Reconciling S-LDSC and LDAK functional enrichment estimates

Steven Gazal[1,2]*, Hilary K. Finucane[2]*, Alkes L. Price[1,2,3]*

1. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.

2. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

3: Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.

*: Equal contribution. Correspondence should be addressed to S.G. (sgazal@hsph.harvard.edu), H.K.F. (finucane@broadinstitute.org), or A.L.P (aprice@hsph.harvard.edu).

Abstract

A recent study proposing a new LDAK method (Speed et al. 2017 Nat Genet) reported that functional enrichments (e.g. coding, conserved, regulatory) estimated by LDAK (largest significant enrichment: 2.51x) were much lower than previous estimates obtained using stratified LD score regression (S-LDSC). To investigate this, we developed a method (S-LDSC+LDAK) that combines our S-LDSC method with annotations constructed from LDAK model weights, and determined that this method produced unbiased estimates both in simulations under the S-LDSC model and in simulations under the LDAK model, unlike existing methods. We applied S-LDSC+LDAK to 16 independent UK Biobank traits, and determined that S-LDSC+LDAK enrichment estimates (largest enrichment: 7.51x) were nearly identical to S-LDSC estimates across 28 main annotations. On the other hand, LDAK enrichment estimates (largest enrichment: 3.96x) were substantially lower than S-LDSC estimates (although the discrepancy was smaller than reported by Speed et al., who did not compare the two methods on the same data set). Our results advocate for using S-LDSC in preference to LDAK to infer functional enrichment and confirm the existence of functional annotations that are highly enriched (>>2.51x) for complex trait heritability, providing strong caveats to the LDAK results reported by Speed et al.

Partitioning heritability by functional annotation can inform disease and trait biology. We previously developed the stratified LD score regression (S-LDSC) method for partitioning heritability by functional annotation using summary association statistics and applied it to a broad set of diseases and complex traits using a "baseline model" that includes coding, conserved and regulatory annotations[1]. A recent study of Speed et al.[2] proposed a new LDAK method that relies on the very strong assumption that marginal association statistics (which include the effects of linked SNPs) are independent of a SNP's linkage disequilibrium (LD) with other SNPs. This study reported that functional enrichments of our baseline model annotations estimated by LDAK (19 diseases; average N = 7K) were much lower than our published estimates from S-LDSC (9 independent diseases and traits; average N = 96K); the largest significant enrichment was 2.51x for LDAK vs. 13.32x for S-LDSC, decreasing to 9.35x for S-LDSC when using a "baseline-LD model" that accounts for LD-dependent architectures[3]. We sought to understand the discrepancy between S-LDSC and LDAK estimates; the reason for this discrepancy was not clear from ref. [2], which did not perform any functional enrichment simulations.

We developed a method (S-LDSC+LDAK) that combines our S-LDSC method with annotations constructed from LDAK model weights (see Supplementary Note). To verify that S-LDSC+LDAK was unbiased under both S-LDSC and LDAK models, we performed simulations in which we simulated effect sizes using either (a) the baseline-LD model with previously estimated parameters[3], including coding enrichment, or (b) coding enrichment under the LDAK model (see Supplementary Note). We compared 4 methods for estimating the proportion of heritability explained by coding variants: S-LDSC using the baseline-LD model (S-LDSC), LDAK using all SNPs (LDAK-nofilters), LDAK with default SNP filtering[2] (LDAK), and S-LDSC+LDAK. Results are reported in Figure 1a-b and Table S1. As expected, S-LDSC and LDAK-nofilters were unbiased when we simulated coding enrichment under their corresponding models. S-LDSC produced unstable estimates under the LDAK model (although S-LDSC with constrained intercept produced stable estimates, see Figure S1; this inconsistency between S-LDSC and S-LDSC with constrained intercept was not observed in analyses of real phenotypes, see Supplementary Note). LDAK was downward biased under its own model (as it restricted analyses to well-imputed SNPs), even more downward biased under the S-LDSC model,

and upward biased in simulations under its own model with no functional enrichment (unlike LDAK-nofilters and S-LDSC+LDAK; Figure S1). Notably, S-LDSC+LDAK was unbiased in simulations under both S-LDSC and LDAK models. We obtained similar results in simulations using conserved and DNase I hypersensitive sites (DHS) (Figure S1). We thus defined the S-LDSC+LDAK method as a "gold standard" for further analyses.

We compared the S-LDSC+LDAK and S-LDSC methods on real data by meta-analyzing their enrichment estimates for 28 main functional annotations from the baseline model (Table S2) across 16 independent UK Biobank quantitative traits[4–6] (average N = 434K; see Supplementary Note and Table S3). S-LDSC+LDAK and S-LDSC estimates were nearly identical (normalized mean square error (nMSE) = 0.002, no annotation significantly different; see Supplementary Note), with consistent high enrichment for the conserved annotation (7.51x, s.e. = 0.49 and 8.11x, s.e. = 0.54x, respectively; Figure 1c and Table S4). These results imply that adding annotations constructed from LDAK model weights did not significantly change S-LDSC estimates.

We next compared LDAK and S-LDSC enrichments (using N = 20K for LDAK due to computational limitations; see Supplementary Note), analogous to the comparison in Speed et al.[2]. Although Speed et al. reported a very large discrepancy for S-LDSC vs. LDAK (nMSE = 1.32 across 28 annotations; Figure S2), our analysis produced a smaller discrepancy (nMSE = 0.23; Figure 1d), including 10 annotations with significant LDAK enrichments larger than 2.51x (largest significant enrichment: 3.96x). We hypothesize that our discrepancy was smaller than reported by Speed et al. because (unlike Speed et al.) we compared these two methods on the same set of traits. Despite the smaller discrepancy, LDAK estimates were consistently lower than S-LDSC estimates (regression slope = 0.71x), with 14 annotations having a nominally significant difference in enrichment ($P < 0.05$), including a large difference for conserved variants (2.62x, s.e. = 0.17x for LDAK; $P = 2.2 \times 10^{-22}$ for difference vs. S-LDSC); these results are consistent with the LDAK downward bias observed in simulations under the S-LDSC model (Figure 1a). Likewise, LDAK estimates were consistently lower than S-LDSC+LDAK estimates (see Supplementary Note and Figure S3). We note that the LDAK

method uses a different enrichment estimand than S-LDSC (and analyses lower sample sizes), but these differences had little impact on our results (see Supplementary Note and Figures S4-S7).

Misspecification of heritability models can bias functional enrichment estimates. The S-LDSC approach of adding degrees of freedom to the model reduces the potential for model misspecification; in particular, the baseline-LD model[3] infers the extent of LD-dependent architectures directly from the data. On the other hand, LDAK relies on the very strong assumption that marginal association statistics (which include the effects of linked SNPs) are independent of a SNP's linkage disequilibrium (LD) with other SNPs (see Supplementary Note), without providing a biological justification for this assumption; in particular, this assumption implies that recent population bottlenecks and founder events (which increase LD) should greatly reduce causal effect sizes, but we are unaware of any reason why this should be the case. Here, we developed and evaluated a method (S-LDSC+LDAK) that flexibly incorporates both S-LDSC and LDAK-based annotations, serving as a "gold standard" for comparison. In analyses of 16 UK Biobank traits, the results of S-LDSC+LDAK closely match the results of S-LDSC; we recommend using S-LDSC in preference to S-LDSC+LDAK in most settings, due to the complexities of computing LDAK model weights and running S-LDSC+LDAK (see Supplementary Note). In conclusion, our results advocate for using S-LDSC with the baseline-LD model[3] in preference to LDAK to infer functional enrichment (as LDAK enrichment results do not match the S-LDSC+LDAK gold standard) and confirm the existence of functional annotations that are highly enriched (>>2.51x) for complex trait heritability, providing strong caveats to the LDAK results reported by Speed et al.[2].

Code and data availability. S-LDSC is available at https://github.com/bulik/ldsc. Baseline-LD model annotations are available at https://data.broadinstitute.org/alkesgroup/LDSCORE/. LDAK version 5 is available at http://dougspeed.com/downloads/. UK Biobank association statistics, computed using BOLT-LMM v2.3 (ref. [6]), are available at http://data.broadinstitute.org/alkesgroup/UKBB/.

Acknowledgements. We are grateful to Po-Ru Loh for assistance with UK Biobank data and Doug Speed and David Balding for helpful discussions. This research was conducted using the UK Biobank Resource under Application 16549 and was supported by NIH grants U01 HG009379, R01 MH101244 and R01 MH107649.

References

1. Finucane, H. K. *et al*. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet*. **47,** 1228–1235 (2015).
2. Speed, D. *et al*. Reevaluation of SNP heritability in complex human traits. *Nat. Genet*. **49,** 986–992 (2017).
3. Gazal, S. *et al*. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet*. **49,** 1421–1427 (2017).
4. 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).
5. Bycroft, C. *et al*. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 166298 (2017). doi:10.1101/166298
6. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed model association for biobank-scale data sets. *bioRxiv* 194944 (2017). doi:10.1101/194944

Figures

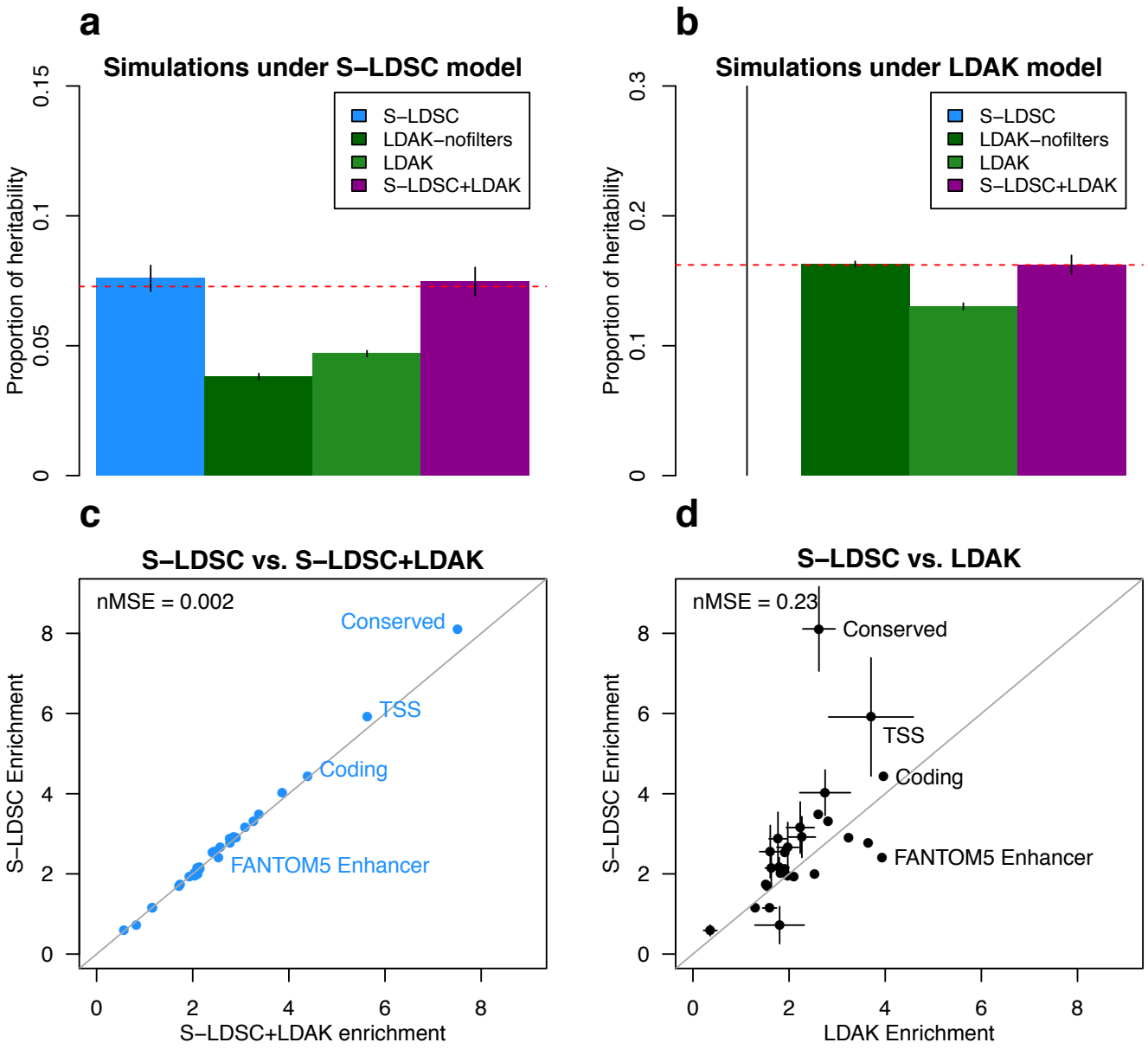

Figure 1: Comparison of S-LDSC and LDAK functional enrichment estimates. (**a**,**b**) We report the estimated proportion of heritability explained by coding variants in simulations in which we simulated effect sizes based on coding enrichment under S-LDSC with the baseline-LD model (**a**) or under the LDAK model (**b**). We report the proportion of heritability explained rather than enrichment due to the different S-LDSC and LDAK enrichment estimands (see Supplementary Note). Dashed red lines indicate the true simulated value. Results are averaged across 500 simulations. Error bars represent 95% confidence intervals. See Figure S1 for other simulation scenarios and Table S1 for numerical results. (**c**,**d**) We report the enrichment of S-LDSC vs. S-LDSC+LDAK (**c**) and S-LDSC vs. LDAK (**d**) for 28 functional annotations, meta-analyzed across 16 independent UK Biobank traits. In each case we report the normalized mean square error (nMSE; see Supplementary Note). Grey lines represent $y = x$. Error bars represent 95% confidence intervals for annotations

for which the estimated enrichment is significantly different ($P < 0.05$) between the two methods. See Table S4 for numerical results.

Supplementary Note

1. S-LDSC, LDAK, and S-LDSC+LDAK methods

S-LDSC method. Stratified LD score regression (S-LDSC)[1,2] is a method for partitioning heritability causally explained by common variants (minor allele frequency (MAF) ≥5%) across overlapping discrete or continuous annotations using genome-wide association study (GWAS) summary statistics and a linkage disequilibrium (LD) reference panel containing both common and low-frequency variants. The method rests on the idea that if an annotation a is associated to increased heritability, LD to variants with large values of a will increase the χ^2 statistic of a variant more than LD to variants with small values of a. More precisely, S-LDSC models the vector β of per normalized genotype effect sizes as a mean-0 vector whose variance depends on D continuous-valued annotations $a_1, \ldots, a_D$:

$$Var(\beta_j) = \sum_{d=1}^{D} a_d(j)\tau_d \quad (1)$$

where $a_d(j)$ is the value of annotation a_d at variant j, and τ_d represents the per-variant contribution of one unit of the annotation a_d to heritability. We can thus estimate the vector τ using the following relationship with the expected χ^2 statistic of variant j:

$$E[\chi_j^2] = N\sum_{d=1}^{D} \tau_d l(j,d) + Nb + 1 \quad (2)$$

where $l(j,d) = \sum_k a_d(k) r_{jk}^2$ is the LD score of variant j with respect to continuous values $a_d(k)$ of annotation a_d, r_{jk} is the correlation between variant j and k in an LD reference panel, N is the sample size of the GWAS study, and b is a term that measures the contribution of confounding biases[3]. Then, the heritability causally explained by a subset of variants S can be estimated as $h_S^2 = \sum_{j\in S}\sum_d a_{j,d}\tau_d$.

The baseline and baseline-LD models. The original set of D functional annotations used by S-LDSC is called the baseline model[1], and contains D = 53 binary functional annotations including 28 main functional annotations such as coding, conserved, and DHS (see Table S2). More recently, we extended this baseline model to the baseline-LD model[2], which contains D = 75 functional annotations including the 53 annotations from the baseline model, 6 new functional annotations, 10 MAF bins to account for MAF-dependent architectures, and 6 LD related annotations to account for LD-dependent architectures (see Table S2).

Application of the S-LDSC method. Previous applications of S-LDSC[1,2] used all SNPs with minor allele count ≥5 in unrelated whole genome sequenced individuals from the 1000 Genomes project[4] as reference data to compute LD scores, estimated τ values using HapMap 3 regression SNPs as a proxy for well-imputed SNPs[1],

and estimated heritability and enrichment values using the set of all reference SNPs with MAF ≥5%. However, here we used UK10K[5] as reference data, owing to its larger sample size (3,567 unrelated individuals) and closer match to the ancestry of UK Biobank samples. In our main S-LDSC analyses of UK Biobank data, we used all SNPs with minor allele count ≥5 as reference SNPs and estimated heritability and enrichment values using the set of all reference SNPs with MAF ≥5% (Table S5), to match previous applications of S-LDSC[1,2]. In all simulations, and in analyses of UK Biobank data using the LDAK estimand, we used all SNPs with MAF ≥1% as reference data and estimated heritability and enrichment values using the set of all reference SNPs with MAF ≥1% (Table S5), to match LDAK[6]. We also ran constrained-intercept S-LDSC[7] (which constrains the intercept, i.e. the $Nb + 1$ term from equation (1), to equal 1) in the regression step when estimating τ values (S-LDSC-nointercept).

LDAK method. LDAK is a method for estimating heritability and functional enrichment from raw genotype-phenotype data[6], modifying the LDAK method of a previous study, which made an important contribution to the literature by highlighting the potential ramifications of LD-dependent architectures[8]. LDAK models the variance of per normalized genotype effect size β as

$$Var(\beta_j) \propto \left(p_j(1-p_j)\right)^{1+\alpha} w_j \tag{3}$$

where p_j is the allele frequency of SNP j, α is a parameter determining the relationship between MAF and heritability (recommended to be fixed at -0.25), and w_j is the LDAK weight defined by minimizing the L1 or L2 norm of

$$\left(1 - \sum_{k=1}^{M} r_{jk}^2 w_k\right)_j \tag{4}$$

The heritability of a binary annotation d is modeled as

$$Var(\beta_j) \propto \left(1_{j \in d} c + 1\right)\left(p_j(1-p_j)\right)^{1+\alpha} w_j \tag{5}$$

where $1_{j \in d}$ is an indicator function with value 1 if SNP j belongs to annotation d and 0 otherwise, and c a constant estimated by restricted maximum likelihood (REML). The enrichment of annotation d is estimated as the proportion of heritability explained by SNPs in annotation d divided by proportion of heritability expected for these SNPs under the LDAK model.

LDAK assumption of equal marginal effect size variance. Both S-LDSC and LDAK are based on a linear model for phenotype, with the causal effect size of SNP j modeled as random and denoted β_j. As described above, under the LDAK model,

$$Var(\beta_j) \propto \left(p_j(1-p_j)\right)^{1+\alpha} w_j \quad (3)$$

where p_j is the allele frequency of SNP j, α is a parameter determining the relationship between MAF and heritability (recommended to be fixed at -0.25), and w_j is the LDAK weight defined by minimizing the L1 or L2 norm of

$$\left(1 - \sum_{k=1}^{M} r_{jk}^2 w_k\right)_j . \quad (4)$$

If w could be chosen so that (4) was equal to the zero vector, then when α = -1, we would have $Var(\beta_j) \propto w_j$ with $\sum_{j=1}^{M} r_{jk}^2 w_k = 1$ for all j.

The marginal effect size of SNP j – i.e., the effect size estimated by a marginal regression as is performed in standard GWAS – is $\sum_{j=1}^{M} r_{jk} \beta_k$, which, assuming independence of SNP effects, has variance $\sum_{j=1}^{M} r_{jk}^2 Var(\beta_k)$. Thus, $Var(\beta_j) \propto w_j$ with $\sum_{j=1}^{M} r_{jk}^2 w_k = 1$ for all j if and only if the variance of the marginal effect sizes is constant. However, the marginal effect size of a SNP has been shown to have a strong linear dependency on its LD score[3,9]. This strong linear dependency is predicted by the S-LDSC model, in which it is the causal effects, rather than the marginal effects, that have equal variance. The effect of LD, which the LDAK weights are designed to undo, is a well-established empirical phenomenon[3].

We note that we have previously inferred LD-dependent architectures and showed that they are consistent with the action of negative selection[2]. Although the direction of effect that we inferred is the same as the direction of effect in the LDAK model (SNPs with lower levels of LD have larger per-SNP heritability), the LD-dependent architectures that we inferred do not imply that marginal SNP effects are independent of LD. In particular, our model uses LD in Africans (LLD-AFR), avoiding the implication (as in the LDAK model) that recent population bottlenecks and founder events (which increase LD) should greatly reduce causal effect sizes.

Application of the LDAK method. The LDAK method recommends that users restrict their analyses to well-imputed SNPs (INFO score[10] ≥0.99) with MAF ≥1% (ref. [6]) and then restrict their analyses to a thinned set of SNPs by removing SNPs in strong LD to each other (default r^2 threshold: 0.99). To investigate the impact of these SNP filtering steps, we performed simulations assessing the performance of LDAK using all SNPs (LDAK-nofilters), as well as LDAK using all well-imputed SNPs with MAF ≥1% (LDAK-noLDthinning).

S-LDSC and LDAK enrichment estimands. The S-LDSC and LDAK methods use different enrichment estimands, which must be carefully accounted for when comparing the two methods. S-LDSC defines the enrichment of a binary annotation as the proportion of heritability causally explained by reference SNPs with

MAF ≥5% (not including tagging of low-frequency SNPs from the reference data) divided by proportion of SNPs in the annotation. LDAK defines enrichment as the proportion of heritability tagged by a thinned set of well-imputed SNPs with MAF ≥1% divided by proportion of heritability expected for these SNPs under the LDAK model. Overall, both the numerator and denominator of S-LDSC and LDAK enrichment estimands capture different information, complicating the direct comparison of these two methods (see Table S5 for the set of SNPs included in S-LDSC and LDAK enrichment estimands in our simulations and UK Biobank analyses). However, the difference between the two estimands has little impact on our UK Biobank results (see below).

S-LDSC+LDAK method. In order to compare S-LDSC and LDAK methods, we developed S-LDSC+LDAK, a method that extends the model fit by S-LDSC to include two annotations constructed using LDAK weights. More precisely, S-LDSC+LDAK partitions heritability of a binary annotation $d' \in D$ as

$$Var(\beta_j) = \left[\sum_{d=1}^{D} a_d(j)\tau_d\right] + \left[1_{j \in d'} LDAK_j . \tau_{d'} + 1_{j \notin d'} LDAK_j . \tau_{\overline{d'}}\right] \quad (6)$$

where $LDAK_j = \left(p_j(1-p_j)\right)^{1+\alpha} w_j$, with p_j and w_j the allele frequency and LDAK weight of SNP j computed on the reference sample, and $1_{j \in d'}$ (resp. $1_{j \notin d'}$) is an indicator function with value 1 if SNP j belongs (resp. does not belong) to annotation d', and 0 otherwise; $1_{j \in d'} LDAK_j$ and $1_{j \notin d'} LDAK_j$ represent the two new annotations added to the S-LDSC model. We note that our use of this method contradicts the statement of Speed et al.[6] that they *cannot envisage how the* (S-LDSC) *method could be modified to accommodate the LDAK SNP weights*.

Application of the S-LDSC+LDAK method. In order to ensure a thorough comparison of S-LDSC+LDAK, S-LDSC and LDAK on real data (see UK Biobank analyses below), we ran two versions of S-LDSC+LDAK, estimating S-LDSC and LDAK estimands, respectively. When comparing S-LDSC+LDAK with S-LDSC, we used default S-LDSC options for S-LDSC+LDAK. When comparing S-LDSC+LDAK with LDAK, we restricted the S-LDSC+LDAK reference panel to SNPs with MAF ≥1%, partitioned the heritability of SNPs with MAF ≥1%, and estimated enrichment using the LDAK enrichment estimand (see Table S5 for the set of SNPs included in S-LDSC+LDAK enrichment estimands in our simulations and UK Biobank analyses). In order to produce the most informative LDAK annotations we used UK10K[5] (3,567 unrelated individuals) rather than 1000 Genomes[4] (489 unrelated individuals) as the reference data for all S-LDSC analyses of UK Biobank data. To investigate the effect of the baseline-LD annotations[2] compared to the two LDAK-derived annotations, we also ran S-LDSC using only the two LDAK-derived annotations (LDAK-sumstats).

2. Simulations

We performed simulations to investigate whether S-LDSC+LDAK was unbiased under both S-LDSC and LDAK models. To overcome the issue of different S-LDSC and LDAK enrichment estimands (see above), we restricted our simulations to SNPs with MAF ≥1%, and modified S-LDSC to partition the heritability of SNPs with MAF ≥1% (see Table S5). We focused on the proportion of heritability explained by the annotations of interest rather than their enrichment, so that S-LDSC, LDAK and S-LDSC+LDAK estimate the same quantity (i.e. the proportion of heritability causally explained by SNPs with MAF ≥1%).

We performed simulations using chromosome 1 of the UK Biobank interim release data set[11] with imputed variants from 1000 Genomes and UK10K. We restricted our simulations to 10,000 unrelated individuals, due to the computational limitations of the LDAK method. We sampled integer-valued genotypes from UK Biobank imputed dosages, restricting to 578,876 SNPs on chromosome 1 with MAF ≥1%. We set trait heritability to $h^2 = 0.5$ for all simulations.

We first simulated effect sizes using the baseline-LD model (including coding, conserved and DHS enrichment) with per-SNP heritability derived from the τ coefficients estimated across 31 independent traits[2]. SNPs with negative per-SNP heritability were set to 0. Second, we performed simulations under LDAK model by selecting M = 100,000 causal variants and simulating effect sizes using a coding or conserved or DHS enriched architecture. Effect sizes were simulated using α = -0.25 (as recommended by LDAK authors[6]) and LDAK weights were computed (in all SNPs polymorphic in the 10,000 individuals) using the LDAK option `--cut-weights --no-thin YES` (to avoid the default LDAK thinning option). We simulated enrichment either by simulating larger effect sizes in the enriched annotation (and the same proportion of causal variants inside and outside the enriched annotation), or by simulating a larger proportion of causal variants in the enriched annotation (and same effect sizes for causal variants inside and outside the enriched annotation). We also performed null simulations without any enrichment. Third, we repeated all LDAK simulations without using LDAK weights by simulating effect sizes using $Var(\beta_j) \propto \left(p_j(1-p_j)\right)^{1+\alpha}$ (instead of $Var(\beta_j) \propto \left(p_j(1-p_j)\right)^{1+\alpha} w_j$) to investigate the impact of LDAK weights on the different methods. We refer to this generative model as the "alpha model". For each simulation scenario, we performed 500 simulations.

S-LDSC was run on summary statistics generated for HapMap3 SNPs (94,467 chromosome 1 SNPs), and by using the baseline-LD model and UK10K[5] (restricting to SNPs with MAF ≥1%) as the reference data (S-LDSC). To overcome the issue of unstable S-LDSC estimates under the LDAK generative model, we also ran S-LDSC with constrained intercept (S-LDSC-nointercept). We ran 3 versions of LDAK on each simulation. First, we ran LDAK on all 578,876 SNPs (LDAK-nofilters). Second, we ran LDAK on the subset of 295,968

well-imputed SNPs (INFO score[10] ≥ 0.99) in all UK Biobank individuals (LDAK-noLDthinning). Third, we ran LDAK on the subset of 89,246 thinned well-imputed SNPs, as recommended[6] (LDAK). To address the complexity of LDAK and S-LDSC models, we also ran S-LDSC using only two LDAK coding-derived annotations (resp. conserved-derived or DHS-derived), with frequency and weights estimated from UK10K (LDAK-sumstats). Finally, we also ran S-LDSC+LDAK by adding the two LDAK coding-derived annotations (resp. conserved-derived or DHS-derived) to the baseline-LD model.

Results for all simulation scenarios and methods are reported Figure S1 and Table S1, and are consistent across coding vs. conserved vs. DHS annotations. As expected, S-LDSC and LDAK-nofilters were unbiased when we simulated coding enrichment under their corresponding models. S-LDSC produced unstable estimates under the LDAK model, although S-LDSC-nointercept produced stable estimates (with a downward bias) and S-LDSC produced stable estimates when we performed simulations with no LDAK weights (Figure S1c). LDAK and LDAK-noLDthinning were downward biased under simulations with functional enrichment under their own model (as they restricted analyses to well-imputed SNPs), and even more downward biased under the S-LDSC model. In addition, LDAK and LDAK-noLDthinning were upward biased in simulations under their own model with no functional enrichment.

3. UK Biobank analyses

UK Biobank data and choice of independent traits. We analyzed data from the full UK Biobank release[12] consisting of 487,409 samples genotyped on ~800,000 markers and imputed to ~93 million SNPs using the Haplotype Reference Consortium dataset[13] (N = 64,976 haplotypes from WGS data). We selected 16 quantitative traits that are heritable and independent, defined as having a heritability Z score >7 with S-LDSC and pairwise phenotypic correlations below 0.1 (Table S3). We restricted our S-LDSC analyses to the set of 459,327 individuals of European ancestry (based on self-reported white ethnicity; average N = 433,751 for the 16 traits), and to the set of 1,187,057 HapMap 3 SNPs with no filter based on imputation accuracy (as recommended[1,2]). We restricted our LDAK analyses to a set of 20,000 unrelated individuals with UK ancestry[14] (due to the computational limitations of LDAK at larger sample sizes) and with no missing information for the 16 selected traits, and to the set of 4,783,589 SNPs with MAF ≥1% and INFO score ≥0.99 (as recommended[6]).

Application of S-LDSC and S-LDSC+LDAK. For each of the 16 selected traits, we computed mixed model association statistics using BOLT-LMM v2.3[15,16] (see URLs) with genotyping array (UK BiLEVE / UK Biobank), assessment center, sex, age, and age squared as covariates. We also included 20 principal components (included with the UK Biobank data release[12]) to correct for ancestry, as recommended by ref. [16]. We included 672,292 directly genotyped SNPs in the mixed model (all autosomal biallelic SNPs with <10% missing data).

S-LDSC and S-LDSC+LDAK were run using 3,567 whole genome sequenced UK10K[5] samples (ALSPAC and TWINSUK cohorts) as the reference data. S-LDSC and S-LDSC+LDAK (with S-LDSC estimand) restricted their analyses to the 13,326,465 SNPs with minor allele count ≥5 (including 5,353,593 SNPs with MAF ≥5%) (Table S5). LDAK annotations for S-LDSC+LDAK (with S-LDSC estimand) were constructed using UK10K allele frequencies and LDAK weights computed for all 13,326,465 UK10K SNPs with allele count ≥5, using LDAK option `--cut-weights --no-thin YES` to avoid the default LDAK thinning option. S-LDSC+LDAK (with LDAK estimand) restricted its analyses to the 7,659,089 SNPs with MAF ≥1% (Table S5). LDAK annotations for S-LDSC+LDAK (with LDAK estimand) were constructed using UK10K allele frequencies and LDAK weights computed for all 7,659,089 UK10K SNPs with MAF ≥1%, using LDAK option `--cut-weights --no-thin YES`. For each trait, S-LDSC was run once using the baseline-LD model version 1.1 (ref. [17]), while S-LDSC+LDAK was run 28 times in turn for each of the functional annotations of interest. We also performed S-LDSC analyses using the baseline model (instead of the baseline-LD model) and using phase 3 of 1000 Genomes[4] (489 individuals) as the reference data (instead of UK10K) to match methods of Finucane et al.[1] (Figure S2 and Table S4).

Application of LDAK. LDAK was run as recommended by Speed et al.[6]. We started our analyses with 4,783,589 SNPs with MAF ≥ 1% and INFO score ≥ 0.99, and computed LDAK weights for a thinned set of 1,282,302 SNPs. We used α = -0.25 to compute kinship matrixes for each of the 28 functional annotations of interest. We ran restricted maximum likelihood (REML) using genotyping array (UK BiLEVE / UK Biobank), assessment center, sex, age and 20 principal components as fixed effect. We also included SNPs with $P < 1 \times 10^{-20}$ from single-SNP analysis (conditioned on the same covariates) and their correlated SNPs as fixed effects. We limited our LDAK analyses to N = 20K individuals due to the computational limitations of LDAK at larger sample sizes.

Comparison of S-LDSC+LDAK, S-LDSC and LDAK results using the nMSE metric. We partitioned the heritability of 16 UK Biobank traits across 28 main functional annotations from the baseline model using S-LDSC+LDAK, S-LDSC and LDAK. We meta-analyzed enrichments using random-effects meta-analyses implemented in the R package *rmeta*. We performed pairwise comparisons of the S-LDSC+LDAK, S-LDSC and LDAK methods by computing the normalized mean square error (nMSE) across the 28 annotations as

$$nMSE(x,y) = \frac{\sum_i (x_i - y_i)^2}{\sqrt{\sum_i x_i^2} * \sqrt{\sum_i y_i^2}} \quad (7)$$

where x_i (resp. y_i) is the estimated enrichment of annotation i for method x (resp. y). (The normalization is analogous to a correlation, but nMSE can account for highly correlated variables with different magnitudes; we note that nMSE can be larger than 1 when x and y have very different magnitudes or are negatively correlated.)

We tested whether two different enrichment estimates for the same annotation were significantly different by computing a Z score based on the respective estimates and standard errors. We note that it is possible for errors to be correlated, as they are computed using the same data; if this is the case, our test for significantly different enrichment estimates is conservative. We describe our pairwise comparisons in greater detail below:

S-LDSC vs. S-LDSC-nointercept: We determined that S-LDSC with constrained intercept (S-LDSC-nointercept) obtained results similar to S-LDSC (Figure S8 and Table S4). This stands in contrast to the inconsistency between the two methods in simulations under the LDAK model (Figure S1).

S-LDSC+LDAK vs. S-LDSC: We compared S-LDSC enrichments to the "gold-standard" S-LDSC+LDAK estimates (using the S-LDSC estimand; Figure 1c and Table S4). Estimates were very similar (nMSE = 0.002, no annotation significantly different), implying that adding annotations constructed from LDAK model weights did not change S-LDSC estimates. We observed significant standardized effect sizes[2] for the LDAK annotations included within the S-LDSC+LDAK model, but these were much lower than the standardized effect sizes for the same annotations when including only the LDAK annotations via LDAK-sumstats (Table S6). On the other hand, including the LDAK annotations only slightly decreased the standardized effect sizes of the LD-related annotations within the baseline-LD model (Figure S9) and slightly increased the proportion of variance in χ^2 statistics (Table S7), suggesting that S-LDSC (using the baseline-LD model) sufficiently models LD-dependent and MAF-dependent architectures. We recommend S-LDSC in preference to S-LDSC+LDAK in most settings, due to the complexities of computing LDAK model weights and running S-LDSC+LDAK.

S-LDSC and LDAK enrichments reported by ref. [1,2,6]: The original S-LDSC analysis of Finucane et al.[1] using the baseline model reported a largest significant enrichment of 13.32x for the conserved annotation (9 independent diseases and traits; average N = 96K). The S-LDSC analysis of Gazal et al.[2] using the baseline-LD model (which accounts for LD-dependent architectures) reported a largest significant enrichment of 9.35x (versus 14.12x using the baseline model) for the conserved annotation (31 independent diseases and traits; average N = 85K). The LDAK analysis of Speed et al.[6] using the annotations from the baseline model reported a largest significant enrichment of 2.51x for the transcription starting site (TSS) annotation and a largest non-significant enrichment of 2.88x for the 5'UTR annotation (19 diseases; average N = 7K). Surprisingly, they also reported non-significant enrichment of 1.34x for the conserved annotation and non-significant enrichment of 2.13x for the coding annotation (conserved and coding not among annotations with green labels in Figure S16 of Speed et al.[6] denoting significant enrichment), despite a considerable body of evidence in the published literature in favor of coding variant enrichment for common diseases and complex traits[5,18]. Speed et al.[6] thus reported a very large discrepancy for S-LDSC vs. LDAK (nMSE = 1.32 across 28 annotations; see Figure S2a and Table S4a).

S-LDSC vs. LDAK: We next compared S-LDSC and LDAK on the same dataset (UK Biobank), and observed a smaller discrepancy (nMSE = 0.23; Figure 1d) than reported by Speed et al.[6]. In addition, we observed 10 annotations with significant LDAK enrichment ($P < 0.05/28$ using a Wald test as in Speed et al.[6]) and with LDAK enrichment estimates larger than 2.51x, including a significant enrichment of 2.62x for conserved variants and a largest significant enrichment of 3.96x for coding variants (Table S4). We hypothesize that our relative results were very different from Speed et al.[6] because (unlike Speed et al.[6]) we compared these two methods on the same set of traits. We also used the baseline-LD model instead of the baseline model (nMSE = 0.44; Figure S2c), and UK10K as the reference data instead of 1000 Genomes (nMSE = 0.54; Figure S2d), two other contributing factors. Despite the nMSE = 0.23 in our main analysis, we observed that LDAK estimates were consistently lower than S-LDSC estimates (regression slope = 0.71x; consistent with the downward bias observed in simulations under the baseline-LD model, Figure 1a), and that 14 annotations had a nominally significant difference in enrichment ($P < 0.05$), including a large difference for conserved variants (2.62x, s.e. = 0.17x for LDAK vs. 8.11x, s.e. = 0.54x for S-LDSC; $P = 2.2 \times 10^{-22}$ for difference).

S-LDSC+LDAK (S-LDSC estimand) vs. S-LDSC+LDAK (LDAK estimand): To assess whether differences between S-LDSC and LDAK could be due to their different estimands, we compared S-LDSC+LDAK enrichments using either S-LDSC or LDAK estimands. We determined that S-LDSC+LDAK results were nearly identical when using the two different estimands (nMSE = 0.004; Figure S4), suggesting that differences between S-LDSC and LDAK enrichment estimates are not driven by their different estimands.

S-LDSC and LDAK sample size limitations: S-LDSC and LDAK analyzed data sets of different sample sizes, due to LDAK computational constraints (N = 434K for S-LDSC vs. N = 20K for LDAK). However, we observed similar results for LDAK when either (i) running LDAK on N = 10K and N = 15K data sets (nMSE ≤ 0.02; Figure S5), (ii) running LDAK on three random subsamples with N = 20K (nMSE ≤ 0.02; Figure S6), or (iii) running LDAK-sumstats on all UK Biobank individuals (average N = 434K; nMSE = 0.04; Figure S7), suggesting that N = 20K is a sample size large enough for LDAK results to converge. On the other hand, S-LDSC (and thus S-LDSC+LDAK) requires GWAS summary statistics computed on a large number of individuals (e.g. heritability Z score >7)[1]; only 3 of 16 UK Biobank traits have a heritability Z score >7 when running S-LDSC on summary statistics computed using the N = 20K LDAK data set (Table S3).

S-LDSC+LDAK vs. LDAK: For completeness, we also compared LDAK (N = 20K) and S-LDSC+LDAK (using LDAK estimand; N = 434K). We observed that the nMSE between S-LDSC+LDAK and LDAK (nMSE = 0.19; Figure S3) is larger than the nMSE between S-LDSC+LDAK and S-LDSC (nMSE = 0.002; Figure 1c). Although we have shown that the nMSE difference cannot be explained by the application of LDAK to a lower sample size (Figures S5-S7), we also compared S-LDSC and LDAK to S-LDSC+LDAK restricting to N = 20K

samples for the three traits that passed the recommended heritability Z score threshold (Figure S10). We observed that the difference between S-LDSC+LDAK (N = 20K) vs. S-LDSC (N = 20K) (nMSE = 0.02) was lower than the difference between S-LDSC+LDAK (N = 20K) vs. LDAK (N = 20K) (nMSE = 0.42). Overall, these experiments advocate for using S-LDSC in preference to LDAK to infer functional enrichment.

We also observed that LDAK tends to overestimate the denominator of the enrichment estimand (i.e. the proportion of heritability expected under the LDAK model) compared to S-LDSC+LDAK (ratio 1.05; Table S4), as these two methods use different set of SNPs to compute this metric (i.e. the thinned set of SNPs with MAF ≥ 1% and INFO score ≥ 0.99 for LDAK, vs. all reference SNPs with MAF ≥ 1% for S-LDSC+LDAK with LDAK estimand). For example, LDAK (N = 20K) and S-LDSC+LDAK (N = 434K; LDAK estimand) estimated similar proportions of heritability explained by the DHS annotation (0.38 and 0.41, respectively; ratio 0.93x), but different proportions of heritability expected under the LDAK model (0.23 and 0.19, respectively; ratio 1.20x), leading to quite different enrichment values (1.63x and 2.11x, respectively; ratio 0.77x). These differences in the denominator also impacted coding enrichment: LDAK (N = 20K) estimated a *higher* proportion of heritability explained by the coding annotation than S-LDSC+LDAK (N = 434K; LDAK estimand) (0.11 and 0.09, respectively; ratio 1.34x) but a much higher proportion of heritability expected under the LDAK model than S-LDSC+LDAK (0.029 and 0.018, respectively; ratio 1.64x)—largely due to LDAK's thinning of SNPs, with the LDAK estimates of heritability expected under the LDAK model equal to 0.019 and 0.027 before and after the thinning step, respectively—leading to lower enrichment (3.96x and 4.85x, respectively; ratio 0.82x).

References

1. Finucane, H. K. *et al*. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat*. *Genet*. **47,** 1228–1235 (2015).
2. Gazal, S. *et al*. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat*. *Genet*. **49,** 1421–1427 (2017).
3. Bulik-Sullivan, B. K. *et al*. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat*. *Genet*. **47,** 291–295 (2015).
4. Auton, A. *et al*. A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
5. The UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature* **526,** 82–90 (2015).
6. Speed, D. *et al*. Reevaluation of SNP heritability in complex human traits. *Nat*. *Genet*. **49,** 986–992 (2017).
7. Bulik-Sullivan, B. *et al*. An atlas of genetic correlations across human diseases and traits. *Nat*. *Genet*. **47,** 1236–1241 (2015).
8. Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved Heritability Estimation from Genome-wide SNPs. *Am*. *J*. *Hum*. *Genet*. **91,** 1011–1021 (2012).
9. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).
10. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat*. *Rev*. *Genet*. **11,** 499–511 (2010).
11. 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).
12. Bycroft, C. *et al*. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 166298 (2017). doi:10.1101/166298
13. The Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat*. *Genet*. **48,** 1279–1283 (2016).

14. Galinsky, K. J., Loh, P.-R., Mallick, S., Patterson, N. J. & Price, A. L. Population Structure of UK Biobank and Ancient Eurasians Reveals Adaptation at Genes Influencing Blood Pressure. *Am. J. Hum. Genet.* **99,** 1130–1139 (2016).

15. Loh, P.-R. *et al*. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47,** 284–290 (2015).

16. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed model association for biobank-scale data sets. *bioRxiv* 194944 (2017). doi:10.1101/194944

17. Hormozdiari, F. *et al*. Leveraging molecular QTL to understand the genetic architecture of diseases and complex traits. *bioRxiv* 203380 (2017). doi:10.1101/203380

18. Hindorff, L. A. *et al*. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci.* **106,** 9362–9367 (2009).

Supplementary Figures

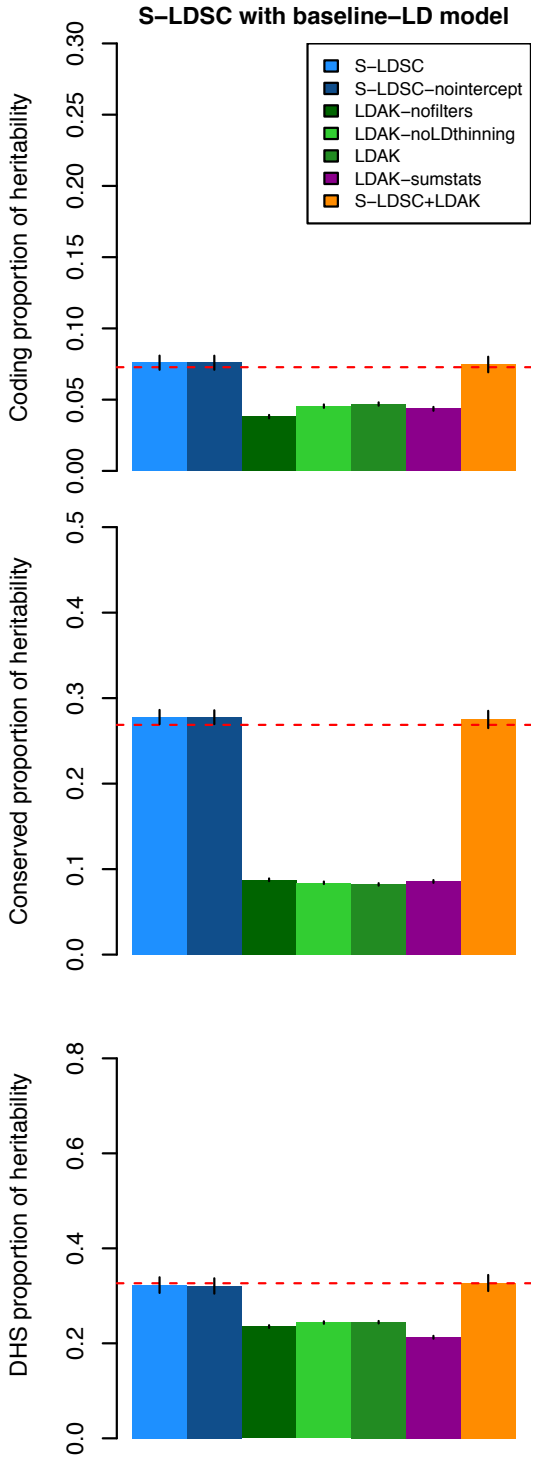

Figure S1a – simulations under S-LDSC model

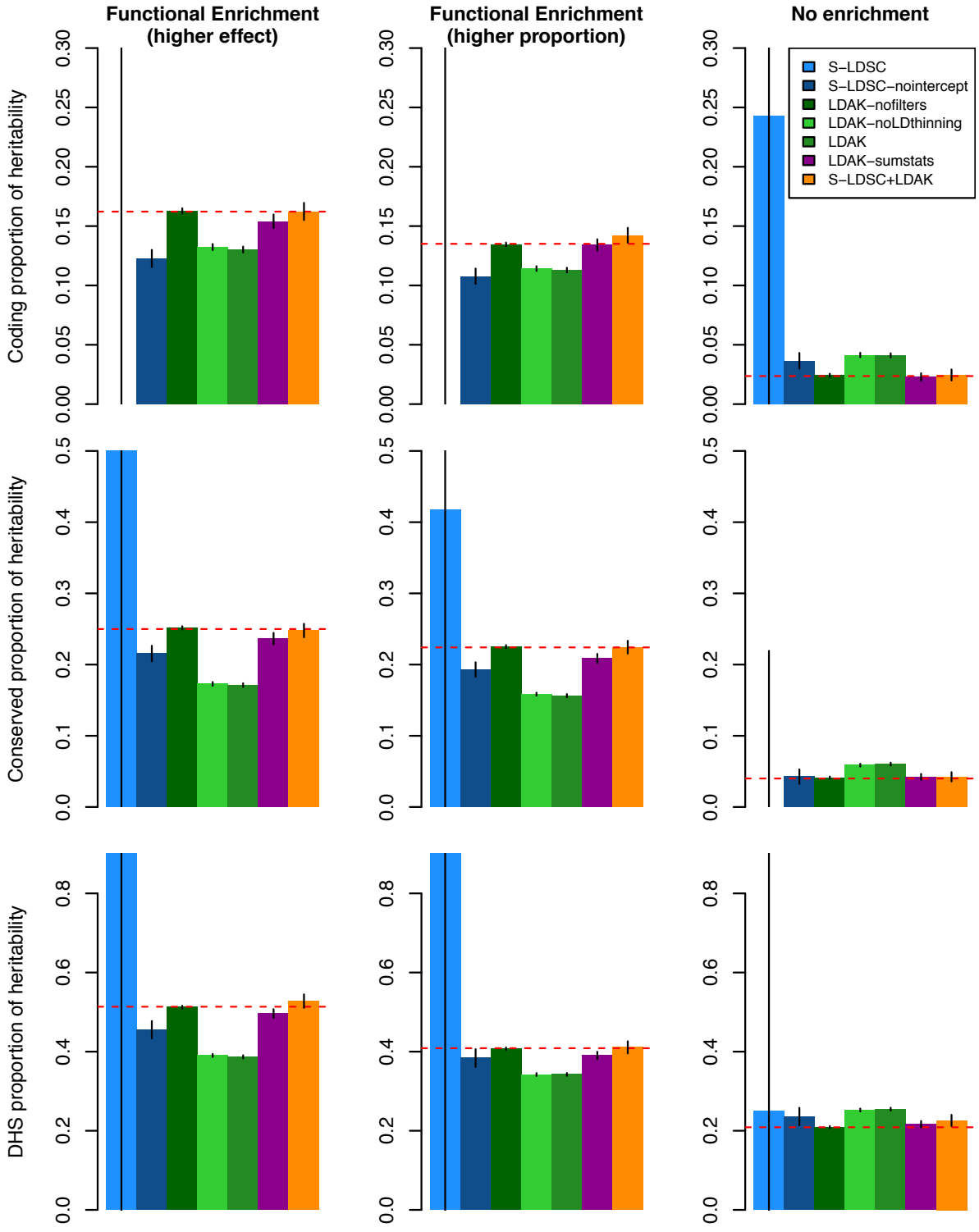

Figure S1b – simulations under LDAK model

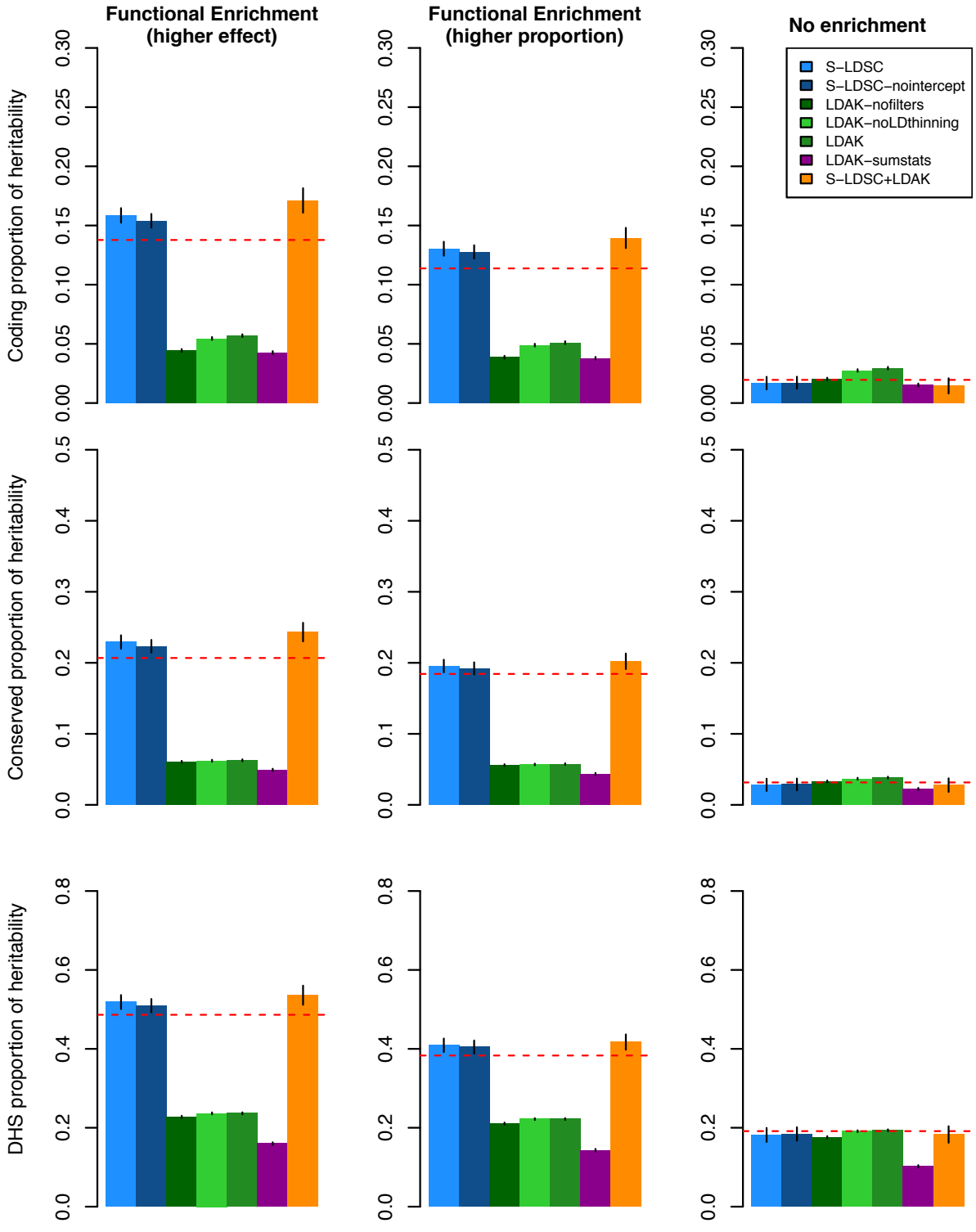

Figure S1c – simulations under alpha model

Figure S1: Simulations to assess LDAK, S-LDSC and S-LDSC+LDAK enrichment accuracy. We report enrichment simulations of 7 different methods, 3 different enriched annotations, and 7 different simulation scenarios involving 3 different generative models. The 7 different methods are S-LDSC with the baseline-LD model (S-LDSC), S-LDSC with intercept constrained to 1 (S-LDSC-nointercept), LDAK using all SNPs (LDAK-nofilters), LDAK restricted to all SNPs with INFO score ≥0.99 but with no LD thinning (LDAK-noLDthinning), LDAK with default SNP filtering, i.e. restricted to SNPs with INFO score ≥0.99 followed by LD thinning (LDAK), LDSC with two annotations based on LDAK weights (LDAK-sumstats), and S-LDSC with the baseline-LD model and two annotations based on LDAK weights (S-LDSC+LDAK). The 3 different enriched annotations are coding (row 1), conserved (row 2), and DHS (row 3). The 3 different simulation scenarios are functional enrichment with same proportion of causal SNPs in the enriched annotation but higher effect size variance (column 1), functional enrichment with higher proportion of causal SNPs in the enriched annotation but same effect size variance (column 2), and no enrichment (column 3). The 3 different generative models are S-LDSC (using the baseline-LD model with previously estimated parameters[2]) (**a**), LDAK (**b**), and the alpha model (**c**). Panel (**a**) shows that in simulations under the S-LDSC model, 1) the S-LDSC, S-LDSC-nointercept and S-LDSC+LDAK methods are unbiased, and 2) other LDAK-related methods are downward biased. Panel (**b**) shows that in simulations under the LDAK model, 1) the LDAK-nofilters and S-LDSC+LDAK methods are unbiased, 2) LDAK is downward biased when simulating functional enrichment and upward biased under no enrichment, and 3) S-LDSC produces unstable estimates while S-LDSC-nointercept does not. Panel (**c**) shows that in simulations under the alpha model, 1) S-LDSC produces stable estimates, and 2) S-LDSC is unbiased in null simulations. Dashed red lines indicate true values. Error bars represent the 95% confidence intervals based on 500 simulations. See Table S1 for numerical results.

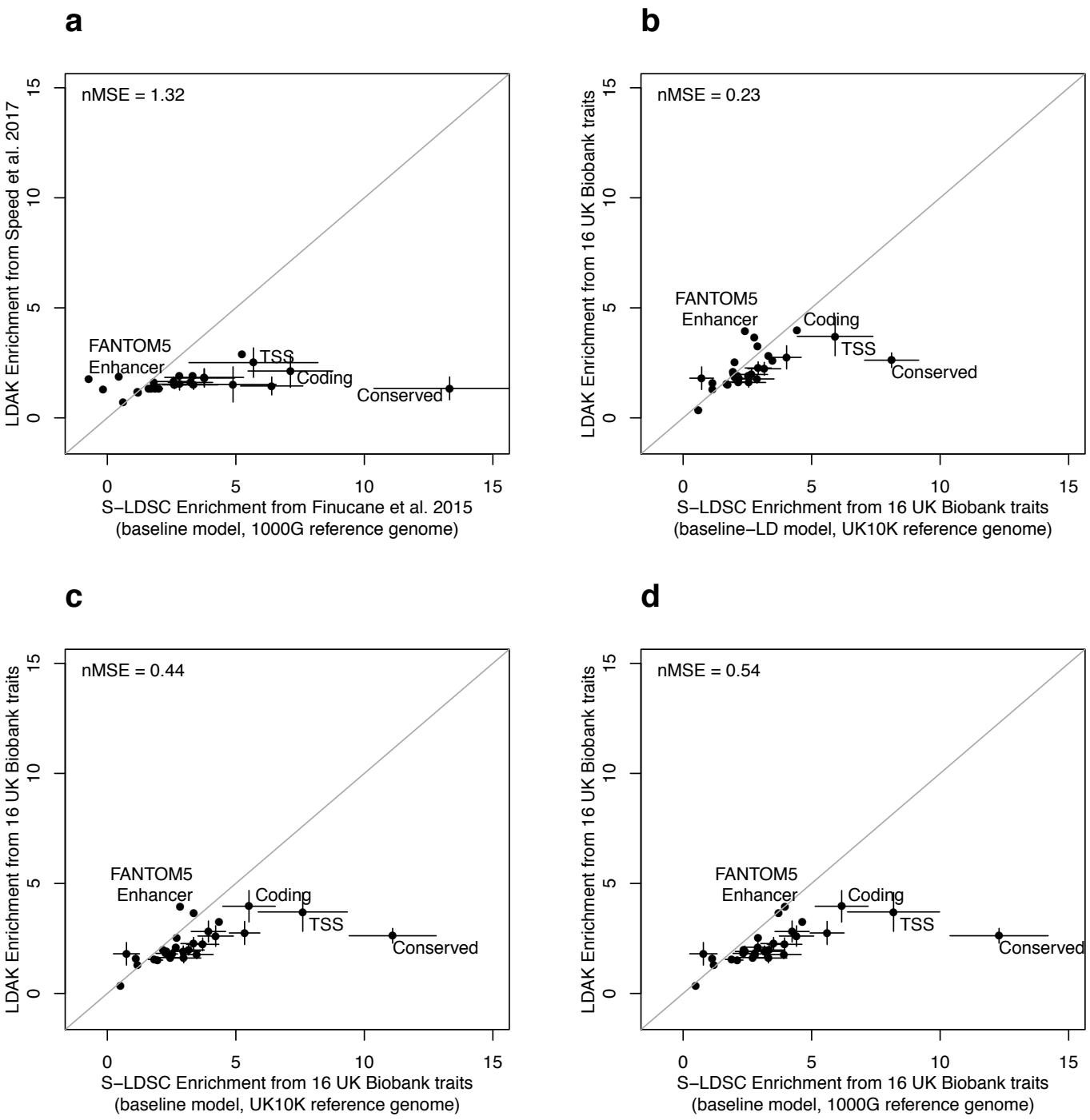

Figure S2: Comparison of S-LDSC and LDAK functional enrichment. (**a**) LDAK functional enrichment reported in Speed et al.[6] vs. S-LDSC functional enrichment reported in Finucane et al.[1] (**b, c, d**) LDAK functional enrichment estimated across 16 independent UK Biobank traits (average N = 20K) vs. S-LDSC functional enrichment estimated across 16 independent UK Biobank traits for various choices of baseline model and reference data set (average N = 434K). In our main analyses, we used the baseline-LD model and UK10K as the reference data set (**b**), and observed a much lower difference between S-LDSC and LDAK than previously reported (nMSE = 0.23, vs. nMSE = 1.32 in (**a**)). Only part of this large difference is due to using the baseline-LD model (nMSE = 0.44 using baseline model, (**c**)) or to using UK10K as the reference data set (nMSE = 0.54 using 1000 Genomes, (**d**)). In each case we report the normalized mean square error (nMSE).

Grey lines represent $y = x$. Error bars represent 95% confidence intervals for annotations for which the estimated enrichment is significantly different between the two methods. See Table S4 for numerical results.

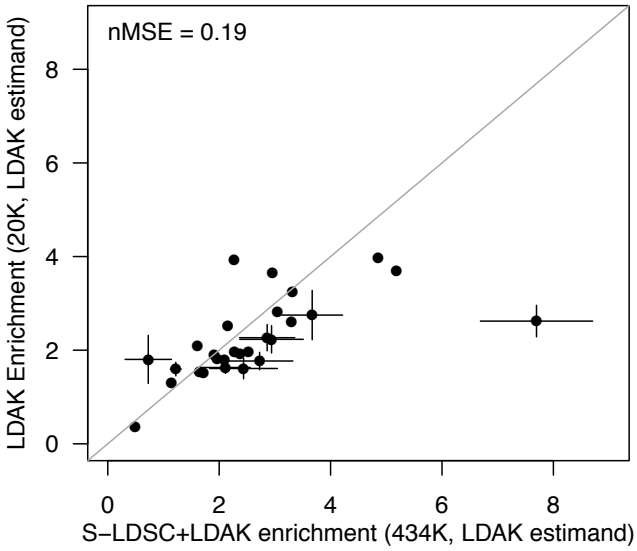

Figure S3: LDAK (N = 20K) vs. S-LDSC+LDAK (N = 434K). We report the enrichment of LDAK (N = 20K) vs. S-LDSC+LDAK (N = 434K; LDAK estimand) for 28 functional annotations, meta-analyzed across 16 independent UK Biobank traits. In each case we report the normalized mean square error (nMSE). The grey line represents $y = x$. Error bars represent 95% confidence intervals for annotations for which the estimated enrichment is significantly different ($P < 0.05$) between the two methods.

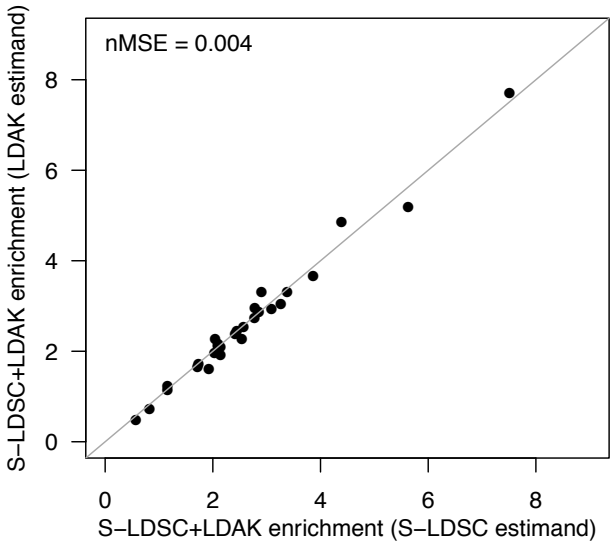

Figure S4: Comparison of S-LDSC+LDAK using different estimands. We report the enrichment of S-LDSC+LDAK using S-LDSC and LDAK estimands for 28 functional annotations, meta-analyzed across 16 independent UK Biobank traits (average N = 434K). We report the normalized mean square error (nMSE). The grey line represents $y = x$.

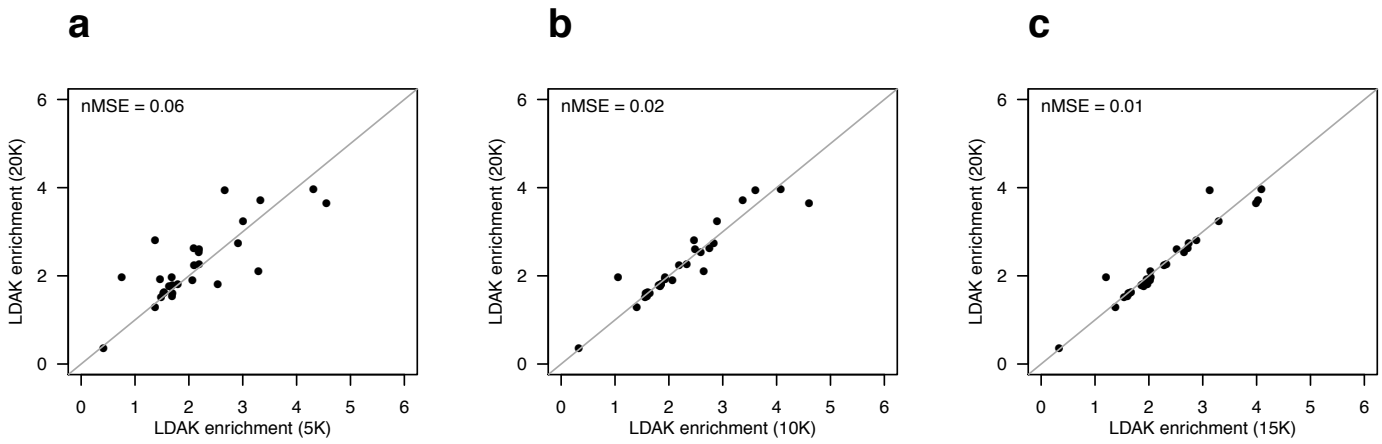

Figure S5: Comparison of LDAK using different sample sizes. We report the enrichment of LDAK using N=20K individuals (as in our main analyses) vs. LDAK using N = 5K individuals (**a**), N = 10K individuals (**b**) and N = 15K individuals (**c**) for 28 functional annotations, meta-analyzed across 16 independent UK Biobank traits. In each case we report the normalized mean square error (nMSE). Grey lines represent $y = x$.

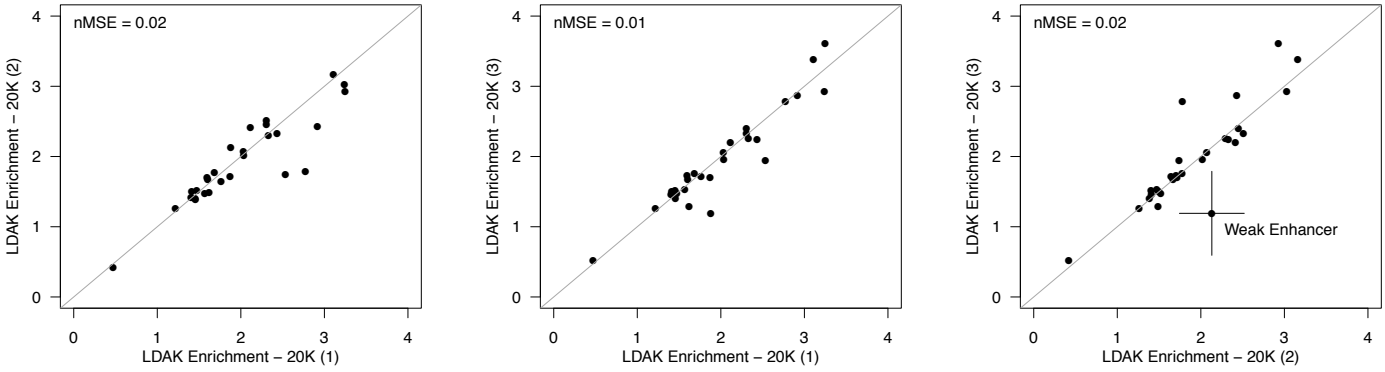

Figure S6: Comparison of LDAK using 3 random data sets with N = 20K individuals. We compared the enrichment of LDAK using 3 random subsets of UK Biobank with N = 20K individuals. In each case we report the normalized mean square error (nMSE). Grey lines represent $y = x$. Error bars represent 95% confidence intervals for annotations for which the estimated enrichment is statistically different between the two analyses.

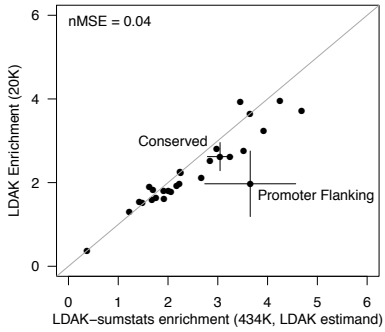

Figure S7: Comparison of LDAK-sumstats using N = 434K and LDAK using N = 20K. We report enrichment estimates for LDAK-sumstats (LDAK estimand) using N = 434K samples and LDAK using N = 20K samples for 28 functional annotations, meta-analyzed across 16 independent UK Biobank traits. We report

the normalized mean square error (nMSE). Grey lines represent $y = x$. Error bars represent 95% confidence intervals for annotations for which the estimated enrichment is significantly different between the two methods.

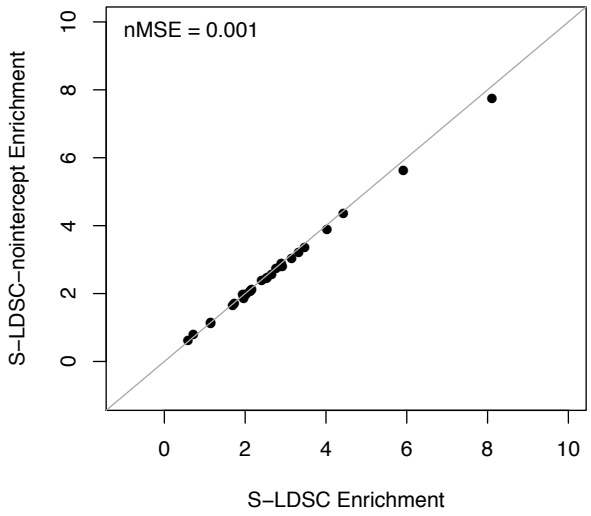

Figure S8: Comparison of S-LDSC and S-LDSC with constrained intercept. We report the enrichment of S-LDSC vs. S-LDSC with constrained intercept (S-LDSC-nointercept) for 28 functional annotations, meta-analyzed across 16 independent UK Biobank traits. We report the normalized mean square error (nMSE). Grey lines represent $y = x$.

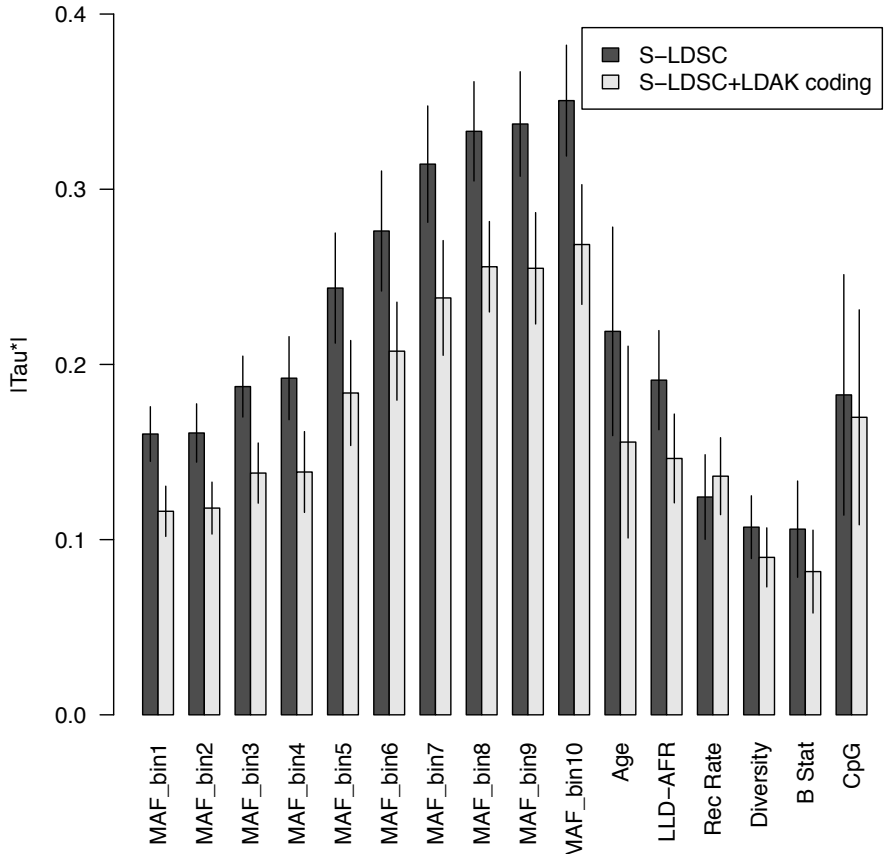

Figure S9: Standardized effect sizes of MAF- and LD-related annotations in S-LDSC and S-LDSC+LDAK. We report the standardized effect sizes[2] (τ^*) of MAF- and LD-related annotations for S-LDSC using the baseline-LD model and S-LDSC+LDAK (S-LDSC estimand) using the baseline-LD model and two coding annotations constructed from LDAK model weights. Error bars represent 95% confidence intervals.

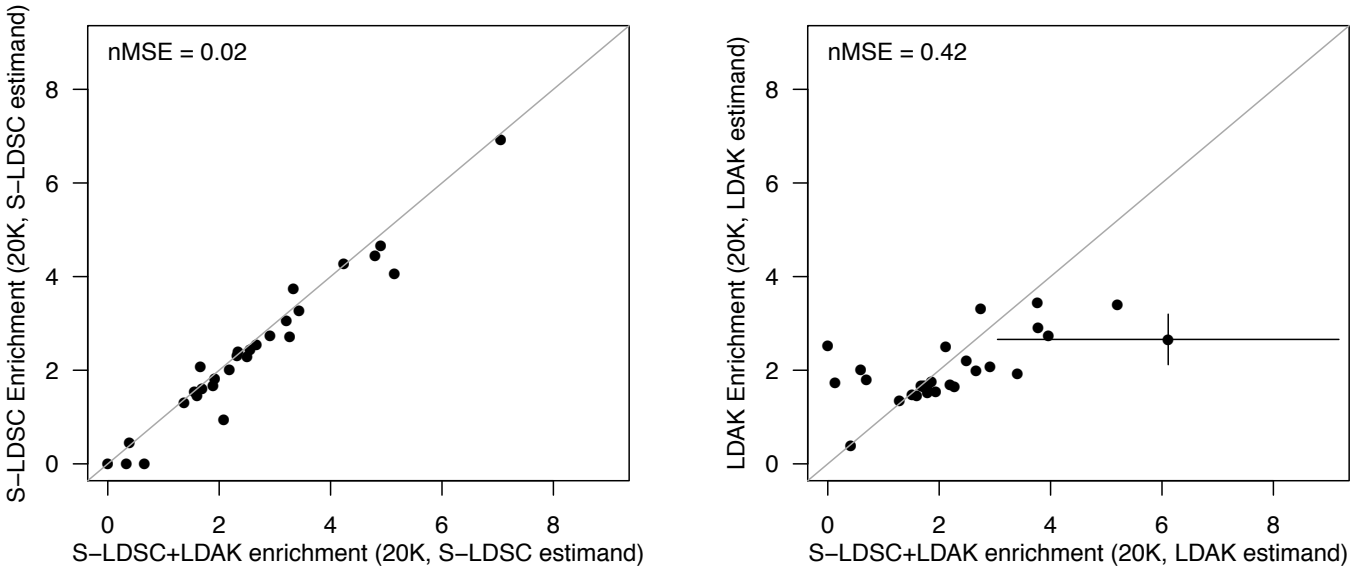

Figure S10: S-LDSC (N = 20K) and LDAK (N = 20K) vs. S-LDSC+LDAK (N = 20K) for 3 highly heritable UK Biobank traits. We report the enrichment of S-LDSC+LDAK (N = 20K; S-LDSC estimand) vs. S-LDSC (N = 20K; S-LDSC estimand) (left panel) and S-LDSC+LDAK (N = 20K; LDAK estimand) vs. LDAK (N = 20K; LDAK estimand) (right panel) for 28 functional annotations, meta-analyzed across 3 independent highly heritable UK Biobank traits. In each case we report the normalized mean square error (nMSE). Grey lines represent $y = x$. Error bars represent the 95% confidence intervals for annotations for which the estimated enrichment is significantly different ($P < 0.05$) between the two methods. Negative enrichment estimates have been set to 0 for visualization purposes but not for nMSE computations.

Supplementary Tables

See Excel file.

Table S1: Simulation results. We report numerical results from **Figure S1**.

See Excel file.

Table S2: Functional annotations of the baseline-LD model. We list the 75 annotations of the baseline-LD model, with accompanying information. The "Main annotation" column indicates whether the annotation is included as one of the 28 main functional annotations.

See Excel file.

Table S3: UK Biobank independent traits. We list the 16 UK Biobank traits analyzed, with accompanying information.

See Excel file.

Table S4: Published and UK Biobank functional enrichment estimates using LDAK, S-LDSC and S-LDSC+LDAK. (**a**) We report enrichment estimates of the 28 main functional annotations published in Finucane et al.[1] (S-LDSC with the baseline model), Speed et al.[6] (LDAK) and Gazal et al.[2] (S-LDSC with the baseline-LD model). (**b**) We report enrichment estimates of the 28 main functional annotations, estimated across 16 independent UK Biobank traits. For each annotation, we report the proportion of SNPs of the annotation; for LDAK we report the proportion of SNPs before thinning, and include in parentheses the proportion of SNPs after thinning. For methods using the LDAK estimand, we report the proportion of heritability expected under the LDAK model.

See Excel file.

Table S5: Reference SNPs, Regression SNPs and Heritability SNPs used by S-LDSC, LDAK and S-LDSC+LDAK. For S-LDSC and S-LDSC+LDAK we list the Reference SNPs (used to compute LD scores), Regression SNPs (included in the regression) and Heritability SNPs (used to compute proportion of heritability and enrichment). For LDAK, we list the Heritability SNPs (used to run LDAK and compute proportion of heritability and enrichment), as References SNPs and Regression SNPs are not applicable. We provide this information for both Simulations (column 1) and Analyses of 16 UK Biobank traits (column 2). In the simulations, we restricted to causal SNPs with MAF $\geq$ 1%, and adapted the S-LDSC and S-LDSC+LDAK Reference SNPs and Heritability SNPs to match LDAK (i.e. SNPs with MAF $\geq$ 1%). In the analyses of 16 UK

Biobank traits, we ran S-LDSC and LDAK using the recommended settings[6], and ran two versions of S-LDSC+LDAK using S-LDSC and LDAK estimands respectively.
See Excel file.

Table S6: Standardized effect sizes of LDAK annotations estimated by S-LDSC+LDAK and LDAK-sumstats. We report the standardized sizes[2] (τ^*), as well as corresponding jackknife standard errors and *P* values[2], for the LDAK annotations in S-LDSC+LDAK (S-LDSC estimand) and LDAK-sumstats (S-LDSC estimand), respectively. Each row of the table represents analyses that include LDAK annotations corresponding to that functional annotation only. *P* values are computed analytically based on (τ^*)/s.e.(τ^*), and do not represent a formal likelihood ratio test quantifying the improvement in model fit.

See Excel file.

Table S7: Proportion of variance in observed χ^2 statistics explained by S-LDSC and S-LDSC+LDAK models. For each model, we report the average across 16 independent UK Biobank traits of the squared correlation between the observed χ^2 statistics and the expected χ^2 statistics (see equation (2)), computed using HapMap 3 SNPs.