $N=1e4$ ) and equal height, and a fourth deme with much smaller population size ( $N=1e3$ ) and much taller height (+1 SD). The migration rate was 0.01 between each pair of adjacent demes, and the small deme was adjacent to the large central deme.

We performed each of these simulations both with and without negative selection, mimicking our analyses of pLoF and synonymous variants respectively. The true heritability was zero, to isolate possible inflation due to population stratification.

We sampled diploid individuals from each deme with independent sites (no LD) and with probability proportional to deme size. We calculated summary statistics for BHR using linear regression with no covariates. We recorded the deme from which each individual was sampled and used these as covariates for GCTA in the “deme-corrected” GCTA analysis. For both BHR and GCTA, we restricted to variants with sample allele frequency at most  $1e-3$  (i.e., a minor allele count of 20).

Our forward simulations involved simulated the allele frequency for each variant in each deme, without simulating the individuals directly, for computational tractability. This choice is appropriate under the assumption of independent sites; a coalescent simulation would be unable to model selection, and an individual-level forward simulation would be slow. In detail, we specified a mutation rate of  $\mu=1e-5$ , a migration rate of 0.01 between adjacent demes, and a population size that differed among demes as described above. We initialized allele frequencies as described in the previous section, and we simulated 1000 generations of selection, mutation, migration, and drift. In each generation, we updated the allele frequency of each variant as follows. Let  $p^{(k)}$  be the number-of-demes by 1 vector of allele frequencies for some variant at generation  $k$ :

1. Migration: set  $p^{(k)} = Gp^{(k-1)}$ , where  $G$  is the migration matrix
2. Mutation: set  $p^{(k)} = \mu(1 - p^{(k)}) + (1 - \mu)p^{(k)}$ , where  $\mu$  is the mutation rate
3. Selection: set  $p^{(k)} = \frac{p^{(k)}(1-s)}{1 - p^{(k)} + p^{(k)}(1-s)}$
4. Drift: sample  $p^{(k)} = \frac{x}{N}$ ,  $x \sim \text{binomial}(N, p^{(k)})$  where  $N$  is the population size.

For efficiency, we approximate Binomial sampling with either a  $\text{Poisson}(Np^{(k)})$  distribution, when  $p^{(k)} < 0.1$ , or a  $N(Np^{(k)}, Np^{(k)}(1 - p^{(k)}))$  distribution, when  $0.1 < p^{(k)} < 0.9$ .

### Observed-scale effect sizes for binary traits

For binary traits, we used raw allele counts in cases in controls as the input to BHR, rather than effect-size estimates from a mixed model. We calculated the observed-scale effect size of SNP  $x$  on the phenotype  $y$  as:

$$\hat{\beta} = \text{corr}(y, x) \quad \#(34)$$

$$= \frac{2 * (\text{MAF in cases} - \text{MAF}) * \text{prevalence}}{\sqrt{2 * \text{MAF} * (1 - \text{MAF}) * \text{prevalence} * (1 - \text{prevalence})}}.$$

We report heritability estimates on an observed scale unless noted otherwise.

### Genes analyzed

We analyzed 17,318 genes, a subset of the 19,407 genes in Genebass<sup>5</sup>. We analyzed genes meeting all of the following criteria: autosomal; LoF observed/expected ratio present in gnomAD<sup>37</sup>; cell type specific t-statistic defined in Finucane 2018 Nature Genetics; and at least one variant present in Genebass.

### Variant annotation and QC

We analyzed a set of 6976410 variants that passed the quality control checks described in Karczewski et al (2022) (Supplementary Table 6). We analyzed variants in four functional categories: predicted loss-of-function (pLoF), missense pathogenic, missense benign, and synonymous variants. pLoF variants were defined as in Genebass<sup>5</sup>, and included stop-gained, essential splice and frameshift variants. Missense functional classes were defined using PolyPhen2: we defined missense pathogenic as a PolyPhen2 variant annotation of “probably damaging” or “possibly damaging” and missense benign as a PolyPhen2 variant annotation of “benign”. Synonymous variants were defined as in Genebass.<sup>5</sup> We excluded a small number of variants that were not annotated as either pLoF, missense or synonymous in Genebass.

### Common-variant heritability estimates

We used GWAS summary statistics from the UK Biobank to facilitate a direct comparison with the phenotypes from exome-sequencing analysis (see URLs). Across 22 core BHR traits, the GWAS had a median effective sample size of 344104 (see Supplementary Table 8).

We used stratified LD Score Regression (S-LDSC)<sup>10,16</sup> to generate common variant heritability and genetic correlation estimates. We elected to use LDSC for direct comparison of heritability estimates because it employs a similar random-effects model to BHR. We used LD scores from the 1000 Genomes project<sup>64</sup> and annotations from the baseline LD model<sup>21</sup> (see URLs).

In order to estimate the fraction of common-variant heritability explained by significant genes (Figure 3), we used HESS, which is able to estimate the local heritability explained by regions with significant associations or significant genes<sup>32</sup>. We used an LD reference

panel from the 1000 Genomes Project<sup>64</sup> and a genome partition composed of approximately LD-independent blocks from Berisa et al<sup>65</sup>.

We used the Abstract Mediation Model (AMM<sup>34</sup>) to estimate the fraction of heritability mediated by gene sets. In brief, AMM estimates the fraction of heritability mediated by a gene set while accounting for uncertainty in SNP-gene mapping. Instead of relying on SNP-to-gene mapping using expression data like eQTLs, AMM first learns a genome-wide probabilistic SNP-to-gene mapping from the decay in heritability across gene proximity (i.e. 27% of heritability mediated by the closest gene). We applied AMM twice: to estimate the fraction of heritability mediated by BHR-significant genes (Figure 3D) and to estimate the enrichment of heritability mediated by constrained genes and gene sets defined by tissue and cell-type expression data (Figure 4A-C). We used a SNP-to-gene probability distributed learned from constrained genes in Weiner et al<sup>34</sup>, which are well-powered across a range of traits.

### Accounting for LD

Rare variants, and to a lesser extent ultra-rare variants, may have within-gene LD. Within-gene LD is a problem for BHR because it causes sampling errors to be correlated among alleles. In particular, if minor alleles have net-positive within gene LD, then their sampling errors will have net-positive correlations, just as true effects are expected to be correlated. This source of bias is potentially strong, as the sampling variance of the effect sizes is large. Net zero LD, which occurs when correlations are nonzero for particular alleles but zero on average, is less of a problem; it leads to decreased power, but not to bias. Outside-of-gene LD is also only a minor concern, as it is not expected to produce net positive correlations between different minor alleles in the same gene.

Net-positive within gene LD can occur as an ascertainment related artefact of binning on the within-sample allele frequency. Suppose that minor alleles within a gene have a mixture of positive and negative LD, such that their net LD is zero: that is,  $1^T R 1 = 0$  where  $1$  is the all ones vector and  $R$  is the population correlation matrix. Suppose that we sample  $n$  haplotypes  $X$ , compute their within-sample LD matrix  $\hat{R}$ , and bin them by their sample minor allele count. For a pair of variants  $ij$ , with correlation  $r_{ij}$ , consider the probability that they are both observed exactly  $n_i = n_j$  times. This probability is low if  $r_{ij}$  is zero or negative, even if their population allele frequencies are equal. It is much higher, however, if  $r_{ij} \approx 1$ , which would cause the sample allele frequencies of the two variants to be highly correlated. Conversely, when variants are observed at similar within-sample minor allele frequency, this ascertainment effect makes them more likely to be in positive LD.

If the amount of within-gene LD is known, it can be incorporated into BHR. Let the within-gene- $g$  LD matrix be  $R_g = n^{-1} E(Z_g^T Z_g)$ . If causal per-allele effect sizes have mean  $\mu_g$ , the mean of the marginal per-normalized-genotype effect size vector  $\beta_g$  is  $E(\beta_g | \mu_g) = R_g \mu_g w_g$ . The burden effect size is:

$$\gamma_g := \frac{E(b_g^T y)}{\sqrt{E(y^T y) E(b_g^T b_g)}} \quad \#(35)$$

$$= \frac{w_g^T \beta_g}{\sqrt{w_g^T R_g w_g}} \quad \#(36)$$

and its mean is:

$$E(\gamma_g | \mu_g) = \mu_g \sqrt{w_g^T R_g w_g}. \quad \#(37)$$

Dropping subscripts, the regression equation becomes:

$$E(\hat{\gamma}^2) = \mu^2 w^T R w + d \bar{p} \bar{q} + n^{-1} + a. \quad \#(38)$$

The intercept is unchanged. Let  $\hat{\beta}$  be the vector of sample correlations; their residuals have covariance

$$E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)^T\right) = (n^{-1} + a)R \quad \#(39)$$

so

$$\begin{aligned} E((\hat{\gamma} - \gamma)^2) &= \frac{w^T E\left((\hat{\beta} - \beta)(\hat{\beta} - \beta)^T\right) w}{w^T R w} \\ &= n^{-1} + a. \end{aligned} \quad \#(40)$$

The overdispersion term behaves the same way.

Equation 37 represents one principled approach to account for within-gene LD, but we were only able to access within-gene LD from UK Biobank for chromosomes 20-22 due to computational limitations of our LD estimation pipeline. Ultra-rare variants on these chromosomes have very little LD (Supplementary Table 3). We calculated the amount of LD related bias that is expected to be observed for each class of variants, assuming that the amount of net positive within-gene LD on chromosomes 20-22 are representative of the rest of the genome. Under the null ( $\beta = 0$ ), the expected burden statistic not accounting for LD is:

$$\begin{aligned} E(\hat{\gamma}^2) &= \frac{w^T E(\hat{\beta} \hat{\beta}^T) w}{w^T w} \\ &= (n^{-1} + a) \frac{w^T R w}{w^T w}. \end{aligned} \quad \#(41)$$

For chromosomes 20-22, we calculated a correction term

$$s = \frac{1}{\text{no.genes}} \sum_{\text{genes } g} \frac{w_g^T R_g w_g}{w_g^T w_g} - 1 \quad \#(42)$$

from synonymous variants in each allele frequency bin. The correction factor was noisy for some individual bins and there was no clear relationship between the allele frequency and the correction factor, so we computed a single precision-weighted mean  $\bar{s} = 4.6$  (s.e.=0.5) across bins (see Supplementary Table 22). Then, we subtracted  $\bar{s}$  from the BHR regression slope in order to obtain LD corrected heritability estimates; the corrected estimate is equal to

$$h_{\text{corrected}}^2 = h_{\text{uncorrected}}^2 - \bar{s} \sum_g | | w_g | |^2. \quad \#(43)$$

The correction is largest for rare synonymous and missense variants; it is much smaller for ultra-rare variants (which have much smaller entries of  $w$ ) and for pLoF variants (which are fewer in number) (see Supplementary Tables 6, 18). It is inconsequential for ultra-rare pLoFs and in analyses of ultra-rare pLoF variants outside of Figure 2b-c, we do not apply this correction to ultra-rare pLoF estimates.

### Alternative burden definitions

Burden heritability can be defined for any choice of burden weights, and in this section, we generalize BHR accordingly. Let  $W_g$  be the diagonal  $m_g \times m_g$  matrix whose diagonal entries are the standard deviations of each variant; under Hardy-Weinberg equilibrium, they are equal to  $\sqrt{2p(1-p)}$  (where  $p$  is the allele frequency). Let  $u_g$  be any vector of burden weights. The burden  $b_g$  is defined as:

$$b_g := X_g u_g = Z_g W_g u_g \quad \#(44)$$

The burden effect sizes are:

$$\gamma := \frac{W_g u_g \beta_g}{\sqrt{(W_g u_g)^T W_g u_g}} \quad \#(45)$$

and if the mean per-allele effect sizes are  $E(\beta_g | u_g) = \mu_g u_g$ , then:

$$E(\gamma_g | v_g) = \frac{W_g u_g E(\beta_g | u_g)}{\sqrt{(W_g u_g)^T W_g u_g}} = \mu_g \frac{(W_g u_g)^T W_g u_g}{\sqrt{(W_g u_g)^T W_g u_g}} = \mu_g \|W_g u_g\|. \quad \#(46)$$

The burden heritability is:

$$h_{\text{burden}}^2 = \sum_g \mu_g^2 | | W_g u_g | |^2.$$

In the definition of burden heritability above, with uniform burden weights,  $u_g$  is the all-ones vector, and  $\mu_g$  is the mean minor allele per-allele effect size.

The BHR regression equation is now:

$$E(\hat{\gamma}^2) = \tau \|W_g u_g\|^2 + d + n^{-1} + a \quad \#(47)$$

where  $\tau = E(\mu^2)$  and the burden heritability estimate is  $\tau \sum_g | |W_g u_g | |^2$ .

In the main text, we use uniform weights (e.g.  $u_g = 1_{m_g \times 1}$ ). This is the natural choice of weights for loss-of-function variants, which are expected to have similar effects; however, for missense variants, a different set of weights may be more appropriate. In particular, variants at higher allele frequencies may have smaller per-allele effect sizes, and one way to model this phenomenon is the “alpha model”<sup>28,66</sup>:

$$E(\beta^2 | p) \sim [2p(1 - p)]^\alpha \quad \#(48)$$

Under this model, the elements of  $u_g$  are  $[2p(1 - p)]^{\frac{\alpha}{2}}$ . Notably, our main BHR analyses correspond to  $\alpha = 0$ , the case where per-allele effect sizes are assumed to be independent of frequency. We estimate heritabilities for two alternative choices of  $\alpha$ :  $\alpha = -1$ , which maximizes the burden heritability when normalized effect sizes (i.e., correlations) are equal across variants, and  $\alpha = -0.5$ , an intermediate value.

### Burden heritability explained by exome-wide significant genes

To compute the heritability explained by exome-wide significant genes, we used significant pLoF burden associations (Bonferroni-corrected  $p < 0.05$ ) from Genebase, which were identified using SAIGE-gene<sup>67</sup>. We computed the ultra-rare pLoF burden statistics for these genes from the SAIGE variant-level effect size estimates (for binary traits, we used the case-control allele frequencies as described above).

When the power to detect a significant gene is smaller than one, its effect size estimate is upwardly biased due to winner’s curse<sup>33</sup>. Similarly, the fraction of heritability explained by significant genes is upwardly biased, especially when most significant genes are close to the significance threshold. We implemented a partial correction for winner’s curse that only depends on the test statistic of each significant gene (and the threshold). In detail, let  $Z_j^2$  be the  $\chi^2$  test statistic for significant gene  $j$ , and let  $T$  be the  $\chi^2$  significance threshold (with 1 degree of freedom). We compute the expected  $\chi^2$  statistic for a gene with non-centrality equal to  $Z_j$  conditional on passing the threshold:

$$X_j^2 = E(X^2 | X > T), \quad X \sim N(Z_j, 1). \quad \#(49)$$

We evaluate the expectation by sampling and computed the winners-curse-corrected test statistic as  $2Z_j^2 - X_j^2$ .



We tested this approach in simulations and determined that it corrects for about half of the observed winner's curse across the whole range of genetic architectures and sample sizes (Supplementary Figure 6). It is less successful in the presence of strong population stratification, which causes excess false positives. In real data, a complication is that the significance test is computed from a statistic that includes not only the ultra-rare pLoFs but other variants as well, and this might overcorrection for some genes.

### Genes sets

We analyzed two existing collections of cell type- or tissue-specific gene sets. First, we analyzed tissue-specific gene sets comprising the top 10% of genes differentially expressed in focal tissue vs. other tissues from GTEx v7 bulk RNA-seq<sup>17</sup>. Second, we analyzed cell type-specific gene sets constructed from single-cell RNAseq data<sup>18</sup>. In brief, genes were ranked based on expression in a given cell type relative to expression of the gene in different cell types in the same tissue. Based on the ranking, each gene-cell type pair was assigned a  $\chi^2$  statistic, the statistics were min-max normalized to the range [0,1], and genes with normalized values of 1 were assigned to the gene set.

### Burden genetic correlation

Between two traits, the burden genetic covariance is defined as:

$$\rho_{\text{burden}} = \sum_{\text{genes } g} |w_g|^2 E(\mu_{1g}\mu_{2g}) \quad \#(50)$$

where  $\mu_{1g}, \mu_{2g}$  are the mean minor-allele effect size for gene  $g$  and traits 1 and 2 respectively. The *burden genetic correlation* is:

$$r_{\text{burden}} = \frac{\rho_{\text{burden}}}{\sqrt{h_{\text{burden}}^2(\text{trait 1}) h_{\text{burden}}^2(\text{trait 2})}} \quad \#(51)$$

To estimate the burden genetic covariance, the cross-trait BHR regression equation is:

$$\begin{aligned} E(\hat{\gamma}_1 \hat{\gamma}_2) &= \tau \|w\|^2 + \text{intercept}, \\ \text{intercept} &= \rho \frac{n_s}{n_1 n_2} + a_{12} + d_{12}. \end{aligned} \quad \#(52)$$

The regression slope is  $\tau = E(\mu_1, \mu_2)$ , and we stratify the regression across gene sets in the same manner as the single-trait case. In the intercept (similar to cross-trait LDSC<sup>12</sup>),  $\rho$  is the phenotypic correlation,  $n_s$  is the number of samples that are shared between the two studies,  $n_1$  and  $n_2$  are the number of individuals in each study,  $a_{12}$  is the covariance of the stratification effects on the two traits, and  $d_{12}$  is the covariance of the overdispersion effects on the two traits.

When the two traits have different sets of variants because there are different individuals for each study,  $\|w\|^2$  is replaced by  $\|w_1\| \|w_2\|$ , where  $w_k$  is the burden weights

vector for trait  $k$ . The same approach is used when computing the correlation between missense and pLoF effects.

With the estimated regression slope  $\tau$ , the estimated genetic covariance is:

$$\hat{\rho}_{\text{burden}} = \sum_{\text{genes } g} A_g \tau \|w_{g1}\| \|w_{g2}\| \quad \#(53)$$

and the genetic correlation is estimated using equation 45.

### Genetic correlation of common coding variants

We computed common variant genetic correlations restricted to coding variants using the implementation of sLDSC implemented in the GenomicSem package<sup>68</sup> noting that the main release of sLDSC does not implement a stratified genetic covariance estimator. We used GenomicSem to compute zero-order genetic covariance and sampling matrices for all annotations in the Baseline LD Model v2.2, and used the zero-order genetic covariance matrix from the “Coding\_UCSC” annotation to compute the common coding genetic correlation.

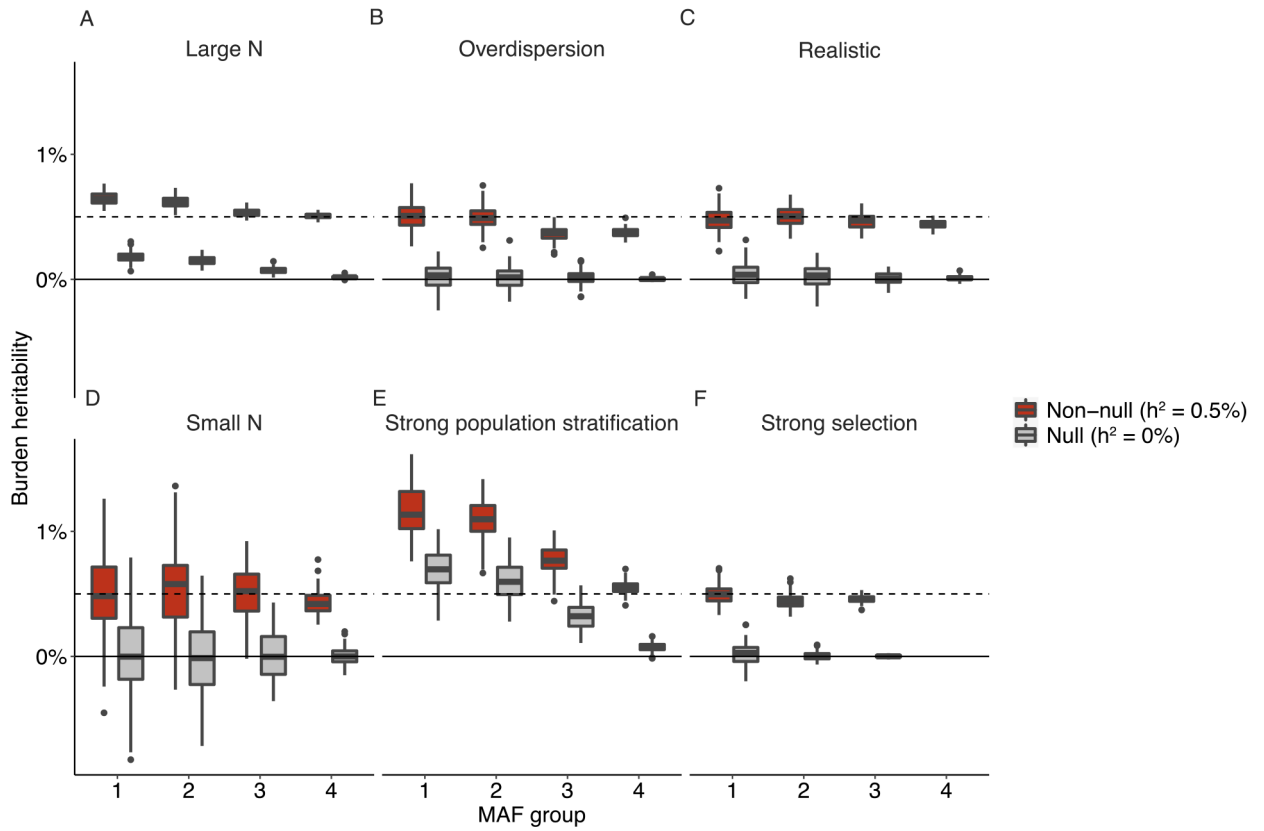
Noting that genetic correlation estimates can be unstable in the setting of low or negative heritability estimates, we only computed coding genetic correlations for pairs of traits where both traits had nominally significant (i.e.,  $z > 1.96$ ) coding heritability estimates. 28 of 37 examined traits had significant coding heritability estimates.

### SCHEMA and BipEx Datasets

We used publicly available variant-level counts data from the SCHEMA<sup>6</sup> and BipEx<sup>7</sup> as input data (URLs). We restricted the SCHEMA analyses to the two study strata with largest sample size: EUR (Exomes, Nextera) and EUR (Exomes, Non-Nextera) (see supplementary information of Singh et al, 2022). For the BipEx dataset, we used the “Bipolar Disorder” group counts. Following Singh et al, 2022, we restricted to variants with minor allele count (MAC) less than 5, and performed separate analyses for pLoF, damaging missense (MPC > 2), and synonymous variants. For each cohort, burden statistics were calculated from allele counts using Equation 33, and burden scores were computed from sample allele frequencies. Then, we used BHR to compute burden heritabilities, enrichments, and genetic correlations separately for the two SCHEMA cohorts. We used this approach to avoid confounding due to differences in the sequencing technology and the sample prevalence between the cohorts.

To produce a single estimate for the schizophrenia heritability, we performed a precision-weighted meta-analysis across the two cohorts. We used BHR to compute the total burden heritability, as well as the burden heritability for constrained genes (the top 5th of genes by observed/expected LoF counts from gnomAD). Within each stratum, we computed the variances for these two estimates, as well as their covariance, using a block jackknife. We used the per-stratum heritability estimates and covariance matrices to perform a precision-weighted meta-analysis. We also computed the jackknife covariance matrix of the heritability estimates for each constraint bin, and used this matrix with the delta method to calculate the standard error for the enrichment of heritability in constrained genes.

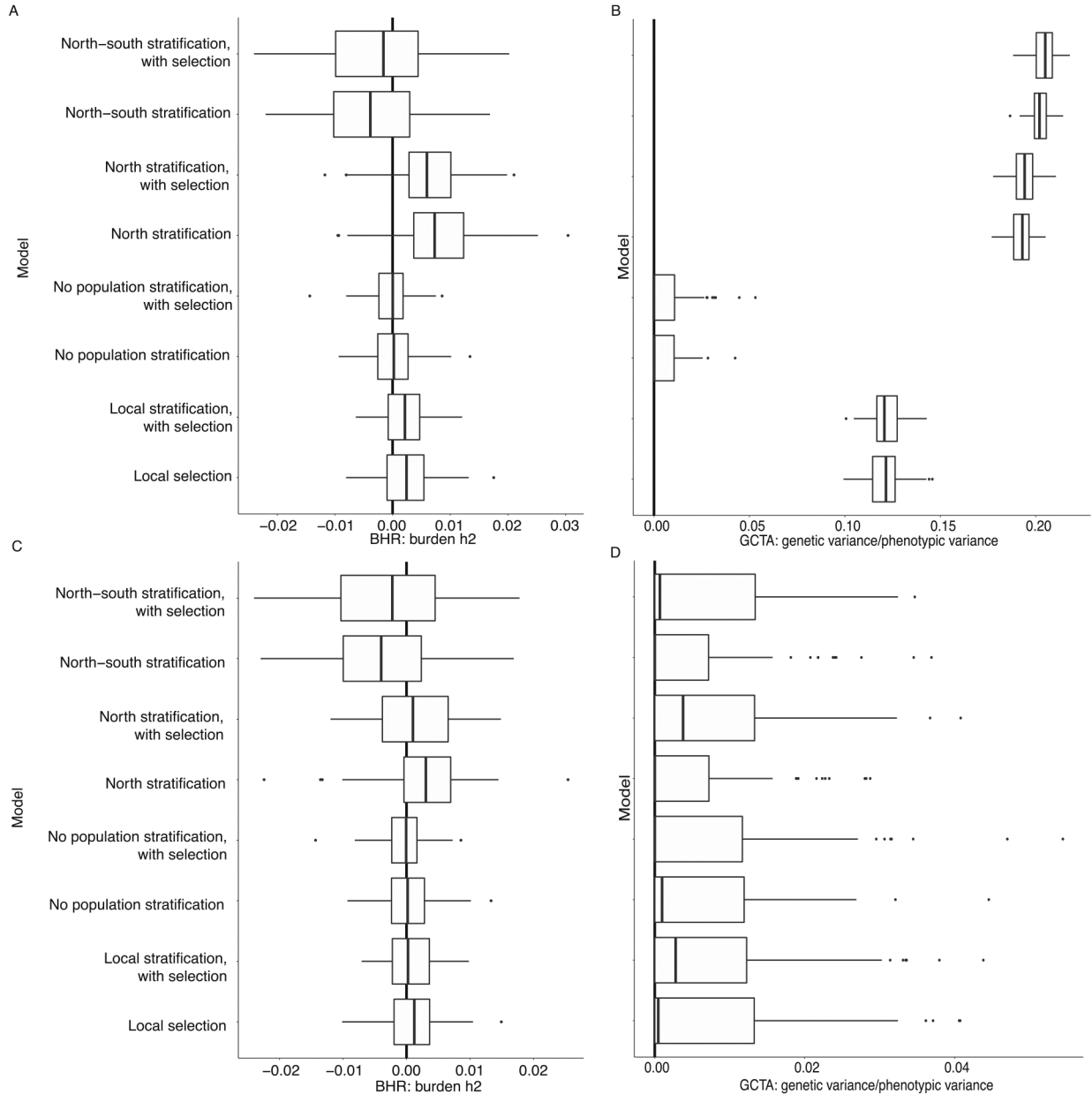
## Extended Data



### Extended Data Figure 1: Performance of BHR in exome-scale simulations with no individual-level data

We performed an extended set of simulations to assess the performance of BHR. The MAF groups are  $< 1e-5$  (group 1),  $1e-5 - 1e-4$  (group 2),  $1e-4 - 1e-3$  (group 3), and  $1e-3 - 1e-4$  (group 4), respectively; the gray and red boxplots indicate the distribution of estimates in null and non-null simulations (true burden  $h^2 = 0\%$ ,  $0.5\%$  respectively). A minor difference in the way that BHR was applied to simulated vs. real data is that in simulated data, significant genes were identified without any attempt to correct for population stratification, whereas in our real-trait analyses, they were identified using SAIGE-GENE.<sup>1</sup> We started with a realistic set of parameters (see Methods) and varied one simulation parameter in each simulation. (A) We increased the sample size from  $5e5$  to  $2e6$ . This increase amplifies the uncorrected population stratification, causing false positive significant genes and upward bias in BHR (no bias is observed in estimates without significant genes). (B) We added overdispersion effects with the same distribution of effect sizes as the burden effects, i.e. with per-allele effect size variance drawn from a discrete mixture distribution (see Methods). This distribution differs from the BHR model, which assumes that overdispersion effects have a constant per-s.d. effect size variance, but this form of misspecification does not lead to bias. (C) We performed simulations with realistic parameters, including stratification and selection (see Methods and Figure 1C). (D) We decreased the sample size from  $5e5$  to  $1e5$ . (E) We increased the strength of population stratification (including the minor-allele biased stratification) by a factor of 10, from a per-s.d. effect size mean of  $1e-7$  and a variance of

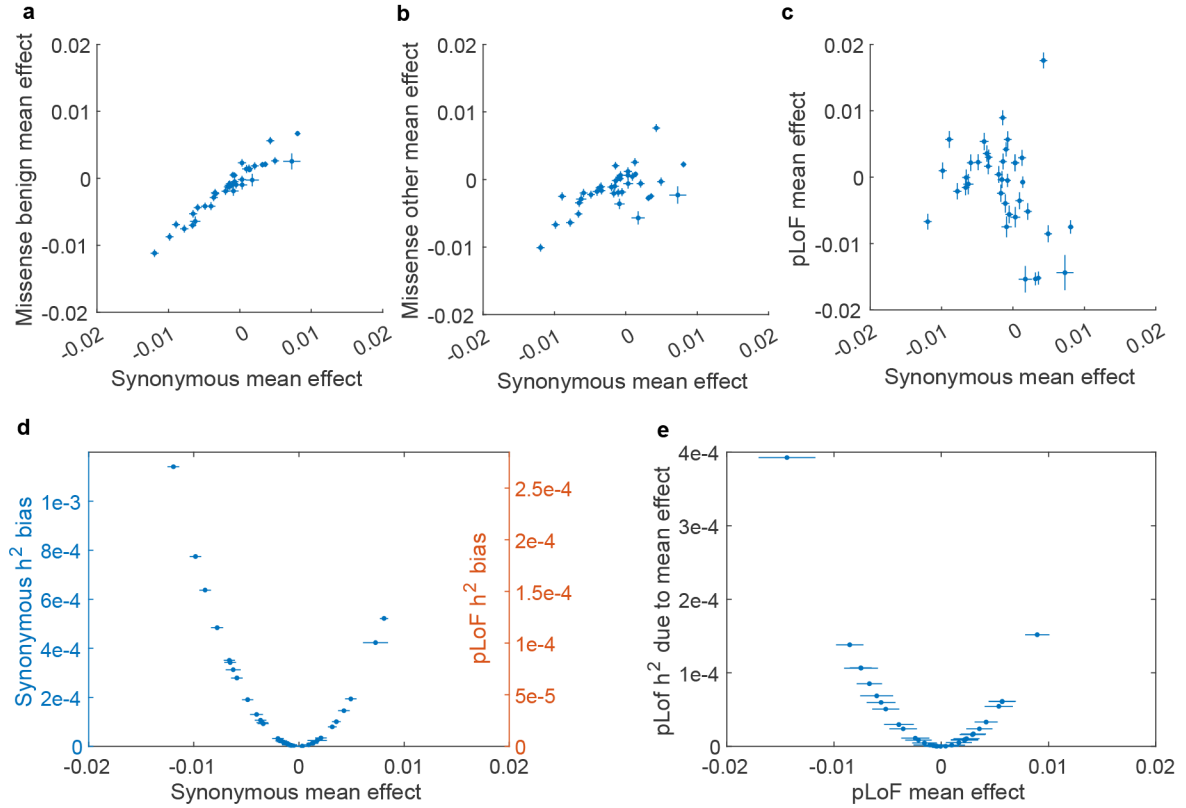
$1e-5$  to a mean of  $1e-6$  and a variance of  $1e-4$ . **(F)** We increased the strength of selection, from mean  $N_s=1$  to mean  $N_s=10$ . There were extremely few variants with allele frequency greater than  $1e-3$ , so MAF group 4 estimates are not shown. Numerical results are contained in Supplementary Table 4. Boxplots denote median, quartiles and range of distribution (excepting outliers).



**Extended Data Figure 2: Comparison of BHR and GCTA in null simulations with individual-level genotypes and phenotypes, and different patterns of population stratification**

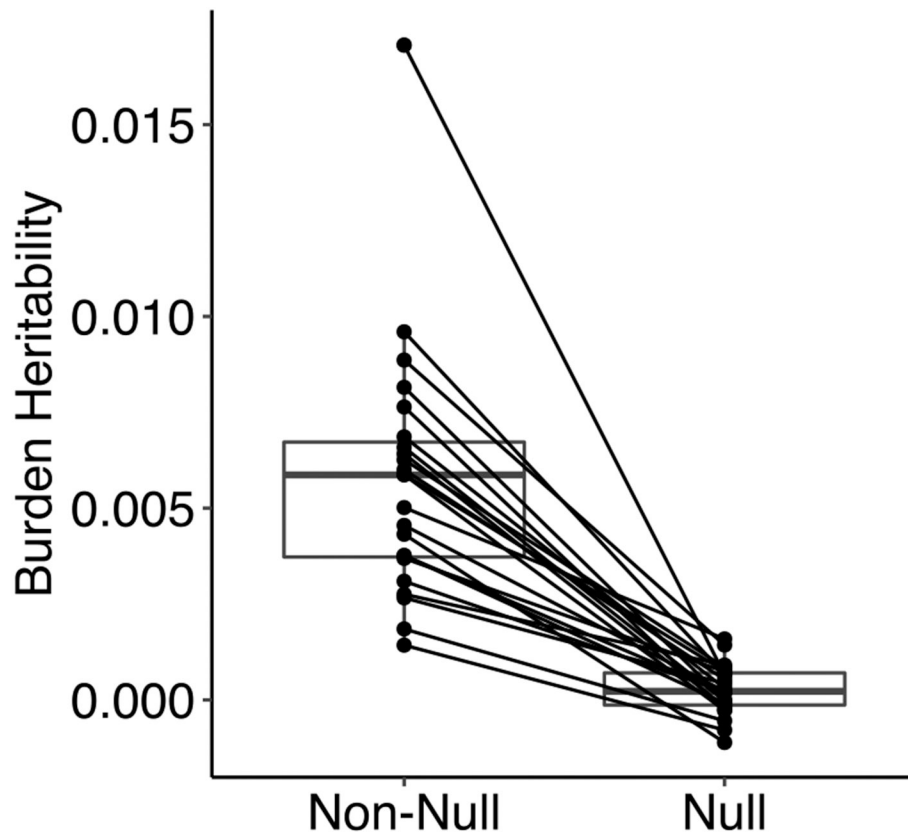
There are four demographic models: no stratification; north-south stratification; north-south stratification with smaller population size in the northern deme; and local stratification with very small population size in one deme (see Methods). Under each model, we

performed simulations with and without selection, mimicking pLoF and synonymous variants respectively. (a) BHR burden heritability estimates with no correction for minor allele-biased stratification. (b) GCTA heritability estimates with no correction for ancestry. (c) BHR burden heritability estimates, correcting for minor allele-biased stratification. (d) GCTA heritability estimates, correcting for ancestry by providing the deme from which each individual was sampled as a covariate. Boxplots denote median, quartiles and range of distribution (excluding outliers).



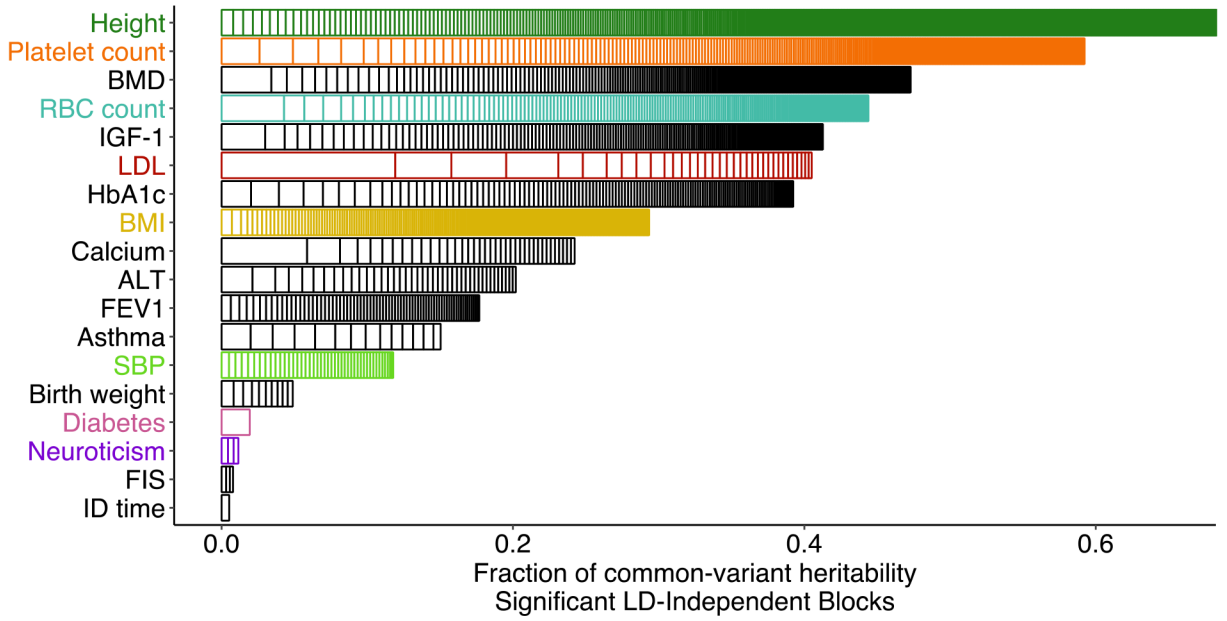
### Extended Data Figure 3: Genome-wide mean minor allele effect sizes

We define the “mean effect” as the effect size of the genome-wide burden, summing all minor alleles across genes within a category, on the phenotype. For synonymous variants, a nonzero mean effect is interpreted as evidence of minor-allele biased population stratification, and this type of stratification produces upward bias in BHR heritability estimates (see Methods). (a-c) Mean effect of synonymous variants vs. mean effect of missense benign, missense other, and pLoF variants respectively. The lack of correlation in (c) suggests that for pLoFs, the nonzero mean effect is mostly biological. (d) Mean effect of synonymous variants vs. the resulting bias in heritability estimates, for synonymous variants (left y axis) or for pLoFs (right y axis). These differ by a constant factor due to the larger number of synonymous variants than pLoFs. (e) Mean effect of pLoF variants vs. the contribution of these effects to burden heritability. These estimates are a small fraction of the total pLoF burden heritability. Error bars represent standard errors, which are computed by assuming independence across genes.



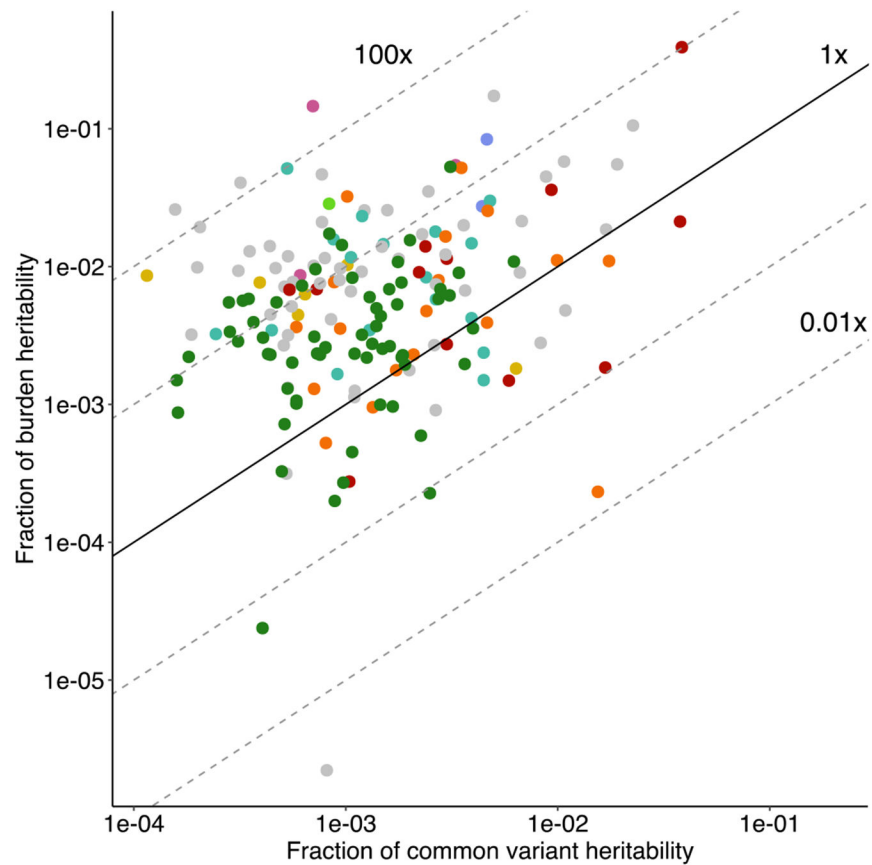
**Extended Data Figure 4: Burden heritability estimates with effect-allele-permuted burden statistics**

We assessed the potential for confounding in our results by repeating our analyses with ultra-rare pLoF burden statistics whose effect alleles were randomly permuted. This permutation is expected to eliminate the burden heritability while not affecting any form of confounding that is symmetrical with respect to the minor vs. major allele. Boxplots indicate the distribution of burden heritability estimates before and after the permutation (non-null and null, respectively), with median, quartiles and range (excepting outliers).



**Extended Data Figure 5: Proportion of common variant heritability explained by LD-independent blocks with significant heritability**

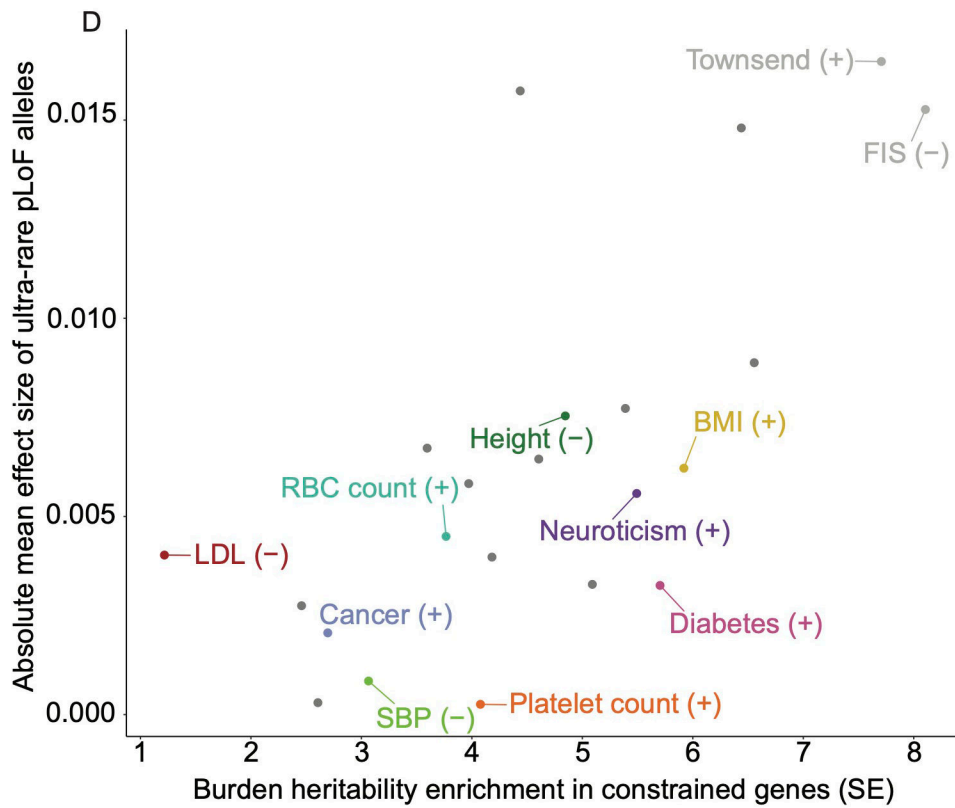
For each trait, we used HESS to identify which of the 1651 LD-independent blocks from Berisa<sup>2</sup> have Bonferroni-significant heritability, and then computed the proportion of the overall HESS heritability mediated by each block. Although these blocks aggregate over many variants in many genes, the proportion of heritability explained by individual significant blocks is still less than the proportion of burden heritability explained by individual significant genes in BHR (Extended Data Figure 4).



**Extended Data Figure 6: Comparison of burden versus common variant heritability explained by exome-wide significant genes**

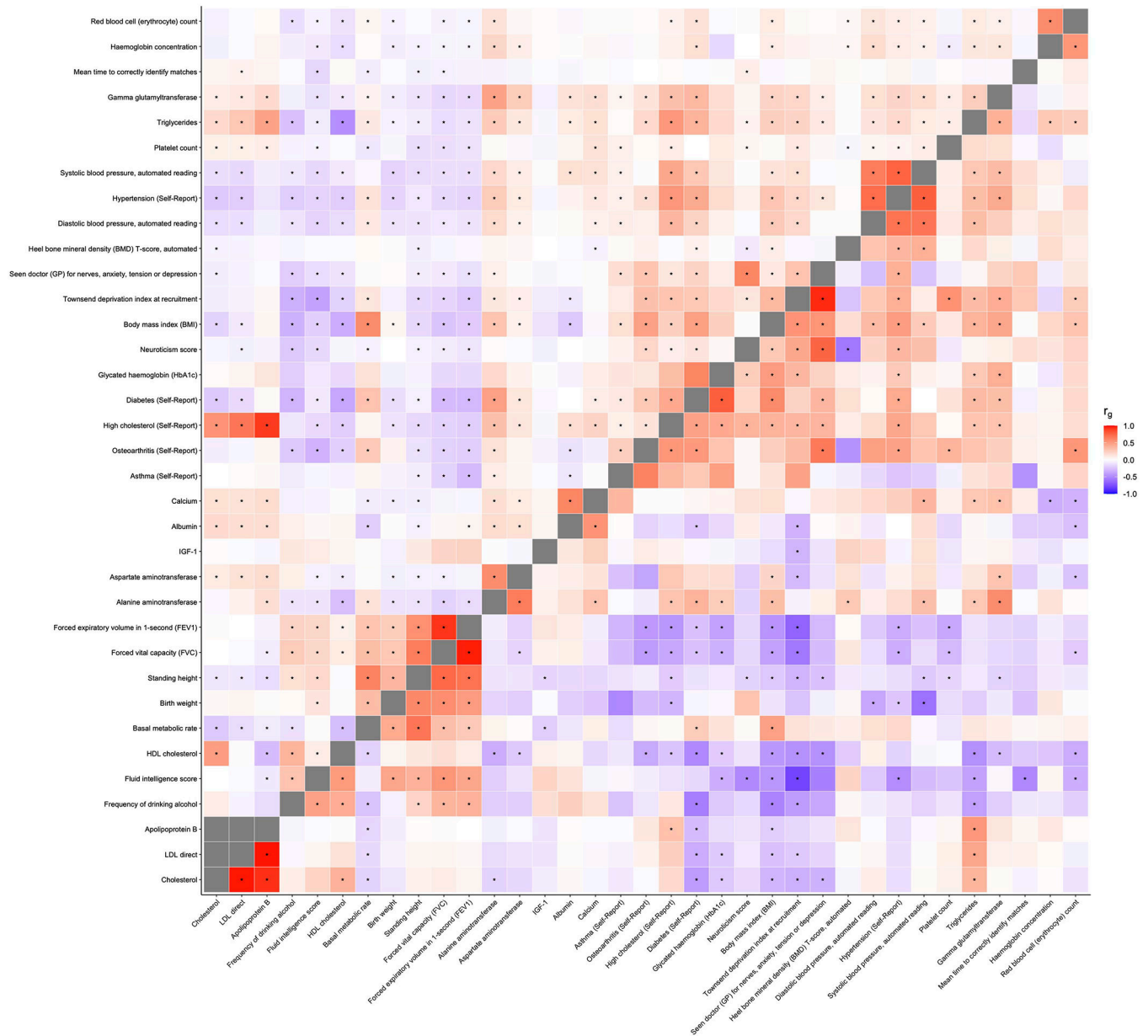
Each point represents a trait-gene significant burden association from the Genebase dataset. X axis values are the fraction of common variant heritability (estimated with HESS) explained by the LD-independent block containing that gene. Y axis values are the fraction of burden heritability (estimated with BHR) explained by the significant gene.





**Extended Data Figure 7: Absolute mean minor allele effect size of ultra-rare pLoF variants genome wide, vs. the constrained gene enrichment of each trait**

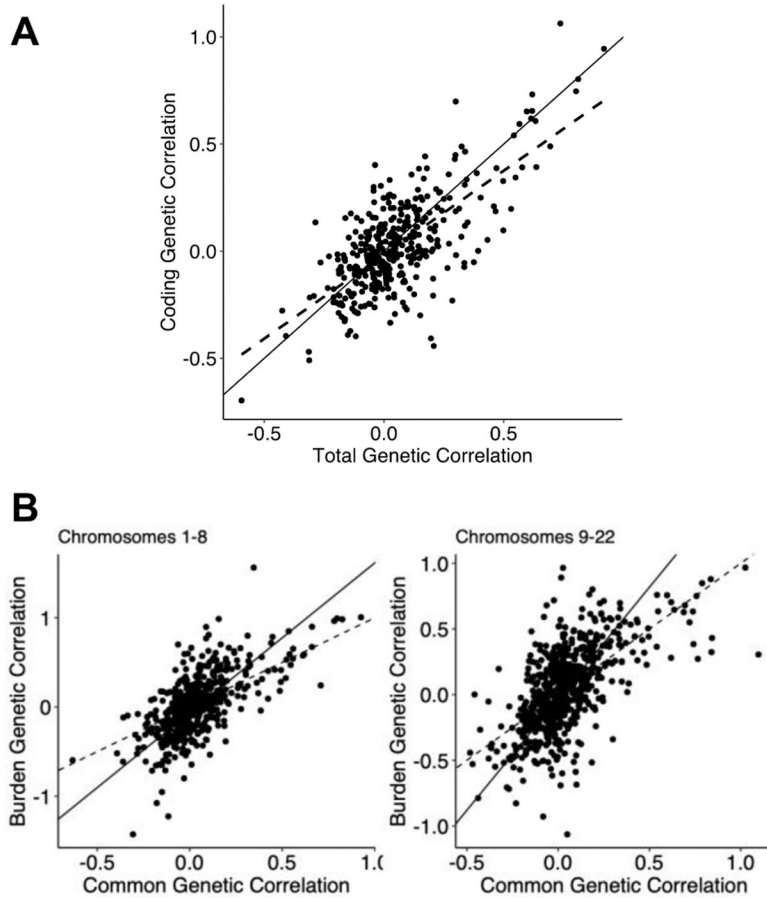
(+) and (-) denote the sign of the mean minor allele effects. For numerical results, see Supplementary Tables 7, 16, and 17.



**Extended Data Figure 8: Genetic correlation estimates across 37 traits, for common variants (upper triangle) and rare coding variants (lower)**

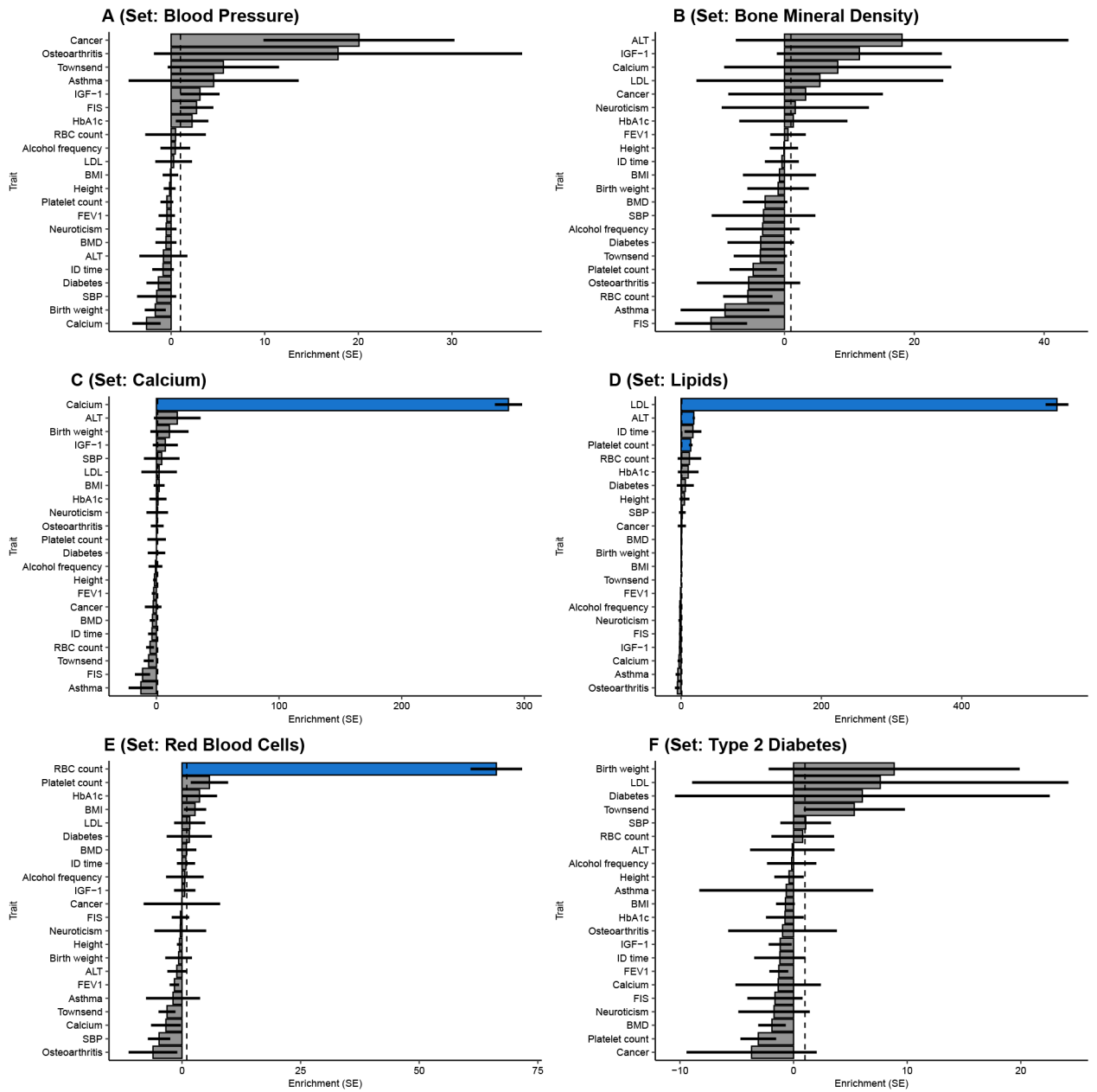
Asterisks indicate nominally significant genetic correlation estimates (two-tailed  $p < 0.05$ ).

Gray boxes not on the diagonal indicate cross-trait LDSC point estimates that are outside of  $[-1.25, 1.25]$ , which cross-trait LDSC does not report by default. For numerical results, see Supplementary Table 19.



**Extended Data Figure 9: Comparison of common coding vs. common whole-genome genetic correlations**

(a) We evaluated whether common coding variants, similar to rare coding variants, have stronger genetic correlations than common variants overall. The fit line indicates the Deming regression slope, which allows for uncertainty in both the X and Y axis values. (b-c) To assess the stability of the Deming regression slope, we separately analyzed chromosomes 1-8 and chromosomes 9-22. (d-e) We also assessed the stability of the Deming regression slope for the burden genetic correlation vs. the common-variant genetic correlation on chromosomes 1-8 and chromosomes 9-22.



**Extended Data Figure 10: Burden heritability enrichments of drug target gene sets**

We used BHR to estimate the ultra-rare loss-of-function burden heritability enrichment in sets of manually curated drug target genes from a previous publication<sup>6</sup>. For all panels, error bars are standard errors, and bars are shaded in blue if the enrichment is significantly greater than 1. (A) Burden heritability enrichment in  $n = 14$  blood pressure drug target genes (union of diastolic and systolic blood pressure gene sets from reference publication). (B) Burden heritability enrichment in  $n = 8$  bone mineral density drug target genes. (C) Burden heritability enrichment in  $n = 6$  calcium drug target genes. (D) Burden heritability enrichment in  $n = 10$  lipid drug target genes (union of LDL and triglyceride gene sets from reference publication). (E) Burden heritability enrichment in  $n = 6$  red blood cell drug target genes. (F) Burden heritability enrichment in  $n = 7$  type 2 diabetes drug target genes.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

The authors are grateful for support from National Institute Mental Health (F30MH129009 to Daniel J. Weiner), National Library of Medicine (T15LM007092 to Daniel Weiner), National Institute of General Medical Science (T32GM007753 to Ajay Nadig), Simons Foundation Autism Research Initiative (704413 to Elise Robinson and Luke O'Connor), and the Broad Institute. We are also grateful to Steven Gazal, Daniel King, Alkes Price, Kaitlin Samocha for analytic assistance and helpful comments on this manuscript. We are grateful to Jinjie Duan for identifying an issue in the first draft of our manuscript.

## Data availability

All data used in this manuscript is publicly available and documented in Supplementary Tables. All results are available in the Supplementary Tables. Neale Lab UKB GWAS summary statistics: <http://www.nealelab.is/uk-biobank/>. Genebass summary statistics: <https://app.genebass.org>. SCHEMA: <https://schema.broadinstitute.org>. BipEx: <https://bipex.broadinstitute.org>. Differentially expressed gene sets: <https://alkesgroup.broadinstitute.org>. Gene-level constraint data: <https://gnomad.broadinstitute.org>. COSMIC cancer gene sets: <https://cancer.sanger.ac.uk/census>.

## Code availability

BHR v0.1.0 is implemented in R, and its source code is publicly available at <https://github.com/ajaynadig/bhr>, with DOI [10.5281/zenodo.7382799](https://doi.org/10.5281/zenodo.7382799). We have also published scripts allowing the results of the manuscript to be reproduced using publicly available data (Data availability); these are implemented in R, Python, Hail, and MATLAB. AMM: <https://github.com/danjweiner/AMM21>. LDSC v1.0.1: <https://github.com/bulik/ldsc>. HESS v0.5.3: <https://huwenboshi.github.io/hess/>. Genomic SEM v0.0.5c: <https://github.com/GenomicSEM/GenomicSEM>. GCTA v1.94.1: <https://yanglab.westlake.edu.cn/software/gcta/#GREMLanalysis>.

## Main Text References

1. Sun BB et al. Genetic associations of protein-coding variants in human disease. *Nature* 603, 95–102 (2022). [PubMed: 35197637]
2. Wang Q et al. Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* 597, 527–532 (2021). [PubMed: 34375979]
3. Backman JD et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* 599, 628–634 (2021). [PubMed: 34662886]
4. Claussnitzer M et al. A brief history of human disease genetics. *Nature* 577, 179–189 (2020). [PubMed: 31915397]
5. Karczewski KJ et al. Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genomics* 2, 100168 (2022). [PubMed: 36778668]
6. Singh T et al. Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* 604, 509–516 (2022). [PubMed: 35396579]
7. Palmer DS et al. Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia. *Nat. Genet* 54, 541–547 (2022). [PubMed: 35410376]

8. International Schizophrenia Consortium et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748–752 (2009). [PubMed: 19571811]
9. Yang J et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet* 42, 565–569 (2010). [PubMed: 20562875]
10. Bulik-Sullivan BK et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet* 47, 291–295 (2015). [PubMed: 25642630]
11. 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 (2015). [PubMed: 26323059]
12. Bulik-Sullivan B et al. An atlas of genetic correlations across human diseases and traits. *Nature Genetics* vol. 47 1236–1241 Preprint at 10.1038/ng.3406 (2015). [PubMed: 26414676]
13. Brainstorm Consortium et al. Analysis of shared heritability in common disorders of the brain. *Science* 360, (2018).
14. Watanabe K et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet* 51, 1339–1348 (2019). [PubMed: 31427789]
15. Gusev A et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet* 95, 535–552 (2014). [PubMed: 25439723]
16. Finucane HK et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet* 47, 1228–1235 (2015). [PubMed: 26414678]
17. Finucane HK et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet* 50, 621–629 (2018). [PubMed: 29632380]
18. Jagadeesh KA et al. Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat. Genet* 54, 1479–1492 (2022). [PubMed: 36175791]
19. O'Connor LJ et al. Extreme Polygenicity of Complex Traits Is Explained by Negative Selection. *Am. J. Hum. Genet* 105, 456–476 (2019). [PubMed: 31402091]
20. Zeng J et al. Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet* 50, 746–753 (2018). [PubMed: 29662166]
21. Gazal S et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet* 49, 1421–1427 (2017). [PubMed: 28892061]
22. Gazal S et al. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat. Genet* 50, 1600–1607 (2018). [PubMed: 30297966]
23. Liu DJ & Leal SM Estimating genetic effects and quantifying missing heritability explained by identified rare-variant associations. *Am. J. Hum. Genet* 91, 585–596 (2012). [PubMed: 23022102]
24. Wainschtein P et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat. Genet* 54, 263–273 (2022). [PubMed: 35256806]
25. Bycroft C et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209 (2018). [PubMed: 30305743]
26. Lee S, Abecasis GR, Boehnke M & Lin X Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet* 95, 5–23 (2014). [PubMed: 24995866]
27. Adzhubei IA et al. A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249 (2010). [PubMed: 20354512]
28. Speed D, Hemani G, Johnson MR & Balding DJ Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet* 91, 1011–1021 (2012). [PubMed: 23217325]
29. Jang S-K et al. Rare genetic variants explain missing heritability in smoking. *Nat Hum Behav* (2022) doi:10.1038/s41562-022-01408-5.
30. Devlin B & Roeder K Genomic control for association studies. *Biometrics* 55, 997–1004 (1999). [PubMed: 11315092]
31. Loh P-R et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet* 47, 1385–1392 (2015). [PubMed: 26523775]
32. Shi H, Kichaev G & Pasaniuc B Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am. J. Hum. Genet* 99, 139–153 (2016). [PubMed: 27346688]

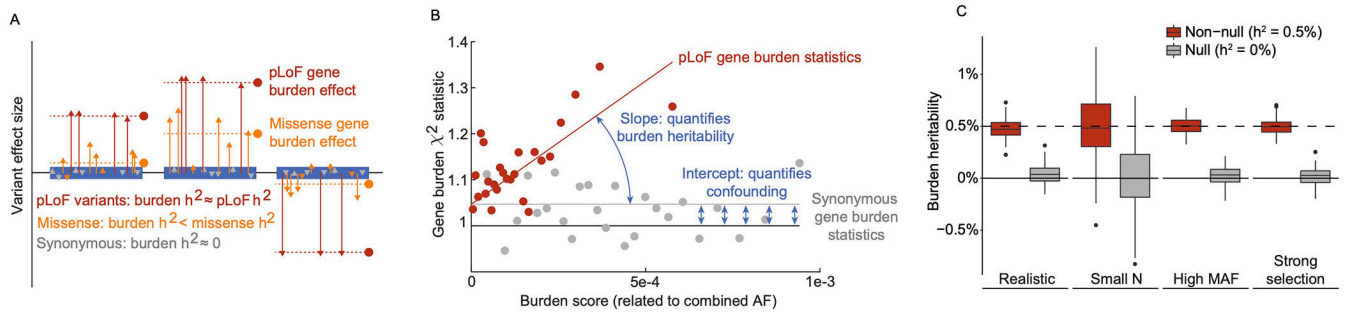
33. Palmer C & Pe'er I Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. *PLoS Genet.* 13, e1006916 (2017). [PubMed: 28715421]
34. Weiner DJ, Gazal S, Robinson EB & O'Connor LJ Partitioning gene-mediated disease heritability without eQTLs. *Am. J. Hum. Genet* 109, 405–416 (2022). [PubMed: 35143757]
35. Sondka Z et al. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* 18, 696–705 (2018). [PubMed: 30293088]
36. Boyle EA, Li YI & Pritchard JK An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169, 1177–1186 (2017). [PubMed: 28622505]
37. Karczewski KJ et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443 (2020). [PubMed: 32461654]
38. Lek M et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291 (2016). [PubMed: 27535533]
39. Mostafavi H, Spence JP, Naqvi S & Pritchard JK Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery. *bioRxiv* 2022.05.07.491045 (2022) doi:10.1101/2022.05.07.491045.
40. Gardner EJ et al. Reduced reproductive success is associated with selective constraint on human genes. *Nature* 603, 858–863 (2022). [PubMed: 35322230]
41. Sebat J et al. Strong association of de novo copy number mutations with autism. *Science* 316, 445–449 (2007). [PubMed: 17363630]
42. Sanders SJ et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485, 237–241 (2012). [PubMed: 22495306]
43. Purcell SM et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 506, 185–190 (2014). [PubMed: 24463508]
44. Simons YB, Bullaughey K, Hudson RR & Sella G A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS Biol.* 16, e2002985 (2018). [PubMed: 29547617]
45. Fu JM et al. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat. Genet* 54, 1320–1331 (2022). [PubMed: 35982160]
46. Border R et al. Cross-trait assortative mating is widespread and inflates genetic correlation estimates. *Science* 378, 754–761 (2022). [PubMed: 36395242]
47. Genovese G et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci* 19, 1433–1441 (2016). [PubMed: 27694994]
48. Kosmicki JA et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat. Genet* 49, 504–510 (2017). [PubMed: 28191890]
49. Baselmans BML, Yengo L, van Rheenen W & Wray NR Risk in Relatives, Heritability, SNP-Based Heritability, and Genetic Correlations in Psychiatric Disorders: A Review. *Biol. Psychiatry* 89, 11–19 (2021). [PubMed: 32736793]
50. Samocha KE et al. Regional missense constraint improves variant deleteriousness prediction. *bioRxiv* 148353 (2017) doi:10.1101/148353.
51. Lefebvre S et al. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 80, 155–165 (1995). [PubMed: 7813012]
52. Mendell JR et al. Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *N. Engl. J. Med* 377, 1713–1722 (2017). [PubMed: 29091557]
53. Pritchard JK Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet* 69, 124–137 (2001). [PubMed: 11404818]
54. Kim SS et al. Genes with High Network Connectivity Are Enriched for Disease Heritability. *Am. J. Hum. Genet* 104, 896–913 (2019). [PubMed: 31051114]
55. Forgetta V et al. An effector index to predict target genes at GWAS loci. *Hum. Genet* 141, 1431–1447 (2022). [PubMed: 35147782]
56. Liu X, Li YI & Pritchard JK Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* 177, 1022–1034.e6 (2019). [PubMed: 31051098]

57. Khera AV et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet* 50, 1219–1224 (2018). [PubMed: 30104762]
58. Fahed AC et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat. Commun* 11, 3635 (2020). [PubMed: 32820175]
59. Khera AV et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell* 177, 587–596.e9 (2019). [PubMed: 31002795]
60. Biddinger KJ et al. Rare and Common Genetic Variation Underlying the Risk of Hypertrophic Cardiomyopathy in a National Biobank. *JAMA Cardiol* (2022) doi:10.1001/jamacardio.2022.1061.
61. Bishop SL, Thurm A, Robinson E & Sanders SJ Prevalence of returnable genetic results based on recognizable phenotypes among children with autism spectrum disorder. *bioRxiv* (2021) doi:10.1101/2021.05.28.21257736.
62. Martin AR et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet* 51, 584–591 (2019). [PubMed: 30926966]
63. Fry A et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am. J. Epidemiol* 186, 1026–1034 (2017). [PubMed: 28641372]

## Methods References

64. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* 526, 68–74 (2015). [PubMed: 26432245]
65. Berisa T & Pickrell JK Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* 32, 283–285 (2016). [PubMed: 26395773]
66. Schoech AP et al. Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nat. Commun* 10, 790 (2019). [PubMed: 30770844]
67. Zhou W et al. Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. *Nat. Genet* 52, 634–639 (2020). [PubMed: 32424355]
68. Grotzinger AD et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav* 3, 513–525 (2019). [PubMed: 30962613]



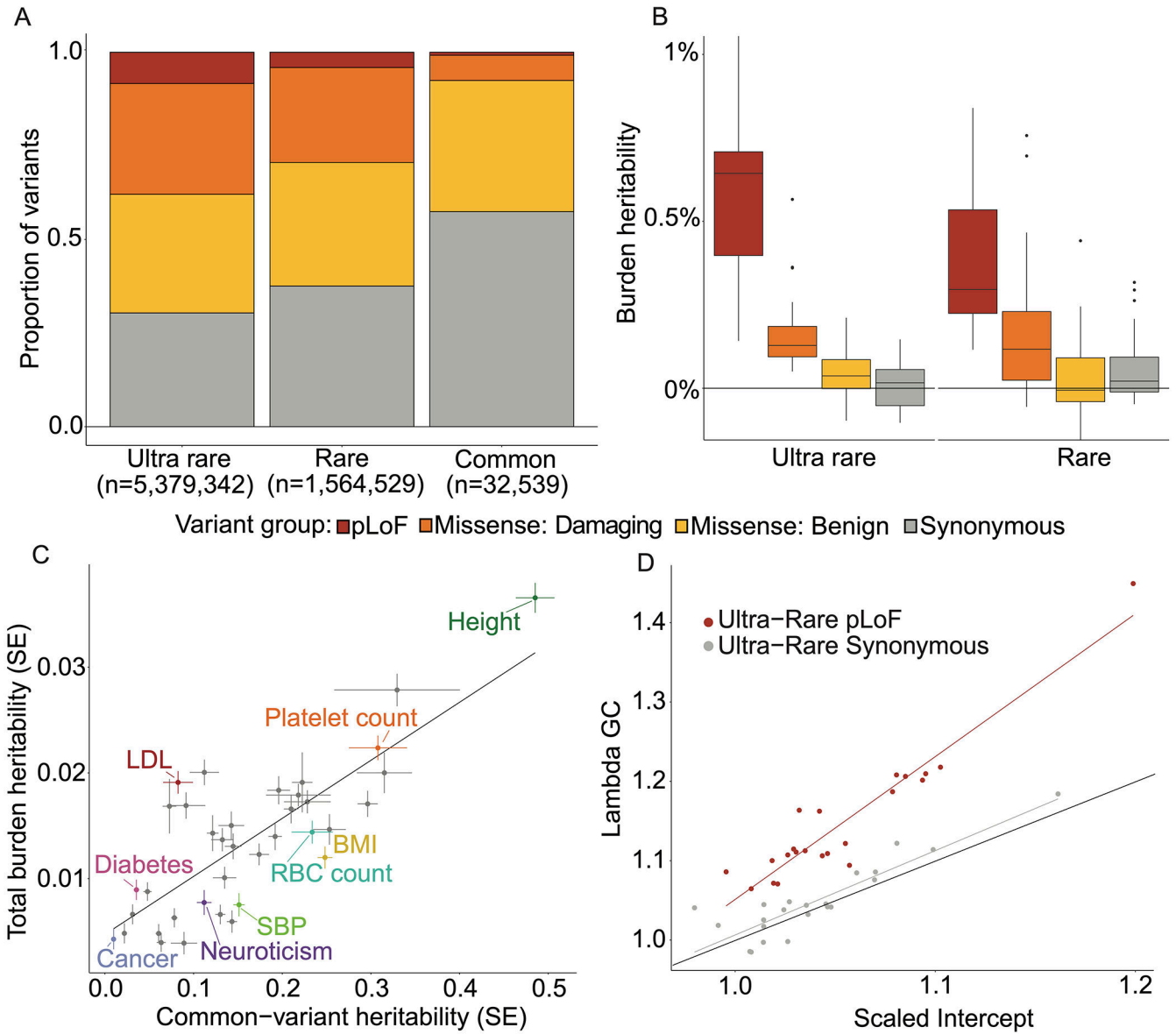


**Figure 1: Overview of Burden Heritability Regression (BHR)**

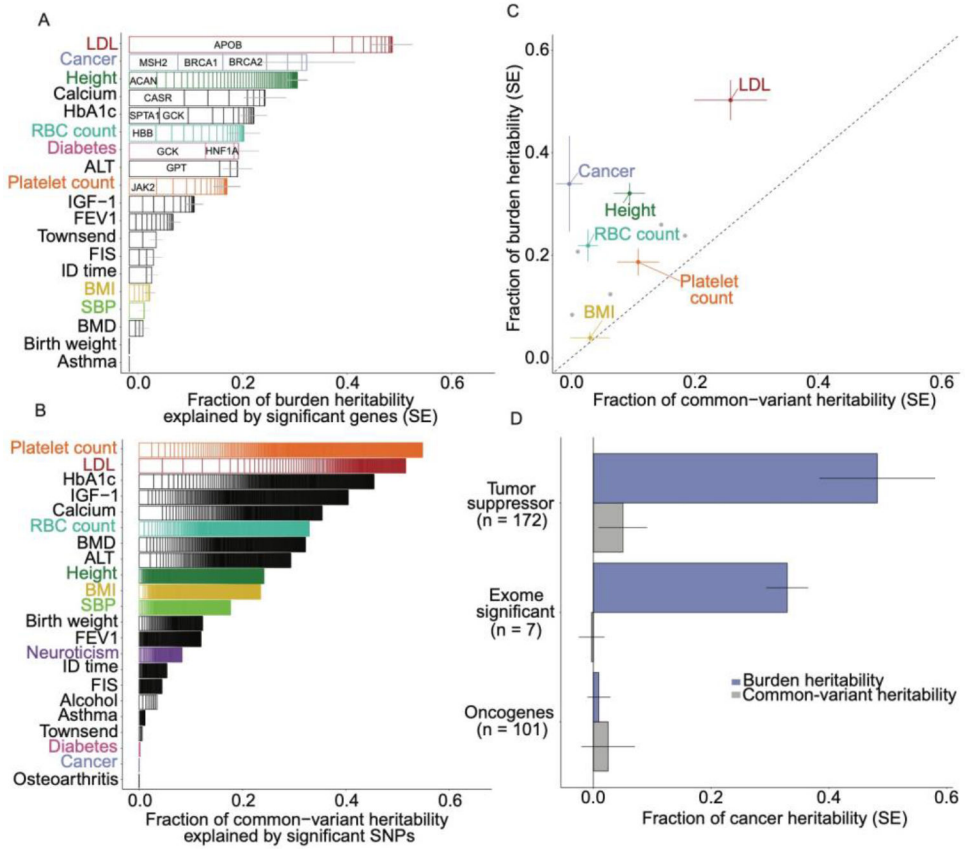
(A) The *burden heritability* of a gene is determined by its mean minor-allele effect size (dashed lines) and its “burden score,” which is approximately the combined allele frequency.

(B) BHR regresses gene burden statistics on gene burden scores, and the burden heritability estimate is proportional to the regression slope. We plot the mean burden statistic within burden score bins for ultra-rare pLoF/synonymous variants and LDL cholesterol levels (Supplementary Tables 1-2).

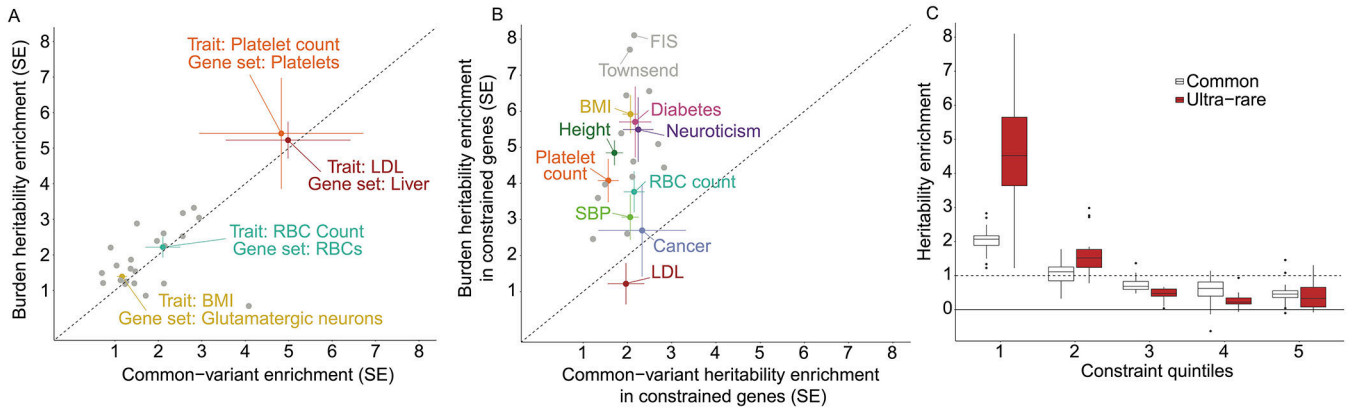
(C) Performance of BHR in simulations. We started with approximately realistic simulations and varied the sample size, the allele frequency of the variants, and the strength of negative selection. The boxplots are the distribution of BHR  $h^2$  estimates across 100 simulation runs, denoting median, quartiles, and range (excepting outliers)



**Figure 2: Burden heritability of 22 complex traits and common diseases in UK Biobank**  
**(A)** Proportions of coding variants by allele frequency and functional consequence in Genebase. Missense variants are categorized as either “benign” or as “possibly damaging/ probably damaging” using PolyPhen2. Ultra-rare is defined as  $AF < 1e-5$ . Rare is defined as  $1e-5 < AF < 1e-3$ . Common is defined as  $AF > 0.05$ . **(B)** Estimates of burden heritability across frequency bins and functional categories. Boxplots show the distribution of heritability estimates across 22 complex traits and common diseases, denoting median, quartiles and range (excepting outliers). Numerical results are contained in Supplementary Table 7. **(C)** Comparison of the total burden heritability (ultra-rare + rare) with the common-variant heritability of each trait (estimated using LDSC<sup>16</sup>). Error bars are standard errors. Numerical results for each trait are contained in Supplementary Tables 8 and 10. **(D)** Comparison of test statistic inflation between ultra-rare pLoF (red) and synonymous variants (gray) across the 22 traits. Lambda GC is the median burden  $\chi^2$  statistic divided by 0.454.<sup>30</sup>



**Figure 3: Burden heritability explained by significant genes**  
**(A)** Fraction of burden heritability explained by exome-wide significant genes from Genebase<sup>5</sup>. Each box represents the fraction of burden heritability explained by one significant gene. For numerical results, see Supplementary Table 12. **(B)** Fraction of common variant heritability explained by genome-wide significant loci. Each box represents the fraction of common variant heritability explained by one significant locus. For numerical results, see Supplementary Table 14. **(C)** The fraction of common variant heritability mediated by exome-wide significant genes, estimated using AMM<sup>34</sup>, compared with the fraction of burden heritability explained by the same genes, for traits with at least 5 exome-wide significant genes. For numerical results, see Supplementary Tables 12 and 16. **(D)** Common- vs. rare-variant cancer heritability mediated by cancer genes. The blue bars are the BHR estimates, and the grey bars are the AMM estimates. For numerical results, see Supplementary Table 16-17. Error bars in A-D are standard errors.



**Figure 4: Common- and rare-variant heritability enrichments**

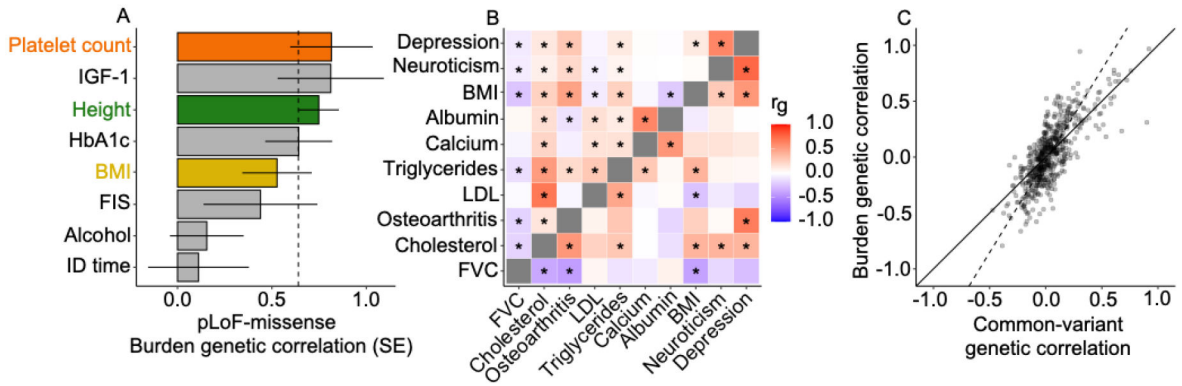
(A) Common and rare variant enrichments across cell type differentially expressed gene sets for selected trait-cell type pairs (see Supplementary Tables 16-17 for numerical results). Error bars are standard errors. (B) Common and rare variant enrichments in constrained genes in the bottom quintile of observed/expected pLoF alleles in gnomAD<sup>37</sup>. Error bars are standard errors. (C) Common and rare variant enrichments for 22 traits across quintiles of constraint. Boxplots denote median, quartiles and range (excepting outliers).

Author Manuscript

Author Manuscript

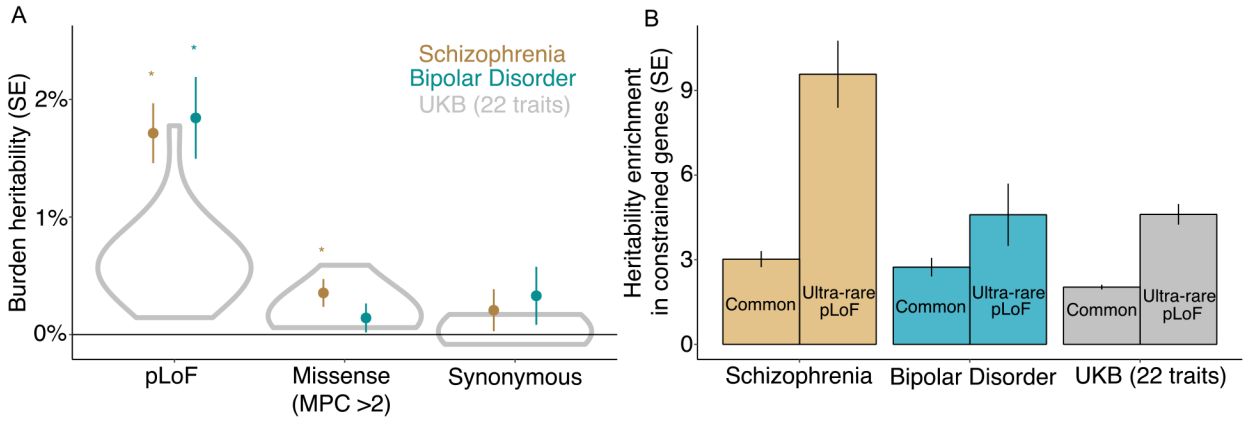
Author Manuscript

Author Manuscript



**Figure 5: Burden genetic correlations between variant classes and traits**

(A) Burden genetic correlations between ultra-rare pLoF and damaging missense variants, across 9 traits that have nominally significant burden heritability for both. The dashed line indicates the mean correlation across all 22 traits, computed as a ratio of averages. Error bars denote standard errors. For numerical results, see Supplementary Table 18. (B) Clustered heatmap of genetic correlations estimated with BHR from ultra-rare pLoF variants (lower triangle) and genetic correlations estimated with LD Score Regression (upper triangle). \* nominal significance (two-tailed  $p < 0.05$ ). For numerical results, see Supplementary Table 19. (C) Comparison of common and burden genetic correlations across trait pairs. The dashed line indicates the least squares regression fit (slope = 1.6).



**Figure 6: Burden heritability of schizophrenia and bipolar disorder**

(A) Burden heritability of ultra-rare pLoF variants, ultra-rare missense variants with MPC > 2, and ultra-rare synonymous variants. Gray violin plots show the distribution of burden heritability estimates in 22 UK Biobank traits (Figure 2B). (B) Constrained gene enrichment of ultra-rare pLoF vs. common variant heritability. Error bars denote standard errors. For numerical results, see Supplementary Table 20.