

# Leveraging LD eigenvalue regression to improve the estimation of SNP heritability and confounding inflation

## Authors

Shuang Song, Wei Jiang, Yiliang Zhang, Lin Hou,  
Hongyu Zhao

## Correspondence

[hongyu.zhao@yale.edu](mailto:hongyu.zhao@yale.edu)



# Leveraging LD eigenvalue regression to improve the estimation of SNP heritability and confounding inflation

Shuang Song,<sup>1,4</sup> Wei Jiang,<sup>2,4</sup> Yiliang Zhang,<sup>2</sup> Lin Hou,<sup>1,3</sup> and Hongyu Zhao<sup>2,\*</sup>

## Summary

Heritability is a fundamental concept in genetic studies, measuring the genetic contribution to complex traits and bringing insights about disease mechanisms. The advance of high-throughput technologies has provided many resources for heritability estimation. Linkage disequilibrium (LD) score regression (LDSC) estimates both heritability and confounding biases, such as cryptic relatedness and population stratification, among single-nucleotide polymorphisms (SNPs) by using only summary statistics released from genome-wide association studies. However, only partial information in the LD matrix is utilized in LDSC, leading to loss in precision. In this study, we propose LD eigenvalue regression (LDER), an extension of LDSC, by making full use of the LD information. Compared to state-of-the-art heritability estimating methods, LDER provides more accurate estimates of SNP heritability and better distinguishes the inflation caused by polygenicity and confounding effects. We demonstrate the advantages of LDER both theoretically and with extensive simulations. We applied LDER to 814 complex traits from UK Biobank, and LDER identified 363 significantly heritable phenotypes, among which 97 were not identified by LDSC.

## Introduction

In genetic studies, heritability is a fundamental quantity measuring the phenotypic variance explained by genetic components.<sup>1,2</sup> Heritability provides an upper bound of genetic risk-prediction performance<sup>3</sup> and acts as a summarizing metric indicating the genetic architecture of complex traits.<sup>4</sup> Accurate estimates of heritabilities help us better understand the degree to which the phenotype is influenced by measured genetic variants and thus provide insights about the genetic mechanisms of complex diseases.<sup>5</sup> We can distinguish the heritability concept into the broad-sense heritability ( $H^2$ ) and the narrow-sense heritability ( $h^2$ ). The former accounts for the genetic variance explained by all genetic factors, including additive effects, dominance effects, and epistatic effects, whereas the latter chip-based heritability evaluates only the additive genetic effects and is our main focus in this study.

In the pre-genomic era, heritability estimation was mainly based on family/pedigree data with the linear mixed model (LMM).<sup>6</sup> With the advance of high-throughput technologies, such as microarray/sequencing, we can accurately measure genotypes on millions of single-nucleotide polymorphisms (SNPs) for individuals with moderate cost. The advances have enabled thousands of genome-wide association studies (GWASs) exploring the genetic basis of various diseases, providing many resources for estimating heritability captured among SNPs. In recent years, a number of methods have been proposed to estimate heritability based on GWAS data, such as GCTA (genome-wide complex trait

analysis),<sup>7,8</sup> BOLT-LMM (efficient large cohorts genome-wide Bayesian mixed-model association testing),<sup>9</sup> FaST-LMM (factored spectrally transformed linear mixed models),<sup>10</sup> and LDAK (LD-adjusted kinships).<sup>11,12</sup> These methods are based on individual-level genotype data and provide accurate estimates of heritability captured among SNPs, namely SNP heritability. However, the sample sizes of GWASs continue to grow, and individual-level-data-based methods are not as scalable to biobank projects that assay hundreds of thousands of individuals (e.g., UK Biobank, UKBB<sup>13</sup>) compared with summary-statistics-based methods.<sup>14</sup> Due to computational and privacy issues, the summary-statistics-based methods become more attractive, as they require only publicly available GWAS summary statistics and the computational efficiency will not be influenced by the increasing sample sizes. The most widely used tool to estimate heritability based on summary statistics is linkage disequilibrium (LD) score regression (LDSC),<sup>15</sup> which leverages LD between SNPs to estimate heritability.<sup>16–18</sup> Extended from LDSC, SumHer relaxes the polygenic assumption and allows users to specify the heritability model.<sup>19</sup> Despite the benefits of summary-statistics-based methods, such as decent data availability and increasing sample sizes, individual-level-data-based methods have more accurate estimates, as they are provided with more information from individual data. In this study, we focus on the summary-statistics-based methods and use the estimates of individual-level-data-based methods as the gold standard for the evaluation of different summary-statistics-based methods in real data analysis.

<sup>1</sup>Center for Statistical Science, Department of Industrial Engineering, Tsinghua University, Beijing 100084, China; <sup>2</sup>Department of Biostatistics, Yale School of Public Health, New Haven, CT 06510, USA; <sup>3</sup>MOE Key Laboratory of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing 100084, China

<sup>4</sup>These authors contributed equally

\*Correspondence: [hongyu.zhao@yale.edu](mailto:hongyu.zhao@yale.edu)

<https://doi.org/10.1016/j.ajhg.2022.03.013>

© 2022 American Society of Human Genetics.



One challenge in analyzing GWAS summary statistics is to address the inflation attributed to confounding effects, such as cryptic relatedness and population stratification. The unadjusted confounding biases can generate spurious signals and lead to false positives in association mapping and upward bias of heritability.<sup>20,21</sup> In practice, genomic control is the most widely used method to address the inflation.<sup>22–25</sup> The intuition behind this method is that other than a small number of SNPs associated with the trait or disease, the test statistics for SNPs should follow the distribution under the null hypothesis. More specifically, the median  $\chi^2$  value over all SNPs should be around 0.455 under the null hypothesis. Inflation due to artificial confounding biases can be detected and corrected by comparing the median of test statistics with 0.455. Despite the simplicity of this approach, the rationale of genomic inflation factors relies on the assumption of sparsity, which is very likely violated due to the polygenic property of many diseases/traits.<sup>26,27</sup> In fact, polygenicity from underlying genetic architectures can also yield inflation in test statistics. Therefore, we need to distinguish the inflation attributed to polygenicity and confounding effects from the corrected inflation caused by the latter.

To date, LDSC is the most commonly used method to address the estimation of both heritability and confounding inflation. LDSC builds the model on the relationship between LD scores and the variance of the test statistics. The basic idea is that the SNPs in higher LD with other SNPs tend to have larger test statistics on average for a polygenic trait, because of more causal variants being tagged. Though LDSC is widely used, compared to individual-level-data-based methods, substantially larger standard errors (SEs), which influence accuracy and precision in the estimation of both heritability and confounding inflation factors, are observed.<sup>28</sup> One reason for the loss of precision is that LDSC utilizes the information of only a small part of the LD matrix.<sup>28,29</sup> To be more precise, only the diagonal elements of the squared LD matrix are utilized in LDSC. Some recent methods based on summary statistics utilize more complete LD information, such as HESS (heritability estimation from summary statistics), which estimates local SNP heritability,<sup>14,26</sup> and HDL (high-definition likelihood), which focuses on genetic correlation.<sup>28</sup>

In this paper, we introduce LD eigenvalue regression (LDER), which extends the LDSC method and provides more accurate estimates of heritability and confounding inflation. The key difference between LDER and LDSC is that LDER makes full use of the information from the LD matrix. The regression equation of LDER is similar to LDSC, while LDER uses eigen-decomposition to diagonalize the LD matrix and aggregates the information onto the diagonal of the transformed LD matrix. Our method needs only GWAS summary statistics and consequently can be applied to large-scale datasets with high computational efficiency. Through simulations, we compared LDER with state-of-the-art estimation methods including LDSC, HESS, and HDL and found that LDER could provide more accurate

estimates than all other methods, especially under small-sample-size scenarios. In addition, LDER achieves higher precision than LDSC. Real data applications on 10 common traits from UKBB also showed that the estimates of LDER are closer to the estimates of the individual-level-data-based method BOLT-LMM than other methods. We also applied LDER to 814 phenotypes from UKBB, including 221 continuous and 593 dichotomous traits,<sup>13</sup> among which LDER identified 97 heritable phenotypes that were not significantly heritable in LDSC estimation.

## Material and methods

### LDER modeling

For a GWAS with  $n$  individuals and  $m$  SNPs, we first derive the marginal test statistics for each SNP based on linear regression for the quantitative traits and logistic regression for the dichotomous traits. We name the test statistic a  $Z$  score if its distribution is standard normal distribution under the null hypothesis. If the null distribution does not follow the standard normal distribution, the inverse cumulative distribution method (CDF) may be used for  $p$  values to obtain  $Z$  scores. We denote the underlying risk effects for all SNPs as  $\beta$ . The distribution of  $\mathbf{z}$  conditioning on  $\beta$  follows  $N(\sqrt{n}\mathbf{R}\beta, \lambda\mathbf{R})$ , where  $\mathbf{R}$  is the LD matrix and  $\lambda$  is the inflation factor due to uncontrolled confounding effects.<sup>30</sup> Let  $h_g^2$  denote the SNP heritability, under the polygenic model,<sup>7</sup>  $\beta \sim N(0, h_g^2/m)$ . Then we have

$$\begin{aligned} \mathbb{E}(\mathbf{z}\mathbf{z}^T) &= \mathbb{E}[\mathbb{E}(\mathbf{z}\mathbf{z}^T|\beta)] = \mathbb{E}[\text{Cov}(\mathbf{z}|\beta) + \mathbb{E}(\mathbf{z}|\beta)\mathbb{E}(\mathbf{z}|\beta)^T] \\ &= \mathbb{E}(\lambda\mathbf{R} + n\mathbf{R}\beta\beta^T\mathbf{R}) = \lambda\mathbf{R} + \frac{nh_g^2}{m}\mathbf{R}^2. \end{aligned} \quad (\text{Equation 1})$$

We denote the LD score for the  $i$ -th SNP as  $l_i$ , which is defined as  $l_i = \sum_{j=1}^m R_{ij}^2$ . By considering the diagonal elements of Equation 1, we have  $\mathbb{E}(z_i^2) = \lambda + nh_g^2 l_i/m$ , which is exactly the main equation used in the LDSC framework.<sup>15</sup> However, this construction considers only the diagonal elements and leads to information loss. Therefore, we extend this model by leveraging the full information in Equation 1. We consider the eigen-decomposition  $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ , where  $\mathbf{U}$  and  $\mathbf{D}$  are orthogonal and diagonal matrices, respectively. By defining  $\tilde{\mathbf{z}} = \mathbf{D}^{-0.5}\mathbf{U}^T\mathbf{z}$ , we obtain:

$$\mathbb{E}(\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T) = \mathbb{E}[\mathbb{E}(\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T|\beta)] = \lambda\mathbf{I} + \frac{nh_g^2}{m}\mathbf{D}. \quad (\text{Equation 2})$$

We note that the right-hand side of Equation 2 becomes a diagonal matrix, and thus for the  $i$ -th variant we have:

$$\mathbb{E}(\tilde{z}_i^2) = \lambda + \frac{nh_g^2}{m}D_{ii}. \quad (\text{Equation 3})$$

The formulation above has the same format to the regression equation of LDSC. Instead of regressing the square of  $Z$  scores on LD scores in LDSC, we regress the square of projected  $Z$  scores on LD eigenvalues. All information of the LD matrix in Equation 1 has been aggregated onto diagonal elements of Equation 2, and then used for the regression. We obtain the estimates of  $h_g^2$  and  $\lambda$  by regressing  $\tilde{z}_i^2$  on the eigenvalues  $D_{ii}$ . We further adopt a two-stage procedure in the estimation of heritability to reduce the variance (supplemental method 3.1). For both LDER and LDSC, we use a delete-block-jackknife procedure for estimating SEs.

For the dichotomous traits, we transform the observed-scale heritability to the liability-scale heritability, with  $h_{liability}^2 = h_{observed}^2 \frac{K^2(1-K)^2}{\varphi(\Phi^{-1}(K))^2 P(1-P)}$ , where  $K$  is the frequency of the dichotomous trait in the population and  $P$  is the frequency of the dichotomous trait in the observed sample.<sup>31</sup> The first denominator component  $\varphi(\Phi^{-1}(K))^2$  is the squared probability density function evaluated at the  $K$ -percentile of the inverse CDF of the standard normal distribution.

### Regression weights

Similar to LDSC, we account for heteroskedasticity in the regression weight. We weight by

$$1 / \left(1 + nh_s^2 D_{ii} / m\right)^2, \quad (\text{Equation 4})$$

the reciprocal of which is proportional to the variance of  $\tilde{z}_i^2$  (supplemental method 3.2). To further ensure the robustness of the estimation, we impose a shrinkage weight,  $\min(D_{ii}, 1)$ , to equations with small eigenvalues resulting from the rank-deficient LD matrix. Combining with Equation 4, the weights  $w_i$  are constructed by:

$$w_i = \frac{\min(D_{ii}, 1)}{\left(1 + nh_s^2 D_{ii} / m\right)^2} \propto \frac{\min(D_{ii}, 1)}{\text{Var}(\tilde{z}_i^2 | \beta)}. \quad (\text{Equation 5})$$

### Simulation settings

We first conducted simulations based on generated GWAS summary statistics under varying genetic architectures. We fixed the number of genetic variants to 100,000. The simulated genetic architectures varied in three aspects: the sparsity of the causal SNPs, the heritability, and the confounding inflation. The effect sizes were simulated with a spike-and-slab distribution:  $\beta_j \sim (1 - \alpha)\delta_0 + \alpha N\left(0, \frac{h_s^2}{am}\right)$ . The proportion of causal SNPs  $\alpha$  varied between 0.005, 0.01, 0.05, and 0.1, and the heritability  $h_s^2$  varied between 0.05, 0.1, 0.2, and 0.5. The inflation factor  $\lambda$  was set to 1 (with no confounding inflation) or 1.1 (with confounding inflation). Conditioning on the effect size  $\beta$ , the  $Z$  scores were simulated from  $\mathbf{z} | \beta \sim N(\sqrt{n} \mathbf{R} \beta, \lambda \mathbf{R})$ . We generated LD matrices  $\mathbf{R}$  with block-wise autoregressive LD structures using the R package CorBin,<sup>32</sup> where the correlation matrix  $\mathbf{R}_l$  of the  $l$ -th LD block is

$$\mathbf{R}_l = \begin{pmatrix} 1 & \rho_l & \dots & \rho_l^{m_l-1} \\ \rho_l & 1 & \dots & \rho_l^{m_l-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_l^{m_l-1} & \rho_l^{m_l-2} & \dots & 1 \end{pmatrix}, \quad (\text{Equation 6})$$

where  $m_l$  is the number of SNPs in the  $l$ -th LD block, and  $\rho_l \sim \text{Unif}[0.1, 0.9]$ . The correlations are higher for adjacent variants and decrease with the increase of the distance between the variants in each LD block, which mimic the real LD structures.<sup>33</sup> We equally divided simulated genomes into 1,000 LD blocks, which is in the same scale with the block number of the human genome partitioned by LDetect,<sup>34</sup> and the SNP number in each block was 100.

For the simulations based on real genotypes, we used genotypes from 276,050 independent UKBB European samples and extracted the HapMap 3 SNPs. We simulated effect sizes from the spike-and-

slab distribution and fixed the heritability at 0.5. The sparsity of causal SNPs varied between 0.005 and 0.01. Then we generated continuous phenotypes with the additive model  $\mathbf{y} = \mathbf{X}\beta + \epsilon$ , where  $\mathbf{X}$  is the standardized genotype and the error term  $\epsilon$  satisfies the i.i.d. normal distribution, i.e.,  $\text{Var}(\epsilon) = (1 - h_s^2) \mathbf{I}_n$ . The summary statistics were computed using PLINK software.<sup>35</sup>

### Reference LD matrix construction

The in-sample LD matrix was estimated from 276,050 independent UKBB European samples. The external LD matrix was estimated from European samples of the 1000 Genomes Project (1000G) reference panel. There are 489 individuals and 9,997,231 SNPs in the 1000G after quality control. The SNPs from the 1000G dataset overlapped with the SNPs in the HapMap 3 dataset, and the GWAS summary data were included in the reference panel. In both cases, we employed a linear shrinkage method for the LD matrix estimation.<sup>36,37</sup> We partitioned the genome into 1,703 independent genomic blocks using LDetect,<sup>34</sup> based on the 1000G reference panel with European ancestry. The LD estimation, shrinkage, and eigen-decomposition were performed within each LD block.

### UKBB GWAS summary statistics

The UKBB GWAS summary statistics were from the second round of results released in August 2018 by Neale's group. They performed association studies on 361,194 individuals of White British ancestry and included with covariates including 20 principal components, and age, the square of age ( $\text{age}^2$ ), sex, the product of age and sex ( $\text{age} \times \text{sex}$ ), and  $\text{age}^2 \times \text{sex}$ .

### Quality control of the UKBB genotype data

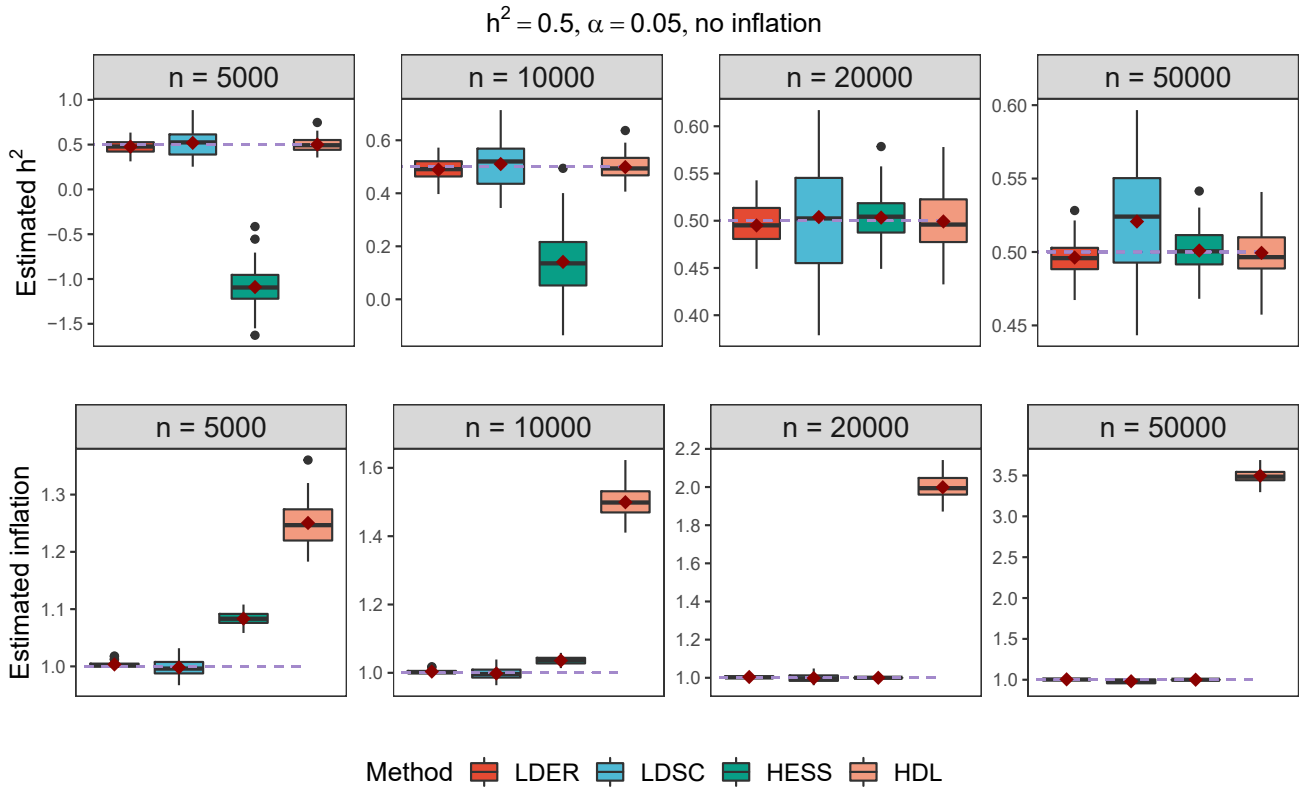
We used phase 3 genotype data released by UKBB wherein the participants underwent genotyping with one of two closely related Affymetrix microarrays (UK BiLEVE Axiom Array or UKBB Axiom Array) for 820,000 variants. Additional genotypes were imputed centrally using the 1000G and Haplotype Reference Consortium (HRC) reference panels, yielding 93 million variants for each individual. We restricted the analysis to 404,892 autosomal variants also presented in the HapMap 3 dataset with a genotype-missing rate per marker  $< 0.01$ , imputation quality score  $> 0.3$ , Hardy-Weinberg  $p$  value  $> 1e - 05$ , and minor allele frequency  $> 0.05$ .

### Compared methods

We compared LDER with three state-of-the-art, summary-statistics-based heritability estimation methods, including LDSC,<sup>38</sup> HESS,<sup>26</sup> and HDL.<sup>28</sup> For LDER, LDSC, and HESS, we computed LD information from both UKBB genotypes and the 1000G reference panel. For HDL, we directly downloaded the pre-computed UKBB LD information of 336,000 British individuals in their website. By default, we let HDL automatically select the number of eigenvalues and eigenvectors used in the estimation.

### Computational time

We compared the CPU time of LDER, LDSC, HESS, and HDL on the analysis of UKBB dataset. The LD preparation time was based on a subset of 10,000 individuals and 404,892 autosomal variants. The computation was performed with an Intel Xeon processor with 2.50 GHz and 48 cores. Among the four methods, LDER and HESS were performed with all 48 cores run in parallel. LDSC and HDL did not provide parallel computing capacity; thus we split



**Figure 1. Comparisons between LDER, LDSC, HESS, and HDL on the estimation of heritability and confounding inflation based on simulated GWAS summary statistics with varying sample sizes**

The number of SNPs was fixed at 100,000. The proportion of causal SNPs was 5%. The effect sizes were sampled from a spike-and-slab distribution with heritability 0.5 and no confounding effects. The simulations were repeated 50 times. Dashed lines represent the true value. Diamonds indicate means in boxplots. The colors of the boxes differentiate the estimation methods.

the data into 22 chromosomes and ran the software on each chromosome in parallel.

## Results

### Method overview

Suppose there are  $n$  individuals and  $m$  SNPs in the GWAS data. For each SNP, we derive the marginal test statistics with linear regression for the quantitative traits and logistic regression for the dichotomous traits. The test statistic is referred to as a  $Z$  score if its distribution is standard normal distribution under the null hypothesis. Otherwise, we use the inverse function of CDF of standard normal distribution to get  $Z$  scores from  $p$  values. We assume the underlying risk effect for the  $m$  SNPs is  $\beta$ . In GWAS, we have  $\mathbf{z}|\beta \sim N(\sqrt{n}\mathbf{R}\beta, \lambda\mathbf{R})$ , where  $\mathbf{R}$  is the LD matrix and  $\lambda$  is the inflation factor due to uncontrolled confounding effects.<sup>30</sup> A larger  $\lambda$  indicates a greater inflation, whereas  $\lambda = 1$  indicates inflation free. We denote the SNP heritability as  $h_g^2$ . Based on the polygenic model,<sup>7</sup> we have  $\beta \sim N(0, h_g^2/m)$  and

$$\mathbb{E}(\mathbf{z}\mathbf{z}^T) = \mathbb{E}[\mathbb{E}(\mathbf{z}\mathbf{z}^T|\beta)] = \lambda\mathbf{R} + \frac{nh_g^2}{m}\mathbf{R}^2. \quad (\text{Equation 7})$$

Equation 7 reduces to LDSC if we only consider the diagonal elements of the matrices on both sides of the

equation. However, utilizing only diagonal elements leads to information loss. Therefore, we first performed eigen-decomposition on the LD matrix, i.e.,  $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ , where  $\mathbf{U}$  is an orthogonal matrix of eigenvectors and  $\mathbf{D}$  is the diagonal eigenvalue matrix. By rotating the GWAS  $Z$  scores with  $\tilde{\mathbf{z}} = \mathbf{D}^{-0.5}\mathbf{U}^T\mathbf{z}$ , we obtain

$$\mathbb{E}(\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T) = \mathbb{E}[\mathbb{E}(\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T|\beta)] = \lambda\mathbf{I} + \frac{nh_g^2}{m}\mathbf{D}. \quad (\text{Equation 8})$$

This transformation aggregates all the information related to  $h_g^2$  to the diagonal elements of Equation 8. For the  $i$ -th element in vector  $\tilde{\mathbf{z}}$ , we have

$$\mathbb{E}(z_i^2) = \lambda + \frac{nh_g^2}{m}D_{ii}, \quad (\text{Equation 9})$$

where  $D_{ii}$  is the  $i$ -th diagonal element of  $\mathbf{D}$ . Despite the similar formulation with LDSC, this expression uses full information of  $h_g^2$  from the LD matrix  $\mathbf{R}$ . Similar to LDSC, we use the iterative reweighted least square to increase estimation efficiency (material and methods), which increases precision when the sample sizes are large (Figure S1). As the dimension of genotype matrices is usually high, we partition the genome into independent genomic blocks with LDetect<sup>34</sup> and perform eigen-decomposition in each block. We also employ a linear shrinkage method to the

**Table 1. Precision and accuracy of the heritability estimates with LDER, LDSC, HESS, and HDL**

|          | Precision (1/SD)   |       |                    |                     | RMSE               |       |       |       |
|----------|--|-------|--------------------|---------------------|--------------------|-------|-------|-------|
|          | LDER   | LDSC  | HESS               | HDL                 | LDER               | LDSC  | HESS  | HDL   |
| <b>n</b> | <b>Performance with in-sample LD estimated by UKBB European samples</b>                |       |                    |                     |                    |       |       |       |
| 5,000    | 17.82  | 12.51 | 20.72 <sup>a</sup> | N/A                 | 0.073 <sup>a</sup> | 0.095 | 0.082 | 0.461 |
| 10,000   | 23.53 <sup>a</sup>   | 17.82 | 21.71              | N/A                 | 0.034 <sup>a</sup> | 0.053 | 0.076 | 0.500 |
| 20,000   | 40.18 <sup>a</sup>   | 31.14 | 33.14              | N/A                 | 0.030 <sup>a</sup> | 0.041 | 0.035 | 0.499 |
| 50,000   | 73.79  | 53.77 | 54.87              | 154.68 <sup>a</sup> | 0.016 <sup>a</sup> | 0.022 | 0.169 | 0.461 |
| <b>n</b> | <b>Performance with external LD estimated by 1000 Genomes Project European samples</b> |       |                    |                     |                    |       |       |       |
| 5,000    | 14.07  | 12.28 | 21.21 <sup>a</sup> | –                   | 0.077 <sup>a</sup> | 0.102 | 0.152 | –     |
| 10,000   | 20.05  | 17.32 | 20.60 <sup>a</sup> | –                   | 0.047 <sup>a</sup> | 0.056 | 0.060 | –     |
| 20,000   | 33.60 <sup>a</sup>   | 30.84 | 27.99              | –                   | 0.032 <sup>a</sup> | 0.045 | 0.060 | –     |
| 50,000   | 64.16 <sup>a</sup>   | 53.84 | 56.47              | –                   | 0.017 <sup>a</sup> | 0.025 | 0.114 | –     |

Simulations were based on UKBB genotypes and repeated 50 times. Heritability was fixed at 0.5, and the proportion of causal SNPs was 1%. N/A indicates the estimates are too close to zero, yielding infinite 1/SD.

<sup>a</sup>Highest precision and smallest RMSEs

estimation of the LD matrix **R** to ensure the robustness of our algorithm accommodating different reference panels ([material and methods](#)).

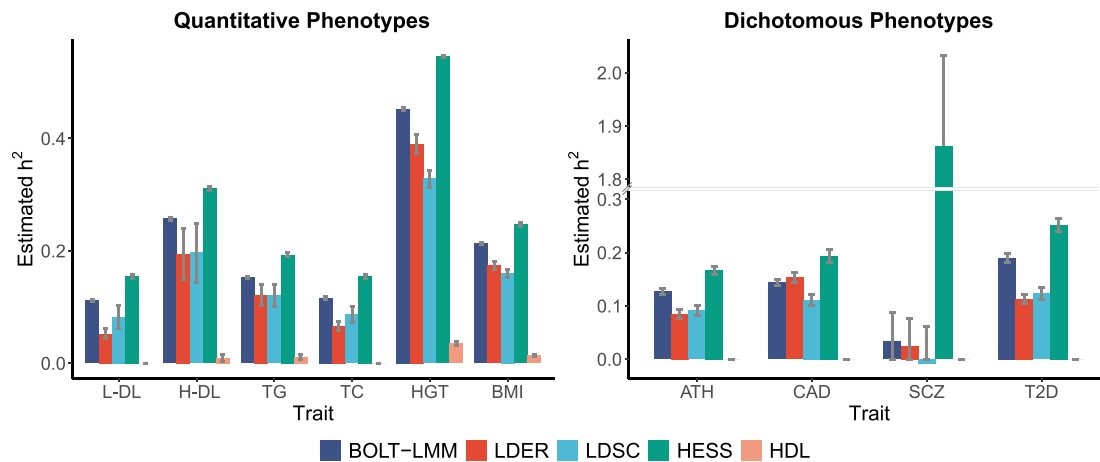
#### Simulation based on generated summary statistics

We first simulated GWAS summary statistics of 100,000 genetic variants to investigate the performance of LDER and other comparable methods including LDSC, HESS, and HDL under different genetic architectures. The simulated genetic architectures varied in three aspects: the sparsity of the causal SNPs, the heritability, and the confounding inflation factor ([material and methods](#)). We used true LD information as the input to all the methods. Simulation experiments were repeated 50 times. [Figure 1](#) and [Figure S2](#) show the heritability and inflation factor estimated by LDER and the three other methods. The accuracies of estimated heritabilities for all methods increased with increased sample sizes. LDER achieved higher accuracy with smaller standard deviations (SDs) in estimating heritability compared with LDSC and HESS. When the sample size was small (i.e.,  $n = 5,000$  and  $10,000$ ), the superiority of LDER was more significant, while the heritability estimator of HESS had downward bias. Despite the comparable estimates of heritability between LDER and HDL, HDL had severe upward bias in estimating inflation factors. Although the SDs of the inflation factor estimation increased with the sample sizes (see [supplemental method 3.3](#)), LDER retained high estimation accuracy. We note that the two-stage estimation procedure led to an underestimate of inflation factor by LDSC, as SNPs with test statistics larger than a certain threshold ( $z_i^2 > 30$  in LDSC as default) were not included in the first step estimating the inflation factor.<sup>15</sup> We further provide a comparison between LDER and LDSC without the two-stage procedure, and LDER still achieved

more accurate and precise estimates than LDSC ([Figure S3](#)).

#### Simulations based on real genotypes

We then analyzed simulated GWAS data with real genotypes from 276,050 independent UKBB European samples ([material and methods](#)). We simulated effect sizes from the spike-and-slab distribution mentioned above. The simulated phenotypes were generated by the sum of all genetic markers weighted by the simulated effect sizes and added by a normally distributed error term fixing the heritability at 0.5. The sparsity of causal SNPs varied from 0.005 to 0.01. We evaluated the estimation accuracy by the root-mean-square error (RMSE) and the precision, which was measured by the inverse of the SDs. [Table 1](#) and [Table S1](#) show that LDER still achieved higher accuracy with smaller RMSE and higher precision than LDSC under scenarios with different sample sizes. Among the four methods, HESS achieved the highest precision when the sample size was 5,000, but its RMSE was larger than LDER. We also notice that HDL severely underestimated the heritability. This may be because the UKBB LD reference panel provided by HDL (either 1,029,876 QCed UKBB imputed HapMap 3 SNPs or 307,519 QCed UKBB Axiom Array SNPs) contains SNPs that were not in the data we analyzed (~50%), and significant deterioration of performance when there are different numbers of SNPs in GWAS and in the reference panel was reported.<sup>28,39</sup> Results in [Table 1](#) and [Table S1](#) demonstrate the robustness of LDER to the external LD by estimating the LD matrix with the 1000G reference panel. Despite that the accuracy and precision for all methods were influenced by external LD reference, LDER still showed the smallest or comparable RMSE among all methods. We also compared the one- and two-stage procedures for



**Figure 2. Comparisons between the estimated heritability by BOLT-LMM and that by LDSC, HESS, and HDL on six quantitative phenotypes and four dichotomous phenotypes in the UKBB**  
Error bars indicate SEs of the estimates, which were derived from a delete-block-jackknife procedure for LDSC.

LDSC. The two-stage estimates yielded smaller RMSE and higher or comparable precision compared with the one-stage estimates for LDSC and for LDSC with large sample sizes (Table S2).

We also simulated a more realistic scenario with both polygenicity and confounding inflation. We used the UKBB genotype data and simulated polygenic phenotypes by drawing causal SNPs only from the chromosomes of odd numbers. All SNPs on chromosomes of even numbers were not causal. This strategy also avoids the influence of LD. We further included the environmental stratification component aligned with the first principal component of the genotype data. The mean  $\chi^2$  statistics in even chromosomes (with no causal SNPs) was regarded as the contribution of population stratification. In all simulation settings with sample sizes varying from 5,000 to 50,000, both LDSC and LDSC accurately estimate the confounding inflation factors (Table S3).

### Real data applications on UKBB phenotypes

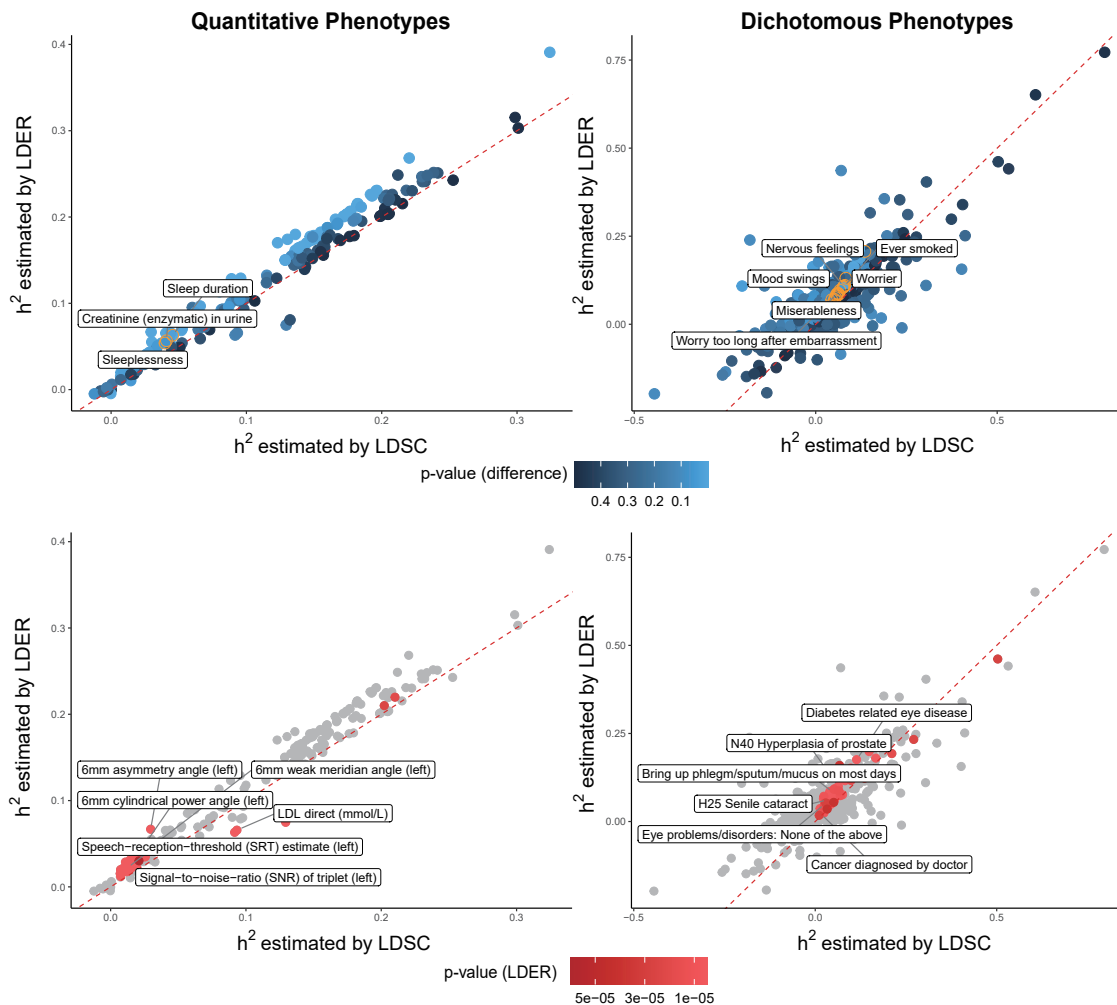
For better calibration and demonstration of the superiority of LDSC, we performed analysis on 10 common UKBB traits with BOLT-LMM, which is an individual-level-data-based method known to provide more accurate estimates than LDSC<sup>39</sup> (Figure 2). Among the 10 common traits, low-density lipoprotein (L-DL), high-density lipoprotein (H-DL), triglyceride (TG), total cholesterol (TC), height (HGT), and body mass index (BMI) are quantitative phenotypes; asthma (ATH), coronary artery disease (CAD), schizophrenia (SCZ), and type 2 diabetes (T2D) are dichotomous phenotypes. We treated estimates from BOLT-LMM as the true values of the heritability and calculated the RMSE of LDSC, LDSC, HESS, and HDL. The precision was measured by the reciprocal of the SEs reported by each method. In particular, the SE of LDSC was estimated through block-jackknife (material and methods). In general, LDSC and LDSC provided estimates lower than

BOLT-LMM, while HESS derived estimates higher than BOLT-LMM. HDL had estimates of heritability close to zero, which was similar to the results in simulations with real genotypes. LDSC still showed the most accurate estimation compared with the other methods, with the smallest RMSE (Figure 2 and Table S4). HESS provided estimates with the highest precision.

We applied LDSC and LDSC to summary statistics of 814 complex traits, including 221 quantitative phenotypes and 593 dichotomous phenotypes (material and methods). On average, LDSC yielded smaller SE than LDSC on both quantitative and dichotomous traits (Figure 3). After Bonferroni correction, LDSC identified 363 significantly heritable phenotypes, among which 97 were not identified by LDSC (Table S5), such as ventral hernia (LDSC p value =  $6.1e-06$ ) and non-insulin-dependent diabetes mellitus (LDSC p value =  $3.4e-05$ ). We found the estimates of heritability by LDSC and LDSC were significantly different for 20 phenotypes (p value < 0.05 after Bonferroni correction; see Table 2). A numerical comparison between the estimates of LDSC and LDSC on the UKBB traits is provided in Tables S6 and S7.

### Computational efficiency

Table S8 shows the computation time of LDSC, LDSC, HESS, and HDL. We divided the total computational time into LD preparation and heritability estimation. HESS and LDSC are more efficient in LD preparation compared with the other two methods. As for the estimation procedure, LDSC and HDL take more time in estimating SEs using a jackknife procedure. We also note that although HESS is most efficient in the estimation step compared with other methods, it is necessary for HESS software to recalculate LD information when it is applied to a new trait. In contrast, LDSC, LDSC, and HDL can be applied to the pre-computed LD information and can be efficient when estimation on multiple traits is required.



**Figure 3. Heritability estimates with LDER and LDSC among 221 quantitative phenotypes and 593 dichotomous phenotypes in the UKBB**

The color of the blue points indicates the significance level ( $p$  value) testing the difference of the estimates using two methods. The orange points highlight phenotypes with significantly different heritability estimates between LDER and LDSC (the upper panels). The red points highlight significantly heritable phenotypes (with Bonferroni correction) estimated by LDER but not identified by LDSC (the lower panels). For clearer visualization, the six smallest  $p$  values for heritability estimation are shown with labels for both quantitative and dichotomous phenotypes. The red dashed line indicates conditions where the heritability estimated by LDER and LDSC is equal. The estimated heritability for dichotomous phenotypes has been transformed to liability scale.

## Discussion

In this article, we propose LDER, a summary-statistics-based method improving the accuracy and precision of estimation of SNP heritability and confounding inflation. As an extension of LDSC, LDER provides more accurate and precise estimates in both simulations and real data applications. The superiority of LDER can be attributed to the fact that it captures more information on the relationship between LD matrix and test statistics, whereas LDSC uses only partial information from the LD matrix. To be more precise, LDSC only utilizes the diagonal elements of the squared LD matrix.

To ensure the robustness of our algorithm accommodating reference panels from different sources, we employ a linear shrinkage method to the estimation of the LD matrices, which is computationally efficient and built in with our software. In addition, as the dimension of genotype matrices is

usually high, estimating and shrinking the LD matrix and performing the eigen-decomposition can be time consuming. In practice, we partition the genome into independent genomic blocks with respect to ethnicity using LDETECT<sup>34</sup> and perform eigen-decomposition for each LD block.

We provide the pre-computed eigen-decomposition information for 276,050 independent samples of European ancestry from UKBB and 489 samples of European ancestry from 1000G. Although the limited sample size and the potential mismatch between the target population and the LD reference panel may threaten the superiority of LDER, LDER remains robust with respect to the external LD reference and superior to LDSC.

For future directions, it would be advantageous to jointly model multiple complex traits to better estimate their genetic correlation, which quantifies the genetic similarity between complex traits. Summary-statistics-based methods



**Table 2. Estimates of SNP heritability on UKBB traits that are significantly different using LDER and LDSC**

| UKBB ID    | Phenotype   | Variable type   | $h^2_{LDER}$  | $h^2_{LDSC}$  | p value (different) |
|------------|---|-----------------|---------------|---------------|---------------------|
| 1210       | snoring   | binary          | 0.086 (0.003) | 0.058 (0.004) | 2.8e−09             |
| 1920       | mood swings   | binary          | 0.097 (0.003) | 0.067 (0.004) | 1.6e−09             |
| 1930       | miserableness                                       | binary          | 0.087 (0.003) | 0.062 (0.004) | 3.1e−07             |
| 1950       | sensitivity/hurt feelings                           | binary          | 0.083 (0.003) | 0.057 (0.004) | 5.7e−09             |
| 1970       | nervous feelings                                    | binary          | 0.108 (0.003) | 0.077 (0.005) | 6.2e−08             |
| 1980       | worrier   | binary          | 0.104 (0.003) | 0.075 (0.004) | 3.8e−08             |
| 2000       | worry too long after embarrassment                  | binary          | 0.088 (0.003) | 0.061 (0.004) | 5.7e−10             |
| 2010       | suffer from “nerves”                                | binary          | 0.071 (0.003) | 0.052 (0.004) | 3.2e−05             |
| 2020       | loneliness, isolation                               | binary          | 0.070 (0.003) | 0.044 (0.004) | 2.9e−08             |
| 2030       | guilty feelings                                     | binary          | 0.076 (0.003) | 0.054 (0.004) | 6.8e−07             |
| 2040       | risk taking   | binary          | 0.085 (0.003) | 0.064 (0.004) | 6.6e−06             |
| 2188       | long-standing illness, disability, or infirmity     | binary          | 0.080 (0.003) | 0.058 (0.004) | 5.7e−07             |
| 2443       | diabetes diagnosed by doctor                        | binary          | 0.207 (0.012) | 0.136 (0.011) | 8.4e−06             |
| 20160      | ever smoked   | binary          | 0.106 (0.003) | 0.077 (0.004) | 1.3e−08             |
| 6149_100   | mouth/teeth dental problems: none of the above      | binary          | 0.053 (0.002) | 0.038 (0.003) | 1.0e−05             |
| 6149_6     | mouth/teeth dental problems: dentures               | binary          | 0.110 (0.004) | 0.080 (0.005) | 7.3e−07             |
| 6150_2     | vascular/heart problems diagnosed by doctor: angina | binary          | 0.131 (0.008) | 0.083 (0.010) | 5.1e−05             |
| 30510_irmt | creatinine (enzymatic) in urine                     | continuous_irmt | 0.057 (0.002) | 0.041 (0.002) | 3.2e−08             |
| 1160       | sleep duration                                      | ordinal         | 0.063 (0.002) | 0.046 (0.002) | 6.9e−09             |
| 1200       | sleeplessness/insomnia                              | ordinal         | 0.054 (0.002) | 0.040 (0.002) | 3.3e−07             |

The estimates of heritability of binary traits have been transformed to liability scale.

for genetic correlation analysis enable study of a wide spectrum of complex human diseases, as the studied phenotypes do not need to be collected from the same individuals. It has also been revealed that SNPs in different functional categories (such as promoters, enhancers, etc.) provide disproportionate contributions to the disease heritability. Therefore, it is also of interest to derive the partitioned heritability within our framework to analyze multiple cell-type-specific functional categories.

#### Data and code availability

LDER software (R package) and analysis scripts are available at <https://github.com/shuangsong0110/LDER>.

#### Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2022.03.013>.

#### Acknowledgments

This work was supported in part by NIH grant NIH GM 134005 and NSF grants DMS 1713120 and 1902903 (H.Z.). This research has been conducted using the UK Biobank Resource under Application Number 29900.

#### Declaration of interests

The authors declare no competing interests.

Received: December 21, 2021

Accepted: March 18, 2022

Published: April 13, 2022

#### Web resources

1000 Genomes, <http://www.1000genomes.org>

HDL, <https://github.com/zhenin/HDL>

HES, [https://huwenboshi.github.io/hess-0.5/local\\_hsqg/](https://huwenboshi.github.io/hess-0.5/local_hsqg/)

LDetect, <https://bitbucket.org/nygresearch/ldetect-data/src/master/>

LDSC, <https://github.com/bulik/ldsc/wiki>

Neale Lab UK Biobank GWAS summary statistics, <http://www.nealelab.is/uk-biobank/>

PLINK software, <https://zzz.bwh.harvard.edu/plink/profile.shtml>

UK Biobank Resource, <https://www.ukbiobank.ac.uk/>

#### References

- de Los Campos, G., Sorensen, D., and Gianola, D. (2015). Genomic heritability: what is it? *PLoS Genet.* 11, e1005048.

2. Yang, J., Zeng, J., Goddard, M.E., Wray, N.R., and Visscher, P.M. (2017). Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* *49*, 1304–1310.
3. Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S.J., and Park, J.-H. (2013). Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat. Genet.* *45*, 400–405, e1–e3.
4. Gusev, A., Lee, S.H., Trynka, G., Finucane, H., Vilhjálmsson, B.J., Xu, H., Zang, C., Ripke, S., Bulik-Sullivan, B., Stahl, E., et al.; Schizophrenia Working Group of the Psychiatric Genomics Consortium; SWE-SCZ Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; and SWE-SCZ Consortium (2014). Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* *95*, 535–552.
5. Zhu, H., and Zhou, X. (2020). Statistical methods for SNP heritability estimation and partition: A review. *Comput. Struct. Biotechnol. J.* *18*, 1557–1568.
6. Eaves, L.J., Last, K.A., Young, P.A., and Martin, N.G. (1978). Model-fitting approaches to the analysis of human behaviour. *Heredity* *41*, 249–320.
7. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* *42*, 565–569.
8. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* *88*, 76–82.
9. Loh, P.-R., Tucker, G., Bulik-Sullivan, B.K., Vilhjálmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B., et al. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* *47*, 284–290.
10. Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., and Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nat. Methods* *8*, 833–835.
11. Speed, D., Hemani, G., Johnson, M.R., and Balding, D.J. (2012). Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* *91*, 1011–1021.
12. Speed, D., Cai, N., Johnson, M.R., Nejentsev, S., Balding, D.J.; and UCLEB Consortium (2017). Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* *49*, 986–992.
13. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* *562*, 203–209.
14. Hou, K., Burch, K.S., Majumdar, A., Shi, H., Mancuso, N., Wu, Y., Sankararaman, S., and Pasaniuc, B. (2019). Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture. *Nat. Genet.* *51*, 1244–1251.
15. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M.; and Schizophrenia Working Group of the Psychiatric Genomics Consortium (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* *47*, 291–295.
16. Robinson, E.B., St Pourcain, B., Anttila, V., Kosmicki, J.A., Bulik-Sullivan, B., Grove, J., Maller, J., Samocha, K.E., Sanders, S.J., Ripke, S., et al.; iPSYCH-SSI-Broad Autism Group (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* *48*, 552–555.
17. Sniekers, S., Stringer, S., Watanabe, K., Jansen, P.R., Coleman, J.R.I., Krapohl, E., Taskesen, E., Hammerschlag, A.R., Okbay, A., Zabaneh, D., et al. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* *49*, 1107–1112.
18. Kranzler, H.R., Zhou, H., Kember, R.L., Smith, R.V., Justice, A.C., Damrauer, S., Tsao, P.S., Klarin, D., Baras, A., Reid, J., et al. (2019). Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat. Commun.* *10*, 1499.
19. Speed, D., and Balding, D.J. (2019). SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat. Genet.* *51*, 277–284.
20. Lander, E.S., and Schork, N.J. (1994). Genetic dissection of complex traits. *Science* *265*, 2037–2048.
21. Lin, D.Y., and Zeng, D. (2011). Correcting for population stratification in genome-wide association studies. *J. Am. Stat. Assoc.* *106*, 997–1008.
22. Devlin, B., and Roeder, K. (1999). Genomic control for association studies. *Biometrics* *55*, 997–1004.
23. Reich, D.E., and Goldstein, D.B. (2001). Detecting association in a case-control study while correcting for population stratification. *Genet. Epidemiol.* *20*, 4–16.
24. Zheng, G., Freidlin, B., and Gastwirth, J.L. (2006). Robust genomic control for association studies. *Am. J. Hum. Genet.* *78*, 350–356.
25. Yang, J., Weedon, M.N., Purcell, S., Lettre, G., Estrada, K., Willer, C.J., Smith, A.V., Ingelsson, E., O'Connell, J.R., Mangino, M., et al.; GIANT Consortium (2011). Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* *19*, 807–812.
26. Shi, H., Kichaev, G., and Pasaniuc, B. (2016). Contrasting the genetic architecture of 30 complex traits from summary association data. *Am. J. Hum. Genet.* *99*, 139–153.
27. Weiner, D.J., Wigdor, E.M., Ripke, S., Walters, R.K., Kosmicki, J.A., Grove, J., Samocha, K.E., Goldstein, J.I., Okbay, A., Bybjerg-Grauholm, J., et al.; iPSYCH-Broad Autism Group; and Psychiatric Genomics Consortium Autism Group (2017). Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.* *49*, 978–985.
28. Ning, Z., Pawitan, Y., and Shen, X. (2020). High-definition likelihood inference of genetic correlations across human complex traits. *Nat. Genet.* *52*, 859–864.
29. Zhang, Y., Lu, Q., Ye, Y., Huang, K., Liu, W., Wu, Y., Zhong, X., Li, B., Yu, Z., Travers, B.G., et al. (2021). SUPERGENOVA: local genetic correlation analysis reveals heterogeneous etiologic sharing of complex traits. *Genome Biol.* *22*, 262.
30. Guan, Y., and Stephens, M. (2011). Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *Ann. Appl. Stat.* *5*, 1780–1815.
31. Lee, S.H., Wray, N.R., Goddard, M.E., and Visscher, P.M. (2011). Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* *88*, 294–305.
32. Jiang, W., Song, S., Hou, L., and Zhao, H. (2021). A set of efficient methods to generate high-dimensional binary data with specified correlation structures. *Am. Stat.* *75*, 310–322.

33. Stephens, M., and Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am. J. Hum. Genet.* 76, 449–462.
34. Berisa, T., and Pickrell, J.K. (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* 32, 283–285.
35. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.
36. Ledoit, O., and Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *J. Multivariate Anal.* 139, 360–384.
37. Ledoit, O., and Wolf, M. (2017). Numerical implementation of the QuEST function. *Comput. Stat. Data Anal.* 115, 199–223.
38. Vilhjálmsson, B.J., Yang, J., Finucane, H.K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P.-R., Bhatia, G., Do, R., et al.; Schizophrenia Working Group of the Psychiatric Genomics Consortium, Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) study (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* 97, 576–592.
39. Zhang, Y., Cheng, Y., Jiang, W., Ye, Y., Lu, Q., and Zhao, H. (2021). Comparison of methods for estimating genetic correlation between complex traits using GWAS summary statistics. *Brief. Bioinform.* 22, bbaa442.