Supplementary information for

"Functional architecture of low-frequency variants

highlights strength of negative selection across coding and noncoding annotations"

Supplementary Notes

Simulations using UK10K target samples to assess extension of S-LDSC to low-frequency variants.

To obtain an initial assessment of the lowest MAF threshold for which S-LDSC with the UK10K LD reference panel might produce unbiased estimates, we first performed simple simulations using UK10K target samples (so that LD in the target samples and LD reference panel perfectly match) to investigate possible biases in h^2_{lf} estimates when using different MAF thresholds to define low-frequency variants. We simulated quantitative phenotypes from chromosome 1 of UK10K data[1] (3,567 individuals and 1,041,378 variants with allele counts greater or equal to 5, i.e. MAF ≥ 0.07%) by setting trait heritability to h^2=0.5, selecting M=100,000 causal variants, and simulating the variance of per-normalized genotype effect sizes proportional to $\left(2p(1-p)\right)^{1+\alpha}$ for a variant of frequency p. We considered three values of α: -0.25 (close to the -0.28 value previously estimated[2]), -0.50 and -0.75 (with the latter values giving higher weight to low-frequency variants), and performed 5,000 simulations for each value of α.

To determine at which MAF thresholds we can accurately estimate heritability above the MAF threshold using UK10K as an LD reference panel, we constructed 6 different annotation models, with heritability variants (used to compute estimates of heritability and enrichment) and reference variants (used to compute LD scores) defined using MAF ≥ 0.1%, 0.2% or 0.5% (see Supplementary Table 3 for the 6 different models). We estimated h^2_c and h^2_{lf} using S-LDSC and three different sets of regression variants (included in the regression, see equation (1)), also defined using MAF ≥ 0.1%, 0.2% or 0.5%. We note that our set of reference variants did not contain causal variants with a MAF between 0.07% and the selected MAF threshold, consistent with a realistic scenario in which causal rare variants may be absent from the reference data set. We determined that in these simulations, restricting to regression variants with MAF ≥ 0.5%, reference variants with MAF ≥ 0.1%, and heritability variants with MAF ≥ 0.5% provided unbiased estimates of h^2_{lf} (Supplementary Figure 1; we observed upward bias when considering MAF thresholds of 0.1% or 0.2% for heritability variants, Supplementary Table 3). We thus used these MAF thresholds for further simulations and applications to real data.

UK Biobank data set and choice of traits for main analyses and meta-analyses.

We analyzed data from the UK Biobank[3,4] consisting of 487,409 samples genotyped on ~800,000 markers and imputed to ~93 million SNPs using the Haplotype Reference Consortium (HRC) dataset[5] (N=64,976 haplotypes from WGS data). We restricted our analyses to 408,963 individuals with UK ancestry[4] and to variants with MAF ≥ 0.1%

that were present in our UK10K LD reference panel (8,523,464 variants in total). Overall, the imputation quality was high even for low-frequency variants (median INFO score larger than 0.9 for variants with MAF between 0.5% and 0.7%; Supplementary Figure 20). For each investigated trait (see below), we computed mixed model association statistics using BOLT-LMM version 2.3 software[6,7] with genotyping array (UK BiLEVE / UK Biobank), assessment center, sex, age, and age squared as covariates. We also included 20 principal components to correct for ancestry (provided with the UK Biobank data release[4]) according to guidelines of ref.[4]. We included 672,292 directly genotyped SNPs in the mixed model (specifically, all autosomal biallelic SNPs with missingness <10%).

We restricted our analyses to UK Biobank traits for which the z-score for nonzero h_c^2 computed using S-LDSC with the baseline-LF model was at least 10, to maximize robustness of h_{lf}^2 (Supplementary Table 5). We selected 40 traits (average N=363,166) with squared phenotypic correlation below 0.5 (except for diastolic and systolic blood pressure phenotypes, which are traditionally analyzed jointly; Supplementary Table 15). In all meta-analyses, we excluded one of each pair of traits with squared phenotypic correlation larger than 0.1, prioritizing traits with larger z-score for nonzero h_c^2 (27 independent traits; average N=354,892). All meta-analyses were performed using random-effects meta-analyses implemented in the R package *rmeta*.

Limitations of this work.

Although our work has provided insights on low-frequency variant architectures of human diseases and complex traits, it has several limitations. First, recent studies have shown that functional enrichment estimates can be sensitive to the choice of LD-dependent heritability model[8,9]; however, this concern is unlikely to impact the results of the current study (see "LD-related heritability models" paragraph in this Supplementary Note). We note that our simulations indicate that S-LDSC (using all variants) produces slightly conservative LFVE estimates for annotations with low levels of LD (Figure 1). However, this does not impact our main conclusions based on the LFVE/CVE ratio, for which S-LDSC produces unbiased estimates in our simulations (Figure 1). We further note that our S-LDSC model assumes a linear effect of each annotation and does not consider the effects of variants with MAF < 0.1% in the UK10K reference panel. Second, all of our analyses are based on UK-ancestry target samples and LD reference samples. Thus, more work is required to extend our approach to GWAS summary statistics from other European or non-European ancestries, together with available LD reference panels (such as the publicly available subset of HRC[5] and TOPMed) that may not provide an exact match. Third, our method requires extremely large GWAS sample sizes with large matching LD reference panel, which are currently available only from the UK Biobank and UK10K data sets. Fourth, our method may sacrifice power by analyzing only GWAS summary statistics rather than the individual-level data that is readily available from UK Biobank. Although functional enrichment analyses of individual-level data can be performed using restricted maximum likelihood (REML) and its extensions[2,10,11], those methods are applicable only to a small number of non-overlapping functional annotations; to our knowledge, all current methods that are applicable to a large number of overlapping functional annotations are based on summary statistics[12], whereas analyzing one annotation at a time can produce severely biased results (see

Figure 2b of ref. [13]). Fifth, many of our analyses were meta-analyzed across 27 independent traits to maximize statistical power; this could mask trait-specific functional architectures, although results for non-CTS annotations were generally consistent across the 40 UK Biobank traits we analyzed (Supplementary Figure 7). Sixth, the set traits that we analyzed includes only a limited number of diseases, due to the population-based design of UK Biobank (which limits power for diseases of lower prevalence); the analysis of larger case-control data sets would be of critical value, particularly for psychiatric disorders. Seventh, our inferences about negative selection and the effect size variance of causal rare variants in each annotation rely both on the Eyre-Walker model[14] and a gamma distribution of selection coefficients[15] (see Methods); large WGS datasets will be necessary to formally validate our extrapolation to rare variant architectures. Finally, our conclusions about CTS and noncoding annotations were limited by the availability and precision of regulatory annotations in relevant cell types. The generation of more comprehensive and more precise CTS annotations will aid our understanding of the noncoding genome and its role in human diseases and complex traits.

LD-related heritability models

Recent studies have shown that functional enrichment estimates can be sensitive to the choice of LD-dependent heritability model. Specifically, including the 6 LD-related annotations of the baseline-LD model leads to a slight decrease in functional enrichment estimates[8], and using the LDAK model (without including the 6 LD-related annotations of the baseline-LD model) leads to a greater decrease in functional enrichment estimates[9]. Although there can be no guarantee that the annotations of the baseline-LD model and/or the LDAK model perfectly model LD-dependent architectures, we believe that inadequate modeling of LD-dependent architectures is unlikely to impact the results of the current study, for three reasons. First, in recent unpublished work[16], we have performed analyses of common variant enrichment in which we included both baseline-LD and LDAK model annotations, via a baseline-LD+LDAK model that produces robust results both in simulations under the baseline-LD model and in simulations under the LDAK model, and we determined that common variant enrichment results on real UK Biobank traits are virtually unchanged vs. analyses using only the baseline-LD model. Second, leveraging a recent unpublished observation that out-of-sample polygenic prediction can be a valuable way to perform formal model comparisons[17], we compared polygenic predictions computed using posterior mean effect sizes under three prior models[18]: an infinitesimal model with equal effect sizes variances (the "GCTA model"[10,19]), the baseline-LD model, and the baseline-LD+LDAK model. We obtained polygenic prediction r^2 values (averaged across 16 UK Biobank traits[4]; average of 365K training samples and 22K validation samples) of 0.1390, 0.1601, and 0.1600, respectively[18]. Thus, a formal model comparison of the baseline-LD and baseline-LD+LDAK models finds no significant difference in model fit for these two models. Going forward, we believe that it will be of high value to perform a comprehensive model comparison of the GCTA, LDAK, baseline-LD, baseline-LD+LDAK and other models using both likelihood methods[9] and out-of-sample polygenic prediction methods[17], complementing previous work comparing the GCTA and LDAK models[9,17]. Third, LD impacts common variants more than low-frequency variants.

Thus, inadequate modeling of LD would be expected to cause our method to overestimate CVE more than LFVE, thus deflating the LFVE/CVE ratio, and would be unlikely to explain the large and statistically significant LFVE/CVE ratios that are the focus of the current study.

Choice of denominator for LFVE and CVE

We emphasize that the choice of LFVE and CVE denominator (e.g. the proportion of low-frequency or common variants in an annotation) can have a large impact on the value of enrichment estimates. For example, coding regions are depleted for common variants (but less depleted for low-frequency variants), so that LFVE/CVE will be lower than (% of h_{lf}^2)/(% of h_c^2); the same may be true for other annotations impacted by negative selection (see Results). Different choices of denominator can be justified, depending on the goal of a particular study; our choice of denominator does not correct for MAF- and LD-dependent architectures (which are consequences of negative selection[8,20–22]), as the goal of our study is to quantify the strength of negative selection across functional annotations.

Application of S-LDSC

Application of S-LDSC was performed using 3,567 unrelated individuals of UK10K data set[1] (ALSPAC and TWINSUK cohorts) as an LD reference panel. This choice was made in order to ensure a close ancestry match between the target sample used to compute summary statistics (UK Biobank) and the LD reference panel (UK10K), as LD patterns of low-frequency variants are expected to vary across European populations[23,24]. We note that there is a difference between the LD reference panel (here UK10K, used to compute LD scores) and the imputation reference panel (here HRC, used to impute variants). Reference variants, used to estimate LD scores, were the set of 11,830,279 biallelic variants with MAF ≥ 0.1%. Regression variants, used to estimate annotation effect sizes τ, were the set of 8,523,464 UK Biobank variants with MAF ≥ 0.5% in UK10K. These MAF thresholds for reference variants and regression variants were assessed via simulations (see below). Variants with very large χ^2 association statistics (larger than $0.001N$), as well as variants in the major histocompatibility complex (MHC) region (chr6:25Mb-34Mb) were removed from all analyses.

We note that BOLT-LMM uses a linear mixed model and achieved an effective sample size (N_{eff}) higher than the true sample size (N) (ref.[7]). This implies that S-LDSC overestimates per-variant heritability when using summary statistics generated by BOLT-LMM. We thus corrected h_c^2 and h_{lf}^2 estimations by dividing them by N_{eff}/N. N_{eff} was estimated by taking ratios of chi-square statistics computed by BOLT-LMM on the N=408,963 individuals with UK ancestry at genome-wide significant SNPs vs. the chi-square statistics computed by linear regression on N=337,539 unrelated individuals with UK ancestry, and by multiplying this ration with the sample size in the N=337,539 unrelated individuals dataset (as in ref.[7]). After this correction, our reported h_c^2 and h_{lf}^2 estimates are only 0.98x and 1.02x higher than when applying S-LDSC to summary statistics computed using linear regression with 20 principal component covariates on N=337,539 unrelated individuals with UK ancestry, and our h_c^2/h_{lf}^2 ratio estimates are nearly identical (h^2_c/h^2_{lf}=6.28x, s.e.=0.23x with BOLT-LMM vs. h^2_c/h^2_{lf}=6.34x, s.e.=0.26x with linear regression; Supplementary Figure 21), confirming that our results are not impacted by the use of BOLT-LMM summary statistics.

Baseline-LF model and functional annotations.

We considered 34 main functional annotations from the baseline-LD model v1.1 (27 binary and 7 continuous annotations, including LD-related annotations; refs.[8,13,25,26]), including coding, UTR, promoter and intronic regions, the histone marks monomethylation (H3K4me1) and trimethylation (H3K4me3) of histone H3 at lysine 4, acetylation of histone H3 at lysine 9 (H3K9ac) and two versions of acetylation of histone H3 at lysine 27 (H3K27ac), open chromatin as reflected by DNase I hypersensitivity sites (DHSs), combined chromHMM and Segway predictions (which make use of many Encyclopedia of DNA Elements (ENCODE) annotations to produce a single partition of the genome into seven underlying chromatin states), three different conserved annotations, two versions of super-enhancers, FANTOM5 enhancers, typical enhancers, and 6 LD-related continuous annotations (predicted MAF-adjusted allele age, level of LD in African populations, recombination rate, nucleotide diversity, a background selection statistic and CpG-content) (see Supplementary Table 1). The 6 LD-related annotations were included in the baseline-LD model of ref.[8] because (i) they all provide statistically significant signals of enriched heritability for real traits after conditioning on all other annotations in the baseline-LD model, and (ii) in forward simulations, the 4 LD-related annotations that do not rely on empirical data are all conditionally informative for the selection coefficient of a variant, with effect sizes roughly proportional to the conditional effect sizes observed in heritability analyses of real traits, implying that the conditionally significant signals of enriched heritability are an expected consequence of negative selection[8].

In order to further dissect the set of coding variants, a major focus of this study, we annotated each coding variant using ANNOVAR[27], and added one synonymous and one non-synonymous annotation (corresponding to "synonymous SNV" and "stopgain" or "stoploss" ANNOVAR outputs, respectively) to our model. (We note that the coding annotation in our baseline model comes from the UCSC coding track and includes UTR variants of protein coding genes, implying that some variants in this annotation do not belong to the synonymous or non-synonymous annotations). We also added three new annotations based on phastCons[28] conserved elements (46 way) in vertebrates, mammals and primates, and one annotation based on flanking bivalent TSS/enhancers from Roadmap data[29] (see URLs); we added the phastCons primate annotation and flanking bivalent TSS/enhancer annotations because they had statistically significant signals of enriched heritability after conditioning on all other annotations in the model, and added the phastCons vertebrate and mammal annotations for comparison purposes (see Supplementary Table 16). These 6 new annotations led to a total of 33 main binary annotations (see Supplementary Table 1).

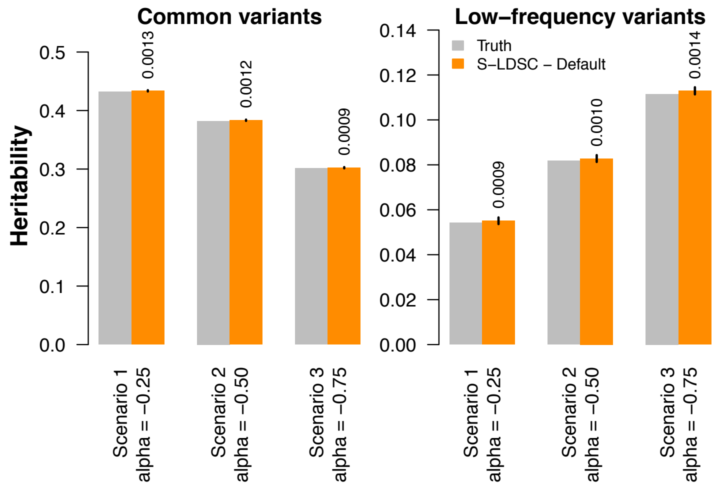
We included 500 bp windows around each binary annotation (except for annotations that are defined separately for each base pair) and 100 bp windows around four of the main annotations (in order to prevent enrichment estimates from being inflated by enriched heritability in flanking regions[13]), leading to a total of 74 main functional annotations. Then, all annotations were duplicated for low-frequency and common variants as described in equation (5), except for the predicted allele age annotation[30] (which had too many missing values for low-frequency variants). Finally, we included one annotation containing all variants, 10 common variant MAF bins (as in the baseline-LD model[7]) and 5 low-frequency variant 5 MAF bins (which were sufficient to produce robust results in simulations under a different generative model; see Results). For both common and low-frequency variants, MAF bin boundaries were chosen so that each MAF bin contained the same proportion of reference variants, as in ref.[8] (see Supplementary Table 17 for boundaries). We thus obtained a set of 163 total annotations. We refer to this set of annotations as the "baseline-LF model" (see Supplementary Table 2), which we used for all of our S-LDSC analyses.

Strategy to obtain realistic distribution of fitness effects for annotations mimicking both non-synonymous and noncoding variants.

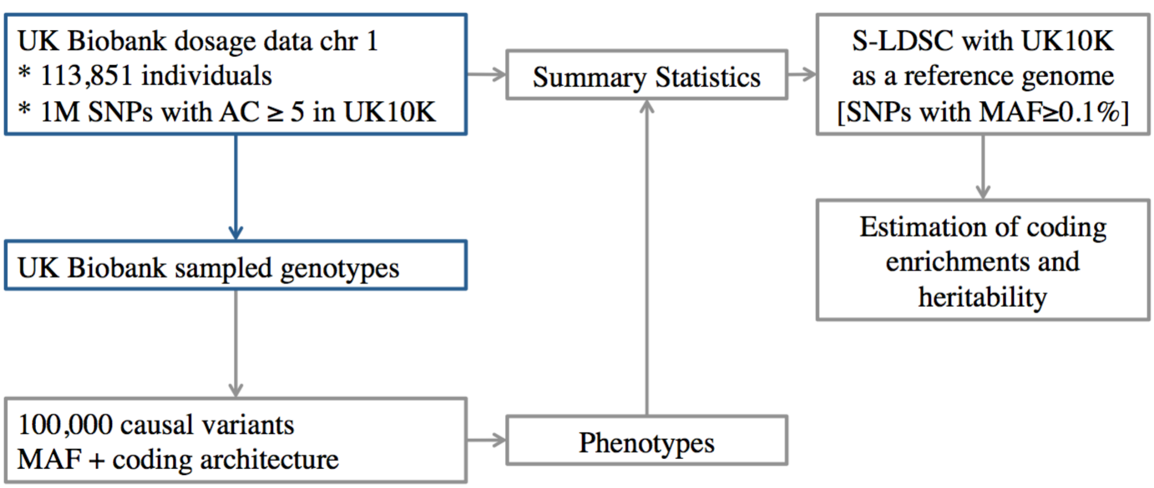
Unlike our previous forward simulation framework[8], we designed these simulations to have a realistic DFE for annotations mimicking both non-synonymous and noncoding variants. We first performed simulations involving non-synonymous and noncoding variants, in order to fit appropriate parameter values for these annotations; subsequent simulations distinguishing functional noncoding and ordinary noncoding variants are described below. We created 50 non-synonymous elements with a realistic length 200bp (10kb in total, 1% of the 1Mb simulated genome) separated by noncoding elements of size 14.9kb (99% of the simulated genome; Supplementary Figure 11a). To mimic non-synonymous elements, we used π_{del}=80%, $\bar{s}_{dn}$=-3.16x10^{-3} and θ=0.32, as previously estimated[31]. As the DFE of noncoding regions has never been investigated (to our knowledge), we performed simulations using a wide range of noncoding DFE to find the best fit to observed patterns of enrichment. We varied π_{del} for noncoding variants between 5% and 80% (16 values in total using 5% incremental steps), $\bar{s}_{dn}$ for noncoding variants between -10^{-5} and -10^{-3} (7 values in total using logarithmic incremental steps), θ for noncoding variants between 0.05 and 1 (5 values in total using logarithmic incremental steps), and τ_{EW} between 0.025 and 1.00 (40 values in total using 0.025 incremental steps), leading to a total of 22,400 different scenarios (Supplementary Table 18). We simulated 100 regions of 1 Mb for each scenario and merged the outputted variants. We considered all deleterious variants as causal, generated per-allele and per normalized genotype effect size using the Eyre-Walker model, and removed scenarios in which our simulated values of per-variant heritability ratio between common and low-frequency variants, non-synonymous LFVE, and non-synonymous CVE were outside the 95% confidence intervals estimated by the meta-analysis of UK Biobank phenotypes (i.e. 3.99x [3.69; 4.29]), 38.25x [33.80; 42.70] and 7.72x [6.05; 9.39], respectively). After those filters, 11 scenarios remained; π_{del} were between 0.30 and 0.75, 10/11 scenarios had $\bar{s}_{dn}$=-0.0001, 9/11 scenarios had θ=0.32, and τ_{EW} between 0.675 and 0.925. We selected the scenario in which the simulated values were closest to the values from the UK Biobank meta-analysis based on Euclidian distance (Supplementary Table 19); more precisely, we retained the scenario where π_{del}=40%, $\bar{s}_{dn}$=-

$1.00 x 10^{-4}$, θ=0.32 for noncoding variants and τ_{EW}=0.75 and obtained the values 4.07x, 36.27x and 7.73x for per-variant heritability ratio between common and low-frequency variants, non-synonymous LFVE, and non-synonymous CVE, respectively. These estimated parameter values suggest that a *de novo* noncoding mutation is only 2 times less likely to be deleterious than a *de novo* coding mutation, but on average has a selection coefficient 30 times smaller. They also predict a non-synonymous rare variant (MAF<0.5%) enrichment of 74.27x (Supplementary Table 19). However, we caution that the diversity patterns in this simulation scenario were not realistic. We observed a ratio between the number of low-frequency and common variants of 0.89, compared to 0.63 in UK10K data (3,398,397 low-frequency variants against by 5,353,593 common variants), and a ratio between the proportion of non-synonymous variants in low-frequency and common variants of 1.38, compared to 1.65 in UK10K data (non-synonymous variants represent 0.45% of low-frequency variants and 0.27% of common variants). Indeed, the Gravel et al. demographic model[32] that we used is based solely on the site frequency spectrum of non-synonymous variants, which becomes unrealistic when considering other variants. However, we believe that this issue does not impact our conclusions, which are based on the per-variant heritability of an annotation rather than the proportion of heritability explained by the annotation.

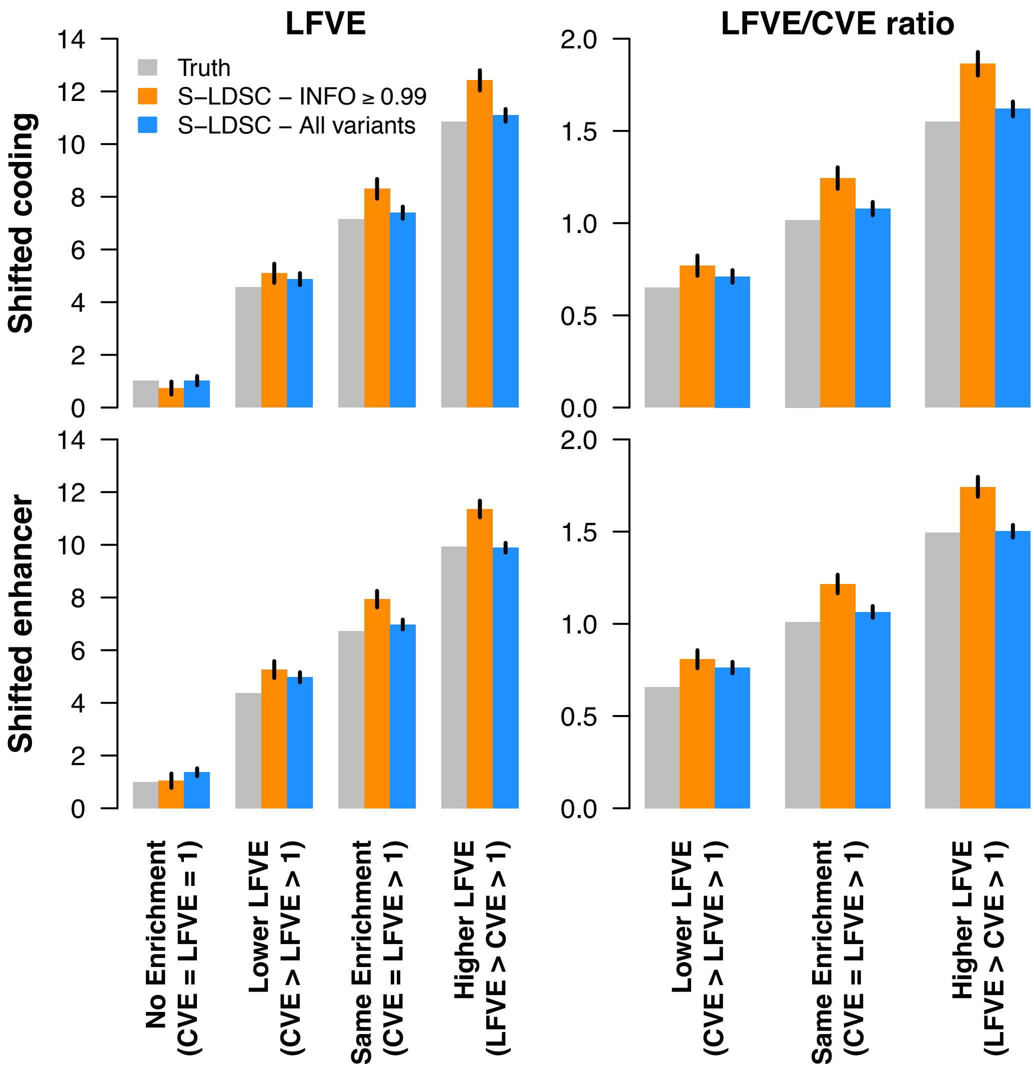
Supplementary Figures



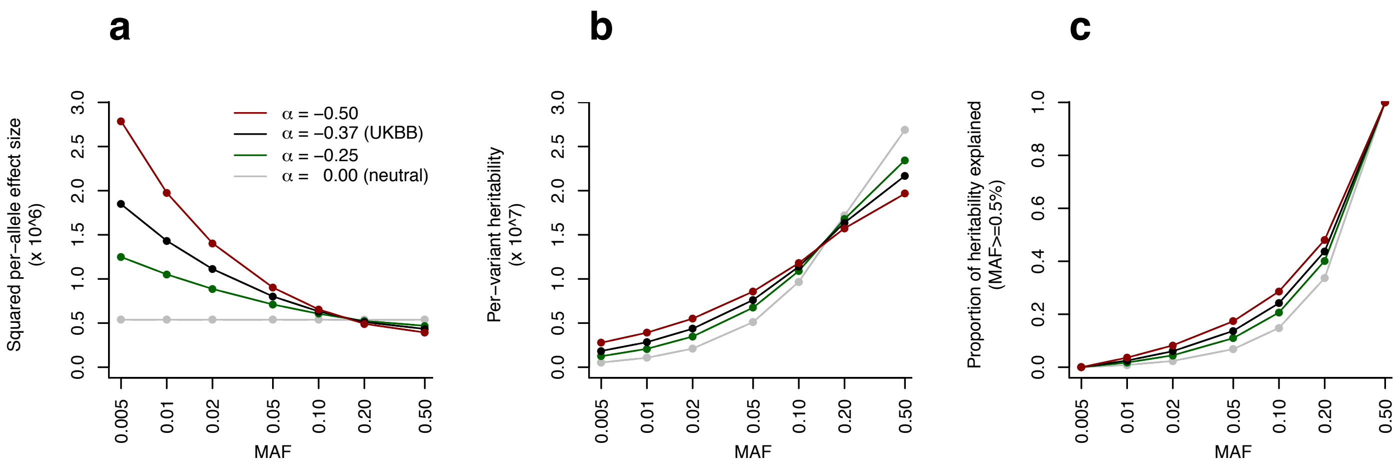
Supplementary Figure 1: Estimation of h_c^2 and h_{lf}^2 in UK10K simulations, in which the target sample and the LD reference panel match perfectly. Here UK10K genotype data has been used both to simulate phenotypes and as an LD reference panel in S-LDSC. Error bars represent 95% confidence intervals; numbers above bars indicate biases.



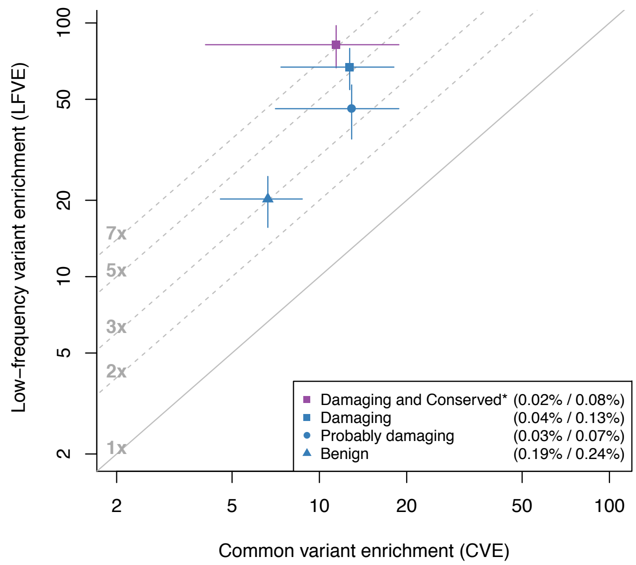
Supplementary Figure 2: UK Biobank simulation workflow, accounting for imputation uncertainty. For each simulation we simulated phenotypes using genotypes sampled from UK Biobank dosage data (i.e. posterior probability of each genotype). We computed corresponding summary statistics using UK Biobank dosage data to take into account imputation uncertainty. Blue rectangles represent fixed parts of the simulations (i.e. UK Biobank sampled genotype data is generated only once), while grey rectangles represent parts that are generated separately for each simulation.



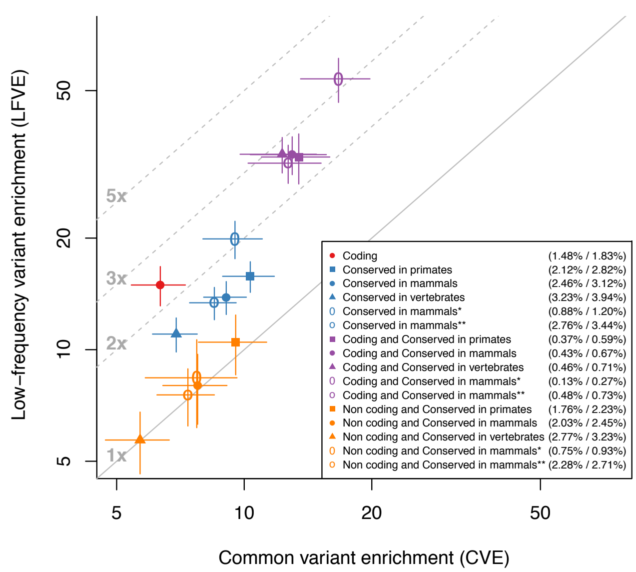
Supplementary Figure 3: Enrichment simulations using annotations with average levels of LD. We report estimates of LFVE and LFVE/CVE ratio in simulations under a "shifted coding"-enriched architecture (first row) or "shifted enhancer"-enriched architecture (second row), where shifted coding and enhancer annotations have been constructed to have average levels of LD. We considered four different simulation scenarios (see main text). S-LDSC was run either by restricting regression variants to accurately imputed variants (S-LDSC – INFO ≥ 0.99), or by including all variants (S-LDSC – All variants). We did not report LFVE/CVE ratio for the No Enrichment simulation (CVE = LFVE = 1) due to unstable estimates; however, all analyses of real traits in this paper focus on annotations with significant CVE. Results are averaged across 1,000 simulations. Error bars represent 95% confidence intervals. Numerical results for h^2_{lf}, h^2_c, LFVE, CVE and LFVE/CVE ratio are reported in Supplementary Table 4.



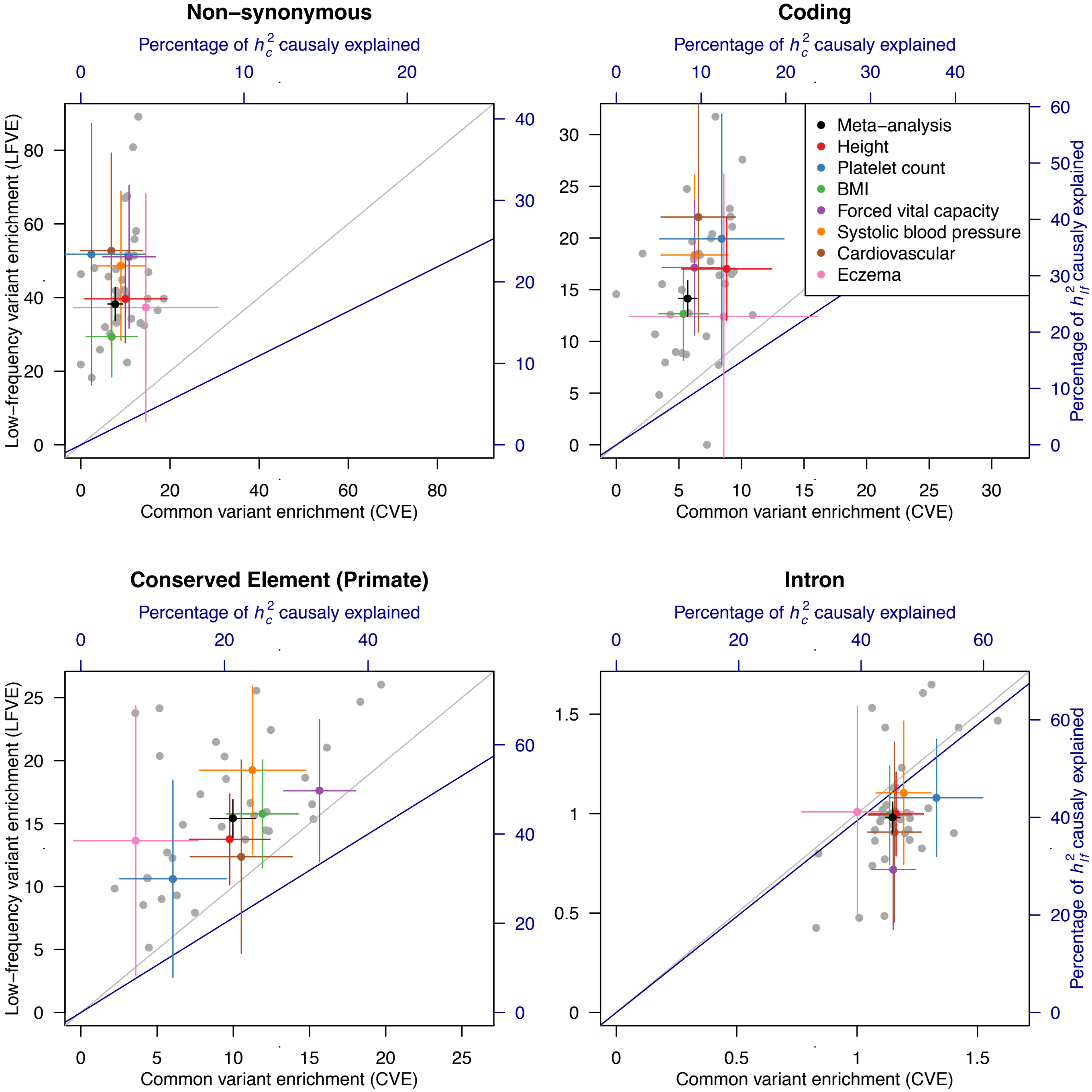
Supplementary Figure 4: Relationship between α parameter, per-allele effect size, per-variant heritability, and proportion of heritability explained. Under a genetic model where per-allele effect size has mean 0 and variance proportional to $\left(p(1-p)\right)^{\alpha}$, we report **(a)** the mean squared per-allele effect size (proportional to $\left(p(1-p)\right)^{\alpha}$), **(b)** the mean squared per-normalized genotype effect size (proportional to $\left(p(1-p)\right)^{1+\alpha}$), and **(c)** the corresponding proportion of low-frequency and common variant heritability explained by variants with MAF below a given threshold using the MAF distribution observed in UK10K. We considered four different α values: *i*) α = 0, corresponding to a neutral model where per-allele effect size is independent of MAF; *ii*) α = -0.25, as reported in Figure 2; *iii*) α = -0.37, corresponding to the value fitting the h_c^2/h_{lf}^2 ratio observed in our meta-analysis of 27 independent UK Biobank traits; and *iv*) α = -0.50, as reported in Figure 2. This figure illustrates that for a model where α = -0.37, low-frequency variants have larger per-allele effects than common variants but lower per-variant heritability, and thus explain less total heritability.



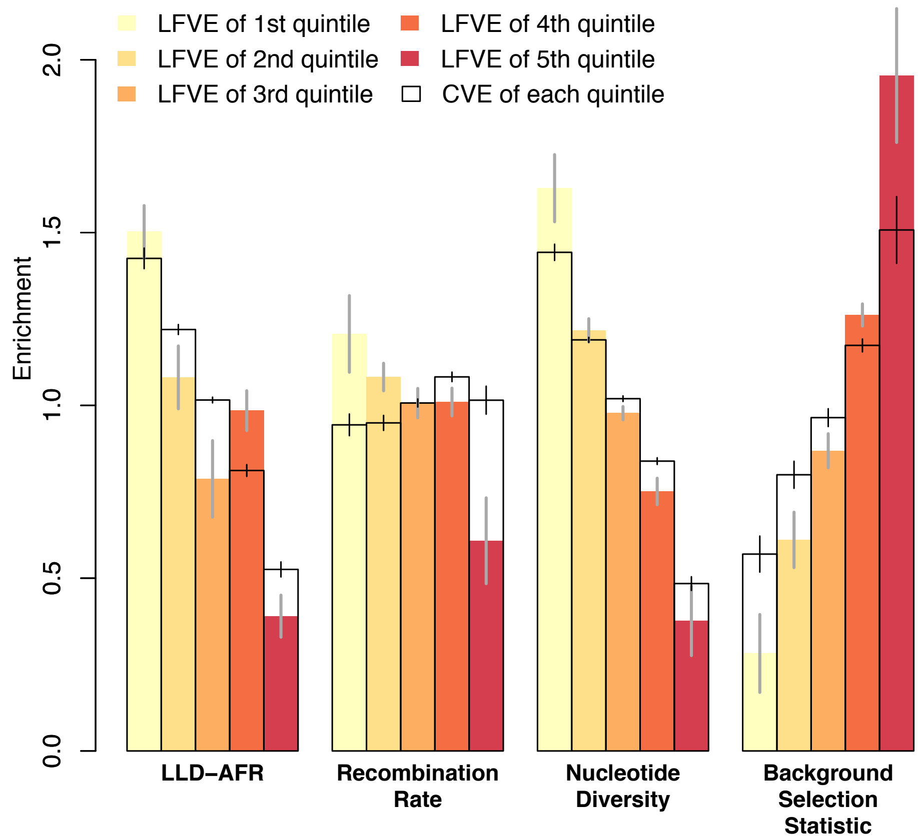
Supplementary Figure 5: Enrichments for non-synonymous variants vary with PolyPhen-2 predicted deleterious effect. We report CVE and LFVE for 4 additional annotations based on PolyPhen-2 (ref. [33]) predicted deleterious effect on non-synonymous variants and conservation score. We observed higher LFVE and LFVE/CVE ratio for the annotations predicted to be more deleterious. Results were obtained by adding the new annotations to the baseline-LF model. Numbers in the legend represent the proportion of common / low-frequency variants inside the annotation, respectively. Conserved* indicates a GERP RS score > 4. Error bars represent 95% confidence intervals.



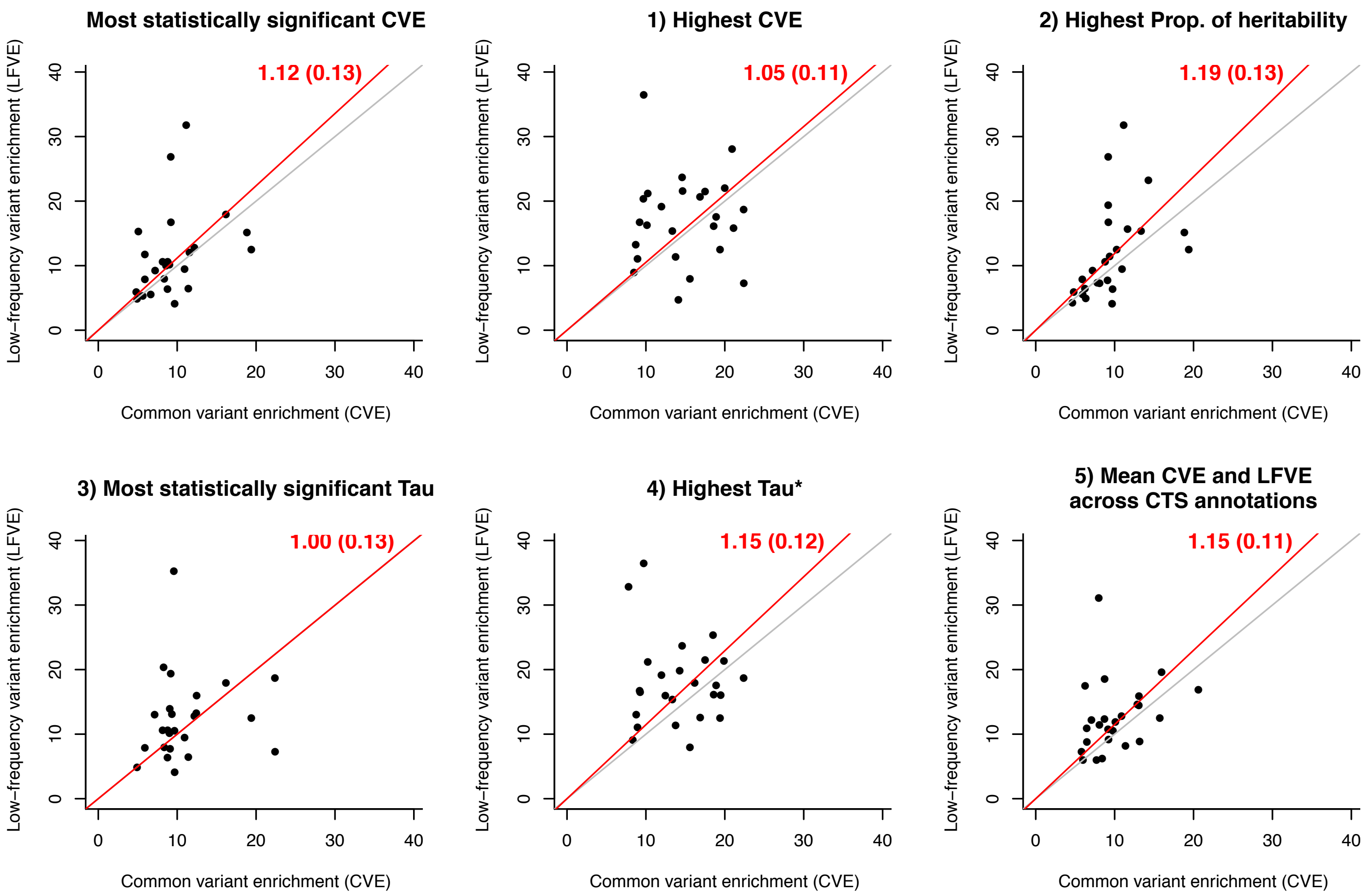
Supplementary Figure 6: Enrichments for conserved coding and conserved noncoding annotations. We report CVE and LFVE for coding (in red), conserved (in blue), conserved coding (in purple), and conserved noncoding (in orange) annotations. We observed that the significantly larger LFVE (compared to CVE) for the conserved annotations is mainly due to their intersection with coding regions, while the significantly larger LFVE and CVE enrichments in primates (compared to more distant species) is entirely due to conserved regions that are noncoding. Results were obtained by adding the new annotations to the baseline-LF model. Numbers in the legend represent the proportion of common / low-frequency variants inside the annotation, respectively. The first three conserved annotations are based on phastCons elements[28], the fourth (denoted *) is based on GERP RS scores[34] (≥4), and the last (denoted **) is based on Lindblad-Toh et al.[35]. Error bars represent 95% confidence intervals.



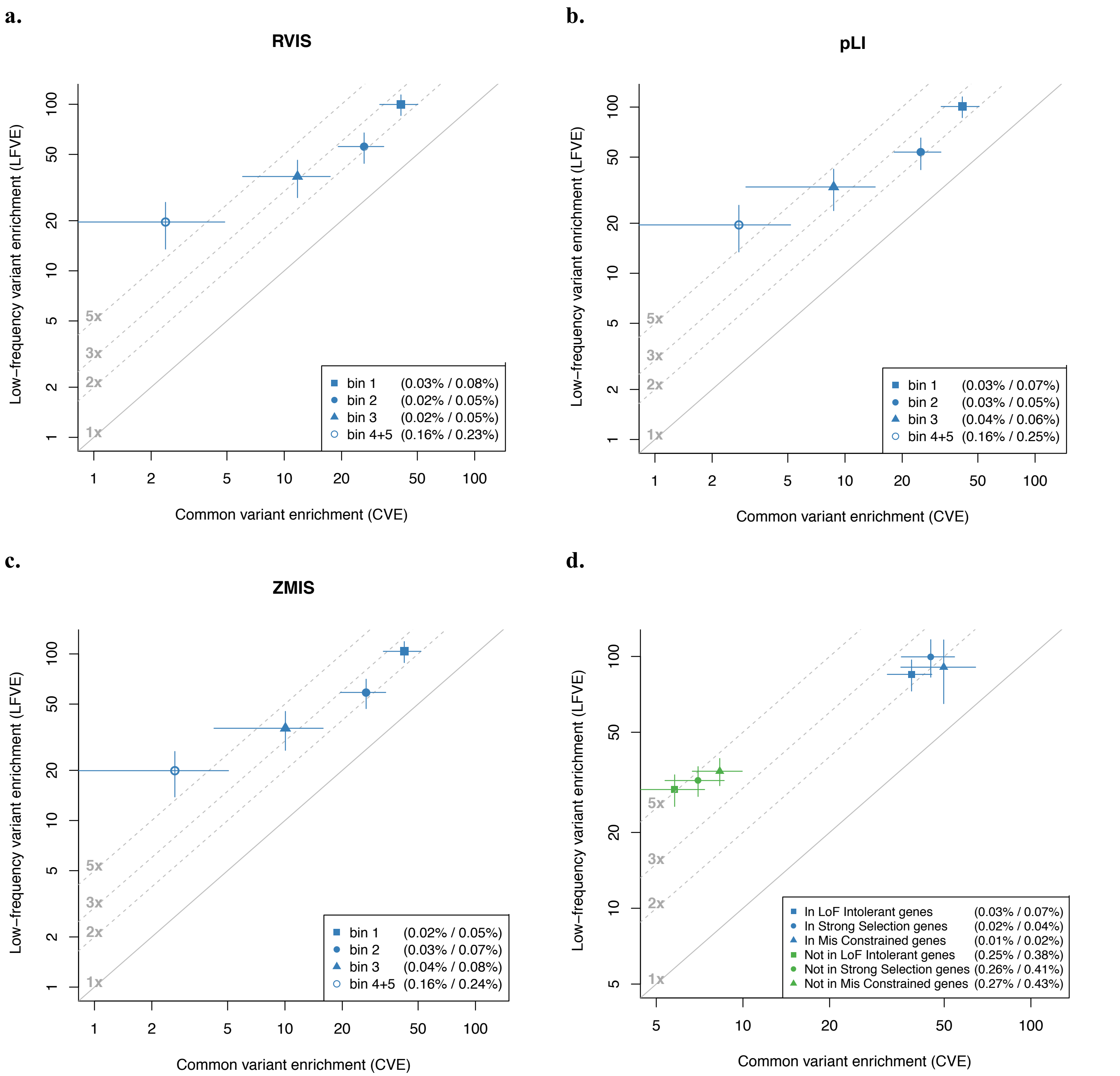
Supplementary Figure 7: Enrichment estimates for 4 functional annotations across 40 UK Biobank traits. We report CVE and LFVE for 4 functional annotations across 40 UK Biobank traits. Black points denote meta-analysis across 27 independent traits; colored points denote 7 representative independent traits. Error bars represent 95% confidence intervals. The grey line represents LFVE = CVE; the blue line represents proportion of h_c^2 causally explained = proportion of h_{lf}^2 causally explained. Numerical results are reported in Supplementary Table 6.



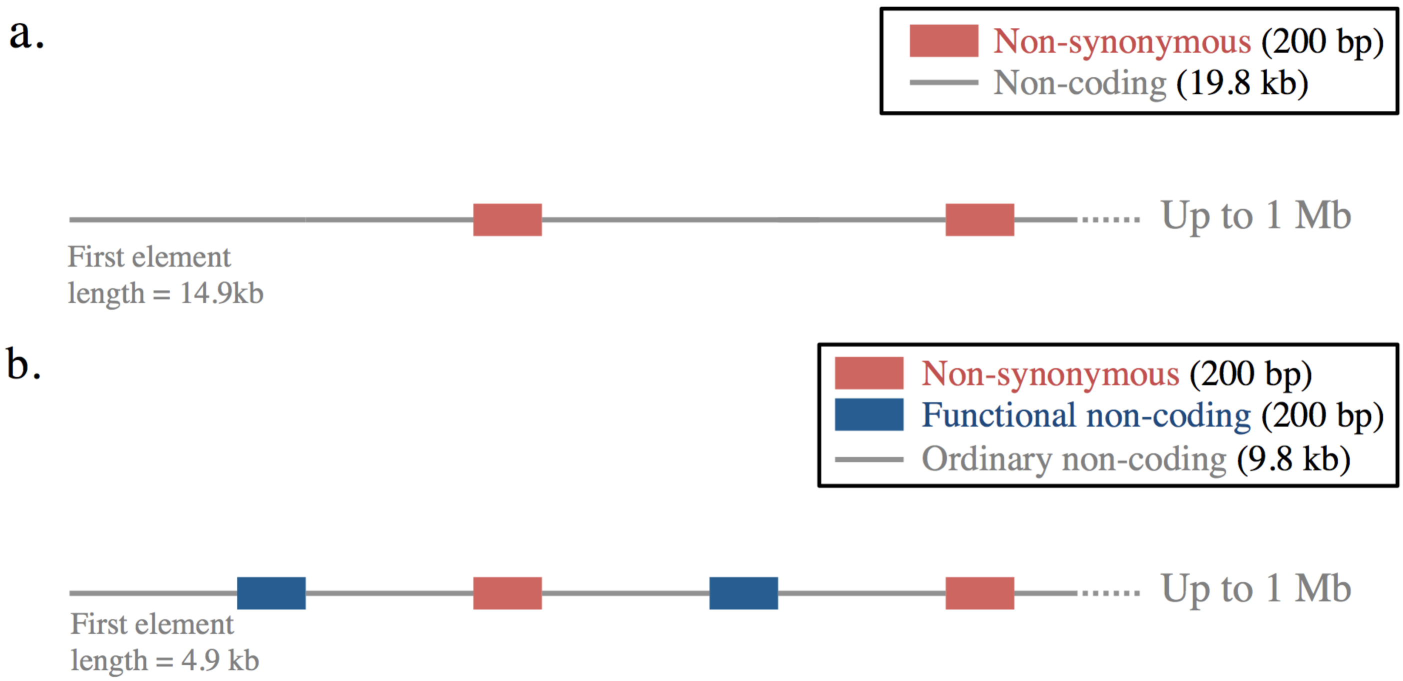
Supplementary Figure 8: Significantly larger enrichment/depletion for LFVE than for CVE for LD-related annotations related to negative selection. We report the LFVE and CVE of each quintile for each of 4 continuous LD-related annotations related to negative selection from ref. [8]. Colored bar plots represent LFVE, while black outline bar plots represent CVE. Error bars represent 95% confidence intervals. Numerical results for all 7 continuous annotations are reported in Supplementary Table 7. For predicted allele age, we were unable to estimate LFVE due to many missing values for low-frequency variants; CVE is reported in Supplementary Table 7. For CpG content and GERP NS scores[34], we report LFVE and CVE results in Supplementary Table 7 only, as these annotations are not related to negative selection[8]. While for common variants the negative effect of recombination rate due to negative selection[36] cancels out the positive effect of recombination rate due to functional annotations[8] (enriched regulatory annotations tend to be in high recombination rate regions), we observed that effect of negative selection exceeds the effect of functional annotations for low-frequency variants, consistent with the observation that recombination coldspots are enriched in rare disease-associated mutations[37].



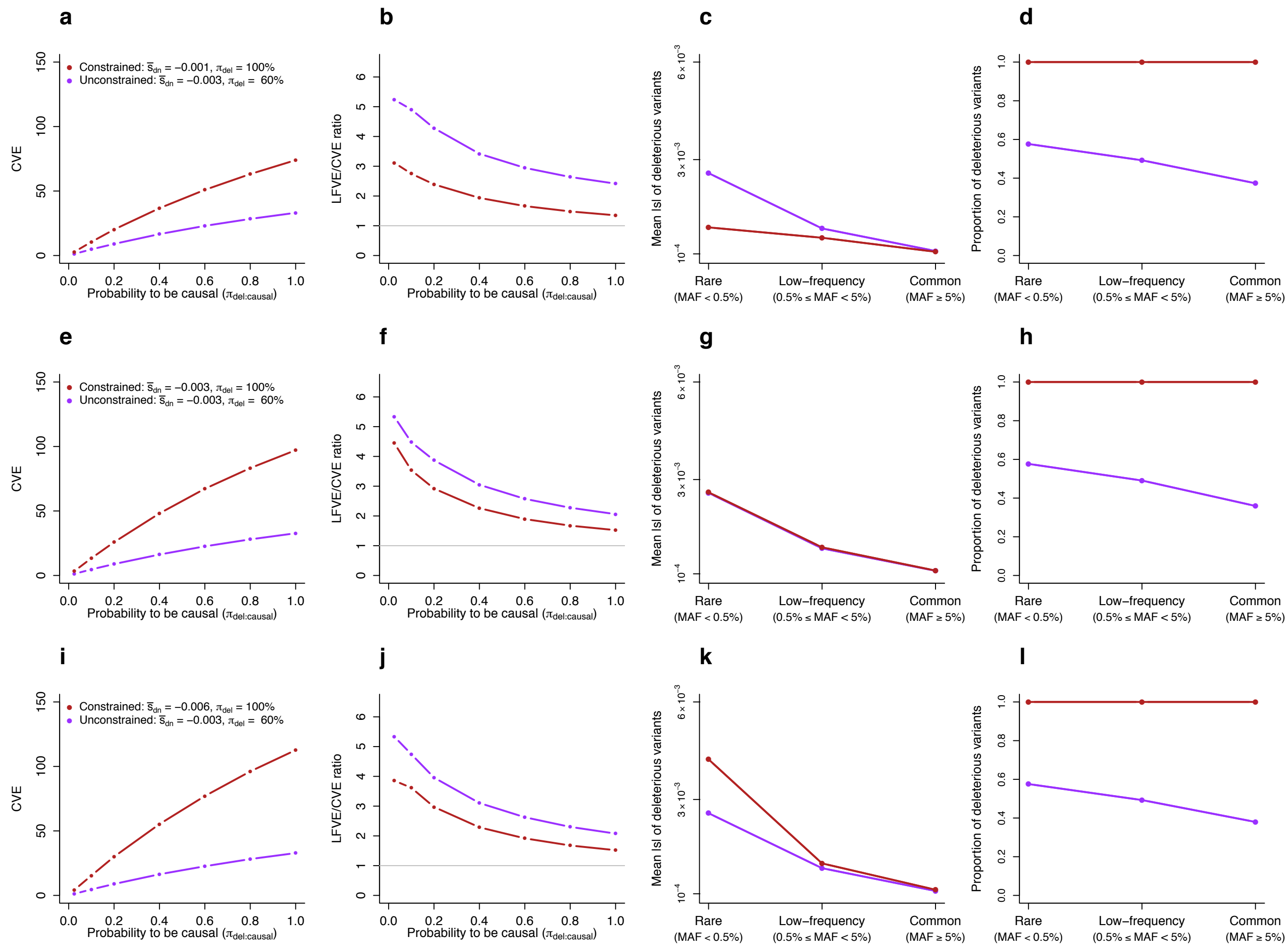
Supplementary Figure 9: Similar relationship between CVE and LFVE for cell-type-specific (CTS) enriched annotations using different definition of "most critical" CTS annotation. We considered 637 pairs of CTS annotations and UK Biobank traits for which the CTS annotation is significantly enriched in common variants. For each independent trait, we selected the "most critical" annotation using our primary definition (top left panel). We also considered 4 alternative definitions of the "most critical" CTS annotation for each trait: 1) the CTS annotation with the highest CVE, 2) the CTS annotation that explains the higher proportion of h_c^2, 3) the CTS annotation with the most significant τ coefficient, or 4) the CTS annotation with the highest standardized τ^* coefficient (defined as the proportionate change in per-variant heritability associated to a one standard deviation increase in the value of the annotation[8]). We also computed 5) the mean LFVE and CVE across 1-53 significant CTS annotations for each independent trait. Grey lines represent LFVE = CVE. Red lines and red numbers (standard errors in parentheses) represent the regression slope. For each of these definitions, LFVE were similar to CVE.



Supplementary Figure 10: Similar LFVE and CVE enrichment results for different criteria for stratifying non-synonymous variants in genes under varying levels of selective constraint. (**a,b,c**) We report LFVE vs. CVE (log scale) for non-synonymous variants in 5 bins based on the residual variation intolerance score of their gene[38] (RVIS; a), on the probability for their genes to be loss-of function intolerant[39] (pLI; b), and on the Z score of their gene to be missense intolerant[39] (ZMIS; c); bins 4+5 are merged for visualization purposes. (**d**) We report LFVE vs. CVE (log scale) for non-synonymous variants inside and outside genes reported to be intolerant to loss-of-function[39] (LoF Intolerant; 3,000 genes), genes reported to be under the strongest selection[40] (Strong Selection; 2,656 genes), and genes reported to be missense-constrained[41] (Mis Constrained; 1,417 genes). Numbers in the legend represent the proportion of common / low-frequency variants inside the annotation, respectively. The solid line represents LFVE = CVE; dashed lines represent LFVE = constant multiples of CVE. Error bars represent 95% confidence intervals.

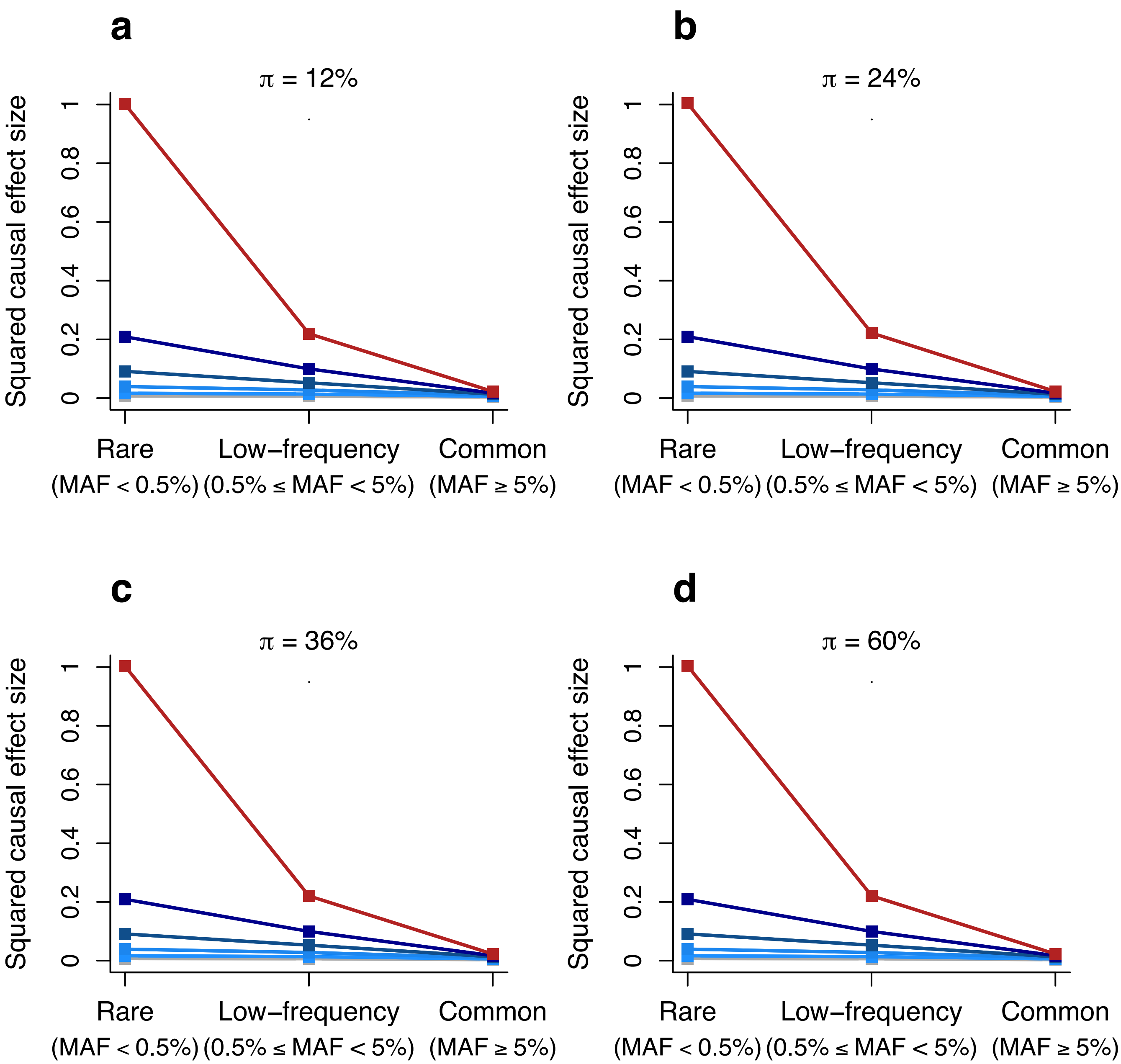


Supplementary Figure 11: Genome organization in forward simulations. (a) We used a genomic organization with only non-synonymous (1% of the genome) and noncoding (99% of the genome) elements to fit the distribution of fitness effects (DFE) of noncoding elements. **(b)** We used a genomic organization with non-synonymous (1% of the genome), "functional noncoding" (1% of the genome) and "ordinary noncoding" (98% of the genome) elements, and varied the DFE and proportion of causal variants in functional noncoding elements.

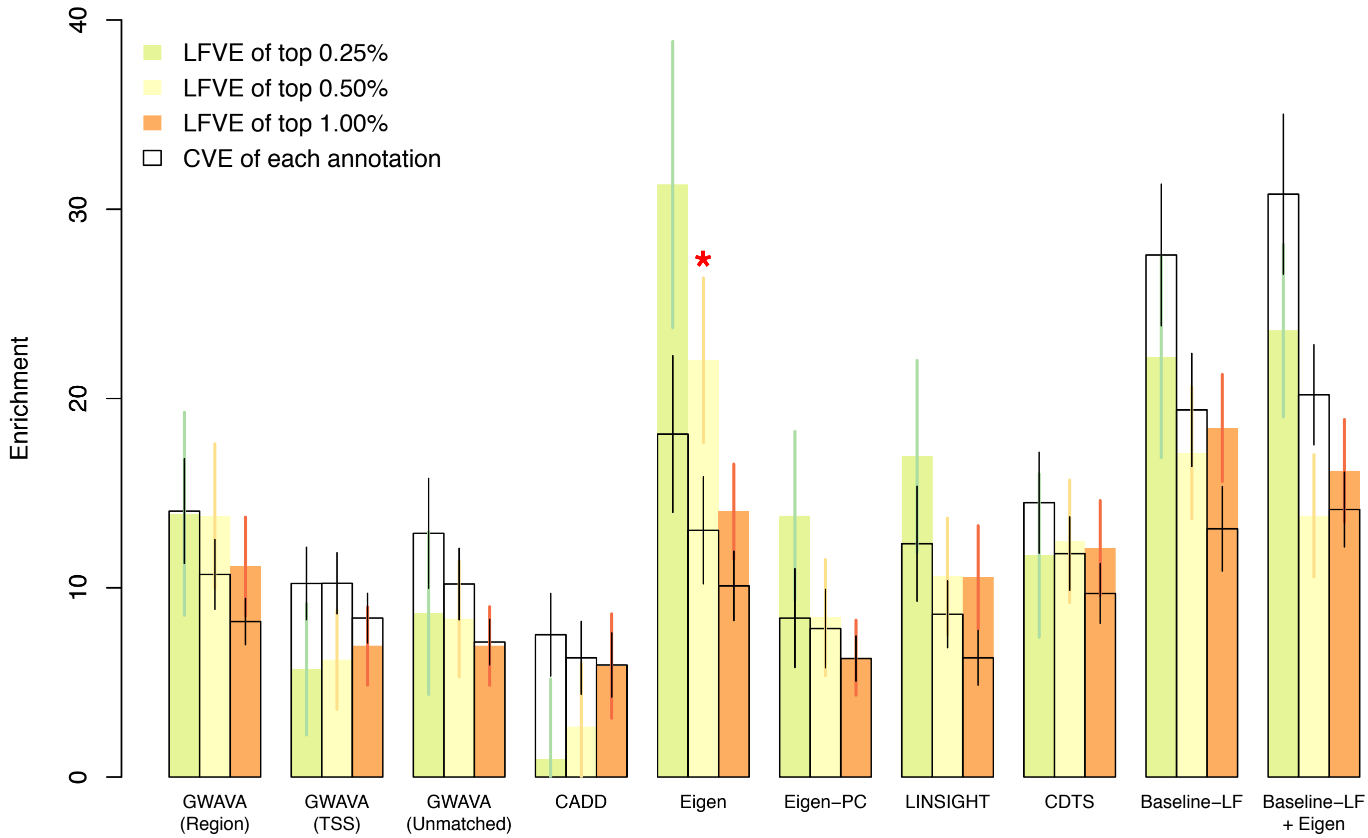


Supplementary Figure 12: Forward simulations show lower LFVE/CVE ratios for annotations with extremely large proportion of deleterious *de novo* variants (π_{del}). We used forward simulations to understand why non-synonymous variants inside genes predicted to be under strong negative selection[40] (large s_{het}; Figure 5) have larger LFVE and CVE but a lower LFVE/CVE ratio (a surprising result that appears to suggest a smaller $|\bar{s}_{dn}|$ (Figure 6b) despite the extremely large value of π_{del}). We hypothesized that our simulation result showing that larger $|\bar{s}_{dn}|$ produces larger LFVE/CVE ratios (Figure 6b) might not hold for annotations with extremely large values of π_{del}. We performed additional forward simulations where we considered a genome with a highly constrained non-synonymous annotation (0.5% of the simulated genome; π_{del} = 100%, different $\bar{s}_{dn}$, different $\pi_{del:causal}$), an unconstrained non-synonymous annotation (0.5%; π_{del} = 60%, $\bar{s}_{dn}$ = -0.003, different $\pi_{del:causal}$), and a noncoding annotation (99%; π_{del} = 40%, $\bar{s}_{dn}$ = -0.0001, $\pi_{del:causal}$ = 10%). **(a,b,c,d)** We report the CVE (a), LFVE/CVE ratio (b), mean selection coefficient of deleterious variants at different MAF (c), and proportion of deleterious variants at different MAF (d) of the highly constrained non-synonymous annotation ($\bar{s}_{dn}$ = -0.001) and the unconstrained non-synonymous annotation. **(e,f,g,h)** We report the CVE (e), LFVE/CVE ratio (f), mean selection coefficient of deleterious variants at different MAF (g), and proportion of deleterious variants at different MAF (h) of the highly constrained non-synonymous annotation ($\bar{s}_{dn}$ = -0.003) and the unconstrained non-synonymous annotation. **(i,j,k,l)** We report the CVE (i), LFVE/CVE ratio (j), mean selection coefficient of deleterious variants at different MAF (k), and proportion of deleterious variants at different MAF (l) of the highly constrained non-synonymous annotation ($\bar{s}_{dn}$ = -0.006) and the unconstrained non-synonymous annotation. In each simulation scenario we observed higher CVE and LFVE for the constrained non-synonymous annotation, but a lower LFVE/CVE ratio. We determined that the smaller LFVE/CVE ratio for the constrained non-synonymous annotation (b,f,j) is due to the smaller ratio of (proportion of low-frequency variants that are deleterious)/(proportion of common variants that are deleterious), which is reduced to 1 at extremely large values of π_{del} (d,h,l), and not to a smaller value of (mean selection coefficient for low-frequency deleterious variants)/(mean selection coefficient for common deleterious variants), which varies across the three simulations (c,g,k).

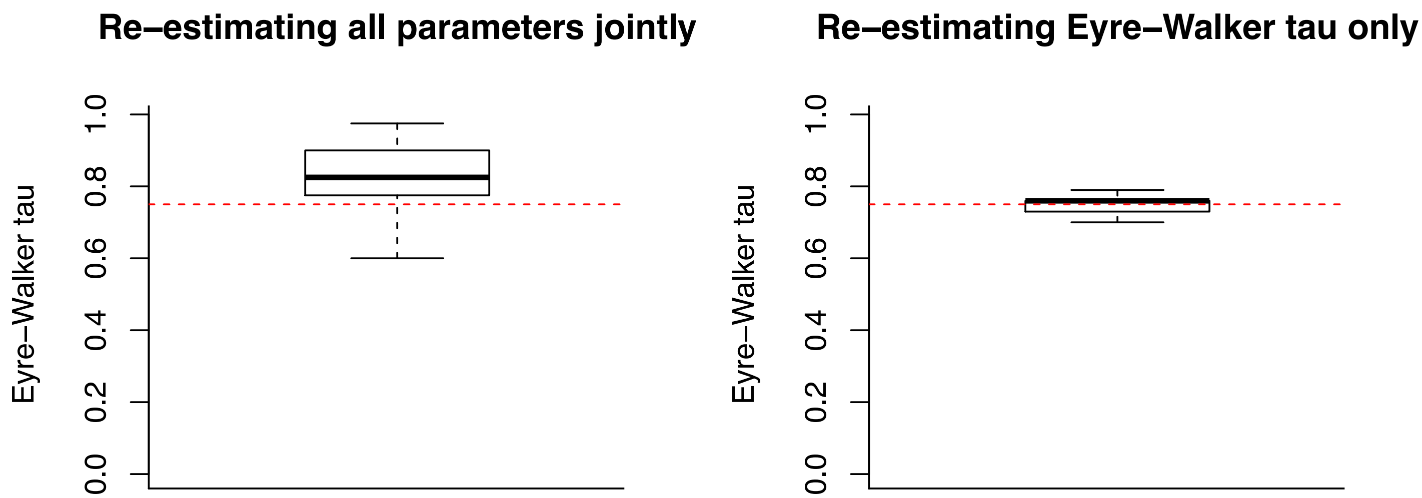
.



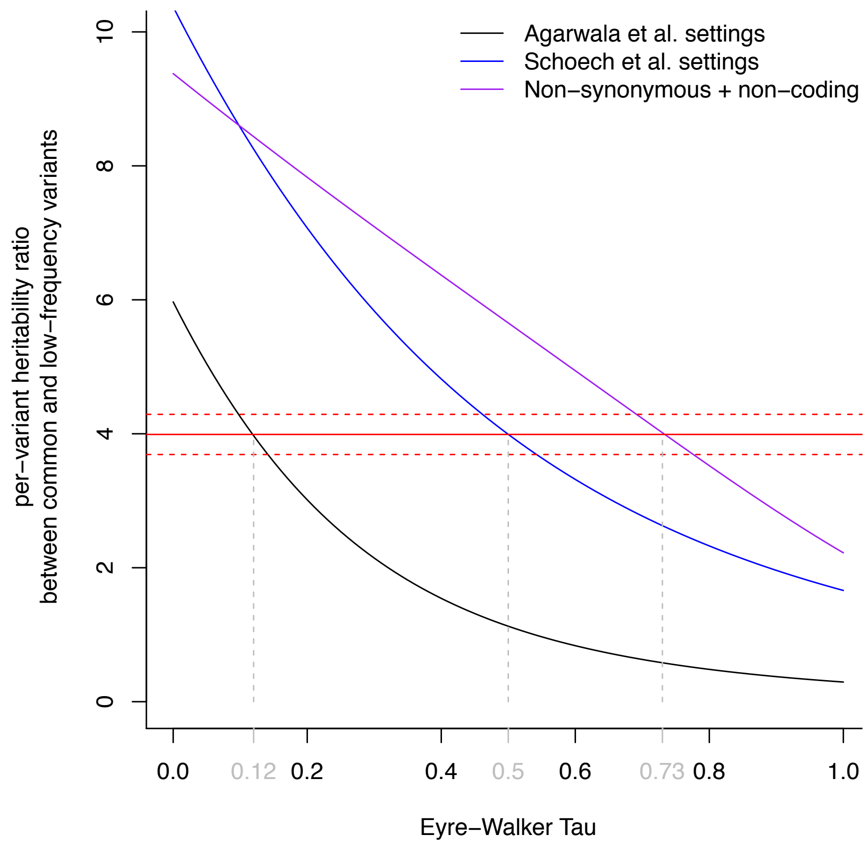
Supplementary Figure 13: Similar inferences about rare variant architectures when using different values of the proportion of causal variants (π). We report the mean squared per-allele effect size of causal variants in the functional noncoding annotation (normalized by the mean squared per-allele effect size of rare causal non-synonymous variants) as a function of $\bar{s}_{dn}$ and MAF (rare, low-frequency and common). We considered four scenarios with $\pi_{del:causal}$ = 20% and π = 12% (**a**), with $\pi_{del:causal}$ = 40% and π = 24% (**b**), with $\pi_{del:causal}$ = 60% and π = 36% (**c**), and with $\pi_{del:causal}$ = 100% and π = 60% (**d**). All results were similar to results obtained with $\pi_{del:causal}$ = 80% and π = 48% (Figure 6d).



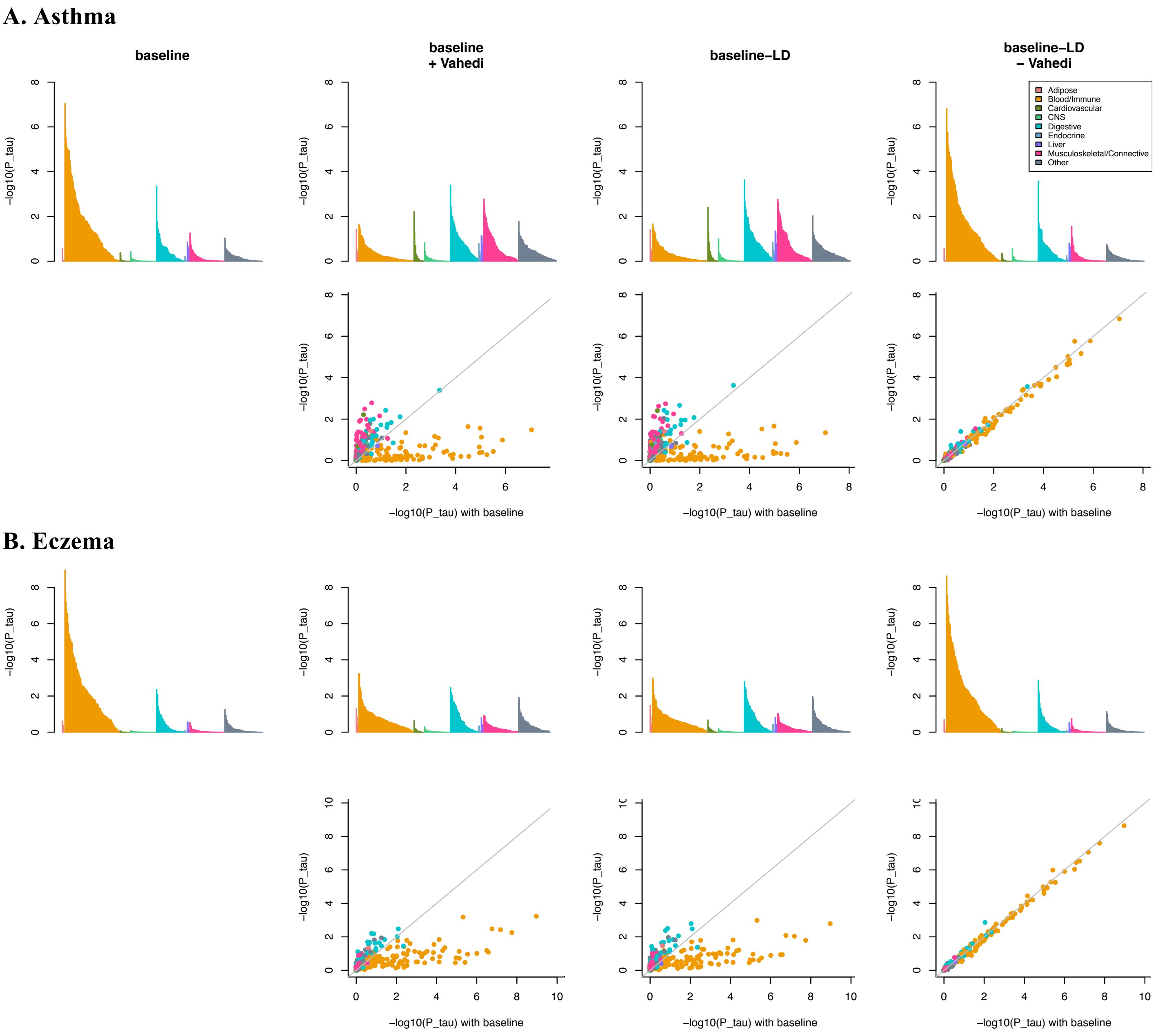
Supplementary Figure 14: Precise noncoding annotations do not have large LFVE/CVE ratio. We report CVE and LFVE for different noncoding annotations based on variant prioritization scores, or per-variant heritability based on S-LDSC outputs, meta-analyzed across 27 independent UK Biobank traits. Each bar represents enrichments of a single annotation keeping the noncoding variants with the highest score (top 0.25%, 0.50% or 1.00%) in different MAF bins, to account for the fact that some methods prioritize variants according to their MAF. For example, the annotation "top 0.25%" includes in each MAF bin (15 MAF bins in total) the 0.25% of noncoding variants with the highest score in that bin. Results were obtained from 10 x 3 = 30 independent S-LDSC analyses conditioned on the baseline-LF model. Colored bar plots represent LFVE, while black bar plots represent CVE. Colored error bars represent LFVE jackknife 95% confidence intervals, while black error bars represent CVE jackknife 95% confidence intervals. Red star indicates statistically significant difference between LFVE and CVE after correction for multiple testing ($P < 0.05/30$). First, we looked at 8 x 3 annotations derived from 5 variant prioritization scores (GWAVA[42], CADD[43], Eigen[44], LINSIGHT[45] and CDTS[46]); we found a significant difference for the annotation built from the top 0.50% of Eigen noncoding scores[44] (1.69x ratio; LFVE = 22.02x, s.e. = 2.22x, vs. CVE = 13.04x, s.e. = 1.44x; $P = 7 \times 10^{-4}$). Second, to compare these results to results of the baseline-LF model alone, we constructed annotations from predicted per-variant heritability under the baseline-LD model fit. More precisely, for each annotation d of the baseline-LF model we meta-analyzed the values $\bar{\tau}_d = \tau_d/\left(h_c^2 + h_{lf}^2\right)$, and computed predicted per-variant heritability of each variant j as $\sum_{d=1}^{D} a_d(j)\bar{\tau}_d$. For each of the 27 independent traits, we meta-analyzed $\bar{\tau}$ on the remaining 26 traits to build annotations independent of the investigated trait. While we observed that annotations based on baseline-LF per-variant heritability maximize the CVE signal, we did not attain an LFVE as large as Eigen. Finally, in an effort to explain the large LFVE obtained with the Eigen scores, we added to our baseline-LF model all the constituent continuous annotations of Eigen (annotations labeled GERP_RS, PhyloPri, PhyloPla, PhyloVer, PhastPri, PhastPla, PhastVer, H3K4Me1, H3K4Me3, H3K27ac, TFBS_max, TFBS_sum, TFBS_num, OCPval, DnaseSig, DnasePval, FaireSig, FairePval, PolIISig, PolIIPval, ctcfSig, ctcfPval, cmycSig, and cmycPval in Eigen raw files). We recomputed predicted per-variant heritability for each trait and added the top noncoding annotations to the baseline-LF model (method labeled Baseline-LF + Eigen). While this procedure helped to increase CVE, we observed a decreased LFVE. These results highlight the Eigen strategy for computing prioritization scores.



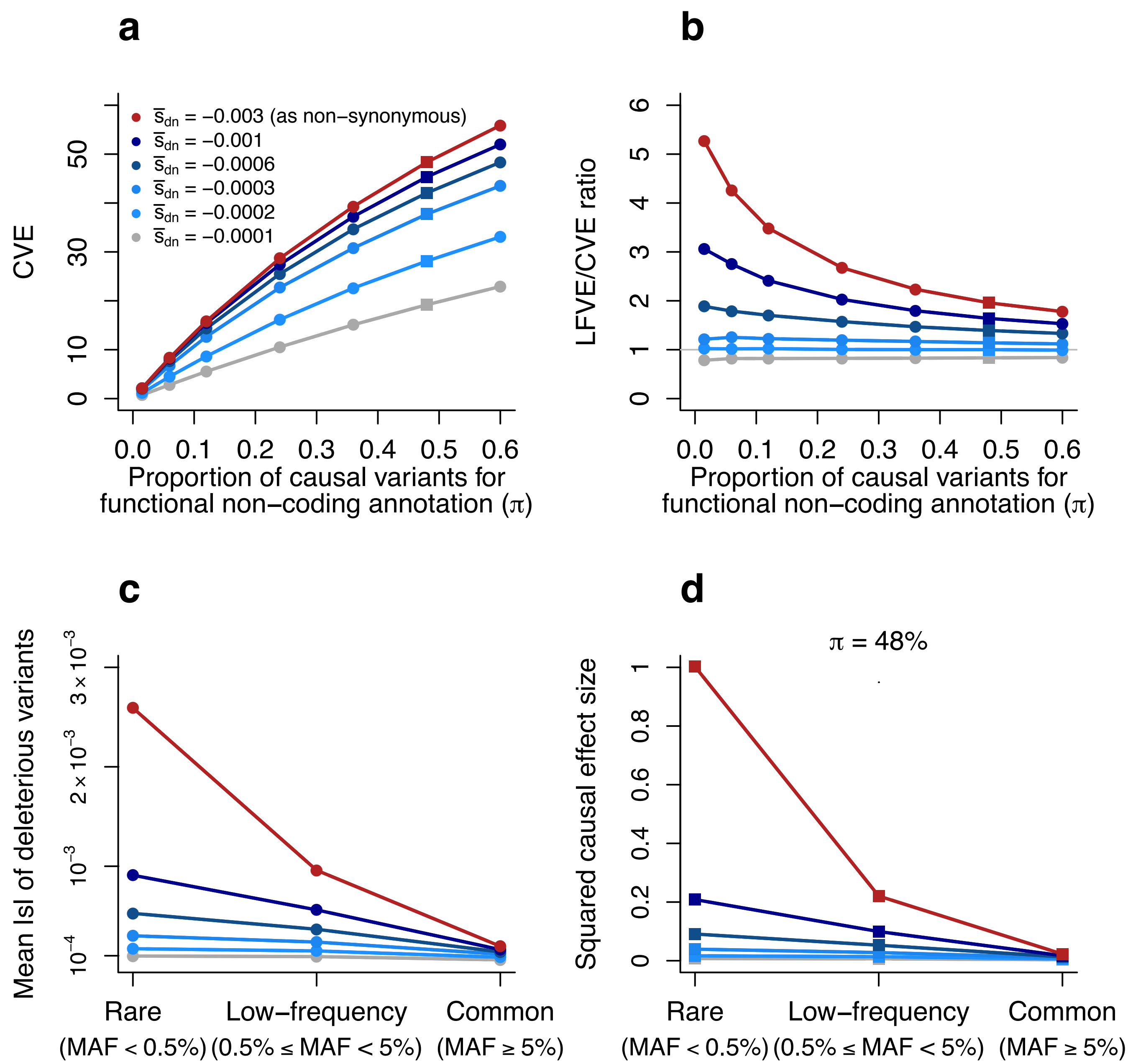
Supplementary Figure 15: τ_{EW} estimate from forward simulations is robust to error bars in UK Biobank LFVE and CVE estimates. We observed that our τ_{EW} estimate (0.75) is much larger than previous estimates[22,31,47,48], primarily due to the fact that we explicitly model different distributions of selection coefficient for non-synonymous and noncoding variants. To assess the robustness of our τ_{EW} estimate, we randomly resampled 1,000 times the values of per-variant heritability ratio between common and low-frequency variants, non-synonymous LFVE, and non-synonymous CVE from their estimated values and standard errors (i.e. 3.99±0.15x, 38.25±2.27x and 7.72±0.85x, respectively) and re-estimated τ_{EW} by choosing the scenario (of 22,400 different scenarios; see Supplementary Table 18) matching the resampled values of per-variant heritability ratio between common and low-frequency variants, non-synonymous LFVE, and non-synonymous CVE. We report the estimated τ_{EW} when re-estimating it at the same time than the π_{del}, $\bar{s}_{dn}$ and θ values for noncoding variants (left), and the estimated τ_{EW} when fixing the π_{del}, $\bar{s}_{dn}$, and θ for noncoding variants to the value initially estimated (right). The horizontal red line represents the τ_{EW} estimate in our primary analysis (0.75). We observed that τ_{EW} estimates tended to be slightly larger than 0.75, confirming that the large estimate of τ_{EW} is robust to error bars in UK Biobank LCVE and CVE estimates.



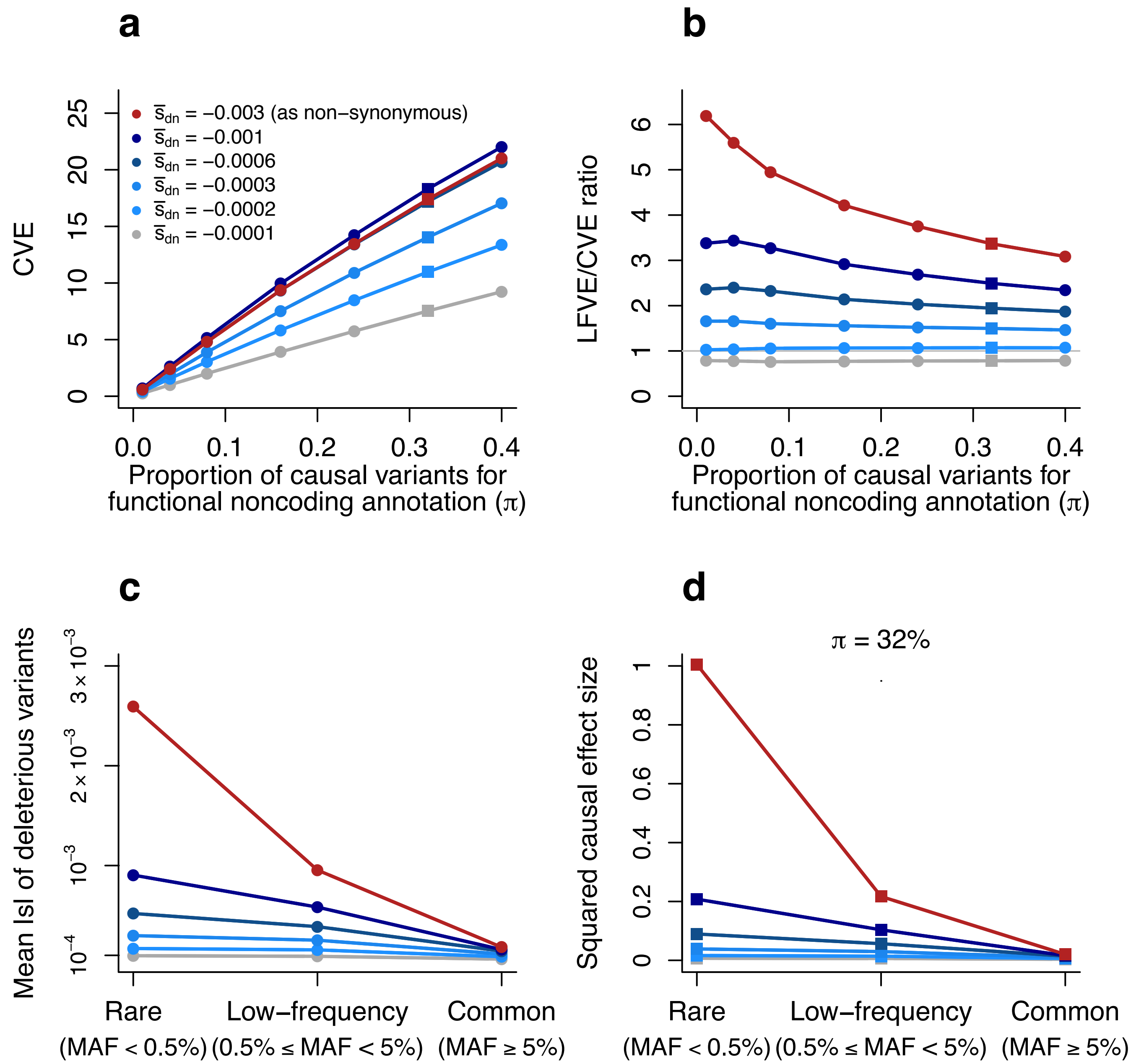
Supplementary Figure 16: Large estimate of τ_{EW} is explained by allowing different distributions of selection coefficients for non-synonymous and noncoding elements. We report the per-variant heritability ratio between common and low-frequency variants generating under 3 forward simulation scenarios with different distribution of selection coefficient, and using different values of Eyre-Walker coupling coefficient between selection and phenotypic effect (τ_{EW}). For each simulation scenario, we highlighted the τ_{EW} value giving the per-variant heritability ratio between common and low-frequency variants closest to the value estimated in UK Biobank data (i.e. 4.0, represented by a redline; 95% confidence intervals are represented by dashed red lines). The first simulation scenario (black curve) mimics Agarwala et al. settings[31] (also used in ref.[48]), where 1% of the genome contains non-synonymous elements (π_{del} = 80% and $\bar{s}_{dn}$ = -0.003) and 99% contains noncoding elements (π_{del} = 0%, i.e. all variants are neutral); τ_{EW} value closest to our UK Biobank results is 0.12, comparable to 0.10 inferred by ref. [48]. The second simulation scenario (blue curve) mimics Schoech et al. settings[22], where every *de novo* mutation is deleterious (π_{del} = 100% and $\bar{s}_{dn}$ = -0.001); τ_{EW} value closest to our UK Biobank results is 0.50, comparable to 0.40 inferred by ref. [22]. The third simulation scenario (purple curve) uses our settings for non-synonymous elements (1% of the genome, π_{del} = 80% and $\bar{s}_{dn}$ = -0.003), and noncoding elements (99% of the genome, π_{del} = 40% and $\bar{s}_{dn}$ = -0.0001); τ_{EW} value closest to our UK Biobank results is 0.73. These results highlight that our larger τ_{EW} estimate (0.75) compared to other studies is explained by the fact that our simulation procedure allows different distributions of selection coefficients for non-synonymous and noncoding variants.



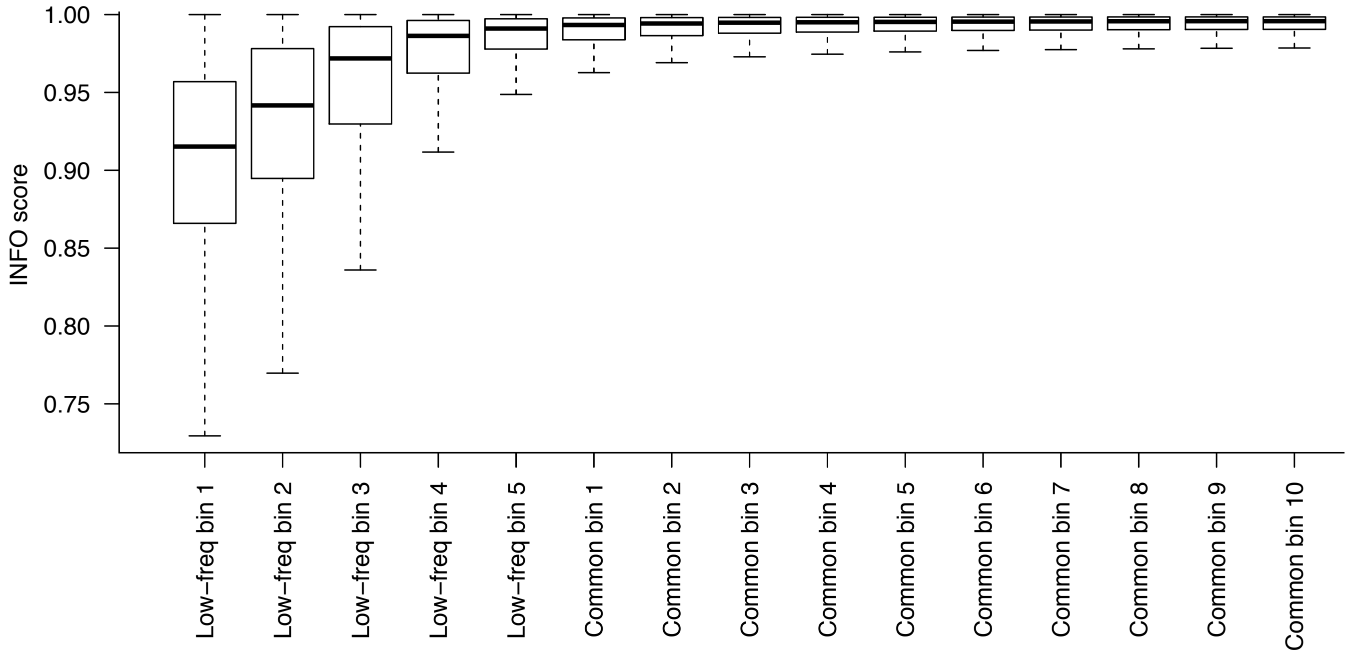
Supplementary Figure 17: Cell-type-specific (CTS) analyses using different versions of the baseline model support removing CTS annotations from the baseline-LD model. We report significance of the τ coefficient of each CTS Roadmap annotation for asthma (**A**) and eczema (**B**) using different versions of the baseline model. Bar plots (top rows) show the significance of each CTS annotation, sorted by cell-type[13,49] group and sorted by significance within each cell-type group. Scatter plots (bottom rows) show the significance of the τ coefficient of CTS annotations for each model vs. the baseline model. Baseline model refers to the set of 53 annotations of Finucane et al.[13]. Baseline + Vadehi refers to the set of annotations in the baseline model plus 4 annotations based on Vahedi et al. T-cell enhancers[50] (i.e. 57 annotations). Baseline-LD refers to the set of 75 annotations of Gazal et al.[8] (which includes the 4 Vahedi et al. annotations). Baseline-LD - Vadehi refers to the set of annotations in the baseline-LD model minus the 4 annotations based on Vahedi et al. T-cell enhancers (i.e. 71 annotations). We observed that adding Vahedi et al. annotations to the baseline model kills the signal of blood related CTS annotations, while removing them from the baseline-LD produces results similar to the original baseline model.



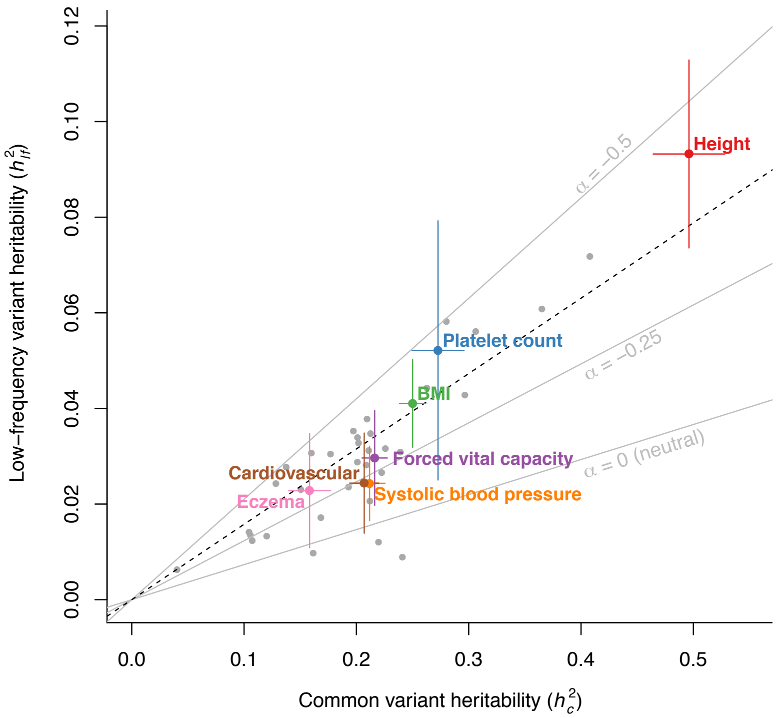
Supplementary Figure 18: Forward simulations conclusions are robust to the choice of $\pi_{del:causal}$ for non-synonymous and ordinary noncoding annotations. This figure replicates Figure 6 by using $\pi_{del:causal}$ = 5% instead of $\pi_{del:causal}$ = 10% for the non-synonymous and ordinary noncoding annotations (see Figure 6 caption for details). While decreasing $\pi_{del:causal}$ increases the CVE of the functional noncoding annotation (a), this does not affect results on the LFVE/CVE ratio, distribution of selection coefficients and squared effect size of causal variants in the functional noncoding annotation (b, c and d).



Supplementary Figure 19: Forward simulations conclusions are robust to the choice of π_{del} for the functional noncoding annotation. This figure replicates Figure 6 by using π_{del} = 40% (same as ordinary noncoding annotation) instead of π_{del} = 60% for the functional noncoding annotation (see Figure 6 caption for details).



Supplementary Figure 20: INFO score within each of the 15 MAF bins shows high imputation accuracy for low-frequency variants. We report INFO score boxplots with each MAF bin of the baseline-LF model (see Supplementary Table 17).



Supplementary Figure 21: Common variant heritability (h_c^2) and low-frequency variant heritability (h_{lf}^2) estimates for 40 UK Biobank traits using summary statistics computed using linear regression with 20 PCs on *N* = 337K unrelated individuals with UK ancestry. We report h_c^2 and h_{lf}^2 estimated by S-LDSC with the baseline-LF model for 40 UK Biobank traits (for binary traits, estimates are on the liability scale), with 7 representative independent traits highlighted. Compared to Figure 2 (generated using summary statistics with BOLT-LMM v2.3 on *N* = 408,963 related individuals with UK ancestry and by correcting h_c^2 and h_{lf}^2 estimates by N_{eff}), S-LDSC was run using summary statistics computing using linear regression with 20 principal component covariates on *N* = 337,539 unrelated individuals with UK ancestry. Error bars represent 95% confidence intervals. The dashed black line represents the meta-analyzed ratio between h_{lf}^2 and h_c^2 for 27 independent traits (1/6.34). Grey lines represent expected ratios for different values of α (see main text). Error bars represent 95% confidence intervals. We observed 1) that the corrected h_c^2 estimates using BOLT-LMM on *N* = 409K (meta-analyzed h_c^2 = 0.166, s.e. = 0.013) is 0.975 times higher than the h_c^2 estimates using linear regression on *N* = 337K (meta-analyzed h_c^2 = 0.171, s.e. = 0.014), 2) that the corrected h_{lf}^2 estimates using BOLT-LMM on *N* = 409K (meta-analyzed h_{lf}^2 = 0.024, s.e. = 0.002) is 1.024 times higher than the h_{lf}^2 estimates using linear regression on *N* = 337K (meta-analyzed h_{lf}^2 = 0.023, s.e. = 0.003), and 3) that the corrected h_c^2/h_{lf}^2 ratio estimates using BOLT-LMM on *N* = 409K (meta-analyzed h_c^2/h_{lf}^2 = 6.285, s.e. = 0.232) is similar to the h_c^2/h_{lf}^2 ratio estimates using linear regression on *N* = 337K (meta-analyzed h_c^2/h_{lf}^2 = 6.343, s.e. = 0.261).

Supplementary Tables

See Excel file.

Supplementary Table 1: Main functional annotations of the baseline-LF model. We report the 40 main annotations of the baseline-LF model, including the 34 main annotations from our "baseline-LD model" (27 binary annotations, 1 continuous conserved annotation, and 6 continuous LD-related annotations) and 6 new functional annotations. We report the number of variants inside each binary annotation.

See Excel file.

Supplementary Table 2: All annotations of the baseline-LF model. We report the 163 annotations of the baseline-LF model. We report the number of variants inside each binary annotation.

See Excel file.

Supplementary Table 3: S-LDSC results using different sets of heritability variants, reference variants, and regression variants in UK10K simulations, in which the target sample and the LD reference panel match perfectly. For each simulation scenario, we report the mean true and estimated values of h^2_{lf} and h^2_c, and corresponding biases. Line in bold indicates the sets of variants retained for all further analyses. Numbers in brackets are standard errors. We note that we observed upward bias when considering MAF thresholds of 0.1% and 0.2% due to 1) different MAF-dependent architecture in simulations and in the S-LDSC model, and/or 2) noisy estimates of LD scores as these MAF thresholds correspond to at least 8 and 15 allele counts in 3,567 individuals of the LD reference genome.

See Excel file.

Supplementary Table 4: Estimation of heritability and enrichment in UK Biobank simulations. For each simulation scenario, we report **(a)** the mean true and estimated values of h^2_c and h^2_{lf}, as well as the CVE, LFVE, and LFVE/CVE ratio of the enriched annotation, and **(b)** the corresponding bias. S-LDSC was run using all variants (regardless of INFO score[51]), by restricting regression variants to variants with INFO score ≥ 0.99, 0.90 or 0.50, and by taking into account INFO scores into the regression. Numbers in brackets are standard errors.

See Excel file.

Supplementary Table 5: UK Biobank traits analyzed. For each of 40 UK Biobank traits analyzed, we report the sample size *N*, the effective sample size N_{eff} from BOLT-LMM summary statistics, the prevalence (for binary traits), the estimated h^2_c and h^2_{lf} (for binary traits, heritability estimates are on the liability scale), the corresponding *z*-scores for non-zero heritability using S-LDSC with the baseline-LF model, and information on the UK Biobank data fields used to construct the phenotypes. We also indicate the 27 independent UK Biobank traits used in meta-analyses.

See Excel file.

Supplementary Table 6: LFVE and CVE of baseline-LF annotations for UK Biobank traits. We report the LFVE, CVE, and related information for all binary annotations of the baseline-LF model (and also for all coding and UTR variants) for the 40 UK Biobank traits analyzed as well as for the meta-analysis across 27 independent traits.

See Excel file.

Supplementary Table 7: Proportion of heritability explained by quintiles of continuous baseline-LF annotations. We report the proportion of h_c^2 and h_{lf}^2 explained by each quintile of the 7 continuous annotations of the baseline-LF model for the 40 UK Biobank traits analyzed as well as for the meta-analysis across 27 independent traits. For predicted allele age, we were unable to estimate LFVE due to many missing values for low-frequency variants; we report CVE only. We note that the LLD-AFR annotation does not have 20% in each low-frequency quintile due to many tied values (for low-frequency variants that are not present in the African populations and thus have a null LD score).

See Excel file.

Supplementary Table 8: Forward simulations confirm larger effects of low-frequency variants in LD-related annotations. We performed similar forward simulations as in Gazal et al.[8]. We used an out-of-Africa demographic model, different recombination rate patterns, segments with 60% or 90% of deleterious variants to model a non-homogeneous distribution of deleterious variants, and a mean selection coefficient for *de novo* variants $\bar{s}$ = -0.05. We outputted a sample of 5,000 European individuals, and a sample of 500 African individuals (to compute LLD-AFR). We simulated 200 regions of 1 Mb. We report the standardized regression coefficients for each of three LD-related annotations in regressions of absolute selection coefficient of common variants (2^{nd} column), and low-frequency variants (3^{rd} column) against these annotations and MAF bins (10 for common variants, 5 for low-frequency variants). Note that we did not consider the background selection statistic and CpG-content annotations as they rely on UK Biobank data. Numbers in parentheses are standard errors. We observed a higher impact of LLD-AFR and nucleotide diversity on the low-frequency architecture, but similar effect for recombination rate. However, while for common variants the negative effect of recombination rate due to negative selection[36] cancels out the positive effect of recombination rate due to functional annotations[8] (enriched regulatory annotations tend to be in high recombination rate regions), we observed that effect of negative selection exceeds the effect of functional annotations for low-frequency variants, consistent with the observation that recombination coldspots are enriched in rare disease-associated mutations[37]. The higher effect observed in simulations for the LLD-AFR annotation is due to the unrealistic $\bar{s}$ value (-0.05, as recommended in the SLiM manual and used in ref. [8]), which is more deleterious to that what we considered in this paper as realistic (i.e. $\bar{s}$ = -0.003 for non-synonymous variants and $\bar{s}$ = -0.0001 for noncoding variants).

See Excel file.

Supplementary Table 9: Cell-type specific (CTS) annotation analysis. For each of the 40 UK Biobank traits and for each of the 396 Roadmap CTS annotations, we report the S-LDSC τ coefficient when conditioned on all of the (non-CTS) annotations of the baseline-LD model. For each trait, we selected CTS annotations for which τ coefficients have $P < 0.05/396$ (637 total trait-annotation pairs).

See Excel file.

Supplementary Table 10: LFVE and CVE of 637 trait-annotation pairs. We report the LFVE, CVE, and related information for all 637 significant trait-annotation pairs (see Supplementary Table 9). For neuroticism and age at first birth, H3K4me3* (in row of table denoted in red) refers to H3K4me3 annotations with all UTR and coding variants removed.

See Excel file.

Supplementary Table 11: LFVE and CVE for non-synonymous variants in 5 bins of s_{het}. We report the LFVE, CVE, and related information for non-synonymous variants in 5 bins of s_{het} (see main text), meta-analyzed across 27 independent UK Biobank traits.

See Excel file.

Supplementary Table 12: Forward simulations results. Each simulation scenario considered different values of $\bar{s}_{dn}$, $\pi_{del:causal}$ and π for the functional noncoding annotation (see main text). We report the CVE, the LFVE, and the rare variant enrichment (RVE) of the functional noncoding, non-synonymous, and ordinary noncoding annotations. We also report 1) the mean squared effect sizes of the common, low-frequency and rare causal variants in these 3 annotations; 2) the proportion of common, low-frequency and rare variants that lie in these annotations; and 3) the proportion of common, low-frequency and rare variants that are deleterious within each annotation.

See Excel file.

Supplementary Table 13: Best-fit α parameter values for non-synonymous and other variants using UK Biobank results. We considered a model in which the variance of per-normalized genotype effect sizes is proportional to $c * \left(2p(1-p)\right)^{1+\alpha_{non-synonymous}}$ for non-synonymous variants, and $\left(2p(1-p)\right)^{1+\alpha_0}$ for other variants, with p the frequency of the variant in UK10K. For a grid of 15 values of c, 24 values of $\alpha_{non-synonymous}$, and 9 values of α_0, we computed the expected LFVE and CVE of the non-synonymous annotation (columns *NonSynonymous_LFVE* and *NonSynonymous_CVE*, respectively), and the per-variant heritability for common vs. low-frequency variants (column *per-h2_ratio*), by using UK10K as a reference genome. The column *in95%CI* indicates whether these values of LFVE, CVE and per-variant heritability ratio lie within the 95% confidence intervals of our corresponding estimates in UK Biobank data (lines highlighted in yellow match these criteria). The line in red indicates the scenario most closely matching our UK Biobank results (i.e. c = 3, $\alpha_{non-synonymous}$ = -1.10, α_0 = -0.30).

	Standard S-LDSC	S-LDSC extended to low-frequency variants
LD reference panel	479 European from 1000 Genomes phase 3	3,567 individuals from UK10K
Reference variants (used to compute LD scores)	MAF ≥ 0.52% (Allele Count ≥ 5) (9,997,231variants)	MAF ≥ 0.1% (11,830,279 variants)
Regression variants (used to estimate τ parameters)	HapMap 3 variants (around 1 million variants)	MAF ≥ 0.5% (8,751,990 variants)
Heritability variants (used to estimate heritability and enrichment)	MAF ≥ 5% (5,961,159 variants)	MAF ≥ 0.5% (8,751,990 variants)
Variance of effect sizes (D_f refers to functional annotations)	$\sum_{d=1}^{D_f} a_d(j)\tau_d$	$\sum_{d=1}^{D_f} a_d(j).\left(1_{j\in(c)}\tau_d^{(c)} + 1_{j\in(lf)}\tau_d^{(lf)}\right)$ $= \sum_{d=1}^{D_f} a_d^{(c)}(j).\tau_d^{(c)} + a_d^{(lf)}(j).\tau_d^{(lf)}$
Additional annotations	1 annotation including all variants 10 MAF bins for common variants	1 annotation including all variants 10 MAF bins for common variants 5 MAF bins for low-frequency variants
Regression* (D refers to all annotations in the model)	$E[\chi_j^2] = N\sum_{d=1}^{D} \tau_d l(j,d) + Nb + 1$	$E[\chi_j^2] := N\sum_{d=1}^{D} \tau_d l(j,d) + Nb + 1$
Outputs of interest (d refers to the D_f functional annotations)	τ_d, h_c^2 and CVE$_d$	$\tau_d^{(c)}$, $\tau_d^{(lf)}$, h_c^2, h_{lf}^2, CVE$_d$ and LFVE$_d$

Supplementary Table 14: Main differences between the standard version of S-LDSC for common variants and its extension to low-frequency variants. Note that D_f refers to the number of functional annotations while D refers to the total number of annotations when MAF bins have been included and when functional annotations have been duplicated for low-frequency and common variants. See main manuscript for more details on the equations.

See Excel file.

Supplementary Table 15: Phenotypic correlations between the 40 UK Biobank traits. We report the squared phenotypic correlation between the 40 UK Biobank traits analyzed.

See Excel file.

Supplementary Table 16: Analyses informing the choice of new conserved and functional annotations selected for inclusion in the baseline-LF model. For each annotation of interest, we reported its per-standardized effect size τ^* (ree ref. [8] for details), τ^* standard error, τ^* P value, CVE and CVE standard error. All analyses were performed using the Gazal et al. framework and baseline-LD model[8]. Meta-analyses were performed on 20 independent traits (the 31 independent traits of Gazal et al. minus the 11 phenotypes from 23andMe). **(a)** We added annotations based on phastCons[28] conserved elements (46 way) in vertebrates, mammals and primates. We observed that phastCons conserved elements in primates were explaining most of the conserved signals. GERP and Lindblad-Toh et al.[35] annotations did not have significant effect anymore, but we decided to keep them in the model, in order to compare results with previous papers using the baseline/baseline-LD models. We also kept phastCons conserved elements in vertebrates and mammals to compare conserved enrichment in different types of species. We added 500 bp windows around each phastCons annotations, consistently with what we did for each binary annotations based on genomic regions[13]. **(b)** We investigated annotations based on ChromHMM 15-state model from Roadmap data (see URLs). For each annotation, we combined the data from the different cell types into a single annotation by taking a union. For each analysis, we added 500 bp windows around each ChromHMM annotations. We retained only flanking bivalent TSS/enhancers (BivFlnk) annotation. Bivalent/Poised TSS (TssBiv) annotation was not retained, as it is not significant after conditioning on BivFlnk.

See Excel file.

Supplementary Table 17: Boundaries of the 15 MAF bins of the baseline-LF model. We report boundaries of each MAF bin, based on the UK10K LD reference panel.

See Excel file.

Supplementary Table 18: Results of forward simulations to assess the distribution of fitness effects (DFE) of noncoding variants and coupling parameter between selection coefficient and trait effect size (τ_{EW}). We report results for 22,400 simulation scenarios (see main text). In all simulations scenarios non-synonymous elements represent 1% of the genome, and the DFE of non-synonymous variants was set using π_{del} = 80%, $\bar{s}$ = -3.16 x 10^{-3} and θ = 0.32. The DFE of noncoding variants was set using a grid of parameters of π_{del}, $\bar{s}$ and θ (columns *Prop_deleterious*, *Mean_s_deleterious*, and *Shape_gamma*, respectively); logarithmic grids were used for $\bar{s}$ and θ, as in ref[31]. Effect sizes were simulated using a grid of values τ_{EW}. We report for each simulation *i*) the ratio between the proportion of coding variants in low-frequency variants and the proportion of coding variants in common variants (column *Diversity_NonSynonymous*); *ii*) the ratio between the number of low-frequency and common variants (column *#low-freq/#common*); *iii*) CVE, LFVE and rare variant enrichment (RVE) for non-synonymous variants (columns *NonSynonymous_CVE*, *NonSynonymous_LFVE*, and *NonSynonymous_RVE*, respectively); and *iv*) the per-variant heritability ratio between common and low-frequency variants (column *per-h2_ratio*). Asterisks indicate columns that were used to match simulation results to UK Biobank results. Rows in yellow denote the 11 scenarios matching the 95% confidence intervals of UK Biobank results (last raws).

See Excel file.

Supplementary Table 19: Results of forward simulations matching 95% confidence intervals of UK Biobank results. We report results for the 11 scenarios matching the 95% confidence intervals of UK Biobank results for non-synonymous LFVE, non-synonymous CVE and per-variant heritability ratio between common and low-frequency variants (values in last row). The row in red font denotes the scenario most closely matching UK Biobank results.

References

1. The UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature* **526,** 82–90 (2015).
2. 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).
3. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med.* **12,** e1001779 (2015).
4. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 166298 (2017). doi:10.1101/166298
5. The Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48,** 1279–1283 (2016).
6. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47,** 284–290 (2015).
7. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50,** 906–908 (2018).
8. Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49,** 1421–1427 (2017).
9. Speed, D. *et al.* Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49,** 986–992 (2017).
10. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42,** 565–569 (2010).
11. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43,** 519–525 (2011).
12. Pasaniuc, B. & Price, A. L. Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.* **18,** 117–127 (2017).
13. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47,** 1228–1235 (2015).
14. Eyre-Walker, A. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc. Natl. Acad. Sci.* **107,** 1752–1756 (2010).
15. Boyko, A. R. *et al.* Assessing the Evolutionary Impact of Amino Acid Mutations in the Human Genome. *PLOS Genet.* **4,** e1000083 (2008).
16. Gazal, S., Finucane, H. K. & Price, A. L. Reconciling S-LDSC and LDAK functional enrichment estimates. *bioRxiv* 256412 (2018). doi:10.1101/256412
17. Speed, D. & Balding, D. Better estimation of SNP heritability from summary statistics provides a new

understanding of the genetic architecture of complex traits. *bioRxiv* 284976 (2018). doi:10.1101/284976

18. Marquez-Luna, C. *et al.* Modeling functional enrichment improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. *bioRxiv* 375337 (2018). doi:10.1101/375337

19. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).

20. 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).

21. Zeng, J. *et al.* Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet.* **50,** 746–753 (2018).

22. Schoech, A. *et al.* Quantification of frequency-dependent genetic architectures and action of negative selection in 25 UK Biobank traits. *bioRxiv* 188086 (2017). doi:10.1101/188086

23. Moore, C. B. *et al.* Low Frequency Variants, Collapsed Based on Biological Knowledge, Uncover Complexity of Population Stratification in 1000 Genomes Project Data. *PLOS Genet.* **9,** e1003959 (2013).

24. Leslie, S. *et al.* The fine-scale genetic structure of the British population. *Nature* **519,** 309–314 (2015).

25. Liu, X. *et al.* Functional Architectures of Local and Distal Regulation of Gene Expression in Multiple Human Tissues. *Am. J. Hum. Genet.* **100,** 605–616 (2017).

26. Hormozdiari, F. *et al.* Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nat. Genet.* **50,** 1041–1047 (2018).

27. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).

28. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15,** 1034–1050 (2005).

29. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).

30. Rasmussen, M. D., Hubisz, M. J., Gronau, I. & Siepel, A. Genome-wide inference of ancestral recombination graphs. *PLoS Genet* **10,** e1004342 (2014).

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

32. Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proc. Natl. Acad. Sci.* **108,** 11983–11988 (2011).

33. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248 (2010).

34. Davydov, E. V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **6,** e1001025 (2010).

35. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478,** 476–482 (2011).

36. Hill, W. G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res.* **8,** 269–294 (1966).

37. Hussin, J. G. *et al.* Recombination affects accumulation of damaging and disease-associated mutations in human populations. *Nat. Genet.* **47,** 400–404 (2015).

38. Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S. & Goldstein, D. B. Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. *PLOS Genet.* **9,** e1003709 (2013).

39. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).

40. Cassa, C. A. *et al.* Estimating the selective effects of heterozygous protein-truncating variants from human exome data. *Nat. Genet.* **49,** 806–810 (2017).

41. Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46,** 944–950 (2014).

42. Ritchie, G. R. S., Dunham, I., Zeggini, E. & Flicek, P. Functional annotation of non-coding sequence variants. *Nat. Methods* **11,** 294–296 (2014).

43. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46,** 310–315 (2014).

44. Ionita-Laza, I., McCallum, K., Xu, B. & Buxbaum, J. D. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.* **48,** 214–220 (2016).

45. Huang, Y.-F., Gulko, B. & Siepel, A. Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. *Nat. Genet.* **49,** 618–624 (2017).

46. di Iulio, J. *et al.* The human noncoding genome defined by genetic diversity. *Nat. Genet.* **50,** 333–337 (2018).

47. Mancuso, N. *et al.* The contribution of rare variation to prostate cancer heritability. *Nat. Genet.* **48,** 30–35 (2015).

48. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536,** 41–47 (2016).

49. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50,** 621–629 (2018).

50. Vahedi, G. *et al.* Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature* **520,** 558–562 (2015).

51. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11,** 499–511 (2010).