# Accurate and Efficient Estimation of Local Heritability using Summary Statistics and LD Matrix

**Hui Li**[1], **Rahul Mazumder**[2], **and Xihong Lin**[1,3*]

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, USA
[2]Operations Research Center and Center for Statistics and Data Science, MIT Sloan School of Management, Cambridge, USA
[3]Department of Statistics, Harvard University, Cambridge, USA
[*]Corresponding authors: xlin@hsph.harvard.edu

## ABSTRACT

Existing SNP-heritability estimation methods that leverage GWAS summary statistics produce estimators that are less efficient than the restricted maximum likelihood (REML) estimator using individual-level data under linear mixed models (LMMs). Increasing the precision of a heritability estimator is particularly important for regional analyses, as local genetic variances tend to be small. We introduce a new estimator for local heritability, "HEELS", which attains comparable statistical efficiency as REML (*i.e.* relative efficiency greater than 92%) but only requires summary-level statistics – Z-scores from the marginal association tests plus the empirical LD matrix. HEELS significantly improves the statistical efficiency of the existing summary-statistics-based heritability estimators– for instance, HEELS produces heritability estimates that are more than 3-fold and 7-times less variable than GRE and LDSC, respectively. Moreover, we introduce a unified framework to evaluate and compare the performance of different LD approximation strategies. We propose representing the empirical LD as the sum of a low-rank matrix and a banded matrix. This approximation not only reduces the storage and memory cost of using the LD matrix, but also improves the computational efficiency of the HEELS estimation. We demonstrate the statistical efficiency of HEELS and the advantages of our proposed LD approximation strategies both in simulations and through empirical analyses of the UK Biobank data.

## Introduction

In the last decade, advances in biotechnology have enabled estimation of genetic variance contributed by genotyped variants without requiring assumptions about the shared environmental effects[1–6]. These methods estimate the so-called SNP-heritability ($h^2_{SNP}$), defined as the proportion of phenotypic variance due to causal variants tagged by genotyped variants, and have led to critical insights into the genetic architectures of complex traits and diseases[7]. The existing $h^2_{SNP}$-estimation methods can be broadly categorized as either based on individual-level genotype and phenotype data[1,2,8–11] or based on summary statistics from genome-wide association studies (GWAS)[3,4,6,12].

Estimators that use individual-level data are generally more precise, but the applicability of these methods can be limiting due to privacy concerns and data-sharing restrictions. Heritability estimation methods that are based on GWAS summary statistics have gained popularity in recent years, but an important drawback of these approaches is low statistical efficiency[11,13]. For instance, studies have shown that the empirical variance of the estimates from LD-score regression (LDSC) is much larger than that of a REML-based estimator[14,15]. Other methods, such as Generalized Random Effects (GRE) and Randomized HE-regression (RHE-reg), are also less precise than REML based on simulations[10,11], To address these limitations, we introduce Heritability Estimation with high Efficiency using LD and association Summary Statistics (HEELS) – an accurate and statistically efficient estimator of local $h^2_{SNP}$ that only requires summary-level statistics.

Our work has two main contributions to the literature of $h^2_{SNP}$ estimation. First, we propose an iterative procedure that uses summary statistics to solve the REML-score equations by transforming the Henderson's algorithm for estimating the variance components in linear mixed models[16,17]. The summary statistics required by HEELS includes the marginal association statistics from GWAS and the correlation statistics estimated in-sample (*i.e.* empirical covariance matrix of the genotypes, "LD"). We show both analytically and through extensive simulations that the summary-statistics-based HEELS estimator attains comparable level of statistical efficiency or precision as the conventional individual-level-data-based REML estimators, such as GREML[1] and BOLT-REML[2]. The relative efficiency of HEELS is significantly higher than that of the state-of-the-art summary-statistics-based methods, *i.e.*, the estimates from HEELS have much smaller variance. Our method is applicable and most suitable to local heritability estimation, as the genetic variance at a specific region tends to be small and thus its statistical efficiency is particularly important. Second, we introduce a novel low-dimensional representation of the LD as the sum of a banded

and a low-rank matrix ("Banded + LR"). This structure helps provide a unified framework where many of the existing LD approximations can be viewed as special cases. We use an optimization approach to solve for the best representation of LD with this "Banded + LR" structure, and evaluate the performance of different low-dimensional approximations in the context of $h^2_{SNP}$ estimation. Compared to existing methods, our proposed decomposing approach not only achieves greater approximation accuracy of the LD matrix, measured in the matrix spectral norm, but also yields heritability estimates with smaller bias.

We applied HEELS to estimate the local SNP-heritability of 30 complex traits and diseases in the UK Biobank (UKB). Consistent with our findings in the simulations, the estimates from HEELS are highly concordant with the estimates from REML. HEELS also attains the highest statistical efficiency compared to the other summary-statistics-based methods. We used the local $\hat{h}^2_{SNP}$ estimates from HEELS to contrast the genetic architectures of traits, and found different distribution patterns of local genetic variance, consistent with previous studies[18,19]. We also used HEELS to prioritize genetic regions that are enriched for related traits (*e.g.* lipid traits, blood-cell and leukocyte traits). The loci we identified not only corroborate known mechanisms of disease-association, but also yield novel insights into the shared genetic basis among correlated traits, pointing to putative pleiotropic hotspots.

Several recently developed methods are related to our work. Here we highlight the their similarities and distinctions compared to HEELS. High-Definition Likelihood ("HDL") has a similar aim as our work in terms of improving statistical efficiency[20] (although the method was proposed for genetic correlation estimation, it necessarily computes heritability as an intermediate step). Notably, our estimator was derived based on the marginal likelihood of the phenotypes (although the estimation does not require accessing individual-level data) while HDL uses the likelihood of the Z-statistics. In terms of LD approximation, our proposed banded + low-rank representation of the LD differs from the truncated SVD approach used by HDL. In simulations, we observed bias in HDL's estimates of heritability, which has been reported in the benchmarking results of another study[21]. Therefore, we do not include HDL in our comparisons among *unbiased* estimators. Another recently published work introduces method called "LDAK-GBAT", using the same likelihood functions as ours[22]. However, LDAK-GBAT focuses on gene-based association testing rather than heritability estimation. The paper does not report or analyze the statistical efficiency of the heritability estimator. LDAK-GBAT also uses a permutation procedure which inherently requires individual-level data, whereas HEELS solely relies on summary-level statistics for both heritability estimation and standard error calculation. Both HDL and LDAK-GBAT maximize their respective likelihood functions directly using a Newton-Raphson type algorithm, whereas our iterative

algorithm operates on the score equations and involves *closed-form* updating equations. Finally, we note that a parallel work – SuSiE-RSS – also exploits the sufficient statistics of the likelihood function to enable "dual" algorithms that are applicable to both individual-level data and summary-level statistics[23]. In other words, the relationship between SuSiE and SuSiE-RSS is akin to the relationship between GREML and HEELS. Nevertheless, SuSiE-RSS is designed for fine-mapping and it performs variable selection under the Bayesian framework, which is distinct from HEELS.

The literature of SNP-heritability estimation is quite rich. From a computational perspective, advances in scalable heritability estimation algorithms have enabled existing estimators, such as those from REML and HE-regression, to be applied to biobank-sized datasets[2,11,24]. From a modeling perspective, researchers have made different assumptions about the distribution of effect sizes[9,25,26]. Although efforts have been made to reconcile these modeling inconsistencies[13,25,27], debates remain unresolved as to which assumptions are more fitting and realistic. We emphasize that the particular focus of our study is to improve the *statistical efficiency* of heritability estimators that are based on summary-level statistics. It is out of the scope of our current work to compare across all or most of the existing heritability estimation methods, nor to evaluate the suitability of any particular heritability model. We choose LMM as the simplest and most widely understood model to demonstrate the merits of our approach both conceptually and empirically. We hope our work can encourage development of methods that produce more precise heritability estimates under potentially different modeling frameworks.

# Results

## Overview of HEELS

HEELS uses summary-level statistics, including marginal association statistics and the LD matrix, to estimate heritability by iteratively solving the REML score equation[17]. HEELS provides heritability estimates that are as precise as those based on REML-based estimates that use individual-level data, such as GREML[1]. We first introduce some notations and define $h^2_{SNP}$ using the linear mixed model based on individual-level data.

Let $\mathbf{y}$ be a length-$n$ vector that denotes the phenotypes of $n$ samples. Denote by $\mathbf{X} \in \mathbb{R}^{n \times p}$ the genotype matrix of $n$ individuals based on $p$ markers or SNPs. We standardize $\mathbf{X}$ and $\mathbf{y}$ such that the variance of the phenotype is 1 and the variance of each marker-specific genotype vector is $1/p$, or $diag(\mathbf{X}^\top \mathbf{X}/n) = 1/p$. Let $\mathbf{S}$ and $\mathbf{R}$ denote the the marginal association statistics and the in-sample LD matrix, *i.e.* $\mathbf{S} = \mathbf{X}^\top \mathbf{y}/\sqrt{n}$

and $\mathbf{R} = \mathbf{X}^\top \mathbf{X}/n$. We use an additive genetic model for the phenotypes as $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, where $\boldsymbol{\beta}$ is a $p \times 1$ vector assumed to follow $N(0, \sigma_g^2 \mathbf{I}_p)$, and $\varepsilon$ is a length-$n$ vector distributed as $\varepsilon \sim N(0, \sigma_e^2 \mathbf{I}_n)$. Under these assumptions, $\mathbf{y} \sim N(0, \mathbf{V})$, where the variance-covariance matrix is $\mathbf{V} \equiv var(\mathbf{y}) = \sigma_g^2 \mathbf{X}\mathbf{X}^\top + \sigma_e^2 \mathbf{I}_n$. Under this model, SNP-heritability is defined as $h^2_{SNP} = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$ (**Online Methods**).

Suppose we do not have access to individual-level data $\mathbf{X}$ and $\mathbf{y}$, and are only provided with the marginal association statistics $\mathbf{S} = \mathbf{X}^\top \mathbf{y}$ and the LD matrix $\mathbf{R} = \mathbf{X}^\top \mathbf{X}$. We show that the REML score equations for $(\sigma_g^2, \sigma_e^2)$ can be solved using these summary-level statistics $\mathbf{S}$ and $\mathbf{R}$ only, by applying the Sherman-Woodbury matrix identity to the Anderson's algorithm for solving the variance components (**Supplementary Notes**). HEELS iterates between updating the Best Linear Unbiased Predictor (BLUP) estimates[17] of the joint effect sizes $\widehat{\beta}$ and updating the variance component estimates $(\sigma_g^2, \sigma_e^2)$ until convergence. Let the superscript $^{(t)}$ denote the value of the parameter at iteration $t$. The HEELS estimation procedure is as follows.

1. Update the BLUP joint effect size estimates using:

$$\widehat{\beta}^{(t)} = [\mathbf{W}^{(t)}]^{-1}\mathbf{S}, \text{ where } \mathbf{W}^{(t)} = \frac{[\widehat{\sigma_e}^2]^{(t)}}{[\widehat{\sigma_g}^2]^{(t)}}\mathbf{I}_p + \mathbf{R} \tag{1}$$

2. Update $\sigma_g^2$ using:

$$[\widehat{\sigma}_g^2]^{(t+1)} = \frac{\widehat{\beta}^{(t)\top}\widehat{\beta}^{(t)}}{p - tr([\mathbf{W}^{(t)}]^{-1})} \tag{2}$$

3. Update $\sigma_e^2$ using:

$$[\widehat{\sigma}_e^2]^{(t+1)} = \frac{\mathbf{y}^\top \mathbf{y} - \mathbf{S}^\top \widehat{\beta}^{(t)}}{n} \tag{3}$$

We initialize $[\sigma_e^2]^{(0)}, [\sigma_g^2]^{(0)}$ with some random values between 0 and 1, and repeat step 1-3 below until convergence or when the change in $\ell_{HEELS}$ between two consecutive iterations is sufficiently small.

We provide derivation details on the updating equations (1)-(3) and further explain the intuition behind these expressions in the **Supplementary Notes**. Notably, the algorithm above does not use the raw individual-level data $\mathbf{X}$ or $\mathbf{y}$, but only utilizes summary statistics $\mathbf{S}$ and $\mathbf{R}$, which are sufficient statistics for the variance component parameters in our model. When sample variance $\mathbf{y}^\top \mathbf{y}/n$ is known, implementing the procedure above is straightforward. When it is not known, we can use the Z-scores, approximate $\mathbf{y}^\top \mathbf{y}/n$ by 1 and rescale $\hat{\sigma}_g^2, \hat{\sigma}_e^2$ at each iteration accordingly. Similar approximation strategies have been adopted by other summary-statistics-based methods, such as SuSiE-RSS.[23]

For a given model, the maximum-likelihood estimator has the minimal variance among unbiased estimators[28]. The *statistical efficiency* of a heritability estimator is calculated using the ratio of the variance of the REML estimator and the variance of the estimator in comparison[29]. The primary reason for the high statistical efficiency of HEELS is that the estimating equations of $(\sigma_g^2, \sigma_e^2)$ solved by HEELS are identical to the REML score equations of $(\sigma_g^2, \sigma_e^2)$[17]. The asymptotic variance of the variance components can be derived using the likelihood theory. Importantly, the variance estimator can also be re-experssed using summary statistics, $\mathbf{S}, \mathbf{R}$, only. We apply the multivariate Delta method to obtain a plug-in estimator of the variance of $\hat{h}^2_{HEELS}$ (**Online Methods**).

## Accuracy and statistical efficiency of the HEELS estimator

To evaluate the performance of the HEELS heritability estimator, we performed simulations using the real genotype array data of $332,430$ unrelated British white individuals in the UK Biobank. We selected GREML, LDSC, GRE and HESS as the representative $h^2_{SNP}$-estimators to be compared with HEELS, and do not include other methods in the comparisons due to our focus on *statistical efficiency*. In the **Supplementary Notes**, we provide a more thorough review of the existing $h^2_{SNP}$-estimation approaches, with their key advantages and limitations summarized in **Supplementary Table** 1, and explain our consideration in selecting the estimators for comparisons.

We first simulated quantitative phenotypes using the LMM assumption (*i.e.* also used by GREML and LDSC), where the causal effect sizes of all markers follow a normal distribution and contribute to the genetic variance equally. We found that HEELS is unbiased in finite samples and attains high statistical efficiency. In particular, we found a high degree of concordance between the estimates from REML-based estimators (such as GREML and BOLT-REML) which use individual-level data and those from HEELS which only relies on summary statistics (**Figure** 1). When the full UKB data is used, the relative statistical efficiency of HEELS is as high as 99.44%, whereas GRE and LDSC have much lower RE's of 19.81% and 8.68%, respectively (**Table** 1).

LDSC's lack of statistical efficiency has been observed in previous studies[12,14,20,30]. We provide a stylized argument to contrast the statistical efficiency of LDSC and HEELS under the framework of moment-matching methods[14] (**Supplementary Notes**). Briefly, the score equations solved by HEELS coincide with the estimating equations of the most efficient generalized method-of-moment estimator. Since LDSC can be viewed as a method-of-moment estimator with sub-optimal weights, its statistical efficiency is expected to be lower than that of HEELS. In simulations, we found that GRE and HEELS are

both unbiased regardless of sample size, but GRE is less statistically efficient that HEELS (**Supplementary Figure** 3). The variability of the GRE estimates is particularly high when sample size is small, as expected from the theory of GRE[10]. Even when the full UKB is used, the variance of $\hat{h}^2_{GRE}$ is still five times as large as that of $\hat{h}^2_{HEELS}$. We also compared the performance of HEELS with that of HESS, which is a state-of-the-art method for local heritability estimation[12]. We found that the HESS estimator is sensitive to the degree of LD regularization applied through the truncated SVD (**Supplementary Figure** 5), a phenomenon that has been reported in the original paper. In contrast, HEELS is unbiased when the full LD matrix is used, and we explicitly minimize the bias in the heritability estimate when we approximate the LD (see below for details). We note that the analytic form of the HESS estimator closely resembles that of the GRE estimator, despite their disparate modeling frameworks, as both GRE and HESS can be viewed as principal-component regression (PCR)-based estimators. In simulations, we indeed observed that the HESS estimates are closer to the GRE estimates as the amount of LD regularization decreases (**Supplementary Figure** 5), matching the theory which implies an alignment of $\hat{h}^2_{HESS}$ and $\hat{h}^2_{GRE}$ when LD is estimated in-sample without regularization.

Next, we considered relaxing the infinitesimal assumption and altering the sparsity of the causal genetic effects. Although this violates the assumptions used in LMM, we consider it an important setting to evaluate the robustness of our estimator, because it is plausible that only a small fraction of the variants have non-zero effects, as has been assumed by many existing methods[31,32]. In simulations, we found that HEELS remains unbiased and statistically efficient in these mis-specified scenarios (**Table** 1). Sparsity of the causal effects generally increases the variability of the heritability estimates, but the RE of HEELS remains the highest among summary-statistics-based methods (**Supplementary Figure** 1). Across settings of various levels of polygenicity and sample sizes, HEELS leads to an increase in precision that is equivalent to increasing the GWAS sample size at least 3 times or 7 times, respectively (**Supplementary Figure** 2). Together, our results are consistent with previous works which showed that REML is unbiased even when the modeling assumption is wrong (in terms of point-normal effects)[10,25], further illustrating the affinity of HEELS and REML. Our findings also corroborate some of the recent theoretical advances in REML estimator analyses, which prove that REML is consistent under non-infinitesimal models[33] and that sparse architecture will result in a larger asymptotic variance[34]

Finally, we evaluated the Type I error rates controlled by different methods. We found that the standard error of HEELS is well-calibrated regardless of sample size when the model is correctly specified (**Supplementary Figure** 6). The standard errors reported by BOLT-REML using individual-level data

generally lead to the correct coverage but can be anti-conservative when sample size is small. The standard errors from GRE can result in under-coverage, as has been reported (see Supplementary Table 4 of Hou *et al.*[10]). When the model is mis-specified (in terms of point-normal effect sizes), we observe under-coverage for all LMM-based estimators. Nevertheless, the calibration of HEELS is still better than GRE and LDSC, and is most comparable to REML across settings (**Supplementary Figure** 7).

## A unified framework to compare LD approximations

It is a well-known challenge in statistical genetics that the LD matrix is expensive to store and compute with. While several low-dimensional representations of the LD matrix have been proposed and used in the literature (*e.g.* banding[35–37], truncated SVD[12,38] and low-rank approximation[20,30]), no previous study has systematically compared the performance of different LD approximation strategies. The impact of LD approximations on heritability estimation is also unclear. Our goal is to construct a low-dimensional representation of the LD matrix that can reduce the storage and computational cost of heritability estimation without incurring large loss of accuracy or efficiency in the $h^2_{SNP}$ estimates.

To this end, we propose representing the in-sample LD matrix, $\mathbf{R}$, as the sum of a banded matrix and a low-rank matrix ("Banded + LR"),

$$\mathbf{R} \approx \mathbf{R}_b + \mathbf{R}_r = \mathbf{R}_b + \sum_{k=1}^{r} \lambda_k \mathbf{u}_k \mathbf{u}_k^\top = \mathbf{R}_b + \mathbf{U}_r \Lambda_r \mathbf{U}_r^\top, \tag{4}$$

where $\mathbf{R}_b$ is a banded matrix with central bandwidth $b$; $\mathbf{R}_r$ is a low-rank matrix with rank $r$; $\lambda_k$ and $\mathbf{u}_k$ (also diagonal entries of $\Lambda_r$ and columns of $\mathbf{U}_r$) are the top-$r$ eigenvalues and eigenvectors of $\mathbf{R}_r$. We consider a total of six strategies to decompose or represent the empirical LD using this "Banded + LR" structure (**Table** 2), all of which take the form of the expression in equation (4), but differ in terms of 1) whether the banded component is a diagonal matrix, in which case the approximation becomes a spiked covariance matrix[39], and 2) whether the banded and low-rank components are constrained to be positive semi-definite (PSD).

Our approximation provides a unified framework for analyzing and evaluating the performance of various LD approximations, as most of the existing LD approximation strategies can be viewed as special cases of decomposition (4). For instance, regularization of the LD matrix through the truncated SVD is equivalent to using "Banded + LR" by setting $b = 0$; the most common approximation of the LD matrix only involves the banded part, so it corresponds to "Banded + LR" by setting $r = 0$. The motivation

behind our proposed "Banded + LR" structure ($b \neq 0, r \neq 0$) is twofold. On one hand, because correlations between two genetic markers are typically induced by physical proximity, the vast majority of the non-zero elements of LD matrix lie on the central band. On the other hand, we want to retain the in-sample correlation structure outside of the central band, because oftentimes there are appreciable non-zero off-central-band elements in finite samples.

We adopt an optimization approach to solve for the best representation of LD with the "Banded + LR" structure, minimizing $||\mathbf{R} - \tilde{\mathbf{R}}||_F^2$, where $\tilde{\mathbf{R}}$ is the working approximating matrix in the form specified in the third column of **Table** 2. We explain the distinctions between different LD approximations in **Online Methods** and provide details on their respective estimation procedures in **Supplementary Table 2**. Briefly, the "Seq_Band_LR" strategy first bands the LD matrix and then performs low-rank decomposition on the residual off-banded matrix; the "PSD_Band_LR" strategy first approximates the banded component of the LD matrix using a PSD matrix and then performs low-rank decomposition on the residual matrix; the "Joint_Band_LR" strategy jointly approximates the banded and the low-rank components using PSD matrices. The PSD assumption can not only reduce the computational cost to solve for the optimal low-dimensional LD representation, but can also improve the computational efficiency of $h^2_{SNP}$ estimation in HEELS (**Supplementary Notes**). We derived the asymptotic variance of our HEELS estimator so that the additional variance incurred by the approximation can be quantified without access to the full LD (**Supplementary Notes**).

## Efficacy of the banded + low-rank representation

In simulations, we found that our proposed "Banded + LR" representations approximate the original full LD with higher accuracy (**Supplementary Figure** 8) and lead to less biased $\hat{h}^2$ than the existing methods (**Supplementary Figure** 9). We attribute the efficiency of our proposed "Banded + LR" representation to two aforementioned motivating factors, namely, retaining the central band of the covariance matrix while exploiting the signals in the residual off-central-band part of the LD matrix.

Relative to the HEELS estimator that employs the full LD ("Exact_LD"), the runtime of HEELS estimation using the "Seq_Band_LR", "PSD_Band_LR" or "Joint_Band_LR" approximation is on average 9.52%, 42.61% and 47.37% less. The improvement of computational efficiency pertains to our application of the Woodbury formula to the low-rank matrices, which helps circumvent the need of direct matrix inversions in HEELS estimation. For certain fine-tuned approximation settings (*e.g.* "Seq_Band_LR" with $b = 600$ and $r = 650$), the bias in heritability estimate based on cross-validation is even smaller than that

when the full LD is used, and MSE is 3.95% lower even though variance is slightly higher. Overall, across the hyperparameter settings, our proposed "Banded + LR" representation improves the computational efficiency of HEELS, without incurring much loss of accuracy and efficiency (**Table** 3).

In comparing the performances of the three "Banded + LR" strategies, we observe that they have varying degrees of sensitivity with respect to changes in hyperparameter values $b$ and $r$ (**Supplementary Figure** 10). Nevertheless, they converge in their approximating abilities (**Supplementary Figure** 11) and lead to small bias when $b$ and $r$ are sufficiently large (**Figure** 2). The bandwidth of the banded component plays a critical role in determining the magnitude of the bias in $h^2_{HEELS}$, whereas the number of the low-rank factors fine-tunes the estimator (**Supplementary Figure** 12). Among the three "Banded + LR" strategies, we recommend using "PSD_Band_LR" as the default, as it is least sensitive to changes in hyperparameter values, significantly reduces the runtime of HEELS estimation, and produces heritability estimates that are minimally biased (**Supplementary Figure** 10-13).

Finally, we developed a data-adaptive procedure to select the optimal hyperparameter values of $b$ and $r$ using cross-validation (CV). We refer to this procedure as "pseudo-validation", as it does not require actual held-out samples but only relies on simulated summary statistics (**Online Methods**). We validated the effectiveness of this procedure, by verifying that the optimal hyperparameter values obtained from this algorithm indeed yield $\hat{h}^2_{HEELS}$ estimates with the highest accuracy and precision (**Supplementary Figure** 14). To improve the computational efficiency of our tuning procedure, we implemented our own version of the incremental SVD algorithm[40], and adopted several computational techniques to speed up LD decomposition and the tuning algorithm (**Supplementary Notes**).

## Precise estimates of local heritability in the UK Biobank

We applied HEELS to estimate the local SNP-heritability of 30 anthropometric, medical and behavioral traits in the UK Biobank (**Online Methods**), using the LD blocks estimated by Berisa and Pickrell[41], which have been widely used to proxy approximately independent loci on the human genome[42–44]. We used Z-statistics for all phenotypes and interpret heritability on the liability scale for binary traits[8,45,46]. In line with our expectation based on the theoretical and simulation results, there is an exceptionally high concordance rate between the local heritability estimates from HEELS and those from REML ($r^2 = 0.98$ on average across the traits), which is the gold standard or the most efficient estimator but requires individual-level data (**Figure** 3). In contrast, the correlations between the GRE or HESS estimates and the REML estimates are weaker ($r^2 = 0.88$).

HEELS yields standard errors that are most comparable to those from individual-level data based REML and its relative efficiency is higher than the other existing summary-statistics based methods (**Supplementary Table** 3). We applied our hyperparameter tuning algorithm to obtain the optimal representations of the LD blocks, and found that these LD approximations generally perform well, yielding local $h^2$ estimates that are similar to those which are based on using the full LD (**Supplementary Figure** 16). The "Joint_Band_LR" strategy leads to lower accuracy loss compared to the other two strategies ("Seq_Band_LR" and "PSD_Band_LR"), plausibly due to its greater capacity of approximating the LD (**Supplementary Figure** 18).

## Applications of HEELS: Contrasting polygenicity of complex traits and diseases; Identifying risk loci with putative pleiotropic effects

Studies have shown that complex traits and diseases have different degrees of polygenicity[18]. Local heritability estimation provides a useful tool for comparing the polygenic architecture of complex traits, as its distribution pattern across the genome can reflect the degree of which the genetic variance is spread among loci[12,47]. Using local heritability estimated with greater precision from HEELS, we found varying degree of polygenicity across a wide range of traits (**Supplementary Figure** 19). For instance, the genetic variances of behavioral traits and anthropometric traits are more evenly spread out across the genome, whereas those of autoimmune diseases and allergic conditions, lipid traits are more localized and disproportionately dispersed on the genome. Alternatively, we evaluated the polygenicity of complex traits by examining the Pearson correlation coefficients between the genetic variances and the sizes of genomic regions. Indeed, we observed a near-perfect linear relationship between chromosome length and the fraction of $h^2_{SNP}$ explained by the chromosome for highly polygenic traits, such as BMI ($r^2 = 0.988$), WHR ($r^2 = 0.919$), educational attainment ($r^2 = 0.991$) and neuroticism ($r^2 = 0.990$). In contrast, the correlations are much lower for less polygenic traits, such as HDL ($r^2 = 0.295$), LDL ($r^2 = 0.537$) and autoimmune diseases ($r^2 = 0.757$) (**Supplementary Figure** 20).

Next, we used the local heritability estimated from HEELS to identify risk loci that are highly enriched for related traits and thus potentially host pleiotropic effects (**Figure** 3). Our results recapitulate known disease-associated loci from previous findings. For example, we found that the Major Histocompatibility Complex (MHC) (Chr6:31Mb-34Mb) shows strong pleiotropic signals for immunologically relevant diseases and leukocytes involved in innate immunity and inflammatory response. This region, often referred as the human leukocyte antigen (HLA), is known to be highly polygenic[48] and plays a key role

in the induction and regulation of immune responses[49–51]. We also identified Chr19:45Mb-46Mb as a heritability hotspot, which is highly enriched for multiple lipid and cardiometabolic traits while being moderately enriched for leukocytes. This locus harbors the gene cluster – *APOE, APOC1, APOC4, APOC2* – which code for apolipoproteins that are known to be responsible for controlling plasma lipid levels, with subsequent implications in cardiovascular pathology[52]. Our results highlight the probable presence of pleiotropic effects regulating lipid metabolism[53,54].

We identified two loci that may provide novel insights into the shared genetic basis between related traits and diseases. First, we found that the Chr19:16-18Mb locus is not only enriched for leukocytes such as platelet, lymphocyte and eosinophil, but also enriched for peak expiratory flow. Recent studies have associated platelet-to-lymphocyte ratio (PLR) with the severity of COVID-19, as the PLR of patients can indicate the degree of cytokine storm[55,56]. Our results suggest a putative link between the inflammatory markers and lung functions on a genetic level, and point to a specific region with pleiotropic effects. Further study of this locus may help illustrate its prognostic value for impaired pulmonary functions due to systemic inflammation.

Another region of particular interest is Chr4:143Mb-146 Mb, which is highly enriched for all three of the lung function traits in our analysis (**Supplementary Figure** 17). Although no evidence has directly implicated this region for lung functions, the 4q31 locus – which contains this block – has been prioritized in three GWAS studies[57–59] and one meta-analysis[60]. One plausible explanation for the pleiotropic effects of this region is that it hosts the Hedgehog (Hh)-interacting protein (*HHIP*) gene, which is associated with multiple pulmonery traits, such as FEV1/FVC ratio, COPD and lung cancer[61–64]. As a regulator of the Hh signalling pathway, *HHIP* is vital for embryonic lung development and is also involved in mature airway epithelial repair[65]. We hypothesize that the observed pleiotropic effects at this locus may be attributable to the *HHIP* gene and its broad influence on lung functionality[66,67].

## Discussion

To summarize, we have introduced in this paper a REML-based approach HEELS to obtain highly efficient variance component and heritability estimators in linear mixed models, using marginal GWAS association summary statistics and in-sample LD matrix (*i.e.* empirical covariance matrix of the genotypes). Our estimator only requires summary-level statistics, but yields a highly efficient heritability estimator that is comparable to the REML estimator based on individual-level data, and outperforms the existing summary statistics based heritability methods such as LDSC, GRE and HESS, in terms of statistical

efficiency. Another important contribution of our work is that we showcase the benefits of approximating the empirical LD matrix using a banded matrix plus a low-rank matrix. We demonstrate in simulations that this low-dimensional representation can improve the computational efficiency of our algorithm while incurring minimal loss of accuracy or efficiency in estimating $h^2_{SNP}$. The pseudo-validation procedure we propose provides a principled way to select hyperparameters, in contrast to existing approaches which make arbitrary assumptions about the structure of the empirical LD. We emphasize that our proposed low-dimensional approximations of the LD matrix can be employed more generally to facilitate the storage and sharing of large LD matrices.

We discuss several limitations of our current work and outline some future directions. First of all, our estimator is derived based on the LMM, which assumes a normal distribution of the causal effect sizes. This assumption can greatly simply our comparison between HEELS and REML and help illustrate our main conceptual advance relating to statistical efficiency, but it can be relaxed in two main ways. One is to allow zero effects and incorporate some "sparse" components in the architecture. This direction has been explored extensively in the Bayesian variable selection literature, leading to the development of BSVR[31], BSLMM[25] and the Bayesian-alphabet models[68]. The other direction is to adopt a more flexible heritability model, where SNP-specific contribution to the genetic variance can vary. This alternative scheme was put forward by authors of the LDAK[9] model and examined carefully in the subsequent works. We hope to extend HEELS such that it can accommodate both sparse effects and marker-specific weights, but this extension is challenging, especially because we hope to retain the high statistical efficiency of HEELS.

Second, we emphasize that HEELS requires and relies on the *in-sample* or empirical LD estimated from the GWAS sample. This restriction relates to our definition of heritability in the LMM where we condition on $\mathbf{X}$. Our preliminary analyses suggest that replacing the empirical LD with the population or out-of-sample LD in HEELS can lead to biased heritability estimates (**Supplementary Figure** 23). Further research is needed to investigate how out-of-sample LD can influence the statistical properties of the HEELS estimator. HEELS can potentially be applied to summary statistics from meta-analysis summary statistics released by large consortia. The key challenge of applying HEELS to meta-analyzed summary statistics is the appropriate integration of the LD information from multiple cohorts. One fruitful direction for future research is to explore ways to combine the low-dimensional representations of the LD matrices from different studies[69]. We note that in rare variant association tests, researchers routinely release both the vector of score statistics and the covariance (LD) matrix when publishing the results[70–73]. While this is not yet standard practice for common variant analysis, we advocate including

the in-sample LD matrix as part of the summary-level statistics when publishing GWAS results. This will enable a more precise characterization of the disease's genetic architecture (*e.g.* through HEELS), facilitate privacy-preserving benchmarking efforts[74], and alleviate challenges that are associated with external LD panels, especially in the context of meta-analyses[75].

Lastly, we note that the HEELS estimator is best suited for studying and contrasting the heritabilities of *local* genomic regions (up to 20,000 markers). Applying HEELS to a larger set of markers (*e.g.* genome-wide estimation) can be difficult for two main reasons. First, the computational cost of running HEELS scales quadratically with the number of markers, so the estimation will require additional computational resources when applied to larger blocks on the genome. Second, we determined in simulations that aggregating local heritability estimates can lead to bias in global heritability estimates, a phenomenon that has previously been observed[10] and may be attributable to LD leakage, *i.e.*, non-zero correlations between SNPs on different blocks[12,38]. To address these challenges, we are actively developing *scalable* algorithms to optimally approximate the LD using a banded + low-rank representations. This advance is not within the scope of our current work, but will make HEELS applicable to estimate the total genetic variance of hundreds of thousands of markers.

# Online Methods

## Statistical model

Let $\mathbf{y}$ be a length-$n$ vector that denotes the phenotypes of $n$ samples. Denote by $\mathbf{X} \in \mathbb{R}^{n \times p}$ the genotype matrix of $n$ individuals based on $p$ markers or SNPs. We standardize $\mathbf{X}$ and $\mathbf{y}$ such that the variance of the phenotype is 1 and the variance of each marker-specific genotype vector is $1/p$, or $diag(\mathbf{X}^\top \mathbf{X}/n) = 1/p$. Let $\mathbf{S}$ and $\mathbf{R}$ denote the the marginal association statistics and the in-sample LD matrix, *i.e.* $\mathbf{S} = \mathbf{X}^\top \mathbf{y}/\sqrt{n}$ and $\mathbf{R} = \mathbf{X}^\top \mathbf{X}/n$. Our goal is to develop a heritability estimator using the two statistics $(\mathbf{S}, \mathbf{R})$, which attains comparable statistical efficiency as the REML estimator based on individual-level data $(\mathbf{X}, \mathbf{y})$. We start by considering the likelihood function, assuming individual-level data can be accessed.

We use an additive genetic model for the phenotypes as $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, where $\boldsymbol{\beta}$ is a $p \times 1$ vector assumed to follow $N(0, \sigma_g^2 \mathbf{I}_p)$, and $\varepsilon$ is a length-$n$ vector distributed as $\varepsilon \sim N(0, \sigma_e^2 \mathbf{I}_n)$. Under these assumptions, $\mathbf{y} \sim N(0, \mathbf{V})$, where the variance-covariance matrix is $\mathbf{V} \equiv var(\mathbf{y}) = \sigma_g^2 \mathbf{X}\mathbf{X}^\top + \sigma_e^2 \mathbf{I}_n$. We define SNP-heritability conditional on $\mathbf{X}$ as the following,

$$h^2_{SNP} := \frac{Var(\mathbf{X}_i^\top \beta | \mathbf{X})}{Var(\mathbf{y}_i | \mathbf{X})} = \frac{Var(\mathbf{X}_i^\top \beta | \mathbf{X})}{Var(\mathbf{X}_i^\top \beta | \mathbf{X}) + \sigma_e^2} = \frac{tr(\sigma_g^2 \mathbf{I_p} \mathbf{X}^\top \mathbf{X})/n}{tr(\sigma_g^2 \mathbf{I_p} \mathbf{X}^\top \mathbf{X})/n + \sigma_e^2} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}.$$

The log-likelihood function for $(\sigma_g^2, \sigma_e^2)$ is,

$$\ell(\mathbf{y}; \sigma_g^2, \sigma_e^2) = -\frac{1}{2} ln|\mathbf{V}| - \frac{1}{2} \mathbf{y}^\top \mathbf{V}^{-1} \mathbf{y}. \tag{5}$$

Henderson developed a set of equations[16,76], known as the mixed model equations (MME), which maximize the joint density of the outcomes and the random effects,

$$\ell(\mathbf{y}, \boldsymbol{\beta}; \sigma_g^2, \sigma_e^2) = -\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \frac{1}{2\sigma_g^2} \boldsymbol{\beta}^\top \boldsymbol{\beta} - \frac{n}{2} log(\sigma_e^2) - \frac{p}{2} log(\sigma_g^2). \tag{6}$$

The Best Linear Unbiased Predictor (BLUP), which are estimates for the random effects from these MMEs, can be plugged into the score equations (2)-(3) to generate an iterative procedure for estimating the variance components.[17,77] We exploit the "dual" form of this algorithm, which gives rise to the HEELS estimator. Note that for simplicity, we have assumed all of the observable environmental factors have been projected out, but covariates (*i.e.* fixed effects) can be easily incorporated into the model by adopting the restricted maximum likelihood approach and using the projection matrix.[78,79]

## The HEELS procedure

HEELS uses the marginal association statistics $\mathbf{S} = \mathbf{X}^\top \mathbf{y}$ and the LD matrix $\mathbf{R} = \mathbf{X}^\top \mathbf{X}$ to solve for the variance component estimates that maximize the likelihood in (7), alternating between updating the BLUP estimates $\hat{\boldsymbol{\beta}}(\widehat{\sigma_g^2}, \widehat{\sigma_e^2}) = \left(\frac{\widehat{\sigma_e^2}}{\widehat{\sigma_g^2}}\mathbf{I}_p + \mathbf{R}\right)^{-1}\mathbf{S}$ and updating the variance component estimates $(\widehat{\sigma_g}^2, \widehat{\sigma_e}^2)$ until convergence. The marginal likelihood in (5) can be expressed using the joint likelihood (6) and the probability of the causal effects, using the partition theorem (**Supplementary Notes**),

$$\ell_{HEELS}(\mathbf{S}, \mathbf{R}; \sigma_g^2, \sigma_e^2) = -\frac{1}{2}log\left|\sigma_e^2 \mathbf{I}_n\right| - \frac{1}{2}log\left|\mathbf{I}_p + \frac{\sigma_g^2}{\sigma_e^2}\mathbf{R}\right| - \frac{1}{2\sigma_e^2}\left(\mathbf{y}^\top\mathbf{y} - \mathbf{S}^\top\left(\frac{\sigma_e^2}{\sigma_g^2}\mathbf{I}_p + \mathbf{R}\right)^{-1}\mathbf{S}\right). \quad (7)$$

It can be shown that the score equations of $(\sigma_g^2, \sigma_e^2)$ derived from the likelihood in (5) are identical to the score equations based on (7), giving rise to our HEELS updating equations (2) and (3) (**Supplemental notes**). The BLUP estimates can be viewed as ridge estimators of the joint effect sizes, where the penalty coefficient is set to the current value of $\sigma_e^2/\sigma_g^2$ at each iteration. We update $\sigma_g^2$ by the mean sum of squares of the BLUP estimates, with the average taken over the effective number of markers, $p - tr([\mathbf{W}]^{-1})$ to account for the bias introduced via the ridge penalty.

We emphasize that the HEELS algorithm does not entail any standalone individual-level data $\mathbf{X}$ or $\mathbf{y}$, but only depend on the summary statistics $\mathbf{S}, \mathbf{R}$. The involvement of phenotype variance $\mathbf{y}^\top\mathbf{y}$ in equation (3) does not necessarily imply a requirement to access $\mathbf{y}$, especially if the phenotypes and genotypes are standardized (see similar discussion in Zou *et al.*[23]). In practice, when only the Z-statistics are available, we scale the estimates of $\sigma_g^2, \sigma_e^2$ at each iteration as: $\sigma_g^{2(t+1)} = \frac{\sigma_g^{2(t)}}{\sigma_g^{2(t)} + \sigma_e^{2(t)}}, \sigma_e^{2(t+1)} = \frac{\sigma_e^{2(t)}}{\sigma_g^{2(t)} + \sigma_e^{2(t)}}$, which is equivalent to approximating $\mathbf{y}^\top\mathbf{y}/n$ by 1. We conducted extensive simulations to validate that the HEELS estimator remains unbiased when such an approximation is applied, although the efficiency of our estimator can be slightly inflated. We note that the HEELS procedure can be viewed as equivalent to an EM algorithm (**Supplementary Notes**), where the resulting estimates are stable numerically and belong to the parameter space[80]. This is an important feature that is not universally shared with other REML estimation algorithms such as the Anderson's algorithm[77,81] and the Newton-Raphson's algorithm[82].

## Standard error of the HEELS estimator

We derive the analytic variance of the HEELS estimator using the Fisher information matrix. For true values of the variance components $\sigma_e^2, \sigma_g^2$ and a positive definite variance-covariance matrix $\mathbf{V}$, the Fisher Information Matrix is (**Supplementary Notes**):

$$I(\sigma_e^2, \sigma_g^2; \mathbf{X}, \mathbf{y}) = \frac{1}{2} \begin{bmatrix} tr(\mathbf{V^{-2}}) & tr(\mathbf{V^{-1}XX^\top V^{-1}}) \\ tr(\mathbf{V^{-1}XX^\top V^{-1}}) & tr(\mathbf{V^{-1}XX^\top V^{-1}XX^\top}) \end{bmatrix}.$$

Using properties of trace and the identity of $\mathbf{W} := \frac{\sigma_e^2}{\sigma_g^2} I + \mathbf{X}^\top \mathbf{X}$, we can rewrite the Fisher information using the summary statistics and the LD matrix as the following[80]:

$$I(\sigma_e^2, \sigma_g^2; \mathbf{S}, \mathbf{R}) = \frac{1}{2} \begin{bmatrix} \frac{n-p}{\sigma_e^4} + \frac{1}{\sigma_g^4} tr(\mathbf{W^{-2}}) & \frac{1}{\sigma_g^4} tr(\mathbf{W^{-1}}) - \frac{\sigma_e^2}{\sigma_g^6} tr(\mathbf{W^{-2}}) \\ \frac{1}{\sigma_g^4} tr(\mathbf{W^{-1}}) - \frac{\sigma_e^2}{\sigma_g^6} tr(\mathbf{W^{-2}}) & \frac{p}{\sigma_g^4} - \frac{2\sigma_e^2}{\sigma_g^6} tr(\mathbf{W^{-1}}) + \frac{\sigma_e^4}{\sigma_g^8} tr(\mathbf{W^{-2}}) \end{bmatrix}, \tag{8}$$

where $\mathbf{W}$ is defined in the same way as in equation (1). To obtain a variance estimator for $h^2_{SNP} \equiv \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} := f(\sigma_g^2, \sigma_e^2)$, we apply the multivariate Delta method, and use the plug-in estimator:

$$\widehat{Var}(\hat{h}^2_{HEELS}) = \nabla f(\hat{\sigma}_e^2, \hat{\sigma}_g^2)^\top I(\hat{\sigma}_e^2, \hat{\sigma}_g^2; \mathbf{S}, \mathbf{R})^{-1} \nabla f(\hat{\sigma}_e^2, \hat{\sigma}_g^2), \tag{9}$$

where $\nabla f(\hat{\sigma}_e^2, \hat{\sigma}_g^2) = \left( -\frac{\hat{\sigma}_g^2}{(\hat{\sigma}_g^2 + \hat{\sigma}_e^2)^2}, \frac{\hat{\sigma}_e^2}{(\hat{\sigma}_g^2 + \hat{\sigma}_e^2)^2} \right)$. Importantly, (9) only involves $\mathbf{S}, \mathbf{R}$ and does not entail $\mathbf{X}, \mathbf{y}$, so we can compute the asymptotic variance of $h^2_{HEELS}$ without resorting to individual-level data.

## Low dimensional representations of the LD matrix

Several low-dimensional representations of the LD matrix have been employed in the literature – for example, LD has been approximated by banded matrices[36,37]; LD reference panel has been regularized via the truncated singular value decomposition (TSVD)[12,38]; other methods have adopted block-diagonal low-rank approximations to account for the LD[20,30]. As far as we know, no previous study has systematically compared the performances of different LD approximation strategies. We close this gap by introducing the "Banded + LR" representation of LD, which also provides a unified framework that facilitates comparisons of the existing LD matrix approximation approaches.

We adopt an optimization approach to solve for the best representation of LD with the "Banded + LR" structure, minimizing $||\mathbf{R} - \tilde{\mathbf{R}}||_F^2$, where $\tilde{\mathbf{R}}$ is the working approximating matrix in the form specified in the third column of **Table** 2. For example, the "Joint_Band_LR" approach simultaneously solves for the banded and the low-rank components of the approximation using PSD matrices,

$$\tilde{\mathbf{L}}^b, \tilde{\mathbf{U}}^r = \arg \min_{\mathbf{L}^b, \mathbf{U}^r \in L_p(\mathbb{R})} ||\mathbf{R} - \mathbf{L}^b \mathbf{L}^{b\top} - \mathbf{U}^r \mathbf{U}^{r\top}||_F^2 \tag{10}$$

where $L_p(\mathbb{R})$ denotes the set of $p \times p$ lower triangular matrices with real entries. The "PSD_Band_LR" strategy involves a two-step procedure,

$$\tilde{\mathbf{L}}^b = \arg \min_{\mathbf{L}^b, \in L_p(\mathbb{R})} ||\mathbf{R} \odot \mathbf{1}_b - \mathbf{L}^b \mathbf{L}^{b\top}||_F^2 \tag{11}$$

$$\tilde{\mathbf{U}}^r = \arg \min_{\mathbf{U}^r \in L_p(\mathbb{R})} ||\mathbf{R} - \tilde{\mathbf{L}}^b \tilde{\mathbf{L}}^{b\top} - \mathbf{U}^r \mathbf{U}^{r\top}||_F^2 \tag{12}$$

where $\mathbf{1}_b$ denotes a square matrix with only the $b$ central band equal to 1 and the rest set to 0, and $\odot$ is the Hadamard product. In both cases, we approximate $\mathbf{R}$ as $\tilde{\mathbf{L}}^b \tilde{\mathbf{L}}^{b\top} + \tilde{\mathbf{U}}^r \tilde{\mathbf{U}}^{r\top}$.

Below we briefly explain the differences between the low-dimensional representation strategies listed in **Table** 2. When the banded component is a diagonal matrix, the representation is a special case of the "Banded + LR" structure, with $b = 1$, which corresponds to a spiked covariance model[39,83]. We estimate the diagonal elements of the covariance matrix either by taking the smallest eigenvalue of a sub-sample of the LD matrix[84] ("Spike_LR"), or via joint optimization which solves for the diagonal and the low-rank matrices simultaneously. In the latter case, we allow elements of the diagonal matrix to be identical ("Spike_PSD") or different ("Spike_PSD_hetero").

The "Seq_Band_LR" strategy first bands the LD matrix and then performs low-rank decomposition on the residual off-banded matrix. It does not make any PSD assumption on the constituents of its solution. The main advantage of this strategy is that it most accurately represents the banded structure of the LD matrix, *i.e.* the banded component matches exactly the central band of the original LD matrix. A drawback of this approach is its lack of flexibility in approximating the off-central-band structure.

In contrast, the "Joint_Band_LR" strategy jointly approximates the banded and low-rank components using PSD matrices. It provides greater flexibility of approximation for all elements indiscriminately, as the elements of the banded matrix and the low-rank matrix are jointly optimized to minimize the Frobenius norm of the error matrix. A benefit of "Joint_Band_LR" over "Seq_Band_LR" is that due to its explicit minimization of the approximation error, its solution is, by definition, closer to the original LD matrix. A potential shortcoming of the "Joint_Band_LR" strategy, however, is that by imposing a PSD assumption on *both* the banded and the low-rank components of the LD representation, we restrict the solution to a particular set of PSD matrices and it is computationally slower.

To strike a balance between accurately representing the central band of the LD matrix and ensuring the computational efficiency of our algorithm, we developed a hybrid strategy, "PSD_Band_LR", which first approximates the banded component of the LD matrix using a PSD matrix followed by low-rank decomposing the residual matrix. On one hand, it differs from "Seq_Band_LR" in that the banded component of the approximation is guaranteed to be PSD. On the other hand, "PSD_Band_LR" differs from "Joint_Band _LR" in that it *sequentially* solves for the banded and low-rank constituents of the representation, as opposed to simultaneously. As a result, "PSD_Band_LR" preserves the original structure of the LD matrix better than "Joint_Band_LR" does, leading to a more accurate representation of the central band. "PSD_Band_LR" also exploits the PSD assumption, making it more computationally stable than "Seq_Band_LR" (**Supplementary Notes**).

## Comparison of different LD approximations

The existing LD approximation methods impose different assumptions on the structure of the LD matrix, and the motivation to induce the low-dimensional LD representation varies from case to case. We introduce a unified framework for contrasting the different LD approximation approaches in the context of heritability estimation. For a given LD representation – characterized by its structure ("Strategy") and the values of its hyper-parameters: the band size $b$ and the rank of the low rank component $r$ – we considered two measures to evaluate its performance: one is the LD approximation accuracy, measured by the ratio of the Frobenius-norms between the approximation matrix and the target matrix, $\frac{||\mathbf{R}_b+\mathbf{R}_r||_F}{||\mathbf{R}||_F}$; the other is the cross-validation (CV) bias in $\hat{h}^2_{HEELS}$, estimated using synthetic phenotypes generated from real in-sample genotypes. The first measure assesses the goodness of the approximation in general, irrespective of the genetic architecture of a trait, whereas the second measure evaluates the approximation performance in the context of heritability estimation.

For low-dimensional representations with smaller $b$ and $r$, the performance of the HEELS estimator is greatly influenced by the assumed structure of the approximation strategy and the hyperparameter values. For example, for each of the "Banded + LR" strategies, we could pinpoint the approximate range of an underlying transition point ($b_{min}$), which marks the minimum optimal value of $b$ – widening the bandwidth up to this value can substantially reduce the bias in $\hat{h}^2_{HEELS}$, but further increasing $b$ beyond this threshold value has a diminishing "de-biasing" effect on $h^2_{HEELS}$ (**Supplementary Figure** 12).

Banding the full LD first and then approximating the residual using a low-rank matrix ("Seq_Band_LR") yields the most unbiased heritability estimates when $b$ and $r$ are sufficiently large, but can produce estimates

with large bias if $b$ and $r$ are not appropriately chosen; approximating the LD matrix by simultaneously solving for the banded and low-rank components ("Joint_Band_LR") leads to less-biased estimates even when $b$ and $r$ are sub-optimal, although solving for the approximation is more computationally expensive than "Seq_Band_LR"; the hybrid strategy which approximates the banded component using a PSD matrix first and finds the low-rank decomposition of the residual ("PSD_Band_LR") is least sensitive to changes in hyperparameter values and produces heritability estimates that are most stable across LD approximation settings.

## Algorithms for hyperparameter tuning

An important aspect of our low-dimensional LD representation algorithm is the tuning of the hyper-parameters. While heuristics or prior knowledge about the structure of the LD can be used to determine the optimal values of $(b, r)$, we used a more principled way to evaluate the performance of low-dimensional LD representation. We propose using a data-adaptive procedure to identify the best low-dimensional representation of the LD matrix, using synthetic phenotypic data and cross-validation (**Supplementary Notes**).

To facilitate the searching of the optimal number of low-rank factors, $r^*$, we implemented our own version of the "incremental SVD algorithm", where the number of low-rank factors increases step by step and the performance of the HEELS estimator is evaluated dynamically. To speed up the low-rank decomposition, we replaced the *exact* solutions based on direct eigen-decomposition with the *approximate* solutions, and compared two approaches – one is the optimization approach ("optim"), where we solve $\arg\min_{U^r \in \mathbb{K}^{p\times p}} ||\mathbf{R}_{resid} - \mathbf{U}^{r\top}\mathbf{U}^r||_F^2$, and $\mathbf{R}_{resid}$ is the residual off-banded component to be decomposed or approximated and $\mathbb{K}^{p\times p}$ is the collection of lower triangular matrices; the other approach is based on random sketching ("random"), where we applied fast PCA to decompose $\mathbf{R}_{resid}$ via SVD[84]. The simulation results indicate that using the optimization approach is less prone to bias and is more robust to variation in the hyperparameter values (**Supplementary Figure** 15). While we observe larger bias when the random sketching SVD approach is used to approximate the low-rank component, it can be more advantageous in applications with larger problem size.

## Simulation framework

We obtained $332,340$ unrelated White British individuals by extracting samples with self-reported British ancestry who are more than third-degree relatives and excluding subjects with putative sex chromosome

aneuploidy. We used genotyped array SNPs only to ensure high quality of measurement, and filtered out variants with genotype missingness > 0.01 or have MAF greater than 0.01. For experiments that involve subsets of the UKB individuals, we recalculated minor allele frequency and re-applied the filters above to the subsample. The environmental effects are drawn independently from the genetic effects for each individual. This ensures that population structure or cryptic relatedness among individuals will have minimal impact on our estimates in the simulations.

We simulated phenotypes with different genetic architectures using real genotypes from the UK Biobank, varying sample size ($n$), the ratio of $n$ to the number of markers ($p$) and the degree of polygenicity ($p_{causal}/p$). Given the raw genotype matrix $\mathbf{G}$, we first standardized the genotype matrix: for each SNP $j$ and individual $i$, we generated $\mathbf{X}_{ij} = (\mathbf{G}_{ij} - 2f_j)/\sqrt{2f_j(1-f_j)}$, where $\mathbf{G}_{ij} \in \{0,1,2\}$ and $f_j$ is the in-sample minor allele frequency of SNP $j$. For a given degree of polygenicity, we randomly sampled $p_{causal}$ markers. Denote the set of causal markers by $C$. We drew standardized effect sizes from the distribution, $\beta_j \sim N(0, \sigma_g^2/p_{causal}), \forall j \in C$, and simulated the phenotype of the $i$-th individual using $y_i = \sum_{j \in C} X_{ij}\beta_j + \varepsilon_i$, where $\varepsilon \sim N(0, (1-\sigma_g^2)\mathbf{I}_n)$. For each genetic architecture, we generated 100 phenotype replicates and obtained 100 estimates using each of the methods we included in the benchmark. Given phenotypes $\mathbf{y} = (y_1, ..., y_N)^\top$ and genotypes $\mathbf{X} = (\mathbf{x}_1^\top, ..., \mathbf{x}_N^\top)^\top$, we computed the marginal association statistics using OLS: $\hat{\beta}_j = \mathbf{X}_j^\top \mathbf{y}/n$. Unless otherwise specified, we used the exact in-sample LD without approximation in the HEELS estimation.

We applied two software to compute the REML estimates. We used GCTA when sample size is small and used BOLT-REML when $n$ exceeds 30,000. For experiments with varying sample size, we selected individuals randomly from the full QC-ed UK Biobank data. To create an imbalance of sample size across causal markers, we ranked the marker-specific sample size and defined the lowest 10 percentile of the SNPs as "low-N" SNPs. It is important to note that we did not include any markers that would have been filtered out by conventional QC steps, such as call rate of genotypes > 0.90, *i.e.* define low-sample size relatively and dynamically *within* the set of high-quality SNPs that have passed QCs, with varying sample size across different settings.

Our primary metric of interest is relative efficiency, which is the ratio of the variance of the REML estimator and the variance of a given heritability estimator. We also compared the bias and the mean squared error of the estimators across different simulation settings.

### UKB empirical analysis

We applied our method to the unrelated white British individuals in the UK Biobank, using common variants with MAF > 1% on the UK Biobank Axiom array ($n = 332,430$, $p = 533,169$). We estimated the local SNP-heritability of 33 phenotypes from different trait domains, including anthropometric traits (height, weight, BMI and waist-to-hip ratio), hematological traits (mean corpuscular volume, red blood cell count, white blood cell count, platelet count, eosinophils, lymphocytes, mean platelet volume and monocytes), lipid or metabolic traits (LDL, HDL, triglycerides, total cholesterol), lung function traits (forced expiratory volume, forced vital capacity, peak expiratory flow rate), behavioral traits (smoking behavior, educational attainment, neuroticism) and immunologically relevant traits (autoimmune conditions, asthma, eczema and dermatologic diseases).

We used PLINK to exclude SNPs with MAF < 1%, genotype missingness > 1%, genotyping rates < 90%, and Hardy-Weinberg disequilibrium p-value > 1e-6. We generated the genome-wide association statistics using the $--$linear association analyses in PLINK, controlling for age, sex, and the top 40 genetic principal components provided by the UKB[85]. We used BOLT-REML to compute the REML estimates in the UK Biobank, and used the HESS software to compute both the HESS estimates (default setting) and the GRE estimates (*i.e.* no regularization of the in-sample LD matrix), after verifying that the estimates from the two methods indeed coincide when the number of eigenvectors used in TSVD equals to the rank of the genotype matrix. We used the approximately independent loci estimated by Berisa and Pickrell[41] to define the regions as units of analysis (1.6 Mb on average or 300-400 markers per block using the genotype array data). The 1,703 non-overlapping regions have been been widely used in previous studies[42–44] and have been demonstrated to capture the true correlation structures among genotypes reasonably well[86].

## Data availability

This work used genotyped and phenotypes from the UK Biobank study (https://www.ukbiobank.ac.uk) and our access was approved under application number 52008.

## Code availability

Our method (HEELS) has been implemented as an open-source Python package, available at: https://github.com/huilisabrina/HEELS. The following software packages were also used in simulation studies and real data analyses: GCTA (https://yanglab.westlake.edu.cn/software/gcta/#Download); LD

Score regression (https://github.com/bulik/ldsc); GRE (https://github.com/bogdanlab/h2-GRE); HESS (https://github.com/huwenboshi/hess); BOLT-LMM v2.3.4 (https://data.broadinstitute.org/alkesgroup/BOLT-LMM/). We also used Python (3.6.3) to perform statistical analyses and used R (3.6.3) for data visualization.

## Acknowledgements

We are very grateful to Luke O'Connor, Huwenbo Shi, Alkes Price, Samuel Kou, Ben Neale and Shamil Sunyaev for their helpful discussions and feedback. This work was supported by grants R35-CA197449, U19-CA203654, R01-HL163560, U01-HG012064, and U01-HG009088 (X. Lin).

## Author contributions statement

H.L., R.M. and X.L. conceived and designed the experiments. H.L. performed the experiments and the statistical analyses. H.L. wrote the manuscript with the participation of R.M. and X.L.

## References


1. Yang, J. *et al.* Common snps explain a large proportion of the heritability for human height. *Nat. genetics* **42**, 565–569 (2010).

2. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. genetics* **47**, 1385 (2015).

3. Bulik-Sullivan, B. K. *et al.* Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. genetics* **47**, 291–295 (2015).

4. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. genetics* **47**, 1228–1235 (2015).

5. Gazal, S. *et al.* Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nat. genetics* **49**, 1421–1427 (2017).

6. Speed, D. & Balding, D. J. Sumher better estimates the snp heritability of complex traits from summary statistics. *Nat. genetics* **51**, 277–284 (2019).

7. Visscher, P. M. *et al.* 10 years of gwas discovery: biology, function, and translation. *The Am. J. Hum. Genet.* **101**, 5–22 (2017).

8. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *The Am. J. Hum. Genet.* **88**, 294–305 (2011).

9. Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved heritability estimation from genome-wide snps. *The Am. J. Hum. Genet.* **91**, 1011–1021 (2012).

10. Hou, K. *et al.* Accurate estimation of snp-heritability from biobank-scale data irrespective of genetic architecture. *Nat. genetics* **51**, 1244–1251 (2019).

11. Wu, Y. & Sankararaman, S. A scalable estimator of snp heritability for biobank-scale data. *Bioinformatics* **34**, i187–i194 (2018).

12. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the genetic architecture of 30 complex traits from summary association data. *The Am. J. Hum. Genet.* **99**, 139–153 (2016).

13. Ma, R. & Dicker, L. H. The mahalanobis kernel for heritability estimation in genome-wide association studies: fixed-effects and random-effects methods. *arXiv preprint arXiv:1901.02936* (2019).

**14.** Zhou, X. A unified framework for variance component estimation with summary statistics in genome-wide association studies. *The annals applied statistics* **11**, 2027 (2017).

**15.** Evans, L. M. *et al.* Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat. genetics* **50**, 737–745 (2018).

**16.** Henderson, C. R. Maximum likelihood estimation of variance components. *Unpubl. manuscript. Ithaca, NY: Dep. Animal Sci. Cornell Univ.* **14850** (1973).

**17.** Harville, D. A. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. statistical association* **72**, 320–338 (1977).

**18.** O'Connor, L. J. *et al.* Extreme polygenicity of complex traits is explained by negative selection. *The Am. J. Hum. Genet.* **105**, 456–476 (2019).

**19.** Weiner, D. J. *et al.* Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. genetics* **49**, 978–985 (2017).

**20.** Ning, Z., Pawitan, Y. & Shen, X. High-definition likelihood inference of genetic correlations across human complex traits. *Nat. genetics* **52**, 859–864 (2020).

**21.** Zhang, Y. *et al.* Comparison of methods for estimating genetic correlation between complex traits using gwas summary statistics. *Briefings bioinformatics* **22**, bbaa442 (2021).

**22.** Berrandou, T.-E., Balding, D. & Speed, D. Ldak-gbat: fast and powerful gene-based association testing using summary statistics. *The Am. J. Hum. Genet.* **110**, 23–29 (2023).

**23.** Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the "sum of single effects" model. *PLoS Genet.* **18**, e1010299 (2022).

**24.** Wu, Y. *et al.* Fast estimation of genetic correlation for biobank-scale data. *The Am. J. Hum. Genet.* **109**, 24–32 (2022).

**25.** Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with bayesian sparse linear mixed models. *PLoS genetics* **9**, e1003264 (2013).

**26.** Zhu, X. & Stephens, M. Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. *The annals applied statistics* **11**, 1561 (2017).

**27.** Gazal, S., Marquez-Luna, C., Finucane, H. K. & Price, A. L. Reconciling s-ldsc and ldak functional enrichment estimates. *Nat. genetics* **51**, 1202–1204 (2019).

**28.** Franklin, C. H. Maximum likelihood estimation for generalized linear models. *Inter-University Consortium for political Soc. Res. (ICPSR). ICPSR, Madison* (2005).

**29.** Serfling, R. J. *Approximation theorems of mathematical statistics* (John Wiley & Sons, 2009).

**30.** Song, S., Jiang, W., Zhang, Y., Hou, L. & Zhao, H. Leveraging ld eigenvalue regression to improve the estimation of snp heritability and confounding inflation. *The Am. J. Hum. Genet.* (2022).

**31.** Guan, Y. & Stephens, M. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. (2011).

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

**33.** Jiang, J., Li, C., Paul, D., Yang, C. & Zhao, H. On high-dimensional misspecified mixed model analysis in genome-wide association study. *The Annals Stat.* **44**, 2127–2160 (2016).

**34.** Bonnet, A., Gassiat, E. & Lévy-Leduc, C. Heritability estimation in high dimensional sparse linear mixed models. *Electron. J. Stat.* **9**, 2099–2129 (2015).

**35.** Wen, X. & Stephens, M. Using linear predictors to impute allele frequencies from summary or pooled genotype data. *The annals applied statistics* **4**, 1158 (2010).

**36.** Weissbrod, O. *et al.* Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* **52**, 1355–1363 (2020).

**37.** Vilhjálmsson, B. J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *The american journal human genetics* **97**, 576–592 (2015).

**38.** Shi, H., Mancuso, N., Spendlove, S. & Pasaniuc, B. Local genetic correlation gives insights into the shared genetic architecture of complex traits. *The Am. J. Hum. Genet.* **101**, 737–751 (2017).

**39.** Johnstone, I. M. On the distribution of the largest eigenvalue in principal components analysis. *The Annals statistics* **29**, 295–327 (2001).

**40.** Brand, M. Incremental singular value decomposition of uncertain data with missing values. In *European Conference on Computer Vision*, 707–720 (Springer, 2002).

**41.** Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283 (2016).

**42.** Sohail, M. *et al.* Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *Elife* **8** (2019).

**43.** Ruderfer, D. M. *et al.* Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705–1715 (2018).

**44.** Mancuso, N. *et al.* Probabilistic fine-mapping of transcriptome-wide association studies. *Nat. genetics* **51**, 675–682 (2019).

**45.** Ojavee, S. E., Kutalik, Z. & Robinson, M. R. Liability-scale heritability estimation for biobank studies of low prevalence disease. *The Am. J. Hum. Genet.* (2022).

**46.** Zhu, H. & Zhou, X. Statistical methods for snp heritability estimation and partition: A review. *Comput. Struct. Biotechnol. J.* **18**, 1557–1568 (2020).

**47.** Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).

**48.** Tumer, G., Simpson, B. & Roberts, T. K. Genetics, human major histocompatibility complex (mhc). *StatPearls [Internet]. Treasure Island (FL): StatPearls Publ.* (2020).

**49.** Mosaad, Y. Clinical role of human leukocyte antigen in health and disease. *Scand. journal immunology* **82**, 283–306 (2015).

**50.** Simmonds, M. & Gough, S. The hla region and autoimmune disease: associations and mechanisms of action. *Curr. genomics* **8**, 453–465 (2007).

**51.** Thorsby, E. & Lie, B. A. Hla associated genetic predisposition to autoimmune diseases: Genes involved and possible mechanisms. *Transpl. immunology* **14**, 175–182 (2005).

**52.** Bennet, A. M. *et al.* Association of apolipoprotein e genotypes with lipid levels and coronary risk. *Jama* **298**, 1300–1311 (2007).

**53.** Pirim, D. *et al.* Apolipoprotein e-c1-c4-c2 gene cluster region and inter-individual variation in plasma lipoprotein levels: a comprehensive genetic association study in two ethnic groups. *PloS one* **14**, e0214060 (2019).

**54.** Fuior, E. V. & Gafencu, A. V. Apolipoprotein c1: its pleiotropic effects in lipid metabolism and beyond. *Int. journal molecular sciences* **20**, 5939 (2019).

**55.** Qu, R. *et al.* Platelet-to-lymphocyte ratio is associated with prognosis in patients with coronavirus disease-19. *J. medical virology* **92**, 1533–1541 (2020).

**56.** Sarkar, S., Kannan, S., Khanna, P. & Singh, A. K. Role of platelet-to-lymphocyte count ratio (plr), as a prognostic indicator in covid-19: A systematic review and meta-analysis. *J. Med. Virol.* **94**, 211–221 (2022).

**57.** Pillai, S. G. *et al.* A genome-wide association study in chronic obstructive pulmonary disease (copd): identification of two major susceptibility loci. *PLoS genetics* **5**, e1000421 (2009).

**58.** Wilk, J. B. *et al.* A genome-wide association study of pulmonary function measures in the framingham heart study. *PLoS genetics* **5**, e1000429 (2009).

**59.** Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat. genetics* **42**, 36–44 (2010).

**60.** Hancock, D. B. *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat. genetics* **42**, 45–52 (2010).

**61.** Van Durme, Y. M. *et al.* Hedgehog-interacting protein is a copd susceptibility gene: the rotterdam study. *Eur. Respir. J.* **36**, 89–95 (2010).

**62.** van der Plaat, D. A. *et al.* Genome-wide association study on the fev1/fvc ratio in never-smokers identifies hhip and fam13a. *J. Allergy Clin. Immunol.* **139**, 533–540 (2017).

**63.** Amos, C. I. *et al.* Genome-wide association scan of tag snps identifies a susceptibility locus for lung cancer at 15q25. 1. *Nat. genetics* **40**, 616–622 (2008).

**64.** Young, R. *et al.* Chromosome 4q31 locus in copd is also associated with lung cancer. *Eur. Respir. J.* **36**, 1375–1382 (2010).

**65.** Zeng, L.-H. *et al.* Hedgehog signaling: Linking embryonic lung development and asthmatic airway remodeling. *Cells* **11**, 1774 (2022).

**66.** Li, Y. *et al.* Hedgehog interacting protein (hhip) represses airway remodeling and metabolic reprogramming in copd-derived airway smooth muscle cells. *Sci. reports* **11**, 1–12 (2021).

**67.** Ortega-Martínez, A. *et al.* Participation of hhip gene variants in copd susceptibility, lung function, and serum and sputum protein levels in women exposed to biomass-burning smoke. *Diagnostics* **10**, 734 (2020).

**68.** Habier, D., Fernando, R. L., Kizilkaya, K. & Garrick, D. J. Extension of the bayesian alphabet for genomic selection. *BMC bioinformatics* **12**, 1–12 (2011).

**69.** Li, X. *et al.* Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole-genome sequencing studies. *Nat. genetics (in press)* (2022).

**70.** Lee, S., Teslovich, T. M., Boehnke, M. & Lin, X. General framework for meta-analysis of rare variants in sequencing association studies. *The Am. J. Hum. Genet.* **93**, 42–53 (2013).

**71.** Liu, D. J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat. genetics* **46**, 200–204 (2014).

**72.** He, Z., Xu, B., Lee, S. & Ionita-Laza, I. Unified sequence-based association tests allowing for multiple functional annotations and meta-analysis of noncoding variation in metabochip data. *The Am. J. Hum. Genet.* **101**, 340–352 (2017).

**73.** Li, X. *et al.* Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nat. genetics* **52**, 969–983 (2020).

**74.** Witteveen, M. J. *et al.* Publicly available privacy-preserving benchmarks for polygenic prediction. *bioRxiv* (2022).

**75.** Yang, Z., Wang, C., Khan, A., Kiryluk, K. & Ionita-Laza, I. Carma: Novel bayesian model for fine-mapping with high-dimensional functional data. *preprint Columbia Univ.* (2022).

**76.** Henderson, C. Estimation of variances in animal model and reduced animal model for single traits and single records. *J. Dairy Sci.* **69**, 1394–1402 (1986).

**77.** Anderson, T. W. Statistical inference for covariance matrices with linear structure. *Multivar. Analysis II* 55–66 (1969).

**78.** Patterson, H. D. & Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554 (1971).

**79.** Patterson, H. Maximum likelihood estimation of components of variance. In *Proceeding Eight International Biometric Conference, 1975* (Biometric Soc., 1975).

**80.** Searle, S. R., Casella, G. & McCulloch, C. E. *Variance components*, vol. 391 (John Wiley & Sons, 2009).

**81.** Miller, J. J. Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals Stat.* 746–762 (1977).

**82.** Hemmerle, W. J. & Hartley, H. O. Computing maximum likelihood estimates for the mixed aov model using the w transformation. *Technometrics* **15**, 819–831 (1973).

**83.** Donoho, D. L., Gavish, M. & Johnstone, I. M. Optimal shrinkage of eigenvalues in the spiked covariance model. *Annals statistics* **46**, 1742 (2018).

**84.** Halko, N., Martinsson, P.-G. & Tropp, J. A. Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions, 2009. *URL http://arxiv. org/abs/0909.4061. oai: arXiv. org* **909** (2009).

**85.** Bycroft, C. *et al.* The uk biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

**86.** Zhao, B., Zheng, S. & Zhu, H. On block-wise and reference panel-based estimators for genetic data prediction in high dimensions. *arXiv preprint arXiv:2203.12003* (2022).

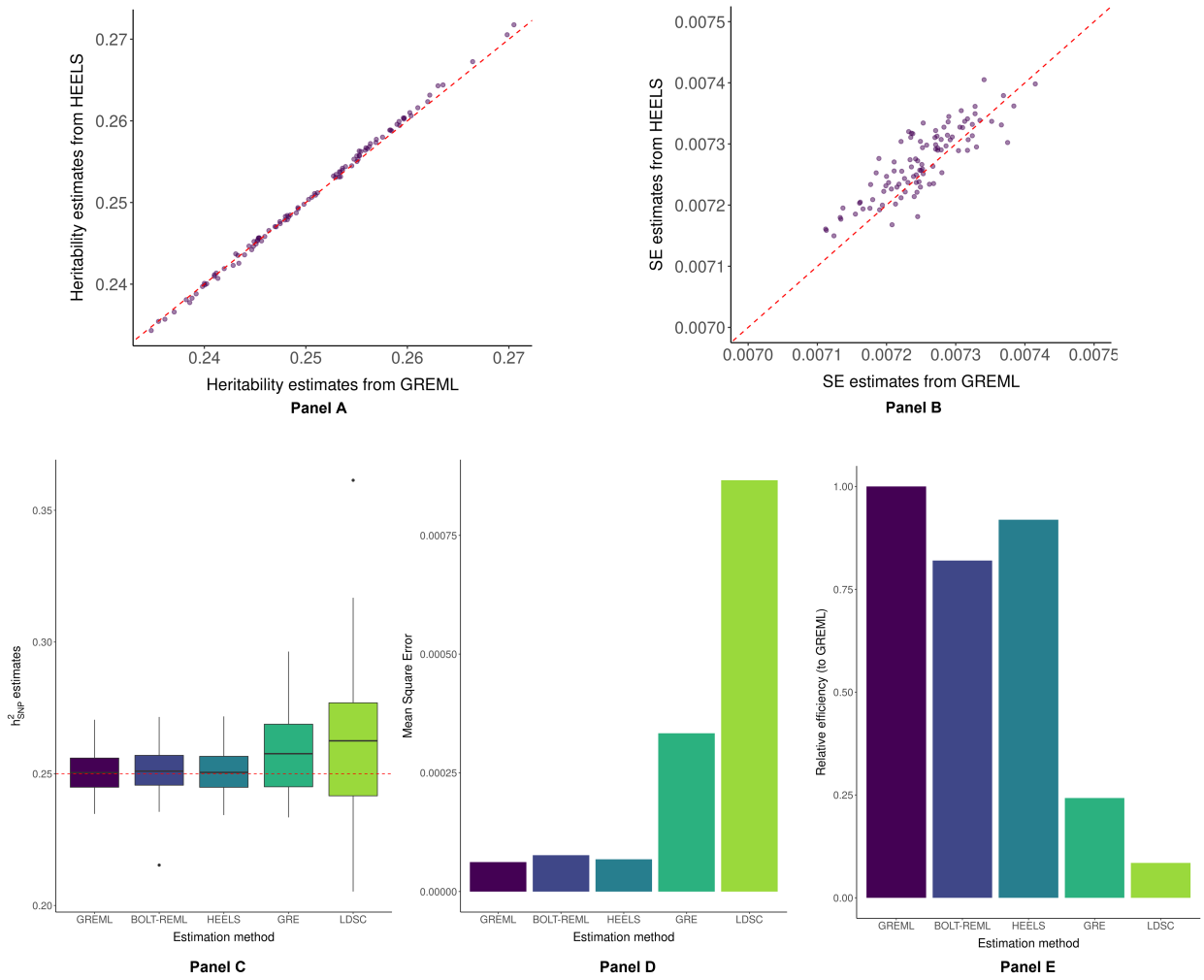


**Figure 1. Comparison of the performance of HEELS with different methods using simulation studies.** Simulated phenotypes using real genotypic data from the UK Biobank, array SNPs on chromosome 22 with MAF > 0.01 ($n = 30,000, p = 9,205$). **Panel A**: Local $h^2_{SNP}$ estimates from HEELS using summary statistics vs GREML using individual-level data. Red-dotted line: $y = x$. **Panel B**: Analytical SE estimates from HEELS vs GREML. Red-dotted line: $y = x$. **Panel C**: Distribution of the $h^2_{SNP}$ estimates using different methods: GREML and BOLT-REML use individual level data; HEELS, GRE and LDSC use summary statistics. Red-dotted line: true $h^2_{SNP}$ of 0.25. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles). The upper (lower) whisker extends from the hinge to the largest (smallest) value no further than $1.5 \times IQR$ from the hinge (where IQR is the inter-quartile range). Data beyond the end of the whiskers are called "outlying" points and are plotted individually. **Panel D**: MSEs of the $h^2_{SNP}$ estimates using different methods. **Panel E**: Relative efficiency of different methods compared to GREML.

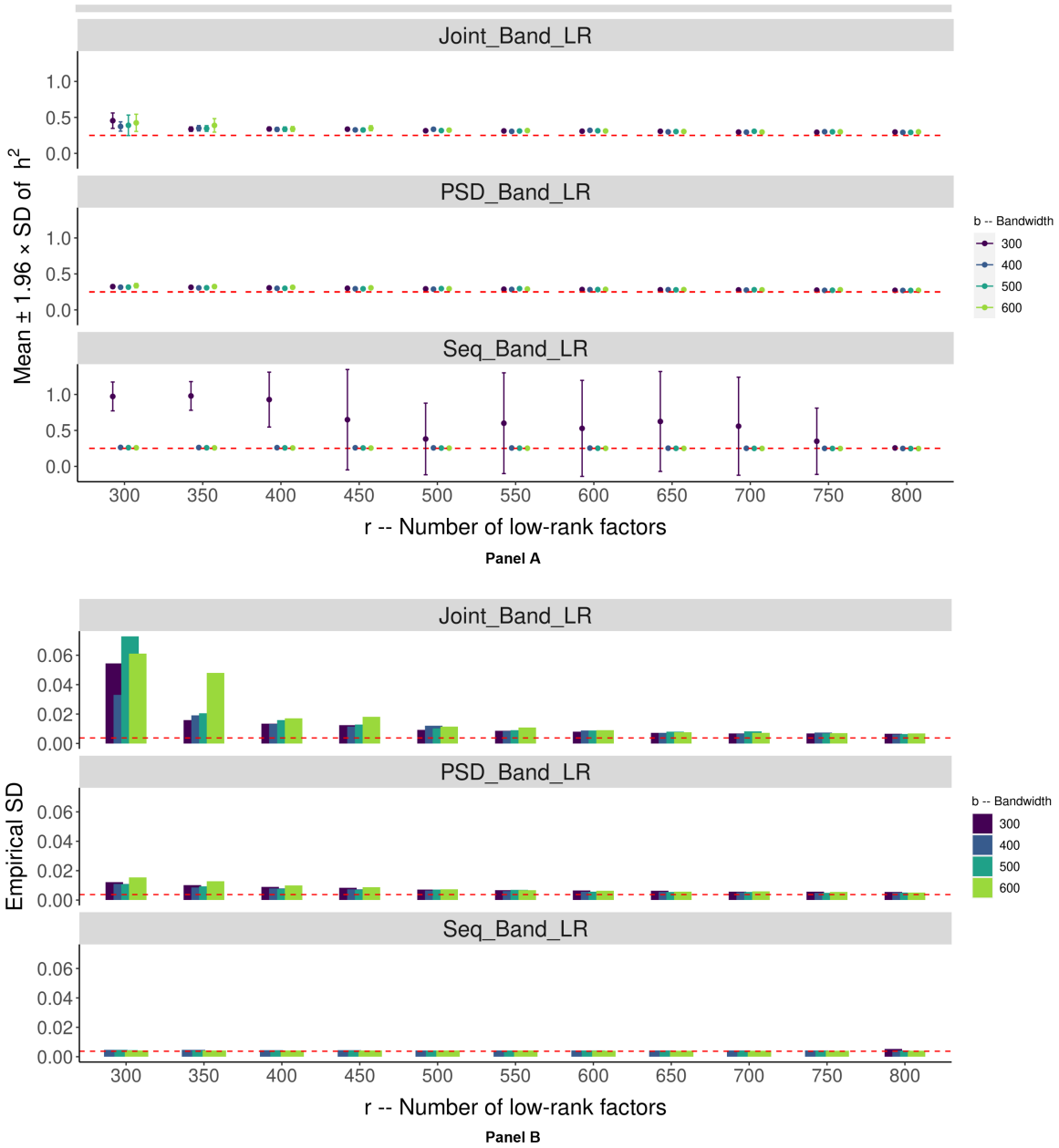


**Figure 2. Comparison of the performance of different LD approximation strategies using simulation studies.** Simulated phenotypes using real genotypic data from UK Biobank, array SNPs on chromosome 22 with MAF > 0.01 ($n = 332,430, p = 9,220$). Three LD approximation methods using the Banded + LR structure with different choices of bandwidths $b$ and ranks $r$ are compared: Joint_Band_LR, PSD_Band_LR, Seq_Band_LR (**Online Methods** and **Supplemental Notes**). **Panel A**: $h^2_{SNP}$ estimates. Red reference line represents the true heritability value of 0.25. The upper (lower) whisker extends from the mean to the values $1.96 \times SE$ above (below) the mean. **Panel B**: Empirical SD. Red reference line corresponds to the empirical SD when no approximation is applied (*i.e.* full LD is used).

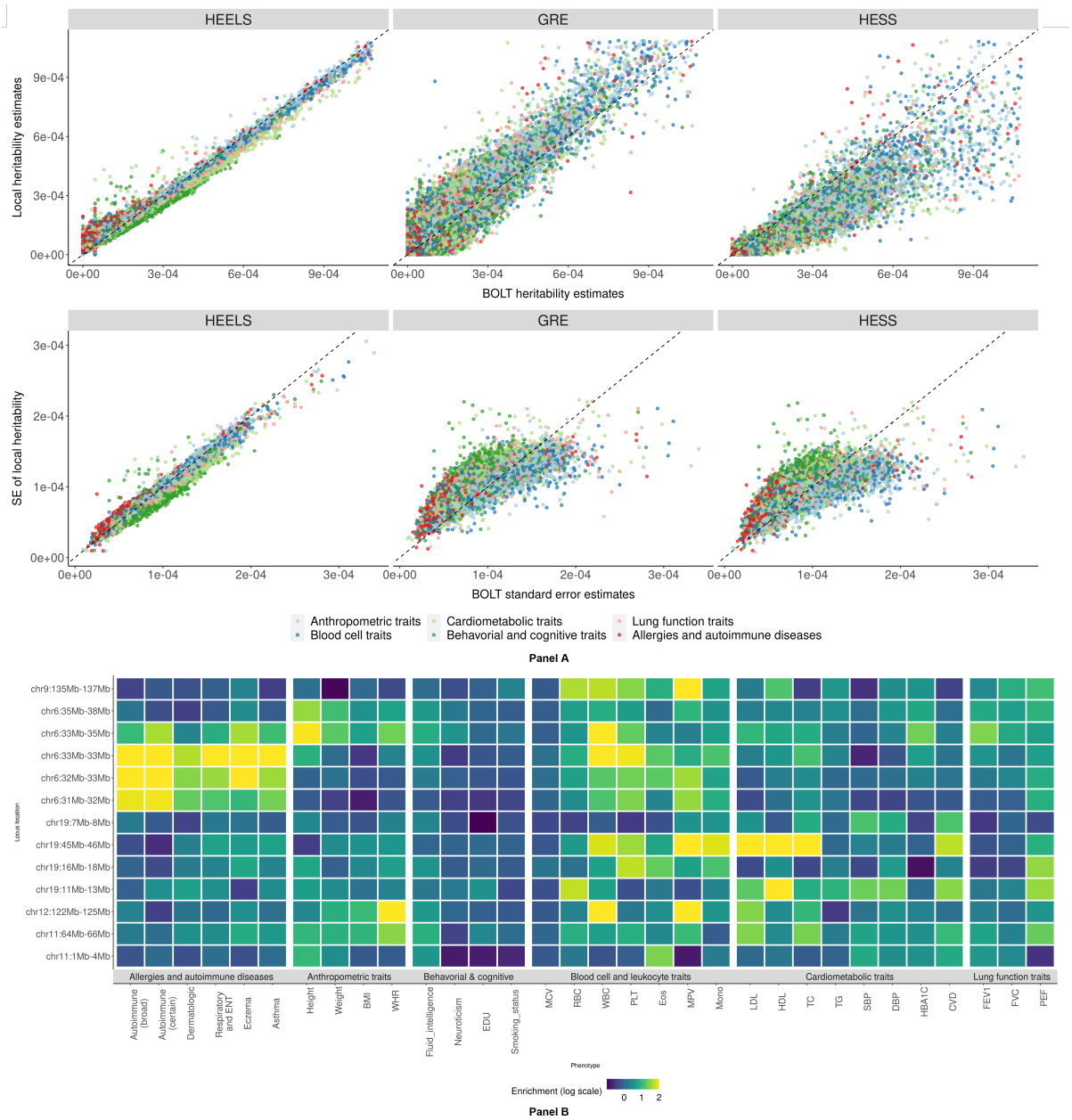


**Figure 3. Analysis results from application of HEELS to UKB. Panel A**: Comparison of heritability estimates and standard errors between BOLT-REML and summary-statistics-based methods (HEELS, GRE, HESS) ($n = 332,340$). Each dot represents a local estimate at one locus for a given trait. Red dashed line: $y = x$. **Panel B**: Heritability enrichment of multiple diseases and traits at putative pleiotropic risk loci. Loci locations on the Y-axis are denoted by the start and end positions (in Mbp). The shade of each box represents the enrichment of heritability for the given region and trait, with log transformation.

| Polygenicity | Sample size | Empirical variance of heritability estimates | | | | Relative efficiency (to REML) | | |
|---|---|---|---|---|---|---|---|---|
| | | REML | HEELS | GRE | LDSC | HEELS | GRE | LDSC |
| 1 | 332k | 1.43E-05 | 1.44E-05 | 7.24E-05 | 1.65E-04 | 99.44% | 19.81% | 8.68% |
| 1 | 140k | 2.15E-05 | 2.20E-05 | 6.53E-05 | 1.77E-04 | 97.82% | 32.94% | 12.17% |
| 1 | 40k | 4.05E-05 | 4.07E-05 | 1.32E-04 | 5.45E-04 | 99.54% | 30.76% | 7.43% |
| 1 | 30k | 6.19E-05 | 6.74E-05 | 2.55E-04 | 7.32E-04 | 91.91% | 24.25% | 8.46% |
| 0.5 | 30k | 6.80E-05 | 7.26E-05 | 2.58E-04 | 8.71E-04 | 93.69% | 26.38% | 7.81% |
| 0.2 | 30k | 9.35E-05 | 1.00E-04 | 2.98E-04 | 8.89E-04 | 93.16% | 31.42% | 10.52% |
| 0.1 | 30k | 1.16E-04 | 1.29E-04 | 4.39E-04 | 8.91E-04 | 90.17% | 26.41% | 13.02% |

**Table 1. Simulation results of the relative statistical efficiencies of different summary-statistics based $h^2_{SNP}$ estimates (HEELS, GRE, HESS) compared with individual-level data based GREML.** Sparsity and sample sizes are varied across settings. REML estimates are computed using GCTA for small sample (30k), and BOLT-LMM for large samples (40k, 140k, 332k). Statistical relative efficiency: ratio of the variance of REML and the variance of an estimator. Polygenicity: The proportion of the markers that are causal.

| Structure | Strategy name | Form of Decomposition | Property of the approximation |
|---|---|---|---|
| Spiked covariance | Spike_LR | $Diag(\sigma^2, ...\sigma^2) + \mathbf{U_r}\mathbf{\Lambda}_r\mathbf{U_r}^\top$ | PSD as long as $r << p$ |
| Spiked covariance | Spike_PSD | $Diag(\sigma^2, ...\sigma^2) + \mathbf{L_r}\mathbf{L_r}^\top$ | PSD guaranteed |
| Spiked covariance | Spike_PSD_hetero | $Diag(\sigma_1^2, ...\sigma_p^2) + \mathbf{L_r}\mathbf{L_r}^\top$ | PSD guaranteed |
| Banded + Low-rank | Seq_Band_LR | $\mathbf{R}_b + \mathbf{U_r}\mathbf{\Lambda}_r\mathbf{U_r}^\top$ | PSD not guaranteed |
| Banded + Low-rank | PSD_Band_LR | $\mathbf{L_b}\mathbf{L_b}^\top + \mathbf{U_r}\mathbf{\Lambda}_r\mathbf{U_r}^\top$ | PSD as long as $r << p$ |
| Banded + Low-rank | Joint_Band_LR | $\mathbf{L_b}\mathbf{L_b}^\top + \mathbf{L_r}\mathbf{L_r}^\top$ | PSD guaranteed |

**Table 2. Summary of the proposed LD approximation methods.** Six methods are proposed to approximate the LD matrix using either 1) a the spiked covariance matrix (Spike_LR, Spike_PSD, Spike_PSD_hetero) or 2) the sum of a banded and a low-rank matrix (Seq_Band_LR, PSD_Band_LR, and Joint_Band_LR). See **Online Methods** and **Supplemental Notes** for further explanations. PSD: positive semi-definite. $b$: bandwidth of the banded component; $r$: rank of the low-rank component. $\{\mathbf{L}, \mathbf{U}, \mathbf{\Lambda}\}$: Cholesky factor, eigenvector and eigenvalue of the target matrix. Subscript of $\mathbf{L}, \mathbf{U}, \mathbf{\Lambda}$ represents the component, $b$ for banded or $r$ for low-rank.

| Strategy | Average across hyperparameter settings | | | | | |
|---|---|---|---|---|---|---|
| | % Approx of LD | Heritability estimates | | | Runtime | |
| | | Bias | Variance | MSE | In seconds | % saved |
| **Exact_LD** | 100% | 0.00134 | 0.00001 | 0.00002 | 1578.24 | - |
| **Seq_Band_LR** | 93.06% | 0.09576 | 0.01808 | 0.06663 | 1428.00 | 9.52% |
| **PSD_Band_LR** | 92.89% | 0.04135 | 0.00006 | 0.00205 | 905.81 | 42.61% |
| **Joint_Band_LR** | 94.73% | 0.07530 | 0.00045 | 0.00733 | 830.67 | 47.37% |
| **Strategy** | **Fine-tuned hyperparameter setting** | | | | | |
| | % Approx of LD | Heritability estimates | | | Runtime | |
| | | Bias | Variance | MSE | In seconds | % saved |
| **Seq_Band_LR** | 93.97% | 0.00009 | 0.00002 | 0.00002 | 959.38 | 39.21% |
| **PSD_Band_LR** | 93.84% | 0.01972 | 0.00002 | 0.00041 | 608.42 | 61.45% |
| **Joint_Band_LR** | 95.96% | 0.04140 | 0.00004 | 0.00175 | 477.79 | 69.73% |

**Table 3. Comparison of the performance of different LD approximation methods from hyperparmaeter tuning.** Top table represents the average performance across hyperparameter settings, with $b$ varying from 300 to 600 in increments of 100, and $r$ varying from 300 to 800 in increments of 50. The second column shows the proportion of the original full LD matrix approximated by the low-dimensional representation, measured in matrix (Frobenius) norm. The fine-tuned hyperparameter settings for different strategies are: $(b,r) = (500,700)$ for "Seq_Band_LR", $(b,r) = (400,800)$ for "PSD_Band_LR", and $(b,r) = (400,800)$ for "Joint_Band_LR", which are selected based on minimizing the CV bias. The runtime reported is for the HEELS heritability estimation.