Published in final edited form as:

Nat Genet. 2019 August ; 51(8): 1202–1204. doi:10.1038/s41588-019-0464-1.

Reconciling S-LDSC and LDK functional enrichment estimates

Steven Gazal^{1,2,*}, Carla Marquez-Luna^{2,3}, Hilary K. Finucane^{2,*,**}, Alkes L. Price^{1,2,3,*,**}

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

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

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

Recent work has highlighted the importance of accounting for linkage disequilibrium (LD)-dependent genetic architectures in analyses of heritability^{1–5}. Two models incorporating LD-dependent architectures have been proposed for analyses of functional enrichment: the baseline-LD model⁴ used by stratified LD score regression^{4,6} (S-LDSC), and the LDK model^{1,3}. Although both models include LD-dependent effects, they produce very different estimates of functional enrichment (e.g. $9.35x \pm 0.80$ in ref.⁴ and $1.34x \pm 0.26$ in ref.³ for conserved regions), leading to different interpretations of the functional architecture of complex traits. We performed a comprehensive set of formal model comparisons and empirical analyses to reconcile these findings. Each of these analyses supports the higher functional enrichment estimates of S-LDSC with the baseline-LD model; each paragraph below is detailed in a corresponding section of the Supplementary Note (also see Supplementary Tables 1–10 and Supplementary Figures 1–23 for detailed analyses).

We defined six heritability models, including the infinitesimal model that ref.³ called the “GCTA model”⁷, the baseline-LD model⁴ combining functional annotations⁶ with LD-dependent and minor allele frequency (MAF)-dependent architectures, and the LDK model³ combining LD-dependent and MAF-dependent architectures; notably, the baseline-LD and LDK models employ very different LD-dependent architectures. For comparison purposes, we also defined the “ α -model”¹ comprising only the MAF-dependent part of the LDK model, the “Gazal-LD model” comprising only the LD-dependent and MAF-dependent parts of the baseline-LD model (analogous to the LDK model), and the

*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).

**These authors co-directed this study.

Authors contribution. S.G., H.K.F. and A.L.P. designed experiments. S.G. performed experiments. S.G., C.M.L. and H.K.F. analyzed data. S.G., H.K.F. and A.L.P. wrote the manuscript with assistance from C.M.L..

Competing Financial Interests Statement. The authors declare no conflict of interest.

Data availability. Baseline-LD model version 1.1, https://data.broadinstitute.org/alkesgroup/LDSCORE/1000G_Phase3_baselineLD_v1.1_ldscores.tgz. UK Biobank association statistics, computed using BOLT-LMM v2.3, are available at <http://data.broadinstitute.org/alkesgroup/UKBB/>.

Code availability. LDSC software, <https://github.com/bulik/ldsc>. LDK version 5, <http://dougsspeed.com/downloads/>.

“baseline-LD+LDAK” model that flexibly combines baseline-LD and LDAK model annotations.

We performed formal model comparisons using the likelihood approach of ref.³, analyzing 16 UK Biobank traits ($N=20K$ unrelated British-ancestry samples). First, we analyzed genotypes imputed using 1000 Genomes (1000G), as in ref.³ (Figure 1a). We confirmed that the LDAK model attained higher likelihoods than the GCTA model and α -model, consistent with ref.³, but observed that it attained lower likelihoods than the Gazal-LD, baseline-LD and baseline-LD+LDAK models. Second, we analyzed genotypes imputed using the Haplotype Reference Consortium (HRC), a more complete SNP set that produced higher likelihoods for each model analyzed (Figure 1b). The LDAK model attained slightly lower likelihoods than the GCTA model and α -model, and much lower likelihoods than the Gazal-LD, baseline-LD and baseline-LD+LDAK models; the latter two models attained similar likelihoods. We obtained similar conclusions using out-of-sample polygenic prediction, as proposed in ref.⁵

We defined three methods for estimating functional enrichment: the S-LDSC method⁴, which uses the baseline-LD model⁴; the LDAK method³, which uses the LDAK model³; and the S-LDSC+LDAK method, which uses the baseline-LD+LDAK model. We emphasize the distinction between heritability models and functional enrichment methods.

We compared the three methods using extensive simulations. The S-LDSC method was unbiased in simulations under the baseline-LD model, and produced unstable estimates under the LDAK model (unlike in real data). The LDAK method was downward biased under the LDAK model (as it restricts analyses to well-imputed SNPs), and even more downward biased under the baseline-LD model. The S-LDSC+LDAK method produced robust enrichment estimates under both models (and the baseline-LD+LDAK model), validating it as a “gold-standard” method.

We compared the gold-standard S-LDSC+LDAK method to the S-LDSC and LDAK methods across 16 UK Biobank traits, analyzing 28 functional annotations. S-LDSC and S-LDSC+LDAK produced nearly identical estimates of enrichment (Figure 2a; 8.11 ± 0.54 and 7.48 ± 0.49 for conserved regions); thus, adding LDAK model annotations did not significantly change S-LDSC enrichment estimates. On the other hand, LDAK enrichment estimates were systematically lower than S-LDSC+LDAK enrichment estimates (Figure 2b) —though higher than LDAK enrichment estimates in ref.³

The LDAK model assigns zero weights to most SNPs (85%, to “thin” SNPs in high LD), hard-coding zero heritability for these SNPs; this may have a substantial impact on functional enrichment, in which an out-of-annotation SNP in perfect LD with a zero-weight in-annotation SNP cannot act as a proxy. We investigated the impact of the proportion of SNPs with zero weights on LDAK enrichment estimates. We ran the LDAK software using a non-default flag that models SNPs in perfect LD differently by assigning non-zero weights to all SNPs (LDAK-nonzeroweights). With this flag, LDAK enrichment estimates increased considerably (Figure 2b), suggesting that assigning zero heritability to a large majority of SNPs may lead to downward bias in LDAK functional enrichment estimates.

Recently, the authors of LDAK introduced the SumHer method, which extends the LDAK model to estimate functional enrichment from summary statistics⁵; this method also produced low functional enrichment estimates⁵ (e.g. $1.95x \pm 0.07$ for conserved regions). We investigated the impact of the proportion of SNPs with zero weights on SumHer enrichment estimates. We ran the SumHer method (LDAK software) using the same non-default flag as above (SumHer-nonzeroweights). Again, SumHer enrichment estimates increased considerably (Figure 2c), suggesting that assigning zero heritability to a large majority of SNPs may lead to downward bias in SumHer functional enrichment estimates. We also determined that a model similar to the SumHer model (but more amenable to our likelihood analyses) attained lower likelihoods than the baseline-LD model in formal model comparisons.

In summary, the baseline-LD model attained higher likelihoods than the LDAK model, the S-LDSC method produced functional enrichment estimates that were nearly identical to the gold-standard S-LDSC+LDAK method (which was unbiased in simulations under both baseline-LD and LDAK models) in empirical analyses of 16 UK Biobank traits, and the lower enrichment estimates for LDAK (and SumHer) could potentially be explained by the assignment of zero weights to most SNPs. S-LDSC enrichment estimates are further corroborated by published results on functional enrichment of low-frequency variants⁸ (which are less impacted by LD) and functional enrichment of fine-mapped SNPs (see Figure 2 of ref.⁹ and Figure 3 of ref.¹⁰). We recommend using the S-LDSC method in preference to the S-LDSC+LDAK method in most settings, due to the complexities of computing LDAK model weights and running S-LDSC+LDAK. We note that our original LD score regression (LDSC) method used the GCTA model to assess confounding in GWAS data, estimate genetic correlations between traits, and estimate the heritability causally explained by all common SNPs. We anticipate that S-LDSC with the baseline-LD model⁴ will improve upon results of LDSC with the GCTA model for these applications, consistent with recent observations⁵.

Accurate estimation of components of heritability relies on accurate modeling of genetic architectures, and we anticipate that new models and corresponding methods will continue to improve our current knowledge. However, our results strongly support S-LDSC with the baseline-LD model as the current state-of-the-art for functional enrichment analyses. Additional points are discussed in the Supplementary Note.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to Po-Ru Loh for assistance with UK Biobank data and to Luke O'Connor, 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, R01 MH107649. S.G. is funded by NIH K99 HG010160-01. H.K.F. is funded by Eric and Wendy Schmidt and NIH DP5-OD024582. Computational analyses were performed on the Orchestra High-Performance Compute Cluster at Harvard Medical School.

References

1. Speed D, Hemani G, Johnson MR & Balding DJ Improved Heritability Estimation from Genome-wide SNPs. *Am. J. Hum. Genet* 91, 1011–1021 (2012). [PubMed: 23217325]
2. Yang J et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet* 47, 1114–1120 (2015). [PubMed: 26323059]
3. Speed D et al. Reevaluation of SNP heritability in complex human traits. *Nat. Genet* 49, 986–992 (2017). [PubMed: 28530675]
4. Gazal S et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet* 49, 1421–1427 (2017). [PubMed: 28892061]
5. Speed D & Balding DJ SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat. Genet* 51, 277–284 (2019). [PubMed: 30510236]
6. Finucane HK et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet* 47, 1228–1235 (2015). [PubMed: 26414678]
7. Yang J et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet* 42, 565–569 (2010). [PubMed: 20562875]
8. Gazal S et al. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat. Genet* 50, 1600–1607 (2018). [PubMed: 30297966]
9. Farh KK-H et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518, 337–343 (2015). [PubMed: 25363779]
10. Huang H et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* 547, 173–178 (2017). [PubMed: 28658209]

Figure 1: Likelihood comparison of different models of per-SNP heritability.

We report the change in log likelihood compared to the LDAK model (ΔLL) of five other per-SNP heritability models, summed across 16 independent UK Biobank traits ($N = 20K$). All six models include one heritability parameter that is maximized in-sample when estimating the likelihood; any other parameters are maximized out-of-sample. **(a)** Analyses using $M = 2,835,699$ well-imputed 1000G SNPs (as in ref. ³). **(b)** Analyses using $M = 4,631,901$ well-imputed HRC SNPs. Numbers between parentheses in figure legends indicate the number of traits with $\Delta LL > 0$. The $M = 4.6$ million well-imputed HRC SNPs consistently attained higher likelihoods than the $M = 2.8$ million well-imputed 1000G SNPs in comparisons using the same model. Further details and numerical results are provided in the Supplementary Note.

Figure 2: Comparison of functional enrichment estimates in analyses of UK Biobank traits.

For 28 functional annotations, we report functional enrichment estimates of S-LDSC with the baseline-LD model ($N = 434K$) (a), the LDAK method ($N = 20K$) (b), and the SumHer method ($N = 434K$) (restricted to the 24 annotations included in the SumHer model) (c) vs. functional enrichment estimates of S-LDSC+LDAK with the corresponding estimand ($N = 434K$); S-LDSC+LDAK (S-LDSC estimand) and S-LDSC+LDAK (LDAK estimand) produced very similar results for these 28 annotations (see Supplementary Note). Results were meta-analyzed across 16 independent UK Biobank traits. For the LDAK method (b) and the SumHer method (c), we also report results for corresponding methods using a non-default flag that models SNPs in perfect LD differently by assigning non-zero heritability to all SNPs (LDAK-nonzeroweights and SumHer-nonzeroweights, respectively). In each case we report the concordance correlation coefficient (ρ_c) with S-LDSC+LDAK. Dashed grey lines represent $y = x$. Error bars represent 95% confidence intervals for annotations for which the estimated enrichment is significantly different ($P < 0.05$; two-sided z test) between the two methods. Further details and numerical results are provided in the Supplementary Note.