Accounting for statistical non-additive interactions enables the recovery of missing heritability from GWAS summary statistics

Samuel Pattillo Smith[1-4,*], Gregory Darnell[1,5,*], Dana Udwin[6], Arbel Harpak[3,4], Sohini Ramachandran[1,2,7,§], and Lorin Crawford[1,6,8,§,†]

1 Center for Computational Molecular Biology, Brown University, Providence, RI, USA

2 Department of Ecology and Evolutionary Biology, Brown University, Providence, RI, USA

3 Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA

4 Department of Population Health, University of Texas at Austin, Austin, TX, USA

5 Institute for Computational and Experimental Research in Mathematics, Brown University, Providence, RI, USA

6 Department of Biostatistics, Brown University, Providence, RI, USA

7 Data Science Initiative, Brown University, Providence, RI, USA

8 Microsoft Research New England, Cambridge, MA, USA

*** Authors Contributed Equally**

§ **Authors Contributed Equally**

† **Corresponding E-mail: lcrawford@microsoft.com**

Abstract

LD score regression (`LDSC`) is a method to estimate narrow-sense heritability from genome-wide association study (GWAS) summary statistics alone, making it a fast and popular approach. The key concept underlying the `LDSC` framework is that there is a positive linear relationship between the magnitude of GWAS allelic effect estimates and linkage disequilibrium (LD) when complex traits are generated under the infinitesimal model — that is, causal variants are uniformly distributed along the genome and each have the same expected contribution to phenotypic variation. We present interaction-LD score (`i-LDSC`) regression: an extension of the original `LDSC` framework that accounts for non-additive genetic effects. By studying a wide range of generative models in simulations, and by re-analyzing 25 well-studied

quantitative phenotypes from 349,468 individuals in the UK Biobank and up to 159,095 individuals in BioBank Japan, we show that the inclusion of a *cis*-interaction score (i.e., interactions between a focal variant and nearby variants) significantly recovers substantial non-additive heritability that is not captured by `LDSC`. For each of the 25 traits analyzed in the UK Biobank and 23 of the 25 traits analyzed in BioBank Japan, `i-LDSC` detects a significant amount of variation contributed by genetic interactions. The `i-LDSC` software and its application to these biobanks represent a step towards resolving further genetic contributions of sources of non-additive genetic effects to complex trait variation.

Introduction

Heritability is defined as the proportion of phenotypic trait variation that can be explained by genetic effects[1–3]. Until recently, studies of heritability in humans have been reliant on typically small sized family studies with known relatedness structures among individuals[4,5]. Due to advances in genomic sequencing and the steady development of statistical tools, it is now possible to obtain reliable heritability estimates from biobank-scale data sets of unrelated individuals[1,3,6,7]. Computational and privacy considerations with genome-wide association studies (GWAS) in these larger cohorts have motivated a recent trend to estimate heritability using summary statistics (i.e., estimated effect sizes and their corresponding standard errors). In the GWAS framework, additive effect sizes and standard errors for individual single nucleotide polymorphisms (SNPs) are estimated by regressing phenotype measurements onto the allele counts of each SNP independently. Through the application of this approach over the last two decades, it has become clear that many traits have a complex and polygenic basis—that is, hundreds to thousands of individual genetic loci across the genome often contribute to the genetic basis of of variation in a single trait[8].

Many statistical methods have been developed to improve the estimation of heritability from GWAS summary statistics[1,3,9,10]. The most widely used of these approaches is linkage disequilibrium (LD) score regression and the corresponding `LDSC` software[1], which corrects for inflation in GWAS summary statistics by modeling the relationship between the variance of SNP-level effect sizes and the sum of correlation coefficients between focal SNPs and their genomic neighbors (i.e., the LD score of each variant). The formulation of the `LDSC` framework relies on the fact that the expected relationship between chi-square test statistics (i.e., the squared magnitude of GWAS allelic effect estimates) and LD scores holds

when complex traits are generated under the infinitesimal (or polygenic) model which assumes: (*i*) all causal variants have the same expected contribution to phenotypic variation and (*ii*) causal variants are uniformly distributed along the genome. Importantly, the estimand of the `LDSC` model is the proportion of phenotypic variance attributable to additive effects of genotyped SNPs. The main motivation behind the `LDSC` model is that, for polygenic traits, many marker SNPs tag nonzero effects. This may simply arise because some of these SNPS are in LD with causal variants[1] or because their statistical association is the product of a confounding factor such as population stratification.

As of late, there have been many efforts to build upon and improve the `LDSC` framework. For example, recent work has shown that it is possible to estimate the proportion of phenotypic variation explained by dominance effects[11] and local ancestry[12] using extensions of the `LDSC` model. One limitation of `LDSC` is that, in practice, it only uses the diagonal elements of the squared LD matrix in its formulation which, while computationally efficient, does not account for information about trait architecture that is captured by the off-diagonal elements. This tradeoff helps `LDSC` to scale genome-wide, but it has also been shown to lead to heritability estimates with large standard error[10,13,14]. Recently, newer approaches have attempted to reformulate the `LDSC` model by using the eigenvalues of the LD matrix to leverage more of the information present in the correlation structure between SNPs[3,10].

In this paper, we show that the LDSC framework can be extended to estimate greater proportions of genetic variance in complex traits (i.e., beyond the variance that is attributable to additive effects) when a subset of causal variants are involved in a gene-by-gene (G×G) interaction. Indeed, recent association mapping studies have shown that G×G interactions can drive heterogeneity of causal variant effect sizes[15]. Importantly, non-additive genetic effects have been proposed as one of the main factors that explains "missing" heritability—the proportion of heritability not explained by the additive effects of variants[16].

The key insight we highlight in this manuscript is that SNP-level GWAS summary statistics can provide evidence of non-additive genetic effects contributing to trait architecture if there is a nonzero correlation between individual-level genotypes and their statistical interactions. We present the "interaction-LD score" regression model or `i-LDSC`: an extension of the `LDSC` framework which recovers "missing" heritability by leveraging this "tagged" relationship between linear and nonlinear genetic effects. To validate the performance of `i-LDSC` in simulation studies, we focus on synthetic trait architectures that have been generated with contributions stemming from second-order and *cis*-acting statistical SNP-by-SNP

interaction effects; however, note that the general concept underlying `i-LDSC` can easily be extended to other sources of non-additive genetic effects (e.g., gene-by-environment interactions). The main difference between `i-LDSC` and `LDSC` is that the `i-LDSC` model includes an additional set of "*cis*-interaction" LD scores in its regression model. These scores measure the amount of phenoytpic variation contributed by genetic interactions that can be explained by additive effects. In practice, these additional scores are efficient to compute and require nothing more than access to a representative pairwise LD map, same as the input required for LD score regression.

Through extensive simulations, we show that `i-LDSC` recovers substantial non-additive heritability that is not captured by `LDSC` when genetic interactions are indeed present in the generative model for a given complex trait. More importantly, `i-LDSC` has a calibrated type I error rate and does not overestimate non-additive genetic contributions to trait variation in simulated data when only additive effects are present. While analyzing 25 complex traits in the UK Biobank and BioBank Japan, we illustrate that pairwise interactions are a significant source of "missing" heritability captured by additive GWAS summary statistics—suggesting that phenotypic variation due to non-additive genetic effects is more pervasive in human phenotypes than previously reported. Specifically, we find evidence of significant tagged non-additive genetic effects contributing to heritability estimates in all of the 25 traits in the UK Biobank, and 23 of the 25 traits we analyzed in the BioBank Japan. We believe that `i-LDSC`, with our development of a new *cis*-interaction score, represents a significant step towards resolving the true contribution of genetic interactions.

Results

Overview of the interaction-LD score regression model

Interaction-LD score regression (`i-LDSC`) is a statistical framework for estimating heritability (i.e., the proportion of trait variance attributable to genetic variance). Here, we will give an overview of the `i-LDSC` method and its corresponding software, as well as detail how its underlying model differs from that of `LDSC`[1]. We will assume that we are analyzing a GWAS dats set $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$ where $\mathbf{X}$ is an $N \times J$ matrix of genotypes with J denoting the number of SNPs (each of which is encoded as $\{0, 1, 2\}$ copies of a reference allele at each locus j) and $\mathbf{y}$ is an N-dimensional vector of measurements of a quantitative trait. The `i-LDSC` framework only requires summary statistics of individual-level data: namely, marginal

effect size estimates for each SNP $\widehat{\boldsymbol{\beta}}$ and a sample LD matrix $\mathbf{R}$ (which can be provided via reference panel data).

We begin by assuming the following generative linear model for complex traits

$$\mathbf{y} = \mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, (1 - H^2)\mathbf{I}), \tag{1}$$

where μ is an intercept term; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)$ is a J-dimensional vector containing the true additive effect sizes for an additional copy of the reference allele at each locus on $\mathbf{y}$; $\mathbf{W}$ is an $N \times M$ matrix of (pairwise) *cis*-acting SNP-by-SNP statistical interactions between some subset of causal SNPs, where columns of this matrix are assumed to be the Hadamard (element-wise) product between genotypic vectors of the form $\mathbf{x}_j \circ \mathbf{x}_k$ for the j-th and k-th variants; $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)$ is an M-dimensional vector containing the interaction effect sizes; $\boldsymbol{\varepsilon}$ is a normally distributed error term with mean zero and variance scaled according to the proportion of phenotypic variation not explained by genetic effects[17], which we will refer to as the broad-sense heritability of the trait denoted by H^2; and $\mathbf{I}$ denotes an $N \times N$ identity matrix. For convenience, we will assume that the genotype matrix (column-wise) and the trait of interest have been mean-centered and standardized. Lastly, we let each individual effect size follow a Gaussian distribution with variances proportional to their individual contributions to the heritability of the trait of interest[17–21]

$$\beta_j \sim \mathcal{N}(0, H^2\rho/J), \qquad \theta_m \sim \mathcal{N}(0, H^2(1 - \rho)/M) \tag{2}$$

where ρ measures the proportion of total genetic effects that is contributed by additive genetic effects. Effectively, we say $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = H^2\rho$ is the proportion of phenotypic variation contributed by additive SNP effects under the generative model, while $\mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = H^2(1 - \rho)$ makes up the remaining proportion of phenotypic variation contributed by genetic interactions.

A central objective in GWAS studies is to infer how much phenotypic variation can be explained by genetic effects. To achieve that objective, a key consideration involves incorporating the possibility of non-additive sources of genetic variation to be correlated with and explained by additive effect size estimates obtained from GWAS analyses[22]. If we assume that the genotype and interaction matrices $\mathbf{X}$ and $\mathbf{W}$ are not completely orthogonal (i.e., such that $\mathbf{X}^\intercal\mathbf{W} \neq \mathbf{0}$) then the following relationship between the moment matrix $\mathbf{X}^\intercal\mathbf{y}$, the observed marginal GWAS summary statistics $\widehat{\boldsymbol{\beta}}$, and the true coefficient

values $\boldsymbol{\beta}$ from the generative model in Eq. (1) holds in expectation (see Materials and Methods)

$$\mathbf{X}^\intercal\mathbf{y} = (\mathbf{X}^\intercal\mathbf{X})\boldsymbol{\beta} + (\mathbf{X}^\intercal\mathbf{W})\boldsymbol{\theta} \qquad \overset{\approx}{\Longleftrightarrow} \qquad \widehat{\boldsymbol{\beta}} = \mathbf{R}\boldsymbol{\beta} + \mathbf{V}\boldsymbol{\theta} \tag{3}$$

where $\mathbf{R}$ is a sample estimate of the LD matrix, and $\mathbf{V}$ represents a sample estimate of the correlation between the individual-level genotypes $\mathbf{X}$ and the span of genetic interactions between causal SNPs in $\mathbf{W}$. Intuitively, the term $\mathbf{V}\boldsymbol{\theta}$ can be interpreted as the non-additive effects that are tagged by the additive effect estimates from the GWAS study. Note that, when (i) non-additive genetic effects play a negligible role on the overall architecture of a trait (i.e., such that $\boldsymbol{\theta} = \mathbf{0}$) or (ii) the genotype and interaction matrices $\mathbf{X}$ and $\mathbf{W}$ do not share the same column space (i.e., such that $\mathbf{X}^\intercal\mathbf{W} = \mathbf{0}$), the equation above simplifies to a relationship between LD and summary statistics that is assumed in many GWAS studies and methods[23–29].

The goal of `i-LDSC` is to increase estimates of genetic variance by accounting for sources of non-additive genetic effects that can be explained by additive GWAS summary statistics. To do this, we extend the LD score regression framework and the corresponding `LDSC` software[17]. Here, according to Eq. (3), we note that $\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\mathbf{R}\boldsymbol{\beta} + \mathbf{V}\boldsymbol{\theta}, \lambda\mathbf{R})$ where λ is a scale variance term due to uncontrolled confounding effects[10,30]. Next, we condition on $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \boldsymbol{\theta})$ and take the expectation of chi-square statistics $\boldsymbol{\chi}^2 = N\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^\intercal$ to yield

$$\begin{aligned}
\mathbb{E}[\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^\intercal] = \mathbb{E}\left[\mathbb{E}\left[\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^\intercal \mid \boldsymbol{\Theta}\right]\right] &= \mathbb{E}\left[\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid \boldsymbol{\Theta}\right] + \mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid \boldsymbol{\Theta}\right]\mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid \boldsymbol{\Theta}\right]^\intercal\right] \\
&= \mathbb{E}\left[\lambda\mathbf{R} + (\mathbf{R}\boldsymbol{\beta} + \mathbf{V}\boldsymbol{\theta})(\mathbf{R}\boldsymbol{\beta} + \mathbf{V}\boldsymbol{\theta})^\intercal\right] \\
&= \mathbb{E}\left[\lambda\mathbf{R} + \mathbf{R}\boldsymbol{\beta}\boldsymbol{\beta}^\intercal\mathbf{R} + 2\mathbf{R}\boldsymbol{\beta}\boldsymbol{\theta}^\intercal\mathbf{V} + \mathbf{V}\boldsymbol{\theta}\boldsymbol{\theta}\mathbf{V}^\intercal\right] \\
&= \lambda\mathbf{R} + \left(\frac{H^2\rho}{J}\right)\mathbf{R}^2 + \left(\frac{H^2(1-\rho)}{M}\right)\mathbf{V}^2.
\end{aligned} \tag{4}$$

We define $\ell_j = \sum_k r_{jk}^2$ as the LD score for the additive effect of the j-th variant[17], and $f_j = \sum_m v_{jm}^2$ represents the "*cis*-interaction" LD score which encodes the interaction between the j-th variant and all other variants within a genomic window that is a pre-specified number of SNPs wide[21], respectively. By considering only the diagonal elements of LD matrix in the first term, similar to the original `LDSC` approach[10,17], we get the following simplified regression model

$$\mathbb{E}[\boldsymbol{\chi}^2] \propto \mathbf{1} + \boldsymbol{\ell}\tau + \boldsymbol{f}\sigma \tag{5}$$

where $\boldsymbol{\chi}^2 = (\chi_1^2, \ldots, \chi_J^2)$ is a J-dimensional vector of chi-square summary statistics, and $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_J)$ and $\boldsymbol{f} = (f_1, \ldots, f_J)$ are J-dimensional vectors of additive and *cis*-interaction LD scores, respectively. Furthermore, we define the variance components $\tau = NH^2\rho/J$ and $\sigma = NH^2(1-\rho)/M$ as the additive and interaction regression coefficients of the model, and $\mathbf{1}$ is the intercept meant to model the bias factor due to uncontrolled confounding effects (e.g., cryptic relatedness structure). In practice, we efficiently compute the *cis*-interaction LD scores by considering only a subset of interactions between each j-th focal SNP and SNPs within a *cis*-proximal window around the j-th SNP. In our validation studies and applications, we base the width of this window on the observation that LD decays outside of a window of 1 centimorgan (cM); therefore, SNPs outside the 1 cM window centered on the j-th SNP will not significantly contribute to its LD scores. Note that the width of this window can be relaxed in the `i-LDSC` software when appropriate. We fit the `i-LDSC` model using weighted least squares to estimate regression parameters and derive P-values for identifying traits that have significant statistical evidence of tagged *cis*-interaction effects by testing the null hypothesis $H_0 : \sigma = NH^2(1-\rho)/M = 0$. Importantly, under the null model of a trait being generated by only additive effects, the `i-LDSC` model in Eq. (5) reduces to the infinitesimal model[31].

Lastly, we want to note the empirical observation that the additive ($\boldsymbol{\ell}$) and interaction ($\boldsymbol{f}$) LD scores are lowly correlated. This is important because it indicates that the presence of *cis*-interaction LD scores in the model specified in Eq. (5) has little-to-no influence over the estimate for the additive coefficient τ. Instead, the inclusion of $\boldsymbol{f}$ creates a multivariate model that can identify the proportion of variance explained by both additive and non-additive effects in summary statistics. In other words, we can interpret σ as the phenotypic variation explained by tagged *cis*-acting interaction effects, and we use the sum of coefficient estimates $\hat{\tau} + \hat{\sigma}$ to construct `i-LDSC` heritability estimates. A full derivation of the *cis*-interaction regression framework and details about its corresponding implementation in our software `i-LDSC` can be found in Materials and Methods.

Detection of tagged non-additive effects using `i-LDSC` in simulations

We illustrate the power of `i-LDSC` across different genetic trait architectures via extensive simulation studies (Materials and Methods). We generate synthetic phenotypes using real genome-wide genotype data from individuals of self-identified European ancestry in the UK Biobank. To do so, we first assume that traits have a polygenic architecture where all SNPs have a nonzero additive effect. Next, we randomly

select a set of causal *cis*-interaction variants and divide them into two interacting groups (Materials and Methods). One may interpret the SNPs in group #1 as being the "hubs" in an interaction map[21]; while, SNPs in group #2 are selected to be variants within some kilobase (kb) window around each SNP in group #1. We assume a wide range of simulation scenarios by varying the following parameters:

- heritability: $H^2 = 0.3$ and 0.6;
- proportion of phenotypic variation that is generated by additive effects: $\rho = 0.5$, 0.8, and 1;
- percentage of SNPs selected to be in group #1: 1%, 5%, and 10%;
- genomic window used to assign SNPs to group #2: $\pm$10 and $\pm$100 kb.

We also varied the correlation between SNP effect size and minor allele frequency (MAF) (as discussed in Schoech et al.[32]). All results presented in this section are based on 100 different simulated phenotypes for each parameter combination.

Figure 1 demonstrates that `i-LDSC` robustly detects significant tagged non-additive genetic variance, regardless of the total number of causal interactions genome-wide. Instead, the power of `i-LDSC` depends on the proportion of phenotypic variation that is generated by additive versus interaction effects (ρ), and its power tends to scale with the window size used to compute the *cis*-interaction LD scores (see Materials and Methods). `i-LDSC` shows a similar performance for detecting tagged *cis*-interaction effects when the effect sizes of causal SNPs depend on their minor allele frequency and when we varied the number of SNPs assigned to be in group #2 within 10 kb and 100kb windows, respectively (Figures S1-S5).

Importantly, `i-LDSC` does not falsely identify putative non-additive genetic effects in GWAS summary statistics when the synthetic phenotype was generated by only additive effects ($\rho = 1$). Figure 2 illustrates the performance of `i-LDSC` under the null hypothesis $H_0 : \sigma = NH^2(1-\rho)/M = 0$, with the type I error rates for different estimation window sizes of the *cis*-interaction LD scores highlighted in panel A. Here, we also show that, when no genetic interaction effects are present, `i-LDSC` unbiasedly estimates the *cis*-interaction coefficient in the regression model $\sigma = 0$ (Figure 2B), robustly estimates the heritability (Figure 2C), and provides well-calibrated P-values when assessed over many traits (Figure 2D). This behavior is consistent across different MAF-dependent effect size distributions, and P-value calibration is not sensitive to misspecification of the estimation windows used to generate the *cis*-interaction LD scores (Figures S6-S7).

One of the innovations that `i-LDSC` offers over the traditional `LDSC` framework is increased heritability estimates after the identification of non-additive genetic effects that are tagged by GWAS summary statistics. Here, we applied both methods to the same set of simulations in order to understand how `LDSC` behaves for traits generated with *cis*-interaction effects. Figure 3 depicts boxplots of the heritability estimates for each approach and shows that, across an array of different synthetic phenotype architectures, `LDSC` captures less of phenotypic variance explained by all genetic effects. It is important to note that `i-LDSC` can yield upwardly biased heritability estimates when the *cis*-interaction scores are computed over genomic window sizes that are too small; however, these estimates become more accurate for larger window size choices (Figure S8). In contrast to `LDSC`, which aims to capture phenotypic variance attributable to the additive effects of genotyped SNPs, `i-LDSC` accurately partitions genetic effects into additive versus *cis*-interacting components, which in turn generally leads the ability of `i-LDSC` to capture more genetic variance. The mean absolute error between the true generative heritability and heritability estimates produced by `i-LDSC` and `LDSC` are shown in Tables S1 and S2, respectively. Generally, the error in heritability estimates is higher for `LDSC` than it is for `i-LDSC` across each of the scenarios that we consider.

Lastly, we perform an additional set of simulations where we explore other common generative models for complex trait architecture that involve non-additive genetic effects. Specifically, we compare heritability estimates from `LDSC` and `i-LDSC` in the presence of additive effects, *cis*-acting interactions, and a third source of genetic variance stemming from either gene-by-environment (G×E) or or gene-by-ancestry (G×Ancestry) effects. Details on how these components were generated can be found in Materials and Methods. In general, `i-LDSC` underestimates overall heritability when additive effects and *cis*-acting interactions are present alongside G×E (Figure S9) and/or G×Ancestry effects when PCs are included as covariates (Figure S10). Notably, when PCs are not included to correct for residual stratification, both `LDSC` and `i-LDSC` can yield unbounded heritability estimates greater than 1 (Figure S11). Also interestingly, when we omit *cis*-interactions from the generative model (i.e., the genetic architecture of simulated traits is only made up of additive and G×E or G×Ancestry effects), `i-LDSC` will still estimate a nonzero genetic variance component with the *cis*-interaction LD scores (Figures S12-S14). Collectively, these results empirically show the important point that *cis*-interaction scores are not enough to recover missing genetic variation for all types of trait architectures; however, they are helpful in recovering phenotypic variation explained by statistical *cis*-interaction effects. Recall that the linear relationship between

(expected) χ^2 test statistics and LD scores proposed by the `LDSC` framework holds when complex traits are generated under the polygenic model where all causal variants have the same expected contribution to phenotypic variation. When *cis*-interactions affect genetic architecture (e.g., in our earlier simulations in Figure 3), these assumptions are violated in `LDSC`, but the inclusion of the additional nonlinear scores in `i-LDSC` help recover the relationship between the expectation of χ^2 test statistics and LD.

As a final demonstration of how `i-LDSC` performs when assumptions of the original LD score model are violated, we also generated synthetic phenotypes with sparse architectures using the spike-and-slab model[20]. Here, traits were simulated with solely additive effects, but this time only variants with the top or bottom $\{1, 5, 10, 25, 50, 100\}$ percentile of LD scores were given nonzero effects (see Material and Methods). Breaking the relationship assumed under the `LDSC` framework between LD scores and chi-squared statistics (i.e., that they are generally positively correlated) led to unbounded estimates of heritability in all but the (polygenic) scenario when 100% of SNPs contributed to the phenotypic variation (Figure S15).

Application of `i-LDSC` to the UK Biobank and BioBank Japan

To assess whether non-additive genetic effects are significantly affecting estimates of heritability in empirical biobank data, we applied `i-LDSC` to 25 continuous quantitative traits from the UK Biobank and BioBank Japan (Table S3). Protocols for computing GWAS summary statistics for the UK Biobank are described in the Materials and Methods; while pre-computed summary statistics for BioBank Japan were downloaded directly from the consortium website (see URLs). We release the *cis*-acting SNP-by-SNP interaction LD scores used in our analyses on the `i-LDSC` GitHub repository from two reference groups in the 1000 Genomes: 489 individuals from the European superpopulation (EUR) and 504 individuals from the East Asian (EAS) superpopulation (see also Table S4).

In each of the 25 traits we analyzed in the UK Biobank, we detected significant proportions of estimated genetic variation stemming from tagged pairwise *cis*-interactions (Table 1). This includes many canonical traits of interest in heritability analyses: height, cholesterol levels, urate levels, and both systolic and diastolic blood pressure. Our findings in Table 1 are supported by multiple published studies identifying evidence of non-additive effects playing a role in the architectures of different traits of interest. For example, Li et al.[33] found evidence for genetic interactions that contributed to the pathogenesis of coronary artery disease. It was also recently shown that non-additive genetic effects plays a significant

role in body mass index[10]. Generally, we find that the traditional `LDSC` produces lower estimates of trait heritability because it does not consider the additional sources of genetic signal that `i-LDSC` does (Table 1). In BioBank Japan, 23 of the 25 traits analyzed had a significant nonlinear component detected by `i-LDSC` — with HDL and triglyceride levels being the only exceptions. We performed an additional analysis where the *cis*-interaction scores are included as an annotation alongside 97 other functional categories in the stratified-LD score regression framework and its software `s-LDSC`[34] (Materials and Methods). Here, `s-LDSC` heritability estimates still showed an increase with the interaction scores versus when the publicly available functional categories were analyzed alone (Table S6).

For each of the 25 traits that we analyzed, we found that the `i-LDSC` heritability estimates are significantly correlated with corresponding estimates from `LDSC` in both the UK Biobank ($r^2 = 0.988$, $P = 5.936 \times 10^{-24}$) and BioBank Japan ($r^2 = 0.849$, $P = 6.061 \times 10^{-11}$) as shown in Figure 4A. Additionally, we found that the heritability estimates for the same traits between the two biobanks are highly correlated according to both `LDSC` ($r^2 = 0.848$, $P = 7.166 \times 10^{-11}$) and `i-LDSC` ($r^2 = 0.666$, $P = 6.551 \times 10^{-7}$) analyses as shown in Figure 4B.

After comparing the `i-LDSC` heritability estimates to `LDSC`, we then assessed whether there was significant difference in the amount of phenotypic variation explained by the non-additive genetic effect component in the GWAS summary statistics derived from the the UK Biobank and BioBank Japan (i.e., comparing the estimates of σ; see Figure 4C). We show that, while heterogeneous between traits, the phenotypic variation explained by genetic interactions is relatively of the same magnitude for both biobanks ($r^2 = 0.372$, $P = 0.0119$). Notably, the trait with the most significant evidence of tagged *cis*-interaction effects in GWAS summary statistics is height which is known to have a highly polygenic architecture.

Finally, we show that the intercepts estimated by `LDSC` and `i-LDSC` are highly correlated in both the UK Biobank and the BioBank Japan (Figure 4D). Recall that these intercept estimates represent the confounding factor due to uncontrolled effects. For `LDSC`, this does include phenotypic variation that is due to unaccounted for pairwise statistical genetic interactions. The `i-LDSC` intercept estimates tend to be correlated with, but are generally different than, those computed with `LDSC` — empirically indicating that non-additive genetic variation is partitioned away and is missed when using the standard LD score alone. This result shows similar patterns in both the UK Biobank ($r^2 = 0.888$, $P = 1.962 \times 10^{-12}$) and BioBank Japan ($r^2 = 0.813$, $P = 7.814 \times 10^{-10}$), and it confirms that non-additive genetic effects can be a source of "missing" phenotypic variance explained in heritability estimation.

Discussion

In this paper, we present `i-LDSC`, an extension of the LD score regression framework which aims to recover missing heritability from GWAS summary statistics by incorporating an additional score that measures the non-additive genetic variation that is tagged by genotyped SNPs. Here, we demonstrate how `i-LDSC` builds upon the original `LDSC` model through the development of new "*cis*-interaction" LD scores which help to investigate signals of *cis*-acting SNP-by-SNP interactions (Figures 1 and S1-S5). Through extensive simulations, we show that `i-LDSC` is well-calibrated under the null model when polygenic traits are generated only by additive effects (Figures 2 and S6-S7), and it provides greater heritability estimates over `LDSC` when traits are indeed generated with *cis*-acting SNP-by-SNP interaction effects (Figures 3 and S8, and Tables S1 and S2). Finally, in real data, we show examples of many traits with estimated GWAS summary statistics that tag *cis*-interaction effects in the UK Biobank and BioBank Japan (Figures 4 and S16, and Tables 1 and S3-S6). We have made `i-LDSC` a publicly available command line tool that requires minimal updates to the computing environment used to run the original implementation of LD score regression (see URLs). In addition, we provide pre-computed *cis*-interaction LD scores calculated from the European (EUR) and East Asian (EAS) reference populations in the 1000 Genomes phase 3 data (see Data and Software Availability under Materials and Methods).

The current implementation of the `i-LDSC` framework offers many directions for future development and applications. First, as we showed with our simulation studies (Figures S9-S15), the *cis*-interaction LD scores that we propose are not always enough to recover explainable non-additive genetic effects for all types of trait architectures. While we focus on pairwise *cis*-acting SNP-by-SNP statistical interactions in this work, the theoretical concepts underlying `i-LDSC` can easily be adapted to other types of interactions as well. Second, in our analysis of the UK Biobank and BioBank Japan, we showed that the inclusion of additional categories via frameworks such as stratified LD score regression[35] can be used to provide more refined heritability estimates from GWAS summary statistics while accounting for linkage (see results in Table 1 versus Table S6). A key part of our future work is to continue to explore whether considering functional annotation groups would also improve our ability to identify tagged non-additive genetic effects. Lastly, we have only focused on analyzing one phenotype at a time in this study. However, many previous studies have extensively shown that modeling multiple phenotypes can often dramatically increase power[36,37]. Therefore, it would be interesting to extend the `i-LDSC` framework to multiple traits

to study nonlinear genetic correlations in the same way that LDSC was recently extended to uncover additive genetic correlation maps across traits[38].

URLs

i-LDSC software package for implementing interaction score regression, https://github.com/lcrawlab/i-LDSC; LDSC software package for implementing LD score regression, https://github.com/bulik/ldsc/; UK Biobank, https://www.ukbiobank.ac.uk; BioBank Japan, http://jenger.riken.jp/en/result; 1000 Genomes Project genetic map and haplotypes, http://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html; Database of Genotypes and Phenotypes (dbGaP), https://www.ncbi.nlm.nih.gov/gap; NHGRI-EBI GWAS Catalog, https://www.ebi.ac.uk/gwas/; GRM-MAF-LD package, https://github.com/arminschoech/GRM-MAF-LD; GCTA toolkit, https://yanglab.westlake.edu.cn/software/gcta/.

Acknowledgements

We thank Jeffrey P. Spence, Roshni Patel, Matthew Aguirre, Mineto Ota, and our anonymous referees for insightful comments on an earlier version of this manuscript as well as the Harpak, Ramachandran, and Crawford Labs for helpful discussions. This research was conducted in part using computational resources and services at the Center for Computation and Visualization at Brown University. This research was also conducted using the UK Biobank Resource under Application Numbers 14649 (LC) and 22419 (SR). S.P. Smith and D. Udwin were trainees supported under the Brown University Predoctoral Training Program in Biological Data Science (NIH T32 GM128596). S.P. Smith was also supported by NIH RF1AG073593. G. Darnell was supported by NSF Grant No. DMS-1439786 while in residence at the Institute for Computational and Experimental Research in Mathematics (ICERM) in Providence, RI. This research was supported in part by an Alfred P. Sloan Research Fellowship and a David & Lucile Packard Fellowship for Science and Engineering awarded to L. Crawford. This research was also partly supported by US National Institutes of Health (NIH) grant R01 GM118652, NIH grant R35 GM139628, and National Science Foundation (NSF) CAREER award DBI-1452622 to S. Ramachandran. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not

necessarily reflect the views of any of the funders.

Author Contributions

SPS, GD, SR, and LC conceived the study and developed the methods. SPS, GD, and LC developed the algorithms and software. All authors performed the analyses, interpreted the results, and wrote and revised the manuscript.

Competing Interests

The authors declare no competing interests.

Materials and Methods

Generative statistical model for complex traits

Our goal in this study is to re-analyze summary statistics from genome-wide association studies (GWAS) and estimate heritability while accounting for both additive genetic associations and tagged interaction effects. We begin by assuming the following generative linear model for complex traits and phenotypes

$$\mathbf{y} = \mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, (1 - H^2)\mathbf{I}), \tag{6}$$

where $\mathbf{y}$ denotes an N-dimensional vector of phenotypic states for a quantitative trait of interest measured in N individuals; μ is an intercept term; $\mathbf{X}$ is an $N \times J$ matrix of genotypes, with J denoting the number of single nucleotide polymorphism (SNPs) encoded as $\{0, 1, 2\}$ copies of a reference allele at each locus; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)$ is a J-dimensional vector containing the true additive effect sizes for an additional copy of the reference allele at each locus on $\mathbf{y}$; $\mathbf{W}$ is an $N \times M$ matrix of genetic interactions; $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)$ is an M-dimensional vector containing the interaction effect sizes; $\boldsymbol{\varepsilon}$ is a normally distributed error term with mean zero and variance scaled according to the proportion of phenotypic variation not explained by the broad-sense heritability of the trait, denoted by H^2; and $\mathbf{I}$ denotes an $N \times N$ identity matrix. While in theory, the matrix $\mathbf{W}$ could encode any source of non-additive genetic effects (e.g., gene-by-environmental effects), we limit our focus in this study to trait architectures that have been generated with contributions stemming from *cis*-acting statistical SNP-by-SNP interactions. To that end, we assume that the columns of $\mathbf{W}$ are the Hadamard (element-wise) product between genotypic vectors of the form $\mathbf{x}_j \circ \mathbf{x}_k$ for the j-th and k-th variants.

For convenience, we further assume that the genotype matrix (column-wise) and trait of interest have been mean-centered and standardized. Furthermore, we want to point out that the generative formulation of Eq. (6) can also be easily extended to accommodate other fixed effects (e.g., age, sex, or genotype principal components), as well as other random effects terms that can be used to account for sample non-independence due to other environmental factors. In addition, we choose to assume that $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are fixed effects here, but modeling these coefficients as a random effect is straightforward. Lastly, in this work, we only consider second order (or pairwise) SNP-by-SNP interactions. However, the generalization of the proposed framework to detect genetic effects from higher-order interactions is also straightforward

and only involves manipulating the interaction matrix $\mathbf{W}$ [21,39].

GWA summary statistics and tagged interaction effects

As previously mentioned, the key to this work is that SNP-level GWAS summary statistics can also tag non-additive genetic effects if there is a nonzero correlation between individual-level genotypes and their interactions (as defined in Eq. (6)). Throughout this section, we will use $\mathbf{X}^\intercal\mathbf{X}/N$ to denote the linkage disequilibrium (LD) or pairwise correlation matrix between SNPs. We will then let $\mathbf{R}$ represent an LD matrix empirically estimated from external data (e.g., directly from GWAS study data, or using a pairwise LD map from a population that is representative of the samples analyzed in the GWAS study). The important property here is that

$$\mathbb{E}[\mathbf{X}^\intercal\mathbf{X}] \approx N\mathbf{R}, \qquad \mathbb{E}[\mathbf{x}_j^\intercal\mathbf{x}_j] \approx N, \qquad \mathbb{E}[\mathbf{x}_j^\intercal\mathbf{x}_k] \approx Nr_{jk} \tag{7}$$

where the term r_{jk} is defined as the Pearson correlation coefficient between the j-th and k-th SNPs, respectively, and $\mathbf{x}_j$ denotes the j-th column of the individual-level genotype matrix $\mathbf{X}$.

A central goal in GWAS studies is to jointly infer how much phenotypic variation can be explained by genetic effects. This often amounts to estimating the effect sizes $\boldsymbol{\beta} = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{y}$ for each SNP, given both genotypic and phenotypic measurements for each assayed individual. However, since the generative model in Eq. (6) is an underdetermined linear system (i.e., $J > N$) for many GWAS applications, we need to make additional modeling assumptions on the regression coefficients to make the generative model identifiable. To do so, we follow standard linear modeling approaches [17–21] and assume that each effect size follows a Gaussian distribution with variances proportional to their individual contributions to the heritability of the trait of interest. Namely, we assume that

$$\beta_j \sim \mathcal{N}(0, H^2\rho/J), \qquad \theta_m \sim \mathcal{N}(0, H^2(1-\rho)/M), \qquad j = 1, \ldots, J \qquad m = 1, \ldots, M \tag{8}$$

where ρ measures the proportion of total genetic effects that is contributed by the additive effects in the generative model. Alternatively, we say that $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = H^2\rho$ is the proportion of phenotypic variation contributed by additive SNP effects under the generative model, which then leaves the set of interactions involving some subset of causal SNPs to contribute the remaining $\mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = H^2(1-\rho)$ proportion to the

heritability.

Traditional estimation of additive GWAS summary statistics

In traditional GWAS studies, summary statistics of the true additive effects $\boldsymbol{\beta}$ in Eq. (6) are typically derived by computing a marginal least squares estimate with the observed data

$$\widehat{\beta}_j = (\mathbf{x}_j^\intercal \mathbf{x}_j)^{-1}\mathbf{x}_j^\intercal \mathbf{y} \qquad \Longleftrightarrow \qquad \widehat{\boldsymbol{\beta}} = \text{diag}(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{y}. \tag{9}$$

There are two key identities that may be taken from Eq. (9). The first uses Eq. (7) and is the approximate relationship (in expectation) between the moment matrix $\mathbf{X}^\intercal\mathbf{y}$ and the linear effect size estimates $\widehat{\boldsymbol{\beta}}$:

$$\mathbf{X}^\intercal\mathbf{y} = \text{diag}(\mathbf{X}^\intercal\mathbf{X})\widehat{\boldsymbol{\beta}} \approx N\widehat{\boldsymbol{\beta}}. \tag{10}$$

The second key point combines Eqs. (7) and (10) to describe the asymptotic relationship between the observed marginal GWAS summary statistics $\widehat{\boldsymbol{\beta}}$ and the joint coefficient values $\boldsymbol{\beta}$ where

$$\boldsymbol{\beta} = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{y} \approx (N\mathbf{R})^{-1}N\widehat{\boldsymbol{\beta}} = \mathbf{R}^{-1}\widehat{\boldsymbol{\beta}}. \tag{11}$$

After some algebra, the above mirrors a high-dimensional regression model (in expectation) where $\widehat{\boldsymbol{\beta}} = \mathbf{R}\boldsymbol{\beta}$ with the estimated summary statistics as the response variables and the empirically estimated LD matrix acting as the design matrix[23,26,28,29,40]. Theoretically, the resulting output coefficients from this high-dimensional model are the desired true effect size estimates used to generate the phenotype of interest.

Additive GWAS summary statistics with tagged interaction effects

When interactions contribute to the architecture of complex traits (i.e., $\boldsymbol{\theta} \neq \mathbf{0}$ and $\rho < 1$), the marginal GWAS summary statistics derived using least squares in Eq. (9) can also explain non-additive variation if there is a nonzero correlation between genotypes and their interactions. To see this, we take the joint solution for the true regression coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ from the generative model in Eq. (6)

$$\begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{\theta}\end{bmatrix} = \begin{bmatrix}\mathbf{X}^\intercal\mathbf{X} & \mathbf{X}^\intercal\mathbf{W} \\ \mathbf{W}^\intercal\mathbf{X} & \mathbf{W}^\intercal\mathbf{W}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}^\intercal \\ \mathbf{W}^\intercal\end{bmatrix}\mathbf{y}, \tag{12}$$

where the matrix $\mathbf{X}^\intercal\mathbf{W}$ can be interpreted as the sample correlation between individual-level genotypes and the *cis*-interactions between causal SNPs. By solving for the additive genetic effects (again in expectation using Eqs. (7) and (10)), we get the following alternative relationship between the moment matrix $\mathbf{X}^\intercal\mathbf{y}$, the observed marginal GWAS summary statistics $\widehat{\boldsymbol{\beta}}$, and the true coefficient values $\boldsymbol{\beta}$ where

$$\mathbf{X}^\intercal\mathbf{y} = (\mathbf{X}^\intercal\mathbf{X})\boldsymbol{\beta} + (\mathbf{X}^\intercal\mathbf{W})\boldsymbol{\theta} \qquad \overset{\approx}{\Longleftrightarrow} \qquad \widehat{\boldsymbol{\beta}} = \mathbf{R}\boldsymbol{\beta} + \mathbf{V}\boldsymbol{\theta}. \tag{13}$$

Here, we define $\mathbf{V}$ to represent a sample estimate of the correlation between the individual-level genotypes and the non-additive genetic interaction matrix such that $\mathbb{E}[\mathbf{X}^\intercal\mathbf{W}] \approx N\mathbf{V}$. Similar to the LD matrix $\mathbf{R}$, the correlation matrix $\mathbf{V}$ is also assumed to be computed from reference panel data. Intuitively, when $\mathbf{V}\boldsymbol{\theta} \neq \mathbf{0}$ there is additional phenotypic variation contributed by genetic interactions that can be explained by GWAS effect size estimates. Moreover, when $\mathbf{V}\boldsymbol{\theta} = \mathbf{0}$, then the relationship in Eq. (13) converges onto the conventional asymptotic assumption (in expectation) between GWAS summary statistics and the true additive coefficients in Eq. (11)[23,26,28,29,40].

Full derivation of interaction LD score regression

In order to derive the interaction LD score (`i-LDSC`) regression framework, recall that our goal is to identify evidence of tagged interaction effects within GWAS summary statistics. To do this, we build upon the LD score regression framework and the `LDSC` software[17]. Here, we assume nonzero contributions from *cis*-acting SNP-by-SNP interaction effects in the generative model of complex traits as in Eq. (13), and we use the observed least squares estimates from Eq. (9) to compute chi-square statistics $\chi^2_j = N\widehat{\beta}^2_j$ for every $j = 1, \ldots, J$ variant in the data. Taking the expectation of these statistics yields

$$\mathbb{E}[\chi^2_j] = N\mathbb{E}[\widehat{\beta}^2_j] = N\left[\mathbb{V}[\widehat{\beta}_j] + \left(\mathbb{E}[\widehat{\beta}_j]\right)^2\right]. \tag{14}$$

We can simplify Eq. (14) in two steps. First, by combining the prior assumption in Eq. (8) and the asymptotic approximation in Eq. (13), we can show that marginal expectation (i.e., when not conditioning on the true coefficients) $\mathbb{E}[\widehat{\beta}_j] = 0$ for all variants. Second, by conditioning on the generative model from

Eq. (6), we can use the law of total variance to simplify $\mathbb{V}[\widehat{\beta}_j]$ where

$$\begin{aligned}\mathbb{V}[\widehat{\beta}_j] = \mathbb{E}[\mathbb{V}[\widehat{\beta}_j \,|\, \mathbf{X}]] + \mathbb{V}[\mathbb{E}[\widehat{\beta}_j \,|\, \mathbf{X}]] &\approx \mathbb{E}[\mathbb{V}[\mathbf{x}_j^\intercal \mathbf{y}/N \,|\, \mathbf{X}]] + 0 \\ &= \mathbb{E}\left[\frac{1}{N^2}\mathbf{x}_j^\intercal \left\{\mathbb{V}[\mathbf{y} \,|\, \mathbf{X}]\right\} \mathbf{x}_j\right] \\ &= \mathbb{E}\left[\frac{1}{N^2}\mathbf{x}_j^\intercal \left\{\frac{H^2\rho}{J}\mathbf{X}\mathbf{X}^\intercal + \frac{H^2(1-\rho)}{M}\mathbf{W}\mathbf{W}^\intercal + (1-H^2)\right\} \mathbf{x}_j\right] \\ &= \mathbb{E}\left[\frac{1}{N^2}\left\{\frac{H^2\rho}{J}\mathbf{x}_j^\intercal\mathbf{X}\mathbf{X}^\intercal\mathbf{x}_j + \frac{H^2(1-\rho)}{M}\mathbf{x}_j^\intercal\mathbf{W}\mathbf{W}^\intercal\mathbf{x}_j + N(1-H^2)\right\}\right].\end{aligned}$$

Using the same logic from the original `LDSC` regression framework[17], we can use Isserlis' theorem[41] to write the above in terms of more familiar quantities based on sample correlations

$$\frac{1}{N^2}\mathbf{x}_j^\intercal\mathbf{X}\mathbf{X}^\intercal\mathbf{x}_j = \sum_{k=1}^{J}\widetilde{r}_{jk}^2, \qquad \frac{1}{N^2}\mathbf{x}_j^\intercal\mathbf{W}\mathbf{W}^\intercal\mathbf{x}_j = \sum_{m=1}^{M}\widetilde{v}_{jm}^2 \tag{15}$$

where $\widetilde{r}_{jk}$ is used to denote the sample correlation between additively-coded genotypes at the j-th and k-th variants, and $\widetilde{v}_{jm}$ is used to denote the sample correlation between the genotype of the j-th variant and the m-th genetic interaction on the phenotype of interest (again see Eq. (13)). Furthermore, we can use the delta method (only displaying terms up to $\mathcal{O}(1/N^2)$) to show that (in expectation)

$$\mathbb{E}[\widetilde{r}_{jk}^2] \approx r_{jk}^2 + (1 - r_{jk}^2)/N, \qquad \mathbb{E}\left[\widetilde{v}_{jm}^2\right] \approx v_{jm}^2 + \left(1 - v_{jm}^2\right)/N. \tag{16}$$

Next, we can then approximate the quantities in Eq. (15) via the following

$$\mathbb{E}\left[\sum_{k=1}^{J}\widetilde{r}_{jk}^2\right] \approx \ell_j + (J - \ell_j)/N, \qquad \mathbb{E}\left[\sum_{m=1}^{M}\widetilde{v}_{jm}^2\right] \approx f_j + (M - f_j)/N \tag{17}$$

where ℓ_j is the corresponding LD score for the additive effect of the j-th variant and f_j represents the "interaction" LD score between the j-th SNP and all other variants in the data set[21], respectively. Altogether, this leads to the specification of the univariate framework with the j-th SNP

$$\mathbb{E}[\chi_j^2] \approx N\left[\left(\frac{H^2\rho}{J}\right)\ell_j + \left(\frac{H^2(1-\rho)}{M}\right)f_j + \frac{1}{N}(1-H^2)\right] = \ell_j\tau + f_j\sigma + 1 \tag{18}$$

where we define $\tau = NH^2\rho/J$ as estimates of the additive genetic signal, the coefficient $\sigma = NH^2(1-\rho)/M$ as an estimate of the proportion of phenotypic variation explained by tagged interaction effects, and $\mathbf{1}$ is the intercept meant to model the misestimation due to uncontrolled confounding effects (e.g., cryptic relatedness and population stratification). Similar to the original `LDSC` formulation, an intercept greater than one means significant bias. Note that the simplification for many of the terms above such as $(1-H^2)/N \approx 1/N$ results from our assumption that the number of individuals in our study is large. For example, the sample sizes for each biobank-scale study considered in the analyses of this manuscript are at least on the order of $N \geq 10^4$ observations (see Table S5). Altogether, we can jointly express Eq. (18) in multivariate form as

$$\mathbb{E}[\boldsymbol{\chi}^2] \approx \boldsymbol{\ell}\tau + \boldsymbol{f}\sigma + \mathbf{1} \tag{19}$$

where $\boldsymbol{\chi}^2 = (\chi_1^2, \ldots, \chi_J^2)$ is a J-dimensional vector of chi-square summary statistics, and $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_J)$ and $\boldsymbol{f} = (f_1, \ldots, f_J)$ are J-dimensional vectors of additive and *cis*-interaction LD scores, respectively. It is important to note that, while $\boldsymbol{\chi}^2$ must be recomputed for each trait of interest, both vectors $\boldsymbol{\ell}$ and $\boldsymbol{f}$ only need to be constructed once per reference panel or individual-level genotypes (see next section for efficient computational strategies).

To identify summary statistics that have significant tagged interaction effects, we test the null hypothesis $H_0 : \sigma = NH^2(1-\rho)/M = 0$. The `i-LDSC` software package implements the same model fitting strategy as `LDSC`. Here, we use weighted least squares to fit the joint regression in Eq. (19) such that

$$\widehat{\sigma} = (\boldsymbol{f}^\intercal \boldsymbol{\Psi} \boldsymbol{f})^{-1} \boldsymbol{f}^\intercal \boldsymbol{\Psi} \boldsymbol{\chi}^2, \qquad \psi_{jj} = [\ell_j \widehat{\tau} + f_j \widehat{\sigma} + 1]^{-2} \tag{20}$$

where $\boldsymbol{\Psi}$ is a $J \times J$ diagonal weight matrix with nonzero elements set to values inversely proportional to the conditional variance $\mathbb{V}[\chi_j^2 \,|\, \ell_j, f_j] = \psi_{jj}^{-1}$ to adjust for both heteroscedasticity and over-estimation of the summary statistics for each SNP[17]. Standard errors for each coefficient estimate are derived via a delete-one jackknife over blocks of SNPs in the data[35], and we then use those standard errors to derive P-values with a two-sided test (i.e., testing the alternative hypothesis $H_A : \sigma = NH^2(1-\rho)/M \neq 0$). For all analyses in this paper, we estimate proportion of phenotypic variance explained by genetic effects using a sum of the coefficients $\hat{\tau} + \hat{\sigma}$ (i.e., the estimated additive component plus the additional genetic variance explained by the tagged pairwise interaction effects).

Efficient computation of *cis*-interaction LD scores

In practice, *cis*-interaction LD scores in `i-LDSC` can be computed efficiently through realizing two key opportunities for optimization. First, given J SNPs, the full matrix of genome-wide interaction effects $\mathbf{W}$ contains on the order of $J(J-1)/2$ total pairwise interactions. However, the correlation between the genotype of the j-th SNP and the interactions where its involved (i.e., $\mathbf{x}_j^\intercal(\mathbf{x}_j \circ \mathbf{x}_l)$ for $l \neq j$) is bound to be much larger than the correlation between the genotype of the j-th SNP $\mathbf{x}_j$ and interactions involving some other SNP (e.g., $\mathbf{x}_j^\intercal(\mathbf{x}_k \circ \mathbf{x}_l)$ for $k \neq j$ and $l \neq j$). To that end, we can compute the `i-LDSC` score for each SNP by replacing the full $\mathbf{W}$ matrix with a subsetted matrix $\mathbf{W}_j$ which includes only interactions involving the j-th SNP. Analogous to the original `LDSC` formulation[17], we consider only interactive SNPs within a *cis*-window proximal to the focal j-th SNP for which we are computing the `i-LDSC` score. In the original `LDSC` model, this is based on the observation that LD decays outside of a window of 1 centimorgan (cM); therefore, SNPs outside the 1 cM window centered on the j-th SNP j will not significantly contribute to its LD score.

The second opportunity for optimization comes from the fact that the matrix of interaction effects for any focal SNP, $\mathbf{W}_j$, does not need to be explicitly generated. Referencing Eq. (15), the `i-LDSC` scores are defined as $\mathbf{x}_j^\intercal \mathbf{W}_j \mathbf{W}_j^\intercal \mathbf{x}_j / N^2$. This can be re-written as $\mathbf{x}_j^\intercal(\mathbf{D}_j\mathbf{X}^{(j)})(\mathbf{D}_j\mathbf{X}^{(j)})^\intercal \mathbf{x}_j$, where $\mathbf{D}_j = \text{diag}(\mathbf{x}_j)$ is a diagonal matrix with the j-th genotype as its nonzero elements[21] and $\mathbf{X}^{(j)}$ denotes the subset SNPs within a *cis*-window proximal to the focal j-th SNP. This means that the `i-LDSC` score for the j-th SNP can be simply computed as the following

$$f_j \approx \frac{1}{N^2}(\mathbf{x}_j^\intercal)^2 \mathbf{X}^{(j)} \mathbf{X}^{(j)\intercal}(\mathbf{x}_j)^2. \tag{21}$$

With these simplifications, the computational complexity of generating `i-LDSC` scores reduces to that of computing LD scores — modulo a vector-by-vector Hadamard product which, for each SNP, is constant factor of N (i.e., the number of genotyped individuals).

Coefficient estimates as determined by *cis*-interaction window size

When computing *cis*-interaction LD scores, the most important de*cis*ion is choosing the number of interacting SNPs to include in $\mathbf{X}^{(j)}$ (or equivalently $\mathbf{W}_j$ for each j-th focal SNP in the calculation of f_j in Eq. (21)). The `i-LDSC` framework considers different estimating windows to account for our lack of *a*

priori knowledge about the "correct" non-additive genetic architecture of traits. Theoretically, one could follow previous work[20,25,27,29,30,42] by considering an L-valued grid of possible SNP interaction window sizes. After fitting a series of `i-LDSC` regressions with *cis*-interaction LD scores $\boldsymbol{f}^{(l)}$ generated under the L-different window sizes, we could compute normalized importance weights using their maximized likelihoods via the following

$$\pi^{(l)} = \frac{\mathcal{L}\left(\boldsymbol{\ell}, \boldsymbol{f}^{(l)}; \widehat{\boldsymbol{\beta}}\right)}{\sum_{l'} \mathcal{L}\left(\boldsymbol{\ell}, \boldsymbol{f}^{(l')}; \widehat{\boldsymbol{\beta}}\right)}, \qquad \sum_{l=1}^{L} \pi^{(l)} = 1. \tag{22}$$

As a final step in the model fitting procedure, we could then compute averaged estimates of the coefficients τ and σ by marginalizing (or averaging) over the L-different grid combinations of estimating windows

$$\widehat{\tau} = \sum_{l=1}^{L} \pi^{(l)} \widehat{\tau}^{(l)}, \qquad \widehat{\sigma} = \sum_{l=1}^{L} \pi^{(l)} \widehat{\sigma}^{(l)}. \tag{23}$$

This final step can be viewed as an analogy to model averaging where marginal estimates are computed via a weighted average using the importance weights[43]. In the current study, we explore the utility of *cis*-interaction LD scores generated with different window sizes ±5, ±10, ±25, and ±50 SNPs around each j-th focal SNP. In practice, we find that *cis*-interaction LD scores that are calculated using larger windows lead to the most robust estimates of heritability while also not over representing the total phenotypic variation explained by tagged non-additive genetic effects (see Figure S8). Therefore, unless otherwise stated, we use *cis*-interaction LD scores calculated with a ±50 SNP interaction window for all simulations and real data analyses conducted in this work. For a direct comparison between choosing a single window size versus the model averaging strategy described above, see Tables S1 and S2.

Relationship between minor allele frequency and effect size

The `LDSC` software computes LD scores using annotations over equally spaced minor allele frequency (MAF) bins. These annotations enable the per trait relationship between the MAF and the effect size of each variant in the genome to vary based on the discrete category (or MAF bin) it is placed into. This additional flexibility is intended to help `LDSC` be more robust when estimating heritability. The relationship between MAF and effect size is already implicitly encoded in the `LDSC` formulation since we assume genotypes are normalized. When normalizing by the variance of each SNP (or equivalently its

MAF), we make the assumption that rare variants inherently have larger effect sizes. There exists a true functional relationship between MAF and effect size which is likely to be somewhere between the two extremes of (*i*) normalizing each SNP by its MAF and (*ii*) allowing the variance per SNP to be dictated by its MAF.

Recent approaches have proposed using a single parameter α to better represent the nonlinear relationship between MAF and variant effect size. The main idea is that this α not only provides the same additional flexibility to `LDSC` as the MAF-based discrete annotations, but it also empirically yields even more pre*cis*e heritability estimates[44]. Namely, we use

$$\ell_j(c) := \sum_k L_{jk}(\alpha) a_c(k), \qquad L_{jk}(\alpha) = r_{jk}^2 \mathbb{V}[\mathbf{x}_k]^{1-\alpha} \tag{24}$$

where $a_c(k)$ is the annotation value for the c-th categorical bin. The α parameter is unknown in practice and needs to be estimated for any given trait. While standard ranges for α can be used for heritability estimates, we use a restricted maximum likelihood (REML) based method which was recently developed[32].

In the `i-LDSC` software, we use this α construction to handle the relationship between MAF and variant effect size for two specific reasons. First, by constructing the LD scores using α, we more accurately capture the variation in chi-square test statistics due to additive effects[44]. Second, we note that there is correlation between MAF and (*i*) LD scores, (*ii*) *cis*-interaction LD scores, and (*iii*) trait architecture. To that end, if we do not properly condition on MAF, there becomes additional bias, and we may falsely attribute some amount of variation in the chi-square test statistics to LD or the tagged interaction effects. Therefore, in our formulation, we include an α term on the LD scores to condition on this effect. We demonstrate in simulations that this removes the bias introduced by the relationship between MAF and trait architecture, and it mitigates potential inflation of type I error rates in the `i-LDSC` test.

Estimation of allele frequency parameters

In the main text, we analyzed 25 complex traits in both the UK Biobank and BioBank Japan data sets. In order to account for minor allele frequency (MAF) dependent trait architecture, we calculated α values for each trait that had not been analyzed by previous studies[32]. The α estimates for each of the 25 traits analyzed in this study are shown in Table S4. Intuitively, α parameterizes the weighting of the effects of each individual variant given its frequency in the study cohort and can take on values in the range of

[-1,0]. More negative values of α indicate that lower frequency variants contribute more to the observed variation in a trait of interest, whereas values of α closer to zero indicate that common variants contribute a greater amount of variation to observed trait values.

We took α values for 11 traits (again see Table S4) that had previously been calculated from Schoech et al.[32]. For the remaining 14 traits analyzed in this study, we followed the estimation protocol described in the same manuscript. Specifically, using the variants passing the quality control step in our pipeline for 25,000 randomly selected individuals in the UK Biobank cohort, we constructed MAF-dependent genetic relatedness matrices for values of $\alpha = \{-1, -0.95, -0.9, \ldots, 0\}$ using the `GRM-MAF-LD` software, `https://github.com/arminschoech/GRM-MAF-LD`. We then used the `GCTA` software[45] to obtain heritability and likelihood estimates using REML for each α-trait pairing. We then fit a trait-specific profile likelihood across the range of α values and estimate the maximum likelihood value of α using a natural cubic spline.

Simulation studies

We used a simulation scheme to generate synthetic quantitative traits and SNP-level summary statistics under multiple genetic architectures using real genome-wide data from individuals of self-identified European ancestry in the UK Biobank. Here, we consider phenotypes that have some combination of additive effects, *cis*-acting interactions, and a third source of genetic variance stemming from either gene-by-environment (G×E) or gene-by-ancestry (G×Ancestry) effects. For each scenario, we select some set of SNPs to be causal and assume that complex traits are generated via the following general linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\theta} + \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \kappa^2\mathbf{I}), \tag{25}$$

where $\mathbf{y}$ is an N-dimensional vector containing all the phenotypes; $\mathbf{X}$ is an $N \times J$ matrix of genotypes encoded as 0, 1, or 2 copies of a reference allele; $\boldsymbol{\beta}$ is a J-dimensional vector of additive effect sizes for each SNP; $\mathbf{W}$ is an $N \times M$ matrix which holds all pairwise interactions between the randomly selected subset of the interacting SNPs with corresponding effects $\boldsymbol{\theta}$; $\mathbf{Z}$ is an $N \times K$ matrix of either G×E or G×Ancestry interactions with coefficients $\boldsymbol{\delta}$; and $\boldsymbol{\varepsilon}$ is an N-dimensional vector of environmental noise. The phenotypic variation is assumed to be $\mathbb{V}[\mathbf{y}] = 1$. All additive and interaction effect sizes for SNPs are randomly drawn from independent standard Gaussian distributions and then rescaled so that they explain a fixed proportion of the phenotypic variance $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] + \mathbb{V}[\mathbf{W}\boldsymbol{\theta}] + \mathbb{V}[\mathbf{Z}\boldsymbol{\delta}] = H^2$. Note that we do

not assume any specific correlation structure between the effect sizes $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and $\boldsymbol{\delta}$. We then rescale the random error term such that $\mathbb{V}[\boldsymbol{\varepsilon}] = (1 - H^2)$. In the main text, we compare the traditional `LDSC` to its direct extension in `i-LDSC`. For each method, GWAS summary statistics are computed by fitting a single-SNP univariate linear model via least squares where $\widehat{\beta}_j = (\mathbf{x}_j^\intercal \mathbf{x}_j)^{-1} \mathbf{x}_j^\intercal \mathbf{y}$ for every $j = 1, \ldots, J$ SNP in the data. These effect size estimates are used to derive the chi-square test statistics $\chi_j^2 = N\widehat{\beta}_j^2$. We implement both `LDSC` and `i-LDSC` with the LD matrix $\mathbf{R} = \mathbf{X}^\intercal \mathbf{X}/N$ and the *cis*-interaction correlation matrix $\mathbf{V} = \mathbf{X}^\intercal \mathbf{W}/N$ being computed using a reference panel of 489 individuals from the European superpopulation (EUR) of the 1000 Genomes Project. The resulting matrices $\mathbf{R}$ and $\mathbf{V}$ are used to compute the additive and *cis*-interaction LD scores, respectively.

Polygenic simulations with *cis*-interactions. In our first set of simulations (Figures 1-3 and S1-S8, and Tables S1 and S2), we consider phenotypes with polygenic architectures that are made up of only additive and *cis*-acting SNP-by-SNP interactions. Here, we begin by assuming that every SNP in the genome has at least a small additive effect on the traits of interest. Next, when generating synthetic traits, we assume that the additive effects make up ρ% of the heritability while the pairwise interactions make up the remaining $(1-\rho)$%. Alternatively, the proportion of the heritability explained by additivity is said to be $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = \rho H^2$, while the proportion detailed by interactions is given as $\mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = (1-\rho)H^2$. The setting of $\rho = 1$ represents the limiting null case for `i-LDSC` where the variation of a trait is driven by solely additive effects. Here, we use the same simulation strategy used in Crawford et al.[21] where we divide the causal *cis*-interaction variants into two groups. One may view the SNPs in group #1 as being the "hubs" of an interaction map. SNPs in group #2 are selected to be variants within some kilobase (kb) window around each SNP in group #1. Given different parameters for the generative model in Eq. (25), we simulate data mirroring a wide range of genetic architectures by toggling the following parameters:

- heritability: $H^2 = 0.3$ and 0.6;
- proportion of phenotypic variation that is generated by additive effects: $\rho = 0.5$, 0.8, and 1;
- percentage of SNPs selected to be in group #1: 1% (sparse), 5%, and 10% (polygenic);
- genomic window used to assign SNPs to group #2: $\pm$10 and $\pm$100 kilobase (kb);
- allele frequency parameter: α = -1, -0.5, and 0.

All figures and tables show the mean performances (and standard errors) across 100 simulated replicates.

Polygenic simulations with gene-by-environmental effects. In our second set of simulations (Figures S9 and S12), we continue to consider phenotypes with polygenic architectures that are made up of only additive and *cis*-acting SNP-by-SNP interactions; however, now we also consider each trait to have contributions stemming from nonzero G×E effects. Here, both the additive and *cis*-interaction effects are simulated in the same way as previously described where, for the two groups of interacting variants, 10% of SNPs were selected to be in group #1 and we chose ±10 kb windows to assign SNPs to group #2. To create G×E effects, we follow a simulation strategy implemented by Zhu et al.[46] and split our sample population in half to emulate two subsets of individuals coming from different environments. We randomly draw the effect sizes for the first environment from a standard Gaussian distribution which we denote as $\boldsymbol{\delta}_1$. We then selected an amplification coefficient w and set the effect sizes of the G×E interactions in the second environment to be a scaled version of the first environment effects where $\boldsymbol{\delta}_2 = w\boldsymbol{\delta}_1$. In this paper, we generate traits with heritability $H^2 = \{0.3, 0.6\}$ and amplification coefficients set to $w = [1.1, 1.2, \ldots, 2]$. For the first set of simulations, we hold the proportion of phenotypic variation explained by the different genetic components constant by fixing:

- $H^2 = 0.3$: $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = 0.15$; $\mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = 0.075$; and $\mathbb{V}[\mathbf{Z}\boldsymbol{\delta}] = 0.075$;
- $H^2 = 0.6$: $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = 0.3$; $\mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = 0.15$; and $\mathbb{V}[\mathbf{Z}\boldsymbol{\delta}] = 0.15$;

where $\mathbf{Z} = [\mathbf{X}_1, \mathbf{X}_2]$ is the set of genotypes split according to environment and $\boldsymbol{\delta} = [\boldsymbol{\delta}_1, \boldsymbol{\delta}_2]$. To test the sensitivity of the *cis*-interaction LD scores to other sources of non-additive variation, we also repeated the same simulations where there were only additive and G×E effects contributing equally to trait architecture:

- $H^2 = 0.3$: $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = 0.15$; $\mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = 0$; and $\mathbb{V}[\mathbf{Z}\boldsymbol{\delta}] = 0.15$;
- $H^2 = 0.6$: $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = 0.3$; $\mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = 0$; and $\mathbb{V}[\mathbf{Z}\boldsymbol{\delta}] = 0.3$.

Again all figures show the mean performances (and standard errors) across 100 simulated replicates.

Polygenic simulations with gene-by-ancestry effects. In our third set of simulations (Figures S10, S11, S13, and S14), we consider phenotypes with polygenic architectures that are made up of additive, *cis*-interactions, and G×Ancestry effects. Here, we follow Sohail et al.[47] and first run a matrix decomposition on the individual-level genotype matrix $\mathbf{X} = \mathbf{U}\mathbf{Q}^\intercal$ where $\mathbf{U}$ is a unitary $N \times K$ score matrix, $\mathbf{Q}$ is a

$K \times J$ loadings matrix, and K represents the number of (predetermined) principal components (PCs). To generate G×Ancestry interactions, we then create the matrix $\mathbf{Z}_k = \mathbf{X}\mathbf{q}_k$ where $\mathbf{q}_k$ is a J-dimensional vector of SNP loadings for the k-th principal component. In this paper, we generate traits with heritability $H^2 = \{0.3, 0.6\}$ and interaction effects taken over $k = 1, \ldots, 10$ principal components. For the first set of simulations, we hold the proportion of phenotypic variation explained by the different genetic components constant by fixing:

- $H^2 = 0.3$: $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = 0.15$; $\mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = 0.075$; and $\mathbb{V}[\mathbf{Z}\boldsymbol{\delta}] = 0.075$;
- $H^2 = 0.6$: $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = 0.3$; $\mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = 0.15$; and $\mathbb{V}[\mathbf{Z}\boldsymbol{\delta}] = 0.15$;

To test the sensitivity of the *cis*-interaction LD scores to other sources of non-additive variation, we also repeated the same simulations where there were only additive and G×E effects contributing equally to trait architecture:

- $H^2 = 0.3$: $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = 0.15$; $\mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = 0$; and $\mathbb{V}[\mathbf{Z}\boldsymbol{\delta}] = 0.15$;
- $H^2 = 0.6$: $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = 0.3$; $\mathbb{V}[\mathbf{W}\boldsymbol{\theta}] = 0$; and $\mathbb{V}[\mathbf{Z}\boldsymbol{\delta}] = 0.3$.

Note that, for each case, we generate summary statistics in two ways: (*i*) including the top 10 PCs as covariates in the marginal linear model to correct for population structure and (*ii*) not correcting for any population structure. Again all figures show the mean performances (and standard errors) across 100 simulated replicates.

Sparse simulation study design. In our final set of simulations, we consider phenotypes with sparse architectures[20] (Figure S15). Here, traits were simulated with solely additive effects such that $\mathbb{V}[\mathbf{X}\boldsymbol{\beta}] = H^2$, but this time only variants with the top or bottom $\{1, 5, 10, 25, 50, 100\}$ percentile of LD scores were given nonzero coefficients. We once again generate traits with heritability $H^2 = \{0.3, 0.6\}$. We also want to note that, in each of these specific analyses, synthetic trait architectures were generated using all UK Biobank genotyped variants that passed initial preprocessing and quality control (see next section). Since not all of these SNPs are HapMap3 SNPs, some variants were omitted from the `LDSC` and `i-LDSC` regression. Overall, as shown in the main text with results taken over 100 replicates, breaking the assumed relationship between LD scores and chi-squared statistics (i.e., that they are generally positively correlated) led to unbounded estimates of heritability in all but the (more polygenic) scenario when 100% of SNPs contributed to phenotypic variation.

Preprocessing for the UK Biobank and BioBank Japan

In order to apply the `i-LDSC` framework to 25 continuous traits the UK Biobank[48], we first downloaded genotype data for 488,377 individuals in the UK Biobank using the `ukbgene` tool (`https://biobank.ctsu.ox.ac.uk/crystal/download.cgi`) and converted the genotypes using the provided `ukbconv` tool (`https://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=149660`). Phenotype data for the 25 continuous traits were also downloaded for those same individuals using the `ukbgene` tool. Individuals identified by the UK Biobank as having high heterozygosity, excessive relatedness, or aneuploidy were removed (1,550 individuals). After separating individuals into self-identified ancestral cohorts using data field `21000`, unrelated individuals were selected by randomly choosing an individual from each pair of related individuals. This resulted in $N = 349{,}469$ white British individuals to be included in our analysis. We downloaded imputed SNP data from the UK Biobank for all remaining individuals and removed SNPs with an information score below 0.8. Information scores for each SNP are provided by the UK Biobank (`http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=1967`).

Quality control for the remaining genotyped and imputed variants was then performed on each cohort separately using the following steps. All structural variants were first removed, leaving only single nucleotide polymorphisms (SNPs) in the genotype data. Next, all AT/CG SNPs were removed to avoid possible confounding due to sequencing errors. Then, SNPs with minor allele frequency less than 1% were removed using the `PLINK 2.0`[49] command `--maf 0.01`. We then removed all SNPs found to be out of Hardy-Weinberg equilibrium, using the `PLINK` `--hwe 0.000001` flag to remove all SNPs with a Fisher's exact test P-value $> 10^{-6}$. Finally, all SNPs with missingness greater than 1% were removed using the `PLINK` `--mind 0.01` flag.

We then performed a genome-wide association study (GWAS) for each trait in the UK Biobank on the remaining 8,981,412 SNPs. SNP-level GWAS effect sizes were calculated using `PLINK` and the `--glm` flag[49]. Age, sex, and the first twenty principal components were included as covariates for all traits analyzed[47]. Principal component analysis was performed using `FlashPCA 2.0`[50] on a set of independent markers derived separately for each ancestry cohort using the `PLINK` command `--indep-pairwise 100 10 0.1`. Using the parameters `--indep-pairwise` removes all SNPs that have a pairwise correlation above 0.1 within a 100 SNP window, then slides forward in increments of ten SNPs genome-wide.

In order to analyze data from BioBank Japan, we downloaded publicly available GWAS summary statistics for the 25 traits listed in Table S5 from `http://jenger.riken.jp/en/result`. Summary

statistics used age, sex, and the first ten principal components as confounders in the initial GWAS study. We then used individuals from the East Asian (EAS) superpopulation from the 1000 Genomes Project Phase 3 to calculate paired `LDSC` and `i-LDSC` scores from a reference panel. We pruned the reference panel using the `PLINK` command `--indep-pairwise 100 10 0.5` to limit the computational time of calculating scores[49]. This resulted in reference scores for 1,164,666 SNPs that are included on the `i-LDSC` GitHub repository (see URLs). Using summary statistics from BioBank Japan, with scores calculated from the EAS population in the 1000 Genomes, we obtained `i-LDSC` heritability estimates for each of the 25 traits.

Data and software availability

Source code and tutorials for implementing interaction-LD score regression via the `i-LDSC` package are written in Python and are publicly available online at `https://github.com/lcrawlab/i-LDSC`. Files of LD scores, *cis*-interaction LD scores, and GWAS summary statistics used for our analyses of the UK Biobank and BioBank Japan can be downloaded from the Harvard Dataverse (`https://dataverse.harvard.edu/datsset.xhtml?persistentId=doi:10.7910/DVN/W6MA8J&faces-redirect=true`). All software for the traditional and stratified LD score regression framework with `LDSC` and `s-LDSC` were fit using the default settings, unless otherwise stated in the main text. Source code for these approaches was downloaded from `https://github.com/bulik/ldsc`. When applying `s-LDSC`, we used 97 functional annotations from Gazal et al.[34] to estimate heritability. Data from the UK Biobank Resource[48] (`https://www.ukbiobank.ac.uk`) was made available under Application Numbers 14649 and 22419. Data can be accessed by direct application to the UK Biobank.

Figures and Tables

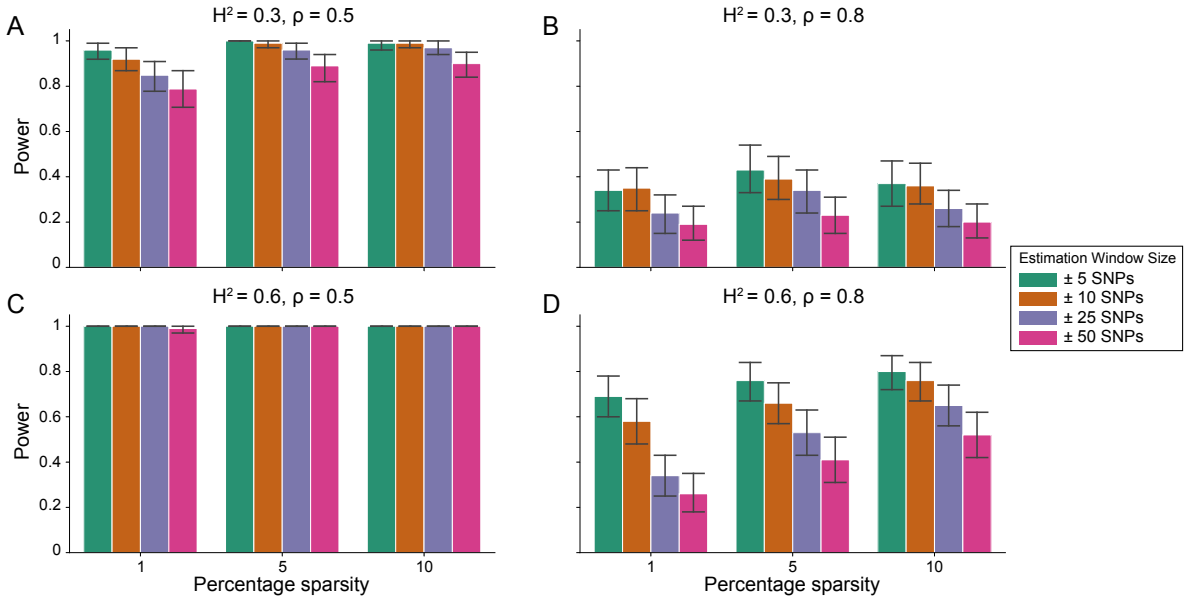

Figure 1. Power of the `i-LDSC` framework to detect tagged non-additive genetic effects on simulated data. Synthetic trait architecture was simulated using real genotype data from individuals of self-identified European ancestry in the UK Biobank. All SNPs were considered to have at least an additive effect (i.e., creating a polygenic trait architecture). Next, we randomly select two groups of interacting variants and divide them into two groups. The group #1 SNPs are chosen to be 1%, 5%, and 10% of the total number of SNPs genome-wide (see the x-axis in each panel). These interact with the group #2 SNPs which are selected to be variants within a ±10 kilobase (kb) window around each SNP in group #1. Coefficients for additive and interaction effects were simulated with no minor allele frequency dependency $\alpha = 0$ (see Materials and Methods). Panels **(A)** and **(B)** are results with simulations using a heritability $H^2 = 0.3$, while panels **(C)** and **(D)** were generated with $H^2 = 0.6$. We also varied the proportion of heritability contributed by additive effects to **(A, C)** $\rho = 0.5$ and **(B, D)** $\rho = 0.8$, respectively. Here, we are blind to the parameter settings used in generative model and run `i-LDSC` while computing the *cis*-interaction LD scores using different estimating windows of ±5 (green), ±10 (orange), ±25 (purple), and ±50 (pink) SNPs. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors. Generally, the performance of `i-LDSC` increases with larger heritability and lower proportions of additive variation. Note that `LDSC` is not shown here because it does not search for tagged interaction effects in summary statistics. Similar plots for a range of α values and generative interacting SNP window sizes are shown in Figures S1-S5.

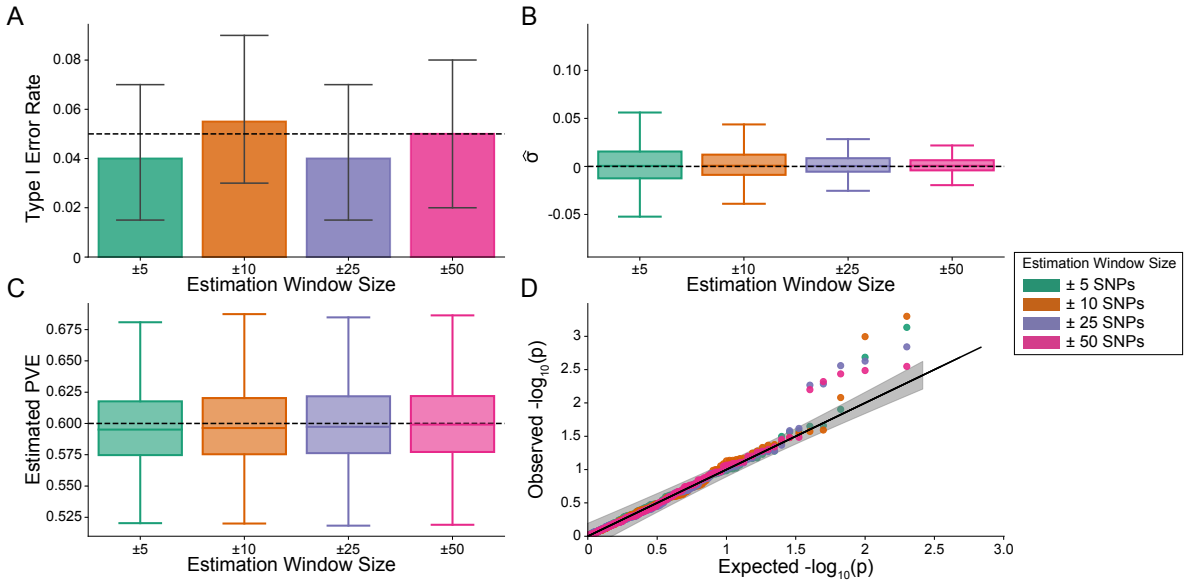

Figure 2. The `i-LDSC` framework is well-calibrated under the null hypothesis and does not identify evidence of tagged non-additive effects when polygenic traits are generated by only additive effects. In these simulations, synthetic trait architecture is made up of only additive genetic variation (i.e., $\rho = 1$). Coefficients for additive and interaction effects were simulated with no minor allele frequency dependency $\alpha = 0$ (see Materials and Methods). Here, we are blind to the parameter settings used in generative model and run `i-LDSC` while computing the *cis*-interaction LD scores using different estimating windows of ±5 (green), ±10 (orange), ±25 (purple), and ±50 (pink) SNPs. **(A)** Mean type I error rate using the `i-LDSC` framework across an array of estimation window sizes for the *cis*-interaction LD scores. This is determined by assessing the P-value of the *cis*-interaction coefficient (σ) in the `i-LDSC` regression model and checking whether $P < 0.05$. **(B)** Estimates of the *cis*-interaction coefficient (σ). Since traits were simulated with only additive effects, these estimates should be centered around zero. **(C)** Estimates of the proportions of phenotypic variance explained (PVE) by genetic effects (i.e., estimated heritability) where the true additive variance is set to $H^2\rho = 0.6$. **(D)** QQ-plot of the P-values for the *cis*-interaction coefficient (σ) in `i-LDSC`. Results are based on 100 simulations per parameter combination and the horizontal bars represent standard errors. Similar plots for a range of α values and generative interacting SNP window sizes are shown in Figures S6-S7.

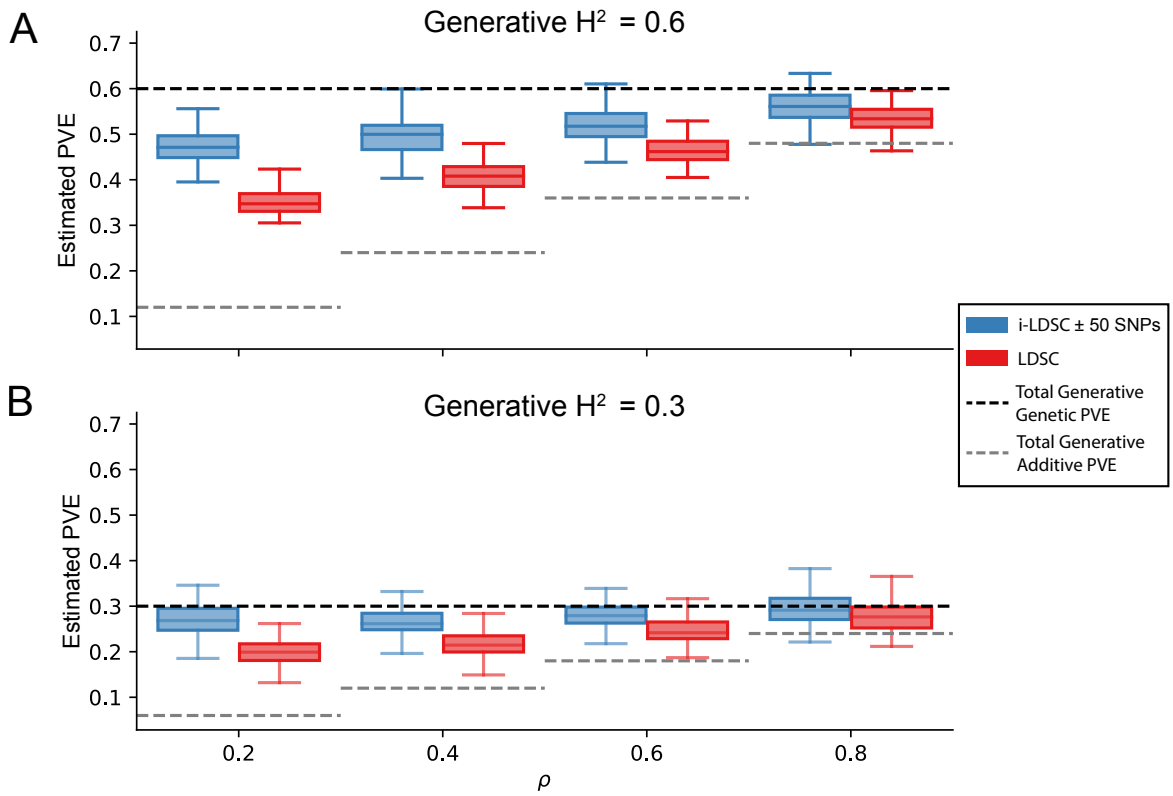

Figure 3. `i-LDSC` robustly and accurately estimates the proportions of phenotypic variance explained (PVE) by genetic effects (i.e., estimated heritability) in simulations in polygenic traits, compared to `LDSC`, due to our accounting for interaction effects tagged in additive GWAS summary statistics. Synthetic trait architecture was simulated using real genotype data from individuals of self-identified European ancestry in the UK Biobank (Materials and Methods). All SNPs were considered to have at least an additive effect (i.e., creating a polygenic trait architecture). Next, we randomly select two groups of interacting variants and divide them into two groups. The group #1 SNPs are chosen to be 10% of the total number of SNPs genome-wide. These interact with the group #2 SNPs which are selected to be variants within a ±100 kilobase (kb) window around each SNP in group #1. Coefficients for additive and interaction effects were simulated with no minor allele frequency dependency $\alpha = 0$ (see Materials and Methods). Here, we assume a heritability **(A)** $H^2 = 0.3$ or **(B)** $H^2 = 0.6$ (marked by the black dotted lines, respectively), and we vary the proportion contributed by additive effects with $\rho = \{0.2, 0.4, 0.6, 0.8\}$. The grey dotted lines represent the total contribution of additive effects in the generative model for the synthetic traits ($H^2\rho$). `i-LDSC` outperforms `LDSC` in recovering heritability across each scenario. Results are based on 100 simulations per parameter combination. `i-LDSC` estimates of heritability partitioned by estimation *cis*-interaction window are shown in Figure S8. The mean absolute error between the true H^2 value and the estimates produced by `i-LDSC` and `LDSC` are shown in Table S1 and S2, respectively.

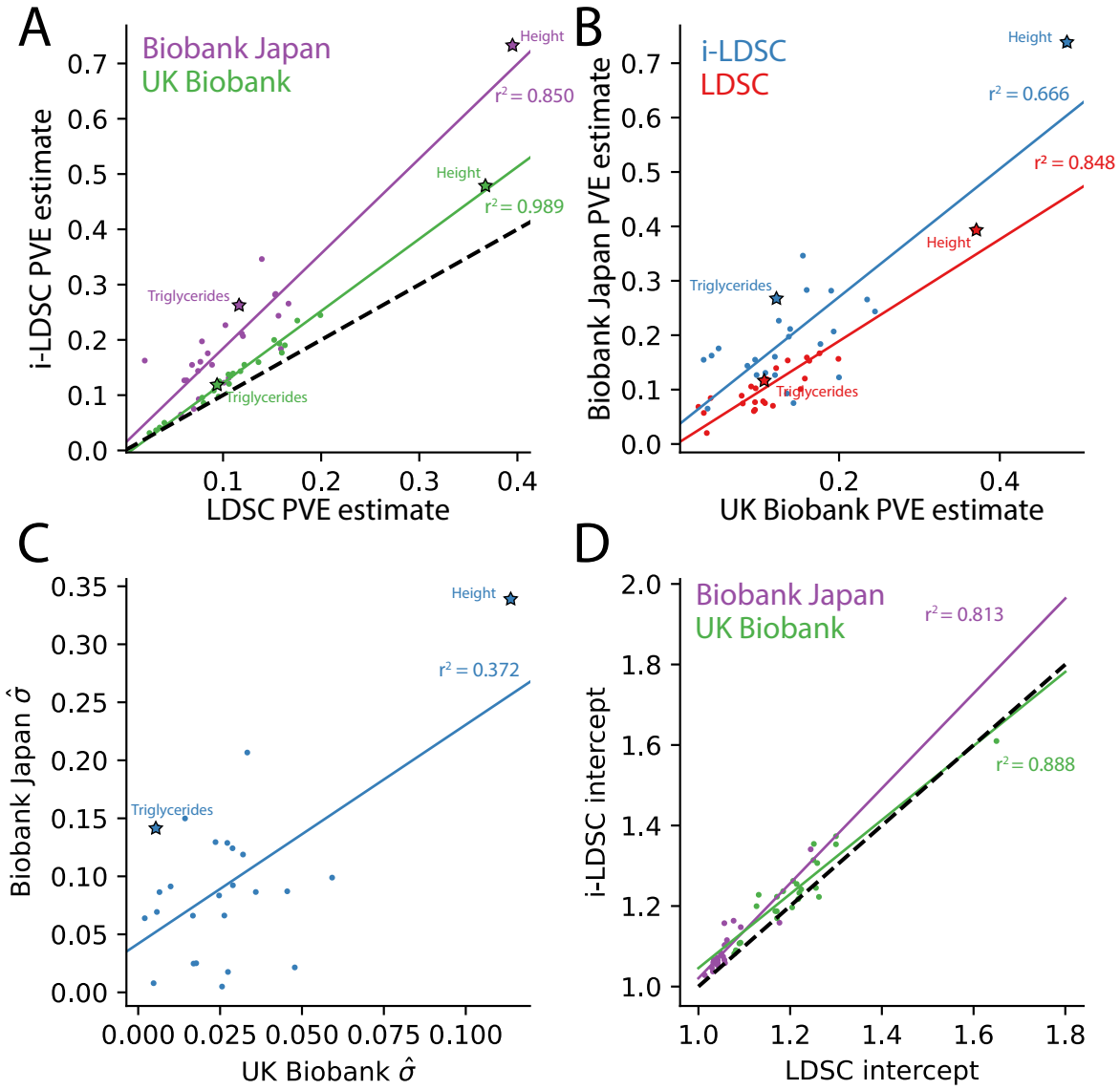

Figure 4. The `i-LDSC` framework recovers heritability and provides estimates of tagged *cis*-interactions in GWAS summary statistics (σ) for 25 quantitiative traits in the UK Biobank and BioBank Japan. (A) In both the UK Biobank (green) and BioBank Japan (purple), estimates of phenotypic variance explained (PVE) by genetic effects from `i-LDSC` and `LDSC` are highly correlated for 25 different complex traits. The Spearman correlation coefficient between heritability estimates from `LDSC` and `i-LDSC` for the UK Biobank and BioBank Japan are $r^2 = 0.989$ and $r^2 = 0.850$, respectively. The $y = x$ dotted line represents the values at which estimates from both approaches are the same. **(B)** PVE estimates from the UK Biobank are better correlated with those from the BioBank Japan across 25 traits using `LDSC` (Spearman $r^2 = 0.848$) than `i-LDSC` (Spearman $r^2 = 0.666$). **(C)** `i-LDSC` estimates of the phenotypic variation explained by tagged non-additive genetic effects using the *cis*-interaction LD score (i.e., estimates of σ) between traits in the UK Biobank and BioBank Japan (Spearman $r^2 = 0.372$). **(D)** Intercept estimates between `i-LDSC` and `LDSC` regression models are highly correlated in the UK Biobank (Spearman $r^2 = 0.888$, slope $= 0.919$) and BioBank Japan (Spearman $r^2 = 0.813$, slope $= 1.179$). When height, an outlier in our UK Biobank analysis is omitted, the slope of the UK Biobank intercept line is closer to that of the Biobank Japan (UKB slope with no outlier $= 1.070$). Note that the heritability estimates displayed in panels **(A)** and **(B)**, and P-values corresponding to panel **(C)**, are given in Table 1.

Trait	UKB (LDSC)	UKB (i-LDSC)	UKB $\hat{\sigma}$	UKB P-value	BBJ (LDSC)	BBJ (i-LDSC)	BBJ $\hat{\sigma}$	BBJ P-value
Basophil	0.0250	0.0315	0.0065	1.572×10^{-12}	0.0684	0.1548	0.0864	0.025
BMI	0.1757	0.2349	0.0592	3.083×10^{-84}	0.1667	0.2656	0.0989	2.438×10^{-18}
Cholesterol	0.0954	0.0974	0.0020	1.821×10^{-16}	0.0629	0.1268	0.0639	2.740×10^{-4}
CRP	0.0354	0.0414	0.0060	9.845×10^{-12}	0.0202	0.1625	0.1423	0.020
DBP	0.0940	0.1203	0.0263	1.118×10^{-65}	0.0605	0.1267	0.0662	1.675×10^{-7}
EGFR	0.1521	0.1999	0.0478	1.187×10^{-46}	0.1010	0.1225	0.0215	4.232×10^{-5}
Eosinophil	0.1055	0.1375	0.0320	1.230×10^{-18}	0.0785	0.1973	0.1188	0.001
HBA1C	0.0906	0.1083	0.0177	1.578×10^{-26}	0.1057	0.1308	0.0251	0.031
HDL*	0.1599	0.1768	0.0169	9.636×10^{-37}	0.1590	0.1838	0.0248	0.081
Height	0.3675	0.4815	0.1140	1.038×10^{-64}	0.2340	0.3941	0.1601	7.433×10^{-33}
Hematocrit	0.1078	0.1352	0.0274	2.479×10^{-25}	0.0752	0.0928	0.0176	3.689×10^{-5}
Hemoglobin	0.1177	0.1433	0.0256	4.284×10^{-27}	0.0702	0.0752	0.0050	9.037×10^{-4}
LDL	0.0802	0.0859	0.0057	5.087×10^{-13}	0.0745	0.1438	0.0693	0.018
Lymphocyte	0.0402	0.0501	0.0099	4.906×10^{-19}	0.0844	0.1757	0.0913	5.479×10^{-5}
MCH	0.1361	0.1597	0.0236	1.785×10^{-25}	0.1536	0.2831	0.1295	1.042×10^{-5}
MCHC	0.0317	0.0364	0.0047	3.730×10^{-12}	0.0571	0.0650	0.0079	0.027
MCV	0.1630	0.1902	0.0272	1.180×10^{-29}	0.1530	0.2818	0.1288	1.042×10^{-5}
Monocyte	0.0788	0.0955	0.0167	5.257×10^{-18}	0.0888	0.1549	0.0661	0.004
Neutrophil	0.1102	0.1391	0.0289	1.777×10^{-33}	0.1191	0.2114	0.0923	5.050×10^{-5}
Platelet	0.1992	0.2447	0.0455	2.303×10^{-37}	0.1565	0.2436	0.0871	7.724×10^{-9}
RBC	0.1574	0.1933	0.0359	3.292×10^{-31}	0.1203	0.2068	0.0865	5.972×10^{-8}
SBP	0.0954	0.1201	0.0247	8.660×10^{-75}	0.0769	0.1604	0.0835	9.075×10^{-10}
Triglycerides*	0.1061	0.1204	0.0143	1.410×10^{-26}	0.1171	0.2670	0.1499	0.110
Urate	0.1217	0.1550	0.0333	9.642×10^{-38}	0.1395	0.3462	0.2067	0.015
WBC	0.0962	0.1250	0.0288	9.866×10^{-34}	0.1024	0.2266	0.1242	1.346×10^{-8}

Table 1. i-LDSC heritability estimates and P-values highlighting statistically significant contributions of tagged non-additive genetic effects for 25 traits in the UK Biobank and BioBank Japan. Here, LDSC heritability estimates are included as a baseline. The difference between the approaches is that the i-LDSC heritability estimates include proportions of phenotypic variation that are explained by tagged non-additive variation (see columns with estimates $\hat{\sigma}$). Note that all 25 traits analyzed in the UK Biobank and 23 of the 25 traits analyzed in BioBank Japan have a statistically significant amount of tagged non-additive genetic effects as detected by the *cis*-interaction LD score ($P < 0.05$). The two traits without significant tagged non-additive genetic effects in BioBank Japan were HDL ($P = 0.081$) and Triglyceride ($P = 0.110$). These traits are indicated by *. The i-LDSC P-values are related to the estimates of the σ coefficients which are also displayed in Figure 4.

References

1. Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Nick Patterson, Mark J Daly, Alkes L Price, and Benjamin M Neale. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature genetics, 47(3):291–295, 2015.

2. Brendan Bulik-Sullivan, Hilary K Finucane, Verneri Anttila, Alexander Gusev, Felix R Day, Po-Ru Loh, Laramie Duncan, John R B Perry, Nick Patterson, Elise B Robinson, Mark J Daly, Alkes L Price, Benjamin M Neale, ReproGen Consortium, Psychiatric Genomics Consortium, and Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3. An atlas of genetic correlations across human diseases and traits. Nature Genetics, 47(11):1236–1241, 2015. doi: 10.1038/ng.3406. URL https://doi.org/10.1038/ng.3406.

3. Huwenbo Shi, Gleb Kichaev, and Bogdan Pasaniuc. Contrasting the genetic architecture of 30 complex traits from summary association data. The American Journal of Human Genetics, 99(1): 139–153, 2016.

4. Noah Zaitlen, Peter Kraft, Nick Patterson, Bogdan Pasaniuc, Gaurav Bhatia, Samuela Pollack, and Alkes L. Price. Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLOS Genetics, 9(5):1–11, 2013. doi: 10.1371/journal.pgen.1003520. URL https://doi.org/10.1371/journal.pgen.1003520.

5. Tinca J C Polderman, Beben Benyamin, Christiaan A de Leeuw, Patrick F Sullivan, Arjen van Bochoven, Peter M Visscher, and Danielle Posthuma. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nature Genetics, 47(7):702–709, 2015. doi: 10.1038/ng.3285. URL https://doi.org/10.1038/ng.3285.

6. Kangcheng Hou, Kathryn S. Burch, Arunabha Majumdar, Huwenbo Shi, Nicholas Mancuso, Yue Wu, Sriram Sankararaman, and Bogdan Pasaniuc. Accurate estimation of snp-heritability from biobank-scale data irrespective of genetic architecture. Nature Genetics, 51(8):1244–1251, 2019. doi: 10.1038/s41588-019-0465-0. URL https://doi.org/10.1038/s41588-019-0465-0.

7. Ali Pazokitoroudi, Yue Wu, Kathryn S. Burch, Kangcheng Hou, Aaron Zhou, Bogdan Pasaniuc, and Sriram Sankararaman. Efficient variance components analysis across millions of genomes.

Nature Communications, 11(1):4020, 2020. doi: 10.1038/s41467-020-17576-9. URL https://doi.org/10.1038/s41467-020-17576-9.

8. Loic Yengo, Julia Sidorenko, Kathryn E Kemper, Zhili Zheng, Andrew R Wood, Michael N Weedon, Timothy M Frayling, Joel Hirschhorn, Jian Yang, Peter M Visscher, et al. Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of european ancestry. Human molecular genetics, 27(20):3641–3649, 2018.

9. Doug Speed and David J Balding. Sumher better estimates the snp heritability of complex traits from summary statistics. Nature genetics, 51(2):277–284, 2019.

10. Shuang Song, Wei Jiang, Yiliang Zhang, Lin Hou, and Hongyu Zhao. Leveraging ld eigenvalue regression to improve the estimation of snp heritability and confounding inflation. The American Journal of Human Genetics, 2022.

11. Duncan S Palmer, Wei Zhou, Liam Abbott, Emilie M Wigdor, Nikolas Baya, Claire Churchhouse, Cotton Seed, Tim Poterba, Daniel King, Masahiro Kanai, et al. Analysis of genetic dominance in the uk biobank. Science, 379(6639):1341–1348, 2023.

12. Tsz Fung Chan, Xinyue Rui, David V. Conti, Myriam Fornage, Mariaelisa Graff, Jeffrey Haessler, Christopher Haiman, Heather M. Highland, Su Yon Jung, Eimear Kenny, Charles Kooperberg, Loic Le Marchland, Kari E. North, Ran Tao, Genevieve Wojcik, Christopher R. Gignoux, PAGE Consortium, Charleston W. K. Chiang, and Nicholas Mancuso. Estimating heritability explained by local ancestry and evaluating stratification bias in admixture mapping from summary statistics. bioRxiv, page 2023.04.10.536252, 2023. doi: 10.1101/2023.04.10.536252. URL http://biorxiv.org/content/early/2023/04/18/2023.04.10.536252.abstract.

13. Zheng Ning, Yudi Pawitan, and Xia Shen. High-definition likelihood inference of genetic correlations across human complex traits. Nature Genetics, 52(8):859–864, 2020. doi: 10.1038/s41588-020-0653-y. URL https://doi.org/10.1038/s41588-020-0653-y.

14. Yiliang Zhang, Qiongshi Lu, Yixuan Ye, Kunling Huang, Wei Liu, Yuchang Wu, Xiaoyuan Zhong, Boyang Li, Zhaolong Yu, Brittany G Travers, Donna M Werling, James J Li, and Hongyu Zhao. Supergnova: local genetic correlation analysis reveals heterogeneous etiologic sharing of complex

traits. Genome Biol, 22(1):262, 2021. ISSN 1474-760X (Electronic); 1474-7596 (Print); 1474-7596 (Linking). doi: 10.1186/s13059-021-02478-w.

15. Roshni A. Patel, Shaila A. Musharoff, Jeffrey P. Spence, Harold Pimentel, Catherine Tcheandjieu, Hakhamanesh Mostafavi, Nasa Sinnott-Armstrong, Shoa L. Clarke, Courtney J. Smith, Peter P. Durda, Kent D. Taylor, Russell Tracy, Yongmei Liu, W. Craig Johnson, Francois Aguet, Kristin G. Ardlie, Stacey Gabriel, Josh Smith, Deborah A. Nickerson, Stephen S. Rich, Jerome I. Rotter, Philip S. Tsao, Themistocles L. Assimes, and Jonathan K. Pritchard. Genetic interactions drive heterogeneity in causal variant effect sizes for gene expression and complex traits. The American Journal of Human Genetics, 2022. doi: 10.1016/j.ajhg.2022.05.014. URL `https://doi.org/10.1016/j.ajhg.2022.05.014`.

16. Evan E. Eichler, Jonathan Flint, Greg Gibson, Augustine Kong, Suzanne M. Leal, Jason H. Moore, and Joseph H. Nadeau. Missing heritability and strategies for finding the underlying causes of complex disease. Nature Reviews Genetics, 11(6):446–450, 2010. doi: 10.1038/nrg2809. URL `https://doi.org/10.1038/nrg2809`.

17. Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick Patterson, Mark J Daly, Alkes L Price, and Benjamin M Neale. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet, 47:291–295, 2015. URL `http://dx.doi.org/10.1038/ng.3211`.

18. Jian Yang, Beben Benyamin, Brian P McEvoy, Scott Gordon, Anjali K Henders, Dale R Nyholt, Pamela A Madden, Andrew C Heath, Nicholas G Martin, and Grant W Montgomery. Common snps explain a large proportion of the heritability for human height. Nat Genet, 42(7):565, 2010.

19. Michael C Wu, Seunggeun Lee, Tianxi Cai, Yun Li, Michael Boehnke, and Xihong Lin. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet, 89(1):82–93, 2011. ISSN 1537-6605 (Electronic); 0002-9297 (Linking). doi: 10.1016/j.ajhg.2011.05.029.

20. Xiang Zhou, Peter Carbonetto, and Matthew Stephens. Polygenic modeling with Bayesian sparse linear mixed models. PLoS Genet, 9(2):e1003264, 2013.

21. Lorin Crawford, Ping Zeng, Sayan Mukherjee, and Xiang Zhou. Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits. PLoS Genet, 13(7): e1006869, 2017. URL https://doi.org/10.1371/journal.pgen.1006869.

22. William G. Hill, Michael E. Goddard, and Peter M. Visscher. Data and theory point to mainly additive genetic variance for complex traits. PLOS Genetics, 4(2):1–10, 2008. doi: 10.1371/journal.pgen.1000008. URL https://doi.org/10.1371/journal.pgen.1000008.

23. Farhad Hormozdiari, Emrah Kostem, Eun Yong Kang, Bogdan Pasaniuc, and Eleazar Eskin. Identifying causal variants at loci with multiple signals of association. Genetics, 198(2):497–508, 2014. doi: 10.1534/genetics.114.167908. URL https://pubmed.ncbi.nlm.nih.gov/25104515.

24. Priyanka Nakka, Benjamin J Raphael, and Sohini Ramachandran. Gene and Network Analysis of Common Variants Reveals Novel Associations in Multiple Complex Diseases. Genetics, 204(2): 783–798, 10 2016. ISSN 1943-2631. doi: 10.1534/genetics.116.188391. URL https://doi.org/10.1534/genetics.116.188391.

25. Xiang Zhu and Matthew Stephens. Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. Ann Appl Stat, 11(3):1561–1592, 2017. doi: 10.1214/17-AOAS1046. URL https://projecteuclid.org:443/euclid.aoas/1507168840.

26. Yan Zhang, Guanghao Qi, Ju-Hyun Park, and Nilanjan Chatterjee. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat Genet, 50(9):1318–1326, 2018.

27. Xiang Zhu and Matthew Stephens. Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. Nat Comm, 9(1):4361, 2018.

28. Wei Cheng, Sohini Ramachandran, and Lorin Crawford. Estimation of non-null snp effect size distributions enables the detection of enriched genes underlying complex traits. PLoS Genet, 16 (6):e1008855, 2020. URL https://doi.org/10.1371/journal.pgen.1008855.

29. Pinar Demetci, Wei Cheng, Gregory Darnell, Xiang Zhou, Sohini Ramachandran, and Lorin Crawford. Multi-scale inference of genetic trait architecture using biologically annotated neu-

ral networks. PLOS Genetics, 17(8):1–53, 2021. doi: 10.1371/journal.pgen.1009754. URL https://doi.org/10.1371/journal.pgen.1009754.

30. Yongtao Guan and Matthew Stephens. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. Ann Appl Stat, 5(3):1780–1815, 2011. doi: 10.1214/11-AOAS455. URL https://projecteuclid.org:443/euclid.aoas/1318514285.

31. Ronald Aylmer Fisher. The genetical theory of natural selection: a complete variorum edition. Oxford University Press, 1999. ISBN 0198504403.

32. Armin P Schoech, Daniel M Jordan, Po-Ru Loh, Steven Gazal, Luke J O'Connor, Daniel J Balick, Pier F Palamara, Hilary K Finucane, Shamil R Sunyaev, and Alkes L Price. Quantification of frequency-dependent genetic architectures in 25 uk biobank traits reveals action of negative selection. Nature communications, 10(1):1–10, 2019.

33. Yabo Li, Hyosuk Cho, Fan Wang, Oriol Canela-Xandri, Chunyan Luo, Konrad Rawlik, Stephen Archacki, Chengqi Xu, Albert Tenesa, Qiuyun Chen, et al. Statistical and functional studies identify epistasis of cardiovascular risk genomic variants from genome-wide association studies. Journal of the American Heart Association, 9(7):e014146, 2020.

34. Steven Gazal, Hilary K Finucane, Nicholas A Furlotte, Po-Ru Loh, Pier Francesco Palamara, Xuanyao Liu, Armin Schoech, Brendan Bulik-Sullivan, Benjamin M Neale, Alexander Gusev, et al. Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. Nature genetics, 49(10):1421–1427, 2017.

35. Hilary K Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, Stephan Ripke, Felix R Day, Shaun Purcell, Eli Stahl, Sara Lindstrom, John R B Perry, Yukinori Okada, Soumya Raychaudhuri, Mark J Daly, Nick Patterson, Benjamin M Neale, and Alkes L Price. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet, 47(11): 1228–1235, 2015. ISSN 1546-1718 (Electronic); 1061-4036 (Print); 1061-4036 (Linking). doi: 10.1038/ng.3404.

36. Daniel Runcie, Hao Cheng, and Lorin Crawford. Mega-scale linear mixed models for genomic predictions with thousands of traits. bioRxiv, page 2020.05.26.116814, 2020. doi:

887 10.1101/2020.05.26.116814. URL http://biorxiv.org/content/early/2020/05/29/2020.05.
888 26.116814.abstract.

889 37. Julian Stamp, Alan DenAdel, Daniel Weinreich, and Lorin Crawford. Leveraging the genetic
890 correlation between traits improves the detection of epistasis in genome-wide association studies.
891 bioRxiv, page 2022.11.30.518547, 2022. doi: 10.1101/2022.11.30.518547. URL http://biorxiv.
892 org/content/early/2022/12/01/2022.11.30.518547.abstract.

893 38. Sahin Naqvi, Yoeri Sleyp, Hanne Hoskens, Karlijne Indencleef, Jeffrey P. Spence, Rose Bruffaerts,
894 Ahmed Radwan, Ryan J. Eller, Stephen Richmond, Mark D. Shriver, John R. Shaffer, Seth M.
895 Weinberg, Susan Walsh, James Thompson, Jonathan K. Pritchard, Stefan Sunaert, Hilde Peeters,
896 Joanna Wysocka, and Peter Claes. Shared heritability of human face and brain shape. Nature
897 Genetics, 53(6):830–839, 2021. doi: 10.1038/s41588-021-00827-w. URL https://doi.org/10.
898 1038/s41588-021-00827-w.

899 39. Yong Jiang and Jochen C. Reif. Modeling epistasis in genomic selection. Genetics, 201:759–768,
900 2015.

901 40. Farhad Hormozdiari, Martijn van de Bunt, Ayellet V. Segrè, Xiao Li, Jong Wha J. Joo, Michael
902 Bilow, Jae Hoon Sul, Sriram Sankararaman, Bogdan Pasaniuc, and Eleazar Eskin. Colocalization
903 of GWAS and eQTL signals detects target genes. Am J Hum Genet, 99(6):1245–1260, 2016. doi:
904 10.1016/j.ajhg.2016.10.003. URL https://doi.org/10.1016/j.ajhg.2016.10.003.

905 41. L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency
906 distribution in any number of variables. Biometrika, 12(1-2):134–139, 1918. ISSN 0006-3444. doi:
907 10.1093/biomet/12.1-2.134. URL https://doi.org/10.1093/biomet/12.1-2.134.

908 42. Peter Carbonetto and Matthew Stephens. Scalable variational inference for bayesian variable
909 selection in regression, and its accuracy in genetic association studies. Bayesian Anal, 7(1):73–108,
910 2012.

911 43. Jennifer A. Hoeting, David Madigan, Adrian E. Raftery, and Chris T. Volinsky. Bayesian model
912 averaging: a tutorial (with comments by m. clyde, david draper and e. i. george, and a rejoinder
913 by the authors. Statist Sci, 14(4):382–417, 1999. doi: 10.1214/ss/1009212519. URL https:
914 //projecteuclid.org:443/euclid.ss/1009212519.

915 44. Shadi Zabad, Aaron P Ragsdale, Rosie Sun, Yue Li, and Simon Gravel. Assumptions about
916 frequency-dependent architectures of complex traits bias measures of functional enrichment.
917 Genetic epidemiology, 45(6):621–632, 2021.

918 45. Jian Yang, S Hong Lee, Michael E Goddard, and Peter M Visscher. Gcta: a tool for genome-
919 wide complex trait analysis. Am J Hum Genet, 88(1):76–82, 2011. ISSN 1537-6605 (Electronic);
920 0002-9297 (Print); 0002-9297 (Linking). doi: 10.1016/j.ajhg.2010.11.011.

921 46. Carrie Zhu, Matthew J. Ming, Jared M. Cole, Michael D. Edge, Mark Kirkpatrick, and Arbel
922 Harpak. Amplification is the primary mode of gene-by-sex interaction in complex human traits.
923 Cell Genomics, 3(100297), 2023. doi: 10.1016/j.xgen.2023.100297. URL `https://doi.org/10.`
924 `1016/j.xgen.2023.100297`.

925 47. Mashaal Sohail, Robert M Maier, Andrea Ganna, Alex Bloemendal, Alicia R Martin, Michael C
926 Turchin, Charleston WK Chiang, Joel Hirschhorn, Mark J Daly, Nick Patterson, et al. Polygenic
927 adaptation on height is overestimated due to uncorrected stratification in genome-wide association
928 studies. Elife, 8:e39702, 2019.

929 48. Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp,
930 Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O'Connell, Adrian Cortes, Samantha
931 Welsh, Alan Young, Mark Effingham, Gil McVean, Stephen Leslie, Naomi Allen, Peter Donnelly,
932 and Jonathan Marchini. The uk biobank resource with deep phenotyping and genomic data.
933 Nature, 562(7726):203–209, 2018. doi: 10.1038/s41586-018-0579-z. URL `https://doi.org/10.`
934 `1038/s41586-018-0579-z`.

935 49. Christopher C Chang, Carson C Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M
936 Purcell, and James J Lee. Second-generation plink: rising to the challenge of larger and richer
937 datasets. Gigascience, 4(1):s13742–015, 2015.

938 50. Gad Abraham, Yixuan Qiu, and Michael Inouye. Flashpca2: principal component analysis of
939 biobank-scale genotype datasets. Bioinformatics, 2017.

940 51. Masahiro Kanai, Masato Akiyama, Atsushi Takahashi, Nana Matoba, Yukihide Momozawa,
941 Masashi Ikeda, Nakao Iwata, Shiro Ikegawa, Makoto Hirata, Koichi Matsuda, et al. Genetic anal-

ysis of quantitative traits in the japanese population links cell types to complex human diseases. Nature genetics, 50(3):390–400, 2018.

52. Masato Akiyama, Yukinori Okada, Masahiro Kanai, Atsushi Takahashi, Yukihide Momozawa, Masashi Ikeda, Nakao Iwata, Shiro Ikegawa, Makoto Hirata, Koichi Matsuda, et al. Genome-wide association study identifies 112 new loci for body mass index in the japanese population. Nature genetics, 49(10):1458, 2017.

53. Masato Akiyama, Kazuyoshi Ishigaki, Saori Sakaue, Yukihide Momozawa, Momoko Horikoshi, Makoto Hirata, Koichi Matsuda, Shiro Ikegawa, Atsushi Takahashi, Masahiro Kanai, et al. Characterizing rare and low-frequency height-associated variants in the japanese population. Nature communications, 10(1):1–11, 2019.