Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A rotated multivariate linear mixed model for dual large-scale genome-wide association study

Qizhai Li

liqz@amss.ac.cn

Chinese Academy of Sciences https://orcid.org/0000-0002-3325-8265

Hongping Guo

Hubei Normal University

Qinyuan Zhang

University of Chinese Academy of Sciences

Xueqin Zheng

Hubei Normal University

Yuan Quan

Huazhong Agricultural University https://orcid.org/0000-0002-5163-2951

Technical Report

Keywords: GWAS, large-scale, multivariate linear mixed model, orthogonal rotation

Posted Date: June 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9922993/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

A rotated multivariate linear mixed model for dual large-scale genome-wide association study

Hongping Guo[1,†], Qinyuan Zhang[2,†], Xueqin Zheng[1], Yuan Quan[3], Qizhai Li[4,5,*]

[1]*School of Mathematics and Statistics, Hubei Normal University, Huangshi, 435002, China*

[2]*School of Advanced Interdisciplinary Sciences, University of Chinese Academy of Sciences, Beijing 101408, China*

[3]*Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China*

[4]*State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*

[5]*School of Mathematical Science, University of Chinese Academy of Sciences, Beijing 101408, China*

[*]Corresponding author. E-mail: liqz@amss.ac.cn

[†] The first two authors should be regarded as Joint First Authors.

Abstract

Modern biobanks have spurred advances in large-scale genome-wide association study (GWAS) and multi-trait GWAS provides valuable insights into the shared genetic architecture of complex traits. However, performing multi-trait GWAS on biobank-scale samples often leads to trade-offs between computational efficiency and statistical power. We develop the **R**otated **m**ulti**v**ariate **L**inear **M**ixed **M**odel (RmvLMM), a powerful and scalable statistical framework for dual large-scale GWAS. RmvLMM employs an orthogonal rotation technique to decorrelate multiple traits and presents a novel fast iterative algorithm for covariance matrix estimation, enabling multi-trait analysis for a large number of phenotypes. For biobank-scale individuals, RmvLMM leverages a parallelizable divided-and-combined strategy to aggregate association signals across groups into a single omnibus test. Simulation results show that RmvLMM achieves higher power and substantially faster runtime than the state-of-the-art tool GEMMA. Furthermore, RmvLMM exhibited robust performances across extensive simulations under varying parameter settings, at scales beyond the computational capacity of GEMMA. In a analysis of 26 blood biochemistry traits from the UK Biobank, RmvLMM identified 1325 significant variants, among which 10 trait-associated genes are validated drug targets, corresponding to 29 developed drugs. RmvLMM is accessible as open source software (https://github.com/amss-stat/RmvLMM).

Keywords: GWAS, large-scale, multivariate linear mixed model, orthogonal rotation.

Introduction

Over the past two decades, a large number of statistical methods have been developed for genome-wide association study (GWAS) [1, 2]. The linear mixed models (LMMs) have been widely applied because of their ability to control for two major confounders of population stratification and sample relatedness [3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. Owing to the substantial decrease in genotyping costs in recent years, sample sizes of GWAS have dramatically increased to over 300,000 in single cohorts such as the UK Biobank (UKB) [13]. Several studies have focused on LMM-based methods for large-scale data, but limited to single-trait analyses [14, 15, 16, 17, 18, 19].

The pleiotropic effect exists when one genetic variant affects multiple traits [20, 21, 22]. There is growing empirical evidence from GWASs that many genetic variants have pleiotropic effects [23]. For example, 44% autoimmune risk single nucleotide polymorphisms (SNPs) have been estimated to be associated with two or more autoimmune diseases [24]. Detecting these effects can provide valuable insights into the shared genetic architecture of complex traits [25]. Multi-trait association analyses, leveraging the pleiotropic effects, have been previously shown to enhance the statistical power to identify trait-associated variants relative to traditional single-trait association analyses [26, 27, 28]. Therefore, we focus on developing a large-scale multi-trait GWAS approach for the identification of pleiotropic variants.

Two main multivariate analysis strategies are commonly used for multi-trait GWAS: dimension reduction methods and regression models. A representive dimension reduction approach is principal component analysis (PCA) [27, 29]. PCA-based analyses that prioritize the first few eigenvectors (also called PCs) derived from PCA of phenotypic data and accounting for the majority of phenotypic variance are inherently suboptimal. This is because, the PCs explaining a small amount of total phenotypic variance have been proved to be as important as those explaining large amount of variance. Discarding the lower-variance PCs can therefore severely decrease the power to detect pleiotropic variants [27]. On the regression side, multivariate linear mixed models (mvLMMs) have been widely applied in GWAS [26, 28, 30], and the genome-wide efficient mixed model association (GEMMA) proposed by Zhou and Stephens [28] represents a state-of-the-art method. However, GEMMA does not scale well, making it computationally intractable for large-scale GWAS (e.g. $\geq$ 50,000 individuals). In addition, computational and statistical barriers remain for even modest numbers of phenotypes (e.g. $\sim$ 10 phenotypes) [28]. As a result, current multi-trait GWAS methods are either highly scalable or highly powered, but not both.

We propose a powerful and scalable method, called RmvLMM (**R**otated **m**ulti**v**ariate **L**inear **M**ixed **M**odel), a statistical framework for dual large-scale GWAS (biobank-scale samples and a large number of phenotypes). RmvLMM uses a series of techniques, especially orthogonal rotation transformation and divided-and-combined strategy, to simultaneously improve association power and computational efficiency. In simulations, RmvLMM achieved over 10% higher power

than GEMMA while dramatically reducing computational cost. For instance, in a genome-wide scan with 20 phenotypes, RmvLMM was over 50 times faster than GEMMA, completing the analysis in less than 2% of the time. In large-scale cohorts and high-dimensional phenotypes, RmvLMM maintained robust power and well-calibrated false discoveries, whereas GEMMA is computationally intractable at equivalent scale. In a UKB application (including 26 blood biochemistry traits and 323,839 individuals), RmvLMM detected 1,325 genome-wide significant variants annotating to 373 protein-coding genes. In total, 309 of these genes have been previously reported in published studies, including 10 trait-associated genes with validated druggable potential that correspond to 29 clinically developed agents. RmvLMM provides an efficient tool for large-scale multi-trait GWAS and can facilitate a deeper understanding of the shared genetic mechanisms at play.

Results

Overview of the RmvLMM

RmvLMM is designed for dual large-scale multi-trait GWAS, which models phenotypes as a combination of covariate effects, SNP genetic effects, random polygenic effects, and random error effects (Methods and Equations 1 and 2). It detects the genetic associations via two steps: model parameter estimation and association test.

In the model parameter estimation step, RmvLMM estimates the polygenic covariance $\boldsymbol{V}_b$ and error covariance $\boldsymbol{V}_e$ under the null hypothesis. By deriving a closed-form expression that characterizes the relationship between $\boldsymbol{V}_b$ and $\boldsymbol{V}_e$, we develop a new fast iterative estimation algorithm, called MoM-REML, which integrates the method-of-moments (MoM) with restricted maximum likelihood (REML) estimation. Subsequently, RmvLMM employs an orthogonal rotation technique to decorrelate multiple phenotypes, thereby transforming the mvLMM into univariate linear regression models for d phenotypes (Methods and Equation 3). This transformation is implemented via three sequential eigendecompositions and plays a key role in improving the power of association tests.

In the association tests step, we perform F-tests on the variables of interest (i.e., the rotated SNP effects) in the univariate models to obtain d significance p-values. Unlike GEMMA that uses the Wald test, likelihood ratio test, etc., the F-test does not require estimation of SNP effects nor additional p-value calibration. Next, we propose a new p-value combination statistic, named TOP2, which combines the two smallest p-values out of d using the Fisher's combination method. The exact distribution of the TOP2 statistic and the corresponding p-value formula can be derived (Supplementary Note 1). For biobank-scale samples, we extend TOP2 to a scalable omnibus version, denoted as TOP2-O, whose null distribution is fitted via the generalized gamma distribution to enable valid statistical inference in large biobank cohorts (Methods).

RmvLMM incorporates multiple technical advancements to achieve greater power and high

scalability. First, based on correcting for population stratification and sample relatedness, the orthogonal rotation transformation used by RmvLMM highlights significant signals while treating non-significant signals as noise, thereby improving the power of association tests (Methods and Supplementary Note 2). Second, when estimating the covariance matrices, RmvLMM adopts the new iterative MoM-REML algorithm and derives the closed-form relationship between $\boldsymbol{V}_b$ and $\boldsymbol{V}_e$. It depends on a computational complexity of $\mathcal{O}(nd^2)$ for n individuals and d phenotypes, wherea GEMMA requires a much higher complexity of $\mathcal{O}(nd^6)$ (Supplementary Notes 3 and 4). Third, during the GWAS scanning phase, RmvLMM improves the computational efficiency of performing F-tests by applying orthogonal projection to the rotated univariate models (Supplementary Note 5). Finally, to accommodate biobank-scale analyses, the TOP2-O statistic adopts a highly parallelizable divided-and-combined strategy, in which association evidence is evaluated across multiple sample groups and summarized into an omnibus statistic, thereby avoiding eigendecomposition of large sample relatedness matrix $\boldsymbol{K}$ while maintaining valid statistical inference (Methods and Supplementary Note 6). More details on orthogonal rotation transformations, algorithm implementations and optimizations are provided in the Methods section and the Supplementary Notes. The methodological framework of RmvLMM is presented in Figure 1.

RmvLMM supports GWAS for a large number of phenotypes

The time complexity of the new MoM-REML iterative algorithm for estimating covariance matrices is

$$\mathcal{O}\big(n^3 + n^2d + n^2c + t(c^3d + d^3 + cd^2 + nd^2 + nc^2d)\big),$$

where t is the number of iterations, c is the number of covariates. Generally, $n \gg c,\ n \gg d$, so, the dominant term in time complexity is $\mathcal{O}\big(n^3 + t(nd^2 + nc^2d)\big)$. A key advantage is that the dependence on the phenotypic dimension d is $\mathcal{O}(nd^2)$, which is a significant reduction from $\mathcal{O}(nd^6)$ required by GEMMA with Newton-Raphson second-order optimization algorithm. This is a huge improvement in computational speed, making it possible to perform analyses on a large number of phenotypes.

After estimating covariance matrices, the remaining computational cost is dominated by the genome-wide association scan over p SNPs, implemented efficiently with a complexity of $\mathcal{O}(p(n^2 + ncd + d^2))$. Apart from the fixed $\mathcal{O}(pn^2)$ cost for rotating the SNP vectors, the computational cost scales linearly with both covariate and phenotypic dimensions. This efficiency stems from the orthogonal projection strategy (Supplementary Note 5), which accelerates the F-tests throughout the GWAS scanning phase.

Table 1 reports both the theoretical time complexity and the empirical runtime of RmvLMM and GEMMA under representative settings. For example, when analyzing moderate-sized samples with a modest number of phenotypes (e.g., $n = 5000$, $d = 20$), RmvLMM reduces the wall-clock time by more than two orders of magnitude compared to GEMMA, achieving a

$\sim$ 150-fold speed-up for fitting a single model and a $\sim$ 50-fold speed-up for performing a GWAS over 524,080 SNPs. Notably, RmvLMM is insensitive to the phenotypic dimension d: with $n = 5000$, $p = 524{,}080$, and $d = 5, 10, 20$, the entire GWAS can be completed in less than 5 minutes, whereas GEMMA typically requires several hours. Detailed information on the computational complexity of RmvLMM can be found in Supplementary Notes 4 and 5.

RmvLMM enables GWAS for biobank-scale sample sizes

Scaling GWAS based on mixed models to biobank-scale cohorts faces a fundamental computational bottleneck: eigendecomposition of the sample relatedness matrix $\boldsymbol{K}$, whose time and memory complexity scale cubically and quadratically with the sample size n. While conventional methods approximate $\boldsymbol{K}$ via low-rank or sparse representations [7, 31, 32], RmvLMM adopts an alternative paradigm by introducing a highly parallelizable divided-and-combined strategy into the framework of multi-trait mixed model association analysis. This strategy facilitates reliable modeling and association testing without explicitly constructing or decomposing the full matrix $\boldsymbol{K}$. Specifically, the n samples are divided into multiple disjoint groups, within which association evidence is evaluated separately using the TOP2 statistic, and the resulting group-level evidence is combined into an omnibus test statistic, TOP2-O (Methods). The null distribution of TOP2-O is empirically calibrated to accommodate the dependence across group-level test statistics without explicit correlation modeling. We construct its empirical null distribution using a large panel of approximately independent SNPs, and further demonstrate that this distribution can be well approximated by the generalized gamma distribution (Supplementary Note 6).

For biobank-scale cohorts, RmvLMM achieves nearly linear scaling with sample size. For example, under the settings of $d = 10$, $p = 100{,}000$ and $n = 50,000, 100,000, 200,000, 300,000$, the runtimes of RmvLMM are 0.52, 1.04, 2.10, and 3.13 hours, respectively. In contrast, GEMMA cannot scale to such large sample sizes.

Statistical power and false positive rate

To evaluate the statistical power and false positive rate (FPR) of RmvLMM and GEMMA, we performed simulations across small ($n = 500, 1{,}000$) and moderate ($n = 5{,}000, 10{,}000$) sample sizes using a modest number ($d = 10$) of phenotypes. All study individuals were randomly sampled from the UKB cohort. For power comparison, SNP genotypes were generated from a binomial distribution across five distinct minor allele frequencies (MAFs) satisfying Hardy–Weinberg equilibrium, with MAF $\in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. For a robust cross-method performance evaluation, we stratified the five MAFs into three groups ($\{0.1\}$, $\{0.2\}$, $\{0.3, 0.4, 0.5\}$) to generate SNP genetic effects for each group. Phenotypes were simulated under the mvLMM. We considered a dense signal scenario, where genetic effects across all phenotypes followed a uniform distribution of $U(-0.5, 0.5)$. For simplicity, only a single covariate was considered.

The design matrix and its coefficient vector had independent entries each following a standard normal distribution $N(0,1)$ and a uniform distribution $U(0.1,0.5)$, respectively. The polygenic covariance matrix $\boldsymbol{V}_b$ was set to be a compound symmetric structure with a pairwise phenotypic correlation of 0.5, i.e., $\boldsymbol{V}_b = (1-\rho)\boldsymbol{I}_d + \rho \mathbf{1}_d \cdot \mathbf{1}_d^\top$ ($\rho = 0.5$), and the random error covariance matrix $\boldsymbol{V}_e$ was configured to an autoregressive structure with a decay correlation of 0.4, i.e., $\boldsymbol{V}_e = (\rho^{|i-j|})_{d\times d}$ ($\rho = 0.4$). The relatedness matrix $\boldsymbol{K}$ was calculated by the product of the centered and scaled SNP genotypes of n individuals randomly selected from the real UKB dataset. We calculated the powers by 1,000 simulation replicates. For FPR comparison, we simulated the phenotypes under the mvLMM for the null hypothesis of no genetic associations. Full details of the simulation procedures are provided in the Methods section.

The results demonstrate that RmvLMM achieves consistently higher power than GEMMA across all simulated scenarios (Figure 2(A) and Supplementary Tables S1–S4). Furthermore, both RmvLMM and GEMMA maintain well-calibrated FPRs across all evaluated simulation settings (Figure 2(B) and Supplementary Tables S5–S6). Notably, at the stringent significance threshold of $\alpha = 1\times10^{-7}$, the power of RmvLMM outperforms GEMMA by over 10% across varying MAF levels. Taking the setting with $n = 1000, d = 10$ as a representative example, the power of GEMMA is 40.5%, 57.9%, and 62.2% for MAFs of 0.3, 0.4, and 0.5, respectively; by contrast, RmvLMM yields substantially elevated powers of 52.4%, 70.2%, and 73.1% under the same conditions (Supplementary Table S2). Moreover, at the same significance threshold of $\alpha = 1\times10^{-7}$, the FPRs of RmvLMM and GEMMA are 8.00×10^{-8} and 6.04×10^{-8}, respectively, across 10^8 simulation replicates (Supplementary Table S6). Both FPR values fall within the nominal tolerance bound, confirming well-calibrated FPR control.

Simulation results for dual large-scale datasets

For high-dimensional phenotypic scenarios ($d = 50, 100, 200$), we further conducted simulations across small ($n = 500, 1,000$) and moderate ($n = 5,000, 10,000$) sample sizes. We adopted a sparse genetic signal pattern, where roughly 61.8% of the d phenotypes had no genetic effects, with the remaining non-zero effects sampled from the uniform distribution $U(-0.5, 0.5)$. Moreover, for large-scale sample sizes of $n = 50,000$, $100,000$, $200,000$ and $300,000$, we curated two phenotypic panels of different dimensionalities ($d = 10$ and 20) derived from the blood biochemistry traits in the real UKB dataset. We simulated dense association signals for the $d = 10$ panel and sparse genetic signals for the $d = 20$ panel. Full details regarding the simulation of SNP genotypes and phenotypes are provided in the Methods section.

Simulation results showed that RmvLMM achieves robust power (Supplementary Tables S7–S18) and well-calibrated FPR control (Supplementary Tables S19–S21) across high-dimensional phenotypic settings (Figure 3). Similar statistical power (Supplementary Tables S22–S29) and FPR performance (Supplementary Tables S30–S31) are further observed in large-scale sample cohorts (Figure 4). By contrast, GEMMA suffers from substantial intrinsic complexity, leading

to severe computational intractability that precludes its application to these scenarios.

For large-scale sample cohorts, we also computed omnibus p-values using the cumulative distribution functions of the five candidate distributions (generalized gamma, gamma, log-normal, chi-square, and Weibull), and evaluated the observed FPR performance. The result demonstrates that the generalized gamma distribution, estimated via our robust log-moment initialization and Nelder-Mead optimization, provides a superior approximation for null distribution of the TOP2-O statistics (Supplementary Note 6 and Supplementary Table S32).

Application of RmvLMM to multi-trait GWAS of UK Biobank

We applied RmvLMM to perform a multi-trait GWAS of 454,296 SNPs for 26 blood biochemistry phenotypes in 323,839 European-ancestry individuals from the UKB. The brief descriptions of the 26 phenotypes are presented in Supplementary Table S33, including high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), total cholesterol (TC), triglycerides (TG), C-reactive protein (CRP), and others. Detailed preprocessing of genotype and phenotype data is provided in the Methods section.

The 323,839 individuals were divided into 33 groups (32 groups of 10,000 individuals and one remaining group). Using the divided-and-combined strategy, we constructed the TOP2-O statistic and derived its empirical distribution via tens of thousands of independent SNPs. We then evaluated the goodness-of-fit of five candidate distributions (generalized gamma, gamma, log-normal, chi-square, and Weibull) using three standard metrics: the Akaike information criterion (AIC), Bayesian information criterion (BIC), and Kullback–Leibler (KL) divergence, (Supplementary Note 6). Goodness-of-fit results are presented in (Supplementary Figure S1), which demonstrate that the generalized gamma distribution provides the optimal fit.

All computations were conducted on a server equipped with a Hygon 7285 2.50GHz CPU. Both the eigendecomposition of the sample relatedness matrices and the genome-wide association scanning were implemented in a basic serial framework without additional parallel optimization. The eigendecomposition step was completed in approximately 26 minutes using 15 cores (45 GB memory), while the GWAS analysis for all 454,296 SNPs took about 12.5 hours using 60 cores (180 GB memory).

RmvLMM identified a total of 1,325 significant SNPs (p-value $< 5 \times 10^{-8}$; Supplementary Table S34), which were mapped to 373 genes using the MAGMA software [33]. Among these 373 genes, 309 have been previously associated with at least one of the 26 blood biochemistry phenotypes analyzed in this study, based on literature searches of the PubMed database via the R package 'rentrez' (Supplementary Table S35). Manhattan and QQ plots for the GWAS results are shown in Figure 5(A)-(B). Detailed procedures for SNP-to-gene annotation and literature validation of known gene-phenotype associations are provided in the Methods section. Pathway enrichment analyses further revealed that the 373 candidate genes are predominantly enriched in functional modules governing lipid metabolism, cholesterol homeostasis and the pathogenesis

of familial hyperlipidemia (Supplementary Figure S2). The enrichment analysis was performed using the WikiPathways 2024 Human database via the Enrichr web server [34].

To validate the accuracy of trait-associated SNPs identified via RmvLMM, we mapped their corresponding genes against the DrugBank and Therapeutic Target Database (TTD) (see Methods). We identified 10 pleiotropic genes associated with blood biochemistry phenotypes that have been validated as drug targets, with 29 therapeutic drugs already developed for these targets. For instance, *PCSK9* and *HMGCR* act as core regulators of lipid metabolism. *PCSK9* modulates cholesterol clearance by promoting the degradation of LDL receptors, and its targeted therapies (e.g., alirocumab, evolocumab) significantly reduce circulating LDL-C levels [35, 36]. Mutations in *PCSK9* cause autosomal dominant hypercholesterolemia [37]. As the molecular target of statins (e.g., rosuvastatin, fluvastatin), *HMGCR* alleviates hypercholesterolemia and mixed dyslipidemia by inhibiting the cholesterol synthesis pathway—an effect directly correlated with TC, LDL and other blood lipids analyzed in this GWAS [38, 39]. The detailed relationships between GWAS-identified SNPs, pleiotropic genes, therapeutic drugs, and blood biochemistry phenotypes are shown in Figure 6. The information on target genes, corresponding drugs, impacted blood phenotypes and therapeutic indications is provided in Supplementary Table S36.

Methods

Multivariate linear mixed model

Let $\widetilde{\boldsymbol{Y}}$ be an $n \times d$ phenotypic value matrix with n and d representing the number of individuals and phenotypes, respectively, $\widetilde{\boldsymbol{g}}$ be an n-dimensional genotype score vector at a given SNP locus, and $\widetilde{\boldsymbol{X}}$ be an $n \times c$ covariate observation matrix including a column of ones to model the intercept term. Denote by $\mathbf{0}_d$ the d-dimensional column vector of all zero units. The multivariate linear mixed model (mvLMM) is

$$\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{\gamma}}^{\top} + \widetilde{\boldsymbol{g}}\widetilde{\boldsymbol{\beta}}^{\top} + \widetilde{\boldsymbol{B}} + \widetilde{\boldsymbol{E}}, \tag{1}$$

where $\widetilde{\boldsymbol{\gamma}}$ is the $d \times c$ coefficient matrix, $\widetilde{\boldsymbol{\beta}}$ is the d-dimensional coefficient column vector of genetic effects, $\widetilde{\boldsymbol{B}}$ is the $n \times d$ random polygenic effect matrix, $\widetilde{\boldsymbol{E}}$ is an $n \times d$ random error matrix, and the superscript $\top$ indicates the transpose of a vector or a matrix. The null hypothesis that the d phenotypes are not associated with the SNP is $\widetilde{H}_0 : \widetilde{\boldsymbol{\beta}} = \mathbf{0}_d$.

Let $\mathrm{vec}(\cdot)$ represent the column-wise vectorization of a matrix and $\otimes$ denote the Kronecker product. The vectorized forms of $\widetilde{\boldsymbol{B}}$ and $\widetilde{\boldsymbol{E}}$ have the multivariate normal distributions with $\mathrm{vec}\big(\widetilde{\boldsymbol{B}}\big) \sim \mathcal{N}\big(\mathbf{0}_{nd}, \boldsymbol{V}_b \otimes \boldsymbol{K}\big)$ and $\mathrm{vec}\big(\widetilde{\boldsymbol{E}}\big) \sim \mathcal{N}\big(\mathbf{0}_{nd}, \boldsymbol{V}_e \otimes \boldsymbol{I}_n\big)$, where $\boldsymbol{V}_b$ and $\boldsymbol{V}_e$ are $d \times d$ random polygenic and error covariance matrices, respectively, and $\boldsymbol{K}$ is an $n \times n$ sample relatedness matrix. The covariance matrix of $\mathrm{vec}(\widetilde{\boldsymbol{Y}})$ is

$$\widetilde{\boldsymbol{V}} = \mathrm{Cov}\big(\mathrm{vec}(\widetilde{\boldsymbol{Y}})\big) = \boldsymbol{V}_b \otimes \boldsymbol{K} + \boldsymbol{V}_e \otimes \boldsymbol{I}_n.$$

The rotated mixed model

We employ an orthogonal rotation transformation to decorrelate the phenotypes. The decorrelation procedure involves three sequential eigendecompositions as

$$\begin{aligned}
&\boldsymbol{K} = \boldsymbol{Q}_k \boldsymbol{\Lambda}_k \boldsymbol{Q}_k^\top, \\
&\boldsymbol{V}_e = \boldsymbol{Q}_e \boldsymbol{\Lambda}_e \boldsymbol{Q}_e^\top = \boldsymbol{Q}_e \boldsymbol{\Lambda}_e^{1/2} \big(\boldsymbol{Q}_e \boldsymbol{\Lambda}_e^{1/2}\big)^\top, \\
&\boldsymbol{\Lambda}_e^{-\frac{1}{2}} \boldsymbol{Q}_e^\top \boldsymbol{V}_b \big(\boldsymbol{\Lambda}_e^{-\frac{1}{2}} \boldsymbol{Q}_e^\top\big)^\top = \boldsymbol{U} \boldsymbol{\Xi} \boldsymbol{U}^\top,
\end{aligned}$$

where $\boldsymbol{\Lambda}_k$, $\boldsymbol{\Lambda}_e$ and $\boldsymbol{\Xi}$ are the diagonal eigenvalue matrices, and $\boldsymbol{Q}_k$, $\boldsymbol{Q}_e$ and $\boldsymbol{U}$ are the corresponding orthogonal eigenvector matrices. Subsequently, model rotation is performed by simultaneously left-multiplying matrix $\boldsymbol{Q}_k^\top$ and right-multiplying matrix $\boldsymbol{Q}_e \boldsymbol{\Lambda}_e^{-1/2} \boldsymbol{U}$ on the two sides of Model (1). Then the rotated mixed model can be written as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\gamma}^\top + \boldsymbol{g}\boldsymbol{\beta}^\top + \boldsymbol{B} + \boldsymbol{E}, \tag{2}$$

where

$$\begin{aligned}
&\boldsymbol{Y} = \boldsymbol{Q}_k^\top \widetilde{\boldsymbol{Y}}\, \boldsymbol{Q}_e \boldsymbol{\Lambda}_e^{-1/2} \boldsymbol{U}, \quad \boldsymbol{X} = \boldsymbol{Q}_k^\top \widetilde{\boldsymbol{X}}, \quad \boldsymbol{g} = \boldsymbol{Q}_k^\top \widetilde{\boldsymbol{g}}, \\
&\boldsymbol{\gamma}^\top = \widetilde{\boldsymbol{\gamma}}^\top (\boldsymbol{Q}_e \boldsymbol{\Lambda}_e^{-1/2} \boldsymbol{U}), \quad \boldsymbol{\beta}^\top = \widetilde{\boldsymbol{\beta}}^\top (\boldsymbol{Q}_e \boldsymbol{\Lambda}_e^{-1/2} \boldsymbol{U}), \\
&\boldsymbol{B} = \boldsymbol{Q}_k^\top \widetilde{\boldsymbol{B}}\, \boldsymbol{Q}_e \boldsymbol{\Lambda}_e^{-1/2} \boldsymbol{U}, \quad \boldsymbol{E} = \boldsymbol{Q}_k^\top \widetilde{\boldsymbol{E}}\, \boldsymbol{Q}_e \boldsymbol{\Lambda}_e^{-1/2} \boldsymbol{U}.
\end{aligned}$$

Since $\boldsymbol{Q}_e \boldsymbol{\Lambda}_e^{-\frac{1}{2}} \boldsymbol{U}$ is an invertible matrix, $\widetilde{\boldsymbol{\beta}} = \boldsymbol{0}_d$ is equivalent to $\boldsymbol{\beta} = \boldsymbol{0}_d$. Thus, to test $\widetilde{H}_0$ is equivalent to testing $H_0 : \boldsymbol{\beta} = \boldsymbol{0}_d$. The covariance matrix of $\mathrm{vec}(\boldsymbol{Y})$ is

$$\boldsymbol{V} = \mathrm{Cov}\big(\mathrm{vec}(\boldsymbol{Y})\big) = \boldsymbol{\Xi} \otimes \boldsymbol{\Lambda}_k + \boldsymbol{I}_d \otimes \boldsymbol{I}_n.$$

The detailed derivation of $\boldsymbol{V}$ is provided in Supplementary Note 2.

Test statistic

Since $\boldsymbol{V}$ is diagonal, the rows and columns of the rotated phenotypic matrix $\boldsymbol{Y}$ are thus mutually uncorrelated. We can divided Model (2) into d univariate linear regression models for analysis. Denote $\boldsymbol{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_d)$, $\boldsymbol{\gamma}^\top = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_d)$, $\boldsymbol{\beta}^\top = (\beta_1, \ldots, \beta_d)$, and $\boldsymbol{B} + \boldsymbol{E} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_d)$. For $j = 1, \ldots, d$, the j-th linear regression model can be constructed as

$$\boldsymbol{y}_j = \boldsymbol{X}\boldsymbol{\gamma}_j^\top + \boldsymbol{g}\beta_j + \boldsymbol{\varepsilon}_j, \tag{3}$$

where $\mathrm{Cov}(\boldsymbol{\varepsilon}_j) = \boldsymbol{\Sigma}_j = \xi_j \boldsymbol{\Lambda}_k + \boldsymbol{I}_n$, and ξ_j is the j-th diagonal element of $\boldsymbol{\Xi}$. In Model (3), the null hypothesis of no association between $\boldsymbol{g}$ and $\boldsymbol{y}_j$ is $H_{0j} : \beta_j = 0$. We adopt an F statistic for univariate test, and subsequently develop a new combined statistic to integrate the p-values obtained from individual univariate tests. Denote the residual sum of squares (RSS) under the null hypothesis H_{0j} as

$$\begin{aligned}
\mathrm{RSS}_{H_{0j}} = \; & \boldsymbol{y}_j^\top \Big(\boldsymbol{I}_n - \boldsymbol{X}\big(\boldsymbol{X}^\top (\xi_j \boldsymbol{\Lambda}_k + \boldsymbol{I}_n)^{-1} \boldsymbol{X}\big)^{-1} \boldsymbol{X}^\top (\xi_j \boldsymbol{\Lambda}_k + \boldsymbol{I}_n)^{-1}\Big)^\top (\xi_j \boldsymbol{\Lambda}_k + \boldsymbol{I}_n)^{-1} \\
& \Big(\boldsymbol{I}_n - \boldsymbol{X}\big(\boldsymbol{X}^\top (\xi_j \boldsymbol{\Lambda}_k + \boldsymbol{I}_n)^{-1} \boldsymbol{X}\big)^{-1} \boldsymbol{X}^\top (\xi_j \boldsymbol{\Lambda}_k + \boldsymbol{I}_n)^{-1}\Big) \boldsymbol{y}_j.
\end{aligned}$$

Denote $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{g})$, and the RSS under the alternative hypothesis H_{1j} with $\beta_j \neq 0$ is given by

$$\begin{aligned}\mathrm{RSS}_{H_{1j}} = \ & \boldsymbol{y}_j^\top \Big(\boldsymbol{I}_n - \boldsymbol{Z}\big(\boldsymbol{Z}^\top(\xi_j\boldsymbol{\Lambda}_k + \boldsymbol{I}_n)^{-1}\boldsymbol{Z}\big)^{-1}\boldsymbol{Z}^\top(\xi_j\boldsymbol{\Lambda}_k + \boldsymbol{I}_n)^{-1}\Big)^\top (\xi_j\boldsymbol{\Lambda}_k + \boldsymbol{I}_n)^{-1} \\ & \Big(\boldsymbol{I}_n - \boldsymbol{Z}\big(\boldsymbol{Z}^\top(\xi_j\boldsymbol{\Lambda}_k + \boldsymbol{I}_n)^{-1}\boldsymbol{Z}\big)^{-1}\boldsymbol{Z}^\top(\xi_j\boldsymbol{\Lambda}_k + \boldsymbol{I}_n)^{-1}\Big)\boldsymbol{y}_j.\end{aligned}$$

The F statistic for the j-th linear regression model is then expressed as

$$F_j = \frac{(\mathrm{RSS}_{H_{0j}} - \mathrm{RSS}_{H_{1j}})/1}{\mathrm{RSS}_{H_{1j}}/(n-1-c)},$$

where F_j follows an F-distribution with $(1,\, n-1-c)$ degrees of freedom. F_j can be efficiently computed via orthogonal projection derived from the QR decomposition of the weighted design matrix $\boldsymbol{\Sigma}_j^{-1/2}\boldsymbol{X}$. Detailed implementation for the efficient calculation of F statistics is provided in Supplementary Note 5.

We perform F test on each of the d rotated univariate models, and obtain d corresponding p-values, denoted as $p_1, p_2, \ldots, p_d$. Let the ordered p-values be $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(d)}$, we propose a new p-value combination statistic, named TOP2, which integrates the first two order statistics via Fisher's combination. The expression of TOP2 combination statistic is

$$\mathrm{TOP2} = -2\ln p_{(1)} - 2\ln p_{(2)},$$

and the p-value can be calculated as

$$p_{\mathrm{TOP2}} = F_{\mathrm{Beta}}(a; 2, d-1) - d(d-1)a^2\left(\ln a + \sum_{k=1}^{d-2}\frac{(1-a)^k}{k}\right),$$

where $F_{\mathrm{Beta}}(a, 2, d-1)$ denotes the Beta cumulative distribution function, with $a = e^{-s_{obs}/4}$, shape parameters are 2 and $d-1$, and s_{obs} is the observed value of TOP2 statistic. When $d = 2$, p-value $= a^2(1 - 2\ln a)$. The detailed derivation of the distribution of TOP2 statistic is provided in Supplementary Note 1.

The omnibus test statistic for biobank-scale sample sizes

For large sample sizes, such as the approximately 500,000 individuals in the UK Biobank, performing an eigendecomposition of the full $\boldsymbol{K}$ is computationally infeasible in both time and memory. We adopt a highly parallelizable divided-and-combined strategy, in which the full dataset is divided into q disjoint groups that are processed separately and subsequently combined. Specifically, for each group k, a group-level p-value $p_{\mathrm{TOP2},k}$ is computed, and these p-values are combined using Fisher's method to construct the omnibus test statistic,

$$\text{TOP2-O} = -2\sum_{k=1}^{q}\ln p_{\mathrm{TOP2},k}.$$

The group-level p-values are generally correlated, and the null distribution of TOP2-O consequently deviates from the standard chi-square distribution. We therefore calibrate its null

distribution empirically. We select tens of thousands of approximately independent SNPs and compute the corresponding TOP2-O statistics, which collectively form an empirical null distribution. The empirical null distribution is well approximated by a generalized gamma distribution (Supplementary Note 6). The omnibus p-value is computed as

$$p_{\text{TOP2-O}} = 1 - F_{GG}(\text{TOP2-O}; \mu, \sigma, Q),$$

where $F_{GG}(x; \mu, \sigma, Q)$ denotes the cumulative distribution function of the generalized gamma distribution with parameters (μ, σ, Q) under the Prentice parameterization. This approach avoids explicit modeling of the correlation structure among group-level test statistics while retaining valid FPR control under the null, and more importantly, enables principled and statistically rigorous inference at biobank scale and beyond.

Iterative estimation algorithm of covariance matrices

To calculate the F statistic for each linear regression models, it is necessary to estimate the diagonal matrices $\boldsymbol{\Lambda}_k$ and $\boldsymbol{\Xi}$. The diagonal matrix $\boldsymbol{\Lambda}_k$ is derived from the eigendecomposition of the relatedness matrix $\boldsymbol{K}$. Since $\boldsymbol{K}$ can be directly estimated as the product of the centered and scaled SNP genotype matrix [6, 11, 28, 40], $\boldsymbol{\Lambda}_k$ is straightforward to calculate. In contrast, the diagonal matrix $\boldsymbol{\Xi}$ is obtained by sequentially applying orthogonal rotations of the random error covariance matrix $\boldsymbol{V}_e$ and the polygenic covariance matrix $\boldsymbol{V}_b$. Therefore, covariance matrices $\boldsymbol{V}_b$ and $\boldsymbol{V}_e$ must first be estimated.

For mvLMMs, maximum likelihood estimation (MLE) or restricted maximum likelihood estimation (REML) has been the predominant approach for estimating matrices $\boldsymbol{V}_b$ and $\boldsymbol{V}_e$. However, these methods entail computationally expensive operations (e.g., calculating the inverse and determinant) of the $nd \times nd$ covariance matrix $\widetilde{\boldsymbol{V}}$, making them computationally infeasible for large sample sizes n and high phenotypic dimensions d. Although the state-of-the-art method GEMMA leverages the properties of block-diagonal matrix and sparse matrix to simplify computations of likelihood function optimization with Newton-Raphson's algorithm, it still cannot be applied to large-scale Biobank-level analyses because of the high cost to evaluate the Hessian matrix, especially when the phenotypic dimension $d > 10$. Consequently, we propose a new fast iterative algorithm to estimate $\boldsymbol{V}_b$ and $\boldsymbol{V}_e$ that circumvents the need to calculate the inverse, determinant and Hessian matrix of the nd-dimensional $\widetilde{\boldsymbol{V}}$. Specifically, we first deduce the analytical expression characterizing the relationship between $\boldsymbol{V}_b$ and $\boldsymbol{V}_e$ based on the MoM-REML estimation of the rotated mixed model (2), whose covariance matrix is diagonal. The MoM estimation has a computationally efficient closed-form expression and the rotated model (2) is equivalent to d univariate heteroscedastic linear models (3), which means that evaluations of the restricted likelihood are low-cost and REML optimizations are d separate one-dimensional problems. Leveraging these analytical expressions, we perform iterative updates to obtain estimates of $\boldsymbol{V}_b$ and $\boldsymbol{V}_e$ until the norm of both matrices convergent. In

addition, by concatenating the covariate observation matrix $\widetilde{\boldsymbol{X}}$ and the genotype score vector $\widetilde{\boldsymbol{g}}$, we can obtain the generalized least squares (GLS) estimate of the genetic effect $\widetilde{\boldsymbol{\beta}}$. The GLS estimation of genetic effects is a byproduct of the F-test and can be calculated efficiently. A comprehensive, step-by-step description of the algorithm is provided in Supplementary Notes 3 and 5.

The UK Biobank data

The UK Biobank (UKB) is a large-scale cohort study encompassing approximately 500,000 participants who were aged 40–69 years at the time of recruitment, with comprehensive phenotypic records [13]. All UKB participants provided written informed consent, and the study protocols were approved by the National Research Ethics Service Committee. Raw genetic data was collected by Affymetrix genome-wide genotyping chips on two different arrays: the Applied Biosystems UK Biobank Axiom Array and the Applied Biosystems UK BiLEVE Axiom Array [32]. In this study, 488,378 UKB participants of European ancestry and 784,256 genotyped variants were used for simulation and real data analyses.

For the simulation study, we excluded variants satifying any of the following criteria: minor allele frequency (MAF) < 0.01, Hardy–Weinberg equilibrium (HWE) test p-value $< 10^{-4}$, and those removed via linkage disequilibrium (LD) pruning with a 100 bp window size, 5 bp window increment, and LD threshold of $r^2 > 0.999999$. This filtering step retained 542,688 SNPs for subsequent analyses. All QC procedures described above were implemented using PLINK v1.9 [41].

False positive rate simulations

We conducted simulation studies to evaluate the false positive rate (FPR) performance of the RmvLMM method. Specifically, we considered both a modest number of phenotypes ($d = 10$) and a large number of phenotypes ($d = 50, 100, 200$), with sample sizes ranging from small to moderate ($n = 500,\ 1,000,\ 5,000, 10,000$). We generated phenotype data from mvLMM under the null hypothesis of no genetic association, i.e., $\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{\gamma}}^\top + \widetilde{\boldsymbol{B}} + \widetilde{\boldsymbol{E}}$, where $\mathrm{vec}\big(\widetilde{\boldsymbol{B}}\big) \sim \mathcal{N}\big(\boldsymbol{0}_{nd}, \boldsymbol{V}_b \otimes \boldsymbol{K}\big)$ and $\mathrm{vec}\big(\widetilde{\boldsymbol{E}}\big) \sim \mathcal{N}\big(\boldsymbol{0}_{nd}, \boldsymbol{V}_e \otimes \boldsymbol{I}_n\big)$. For simplicity, only a single covariate was considered. The design matrix $\widetilde{\boldsymbol{X}} = (\widetilde{x}_{ij})_{n\times 1}$ had independent entries each following a standard normal distribution ($\widetilde{x}_{ij} \sim N(0,1)$), and the coefficient vector $\widetilde{\boldsymbol{\gamma}} = (\widetilde{\gamma}_j)_{d\times 1}$ had independent entries each following a uniform distribution ($\widetilde{\gamma}_j \sim U(0.1, 0.5)$). The polygenic covariance matrix $\boldsymbol{V}_b$ was set to be a compound symmetric structure with a pairwise phenotypic correlation of 0.5, i.e., $\boldsymbol{V}_b = (1-\rho)\boldsymbol{I}_d + \rho \boldsymbol{1}_d \cdot \boldsymbol{1}_d^\top$ ($\rho = 0.5$), and the random error covariance matrix $\boldsymbol{V}_e$ was configured to an autoregressive structure with a decay correlation of 0.4, i.e., $\boldsymbol{V}_e = (\rho^{|i-j|})_{d\times d}$ ($\rho = 0.4$). The relatedness matrix $\boldsymbol{K}$ was calculated by the product of the centered and scaled SNP genotypes of n individuals randomly selected from the real UKB dataset. We calculated the FPRs across a set of significance thresholds: $\alpha \in \{5\times 10^{-2}, 1\times 10^{-2}, 5\times 10^{-3}, 1\times 10^{-3}, 5\times$

$10^{-4}, 1\times10^{-4}\}$, using 10^5 simulation replicates. To further benchmark the FPR performance of RmvLMM and GEMMA under more stringent significance thresholds, we carried out additional simulations with $n = 1000$ and $d = 10$. These simulations extended to a stringent significance level of $\alpha = 1 \times 10^{-7}$ with 10^8 replicates, enabling rigorous assessment of FPR behavior.

Empirical power simulations

We next carried out simulation studies under a series of configurations to assess the statistical power of the RmvLMM. The phenotypes were simulated based on the mvLMM under the alternative hypothesis, i.e., $\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{\gamma}}^\top + \widetilde{\boldsymbol{g}}\widetilde{\boldsymbol{\beta}}^\top + \widetilde{\boldsymbol{B}} + \widetilde{\boldsymbol{E}}$, where $\mathrm{vec}\big(\widetilde{\boldsymbol{B}}\big) \sim \mathcal{N}\big(\boldsymbol{0}_{nd}, \boldsymbol{V}_b \otimes \boldsymbol{K}\big)$ and $\mathrm{vec}\big(\widetilde{\boldsymbol{E}}\big) \sim \mathcal{N}\big(\boldsymbol{0}_{nd}, \boldsymbol{V}_e \otimes \boldsymbol{I}_n\big)$. The settings of design matrix $\widetilde{\boldsymbol{X}}$, covariate coefficient component $\widetilde{\boldsymbol{\gamma}}$, polygenic covariance matrix $\boldsymbol{V}_b$, random error covariance matrix $\boldsymbol{V}_e$, and relateness matrix $\boldsymbol{K}$ were kept consistent with those used in the FPR simulation. The SNP genotype vector $\widetilde{\boldsymbol{g}} = (\widetilde{\boldsymbol{g}}_i)_{n\times1}$ for n-individual was generated from a binomial distribution parameterized by five different minor allele frequencies (MAFs), such that $\widetilde{\boldsymbol{g}}_i \sim Binomial(2, maf)$ with $maf \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. For a robust cross-method performance evaluation, these five MAFs were stratified into three groups: $\{0.1\}$, $\{0.2\}$, and $\{0.3, 0.4, 0.5\}$. Corresponding to each group, we generated the SNP effect vector $\widetilde{\boldsymbol{\beta}}^\top = (\widetilde{\boldsymbol{\beta}}_j)_{1\times d}$. Under the setting of $d = 10$, we adopted a dense signal model where the SNP exerts an effect on all phenotypes. Specifically, the effect size of the SNP on the j-th phenotype β_j, was sampled from a uniform distribution, $\beta_j \sim U(-0.5, 0.5)$. Under the setting of $d = 50, 100, 200$, we adopted a sparse signal model, where the SNP exerted an effect on only a small fraction of phenotypes. Specifically, the SNP effect vector $\widetilde{\boldsymbol{\beta}}$ was generated as

$$\widetilde{\boldsymbol{\beta}}^\top = \Big(\underbrace{-\widetilde{\beta}_1,\, -\widetilde{\beta}_2,\, \ldots,\, -\widetilde{\beta}_a}_{a=\left\lfloor \frac{d-\lfloor 0.618d\rfloor}{2}\right\rfloor},\underbrace{0,\ 0,\ \ldots,\ 0}_{a'=\frac{d}{2}-a},\ \underbrace{\widetilde{\beta}_{1'},\, \widetilde{\beta}_{2'},\, \ldots,\, \widetilde{\beta}_{b'}}_{b'=\left\lceil \frac{d-\lfloor 0.618d\rfloor}{2}\right\rceil},\underbrace{0,\ 0,\ \ldots,\ 0}_{b=\frac{d}{2}-b'}\Big).$$

The number of phenotypes with zero effects was $\lfloor 0.618d \rfloor$, and the nonzero elements were independently and identically distributed as $\widetilde{\beta}_j \sim U(0, 0.5)$ for $j = 1, \ldots, a$ and $j = 1', \ldots, b'$. Here, $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the floor and ceiling functions, respectively. Statistical power was estimated as the proportion of significant p-values across 1000 simulation replicates.

Simulations for large-scale sample sizes

For large sample size ($n = 50,000,\ 100,000,\ 200,000$ and $300,000$), we derived two phenotype subsets ($d = 10, 20$) from the 26 UKB blood biochemistry dataset. To circumvent the memory and time constraints of storing and decomposing the full $\boldsymbol{K}$, we adopted the divided-and-combined strategy: the n individuals were partitioned into q groups of 10,000 each. The sample relatedness matrix $\boldsymbol{K}$ for each group was calculated by the product of the centered and scaled SNP genotype matrix of 10,000 individuals randomly selected from the real UKB

dataset. Then p-values were computed within each group and subsequently aggregated using the omnibus statistic TOP2-O, whose null distribution was fitted by a generalized gamma distributin. Specifically, we repeated the simulation 10^5 times, with one SNP randomly generated each time, and the resulting 10^5 SNPs were regarded to be independent. Subsequently, using the above sampling-based empirical distribution idea, 10,000 SNPs were randomly sampled from them, and the values of the corresponding TOP2-O statistics were calculated to derive the null distribution, which was then used to fit the parameters of the generalized gamma distribution via MLE. For power calculation, we simulated SNP genotypes by randomly selecting individuals from the UKB dataset to construct the group-level $\boldsymbol{K}$ matrice. Dense and sparse signal architectures were assigned to $d = 10$ and $d = 20$, respectively, with the proportion and distribution of non-zero effects identical to those in our small-to-moderate sample scenarios. Power was estimated as the proportion of significant hits across 200 replicates.

Real data analyses

We selected the blood biochemistry phenotypes from 502,387 individuals in the UKB database. A range of key biochemistry markers were measured in the blood sample collected at recruitment (for all 500,000 participants) and at repeat assessment approximately 5 years later (for 20,000 participants). The biomarkers were selected for analysis because they represent established risk factors for disease, are established diagnostic measures, or characterised phenotypes not otherwise well assessed and are feasible to measure at scale. First, phenotypes with a sample missing rate $> 20\%$ were excluded, leaving 26 phenotypes. Subsequently, samples with missing values for the remaining phenotypes were removed, and further filtered out samples with a genotypic missing rate $> 5\%$. For each phenotype, we followed the GEMMA method [28] to quantile-transform phenotypic values to a standard normal distribution. For SNP-level QC, we removed SNPs with missingness $> 5\%$, MAF < 0.01, and HWE test p-value $< 10^{-6}$. Additionally, conducted LD pruning on the QC data with SNPs underwent LD pruning 100 bp window size, 5bp window increment, and LD threshold of $r^2 > 0.999999$. Following these procedures, 454,296 SNPs and 323,839 individuals passed all filters and QC steps.

RmvLMM identified 1325 significant SNPs (p-value $< 5 \times 10^{-8}$) in multi-trait GWAS of 26 UKB blood biochemistry phenotypes. We used the MAGMA software to annotate significant SNPs and obtain the mapping relationships between SNPs and genes. Gene coordinates were referenced to the NCBI human genome build 37 gene annotation, with a window of 5 kb upstream and 5 kb downstream of the gene boundaries set for mapping. A total of 373 genes were identified in this analysis.

To determine whether the 373 trait-associated genes have been previously reported, the literature search was performed by accessing the PubMed database via the 'rentrez' R package. For each gene and phenotype, we searched PubMed (title/abstract, human studies) using gene and phenotype names. We collected matching PubMed Identifiers (PMIDs), downloaded XML

records, and parsed the XML to extract target text. A total of 309 genes were found to have prior reports of associations with at least one of these phenotypes. For biological pathway analysis, Enrichr was used. We input the list of 373 candidate genes into the Enrichr online tool. Enrichment analysis was then performed against multiple biological pathway databases. Finally, significantly enriched pathways and related functional terms were identified based on adjusted p-values. Furthermore, we searched the DrugBank database (https://go.drugbank.com) and TTD (https://db.idrblab.net/ttd) using target gene names to obtain the corresponding drug names and information on the diseases they treat.

Discussion

Here, we present RmvLMM, a novel framework for multi-trait GWAS that addresses two long-standing challenges of conventional GWAS methods: low statistical power for association testing and heavy computational burden in dual large-scale GWAS applicable to biobank-scale samples and a large number of phenotypes. By integrating an orthogonal rotation technique with a parallelizable divided-and-combined strategy, this method overcomes key limitations of existing approaches, enables robust detection of pleiotropic variants, and represents a valuable advance for dissecting the genetic architecture of complex traits.

Unlike GEMMA [28], RmvLMM implements an orthogonal rotation transformation on the traditional mvLMM. This transformation accentuates significant genetic signals while treating non-significant signals as noise, thereby improving statistical power in multi-trait GWAS analyses for moderate-scale data. Additionally, GEMMA relies on REML estimation of the covariance matrix, with a dominant time complexity of $\mathcal{O}(nd^6)$, whereas RmvLMM integrates MoM and REML estimation to reduce the complexity to $\mathcal{O}(nd^2)$. While both GEMMA and RmvLMM have a dominant time complexity term of $\mathcal{O}(n^3)$ for core computations, RmvLMM adopts a highly parallelizable divided-and-combined strategy, which partitions samples into multiple subsets and synthesizes association evidence via a novel omnibus statistic. By combining these two methodological strategies, RmvLMM achieves computational efficiency for dual large-scale multi-trait GWAS, which remains computationally infeasible with GEMMA.

Despite its strengths, our study has several limitations that warrant consideration. First, the null hypothesis of RmvLMM is that none of the d phenotypes are associated with the tested SNP, while the alternative hypothesis is that at least one phenotype is associated with the SNP. Thus, RmvLMM enables multi-trait GWAS analysis of high-dimensional phenotypes to detect SNPs significantly associated with at least one of these phenotypes. However, it cannot identify the specific phenotype(s) driving the association. We therefore recommend a two-stage follow-up procedure in practice. In the first stage, RmvLMM performs initial screening to detect SNPs associated with one or more phenotypes. In the second stage, each SNP reaching significance in the first stage is subjected to phenome-wide association study (PheWAS) [42] to precisely delineate its phenotypic association profile.

A second practical consideration for RmvLMM concerns hyperparameter selection in the divided-and-combine strategy, specifically whether to partition the data and how to determine the optimal number of groups. Computationally, partitioning substantially reduces memory usage and accelerates eigendecomposition. For datasets involving hundreds of thousands of SNPs and around 10,000 individuals, analyses often approach the memory limit of standard computing nodes (e.g., 180 GB), yet can typically be completed within one hour. Memory requirements can be further eased by processing SNPs in smaller batches. As a general guideline, we recommend direct analysis without grouping for datasets with fewer than 10,000 individuals, while biobank-scale cohorts should be partitioned into subsets of approximately 10,000 individuals. Statistically, excessive partitioning introduces additional estimation variance, which propagates to the omnibus test statistic. Accurate calibration of the empirical null distribution then requires a larger set of reference SNPs. Future work should systematically investigate these properties to establish formal guidelines for optimal data partitioning, specifying the conditions under which partitioning is beneficial and the ideal number of groups.

The applicability of the current method represents an additional limitation. RmvLMM is designed for multiple continuous traits and is thus not applicable to binary or ordinal phenotypes, nor to mixed phenotypes consisting of two or more variable types. To address these limitations, future work will focus on extending the model to the generalized multivariate linear mixed model framework, such as the logistic mixed model for binary traits [12] and the proportional odds model for ordinal traits [43].

Looking ahead, the development of multi-trait GWAS methodologies will continue to be driven by the need to dissect complex genetic pleiotropy. Building on RmvLMM, future work in this field should focus on integrating multi-omics data, gene-environment (G×E) interaction models, and gene-gene (G×G) epistasis models, so as to move beyond conventional genotype–phenotype association mapping and toward a more mechanistic understanding of the genetic architecture underlying complex traits.

Conclusion

In summary, RmvLMM establishes a powerful and scalable framework for multi-trait GWAS analysis that addresses key limitations of existing methods. We anticipate that this approach will become a widely used tool in the GWAS field, facilitating the dissection of the genetic architecture of complex traits and accelerating the translation of genetic discoveries into functional and clinical insights.

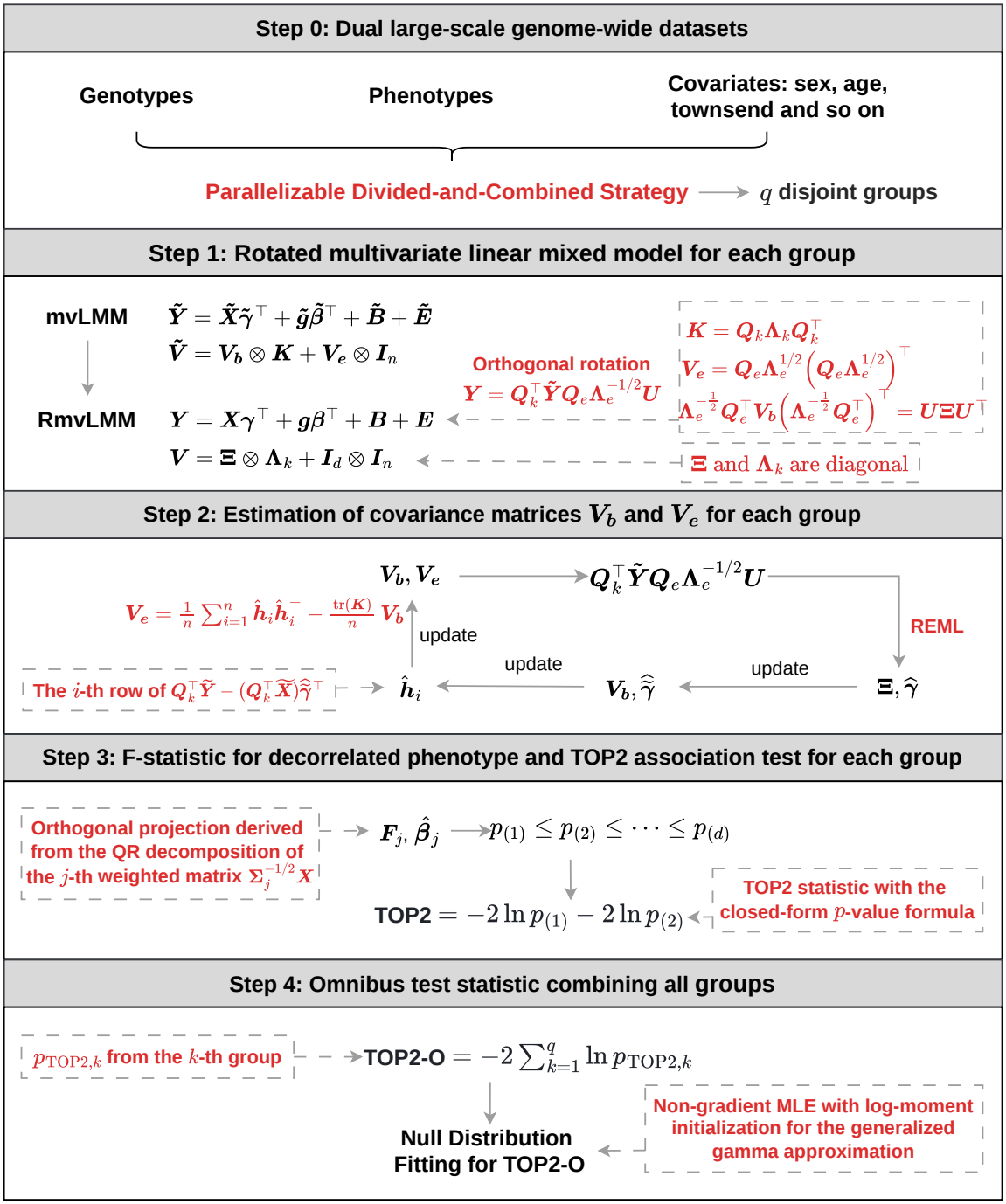


Figure 1: The framewok of RmvLMM method. For dual large-scale genome-wide dataset, a highly parallelizable divided-and-combined strategy is adopted to partition the samples into q disjoint groups. The RmvLMM procedure consists of four steps: (1) Perform orthogonal rotations on the multivariate linear mixed model to decorrelate the phenotypes within each group; (2) Estimate the covariance matrices for each group using the MoM-REML iterative algorithm; (3) Calculate the F statistics and corresponding p-values based on the rotated phenotypes, and aggregate them into the TOP2 statistic for each group; (4) Combine the TOP2 p-values across the q groups using the omnibus statistic TOP2-O. The null distribution of TOP-O is fitted by a generalized gamma distribution, whose parameters are estimated via non-gradient maximum likelihood with log-moment initialization.

Table 1: Computational time comparison of RmvLMM and GEMMA for parameter estimation in a single association test, and for association tests in GWASs. Results are shown for the UK Biobank data sets. All computation was performed on a single core of an Hygon 7285 2.50GHz 64C CPU. n is the number of individuals, p is the number of SNPs, d is the number of phenotypes, c is the number of covariates, t is the number of iterations of our algorithm, t_1 is the iteration count of the Expectation-Maximization (EM) type algorithm in GEMMA, t_2 is the iteration count of the Newton-Raphson (NR) type algorithm in GEMMA.

Method	**Time Complexity**	**Computation Time**					
		$(n = 500, s = 539,852)$			$(n = 5000, s = 542,080)$		
		$d = 5$	$d = 10$	$d = 20$	$d = 5$	$d = 10$	$d = 20$
Parameter estimate and association test for a single SNP							
RmvLMM	$\mathcal{O}\big(n^3 + n^2d + n^2c + t(c^3d + d^3 + cd^2 + nd^2 + nc^2d)\big)$	2.5 s	4.4 s	10.0s	32.4 s	39.1 s	63.7 s
GEMMA	$\mathcal{O}\big(n^3 + n^2d + n^2c + t_1nc^2d^2 + t_2nc^2d^6\big)$	11.0 s	53.0 s	945.0 s	110.0 s	528.0 s	9672.0 s
Parameter estimate and association tests for genome-wide SNPs							
RmvLMM	$\mathcal{O}\big(n^3 + n^2d + n^2c + t(c^3d + d^3 + cd^2 + nd^2 + nc^2d) + p(n^2 + ncd + d^2)\big)$	28.6 s	35.4 s	41.9 s	4.2 m	4.5 m	4.8 m
GEMMA	$\mathcal{O}\big(n^3 + n^2d + n^2c + p(n^2 + t_1nc^2d^2 + t_2nc^2d^6)\big)$	149.0 s	306.0 s	1533.0 s	44.2 m	62.0 m	242.0 m

s denotes seconds and m denotes minutes.

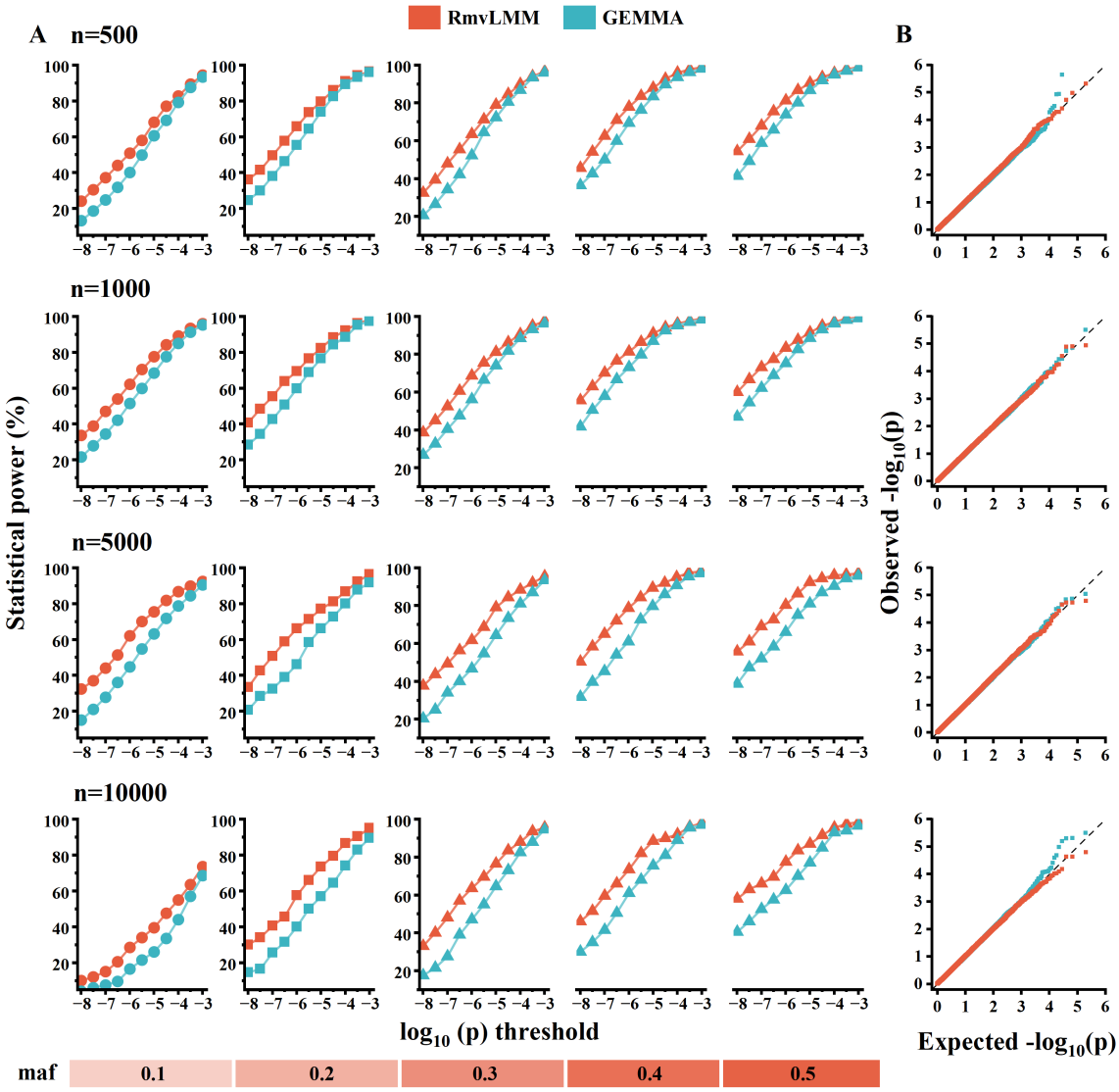


Figure 2: Comparison of statistical power and false positive rate (FPR) between RmvLMM and GEMMA in simulation analyses. Simulations were performed with $d = 10$ dense-signal phenotypes and small-to-moderate sample sizes ($n = 500, 1\,000, 5\,000, 10\,000$). We considered five different minor allele frequencies and divided them into three groups ($maf \in \{0.1\}, \{0.2\}, \{0.3, 0.4, 0.5\}$). Statistical power was quantified as the fraction of significant p-values across 1,000 simulation replicates, evaluated at a series of significance thresholds ($\alpha = 1 \times 10^{-3}$ to 1×10^{-8}). FPR was assessed using additional significance cutoffs ($\alpha \in \{5 \times 10^{-2}, 1 \times 10^{-2}, 5 \times 10^{-3}, 1 \times 10^{-3}, 5 \times 10^{-4}, 1 \times 10^{-4}\}$), defined as the proportion of variants with p-values below the respective α; FPR analyses were conducted with 10^5 independent simulation replicates.

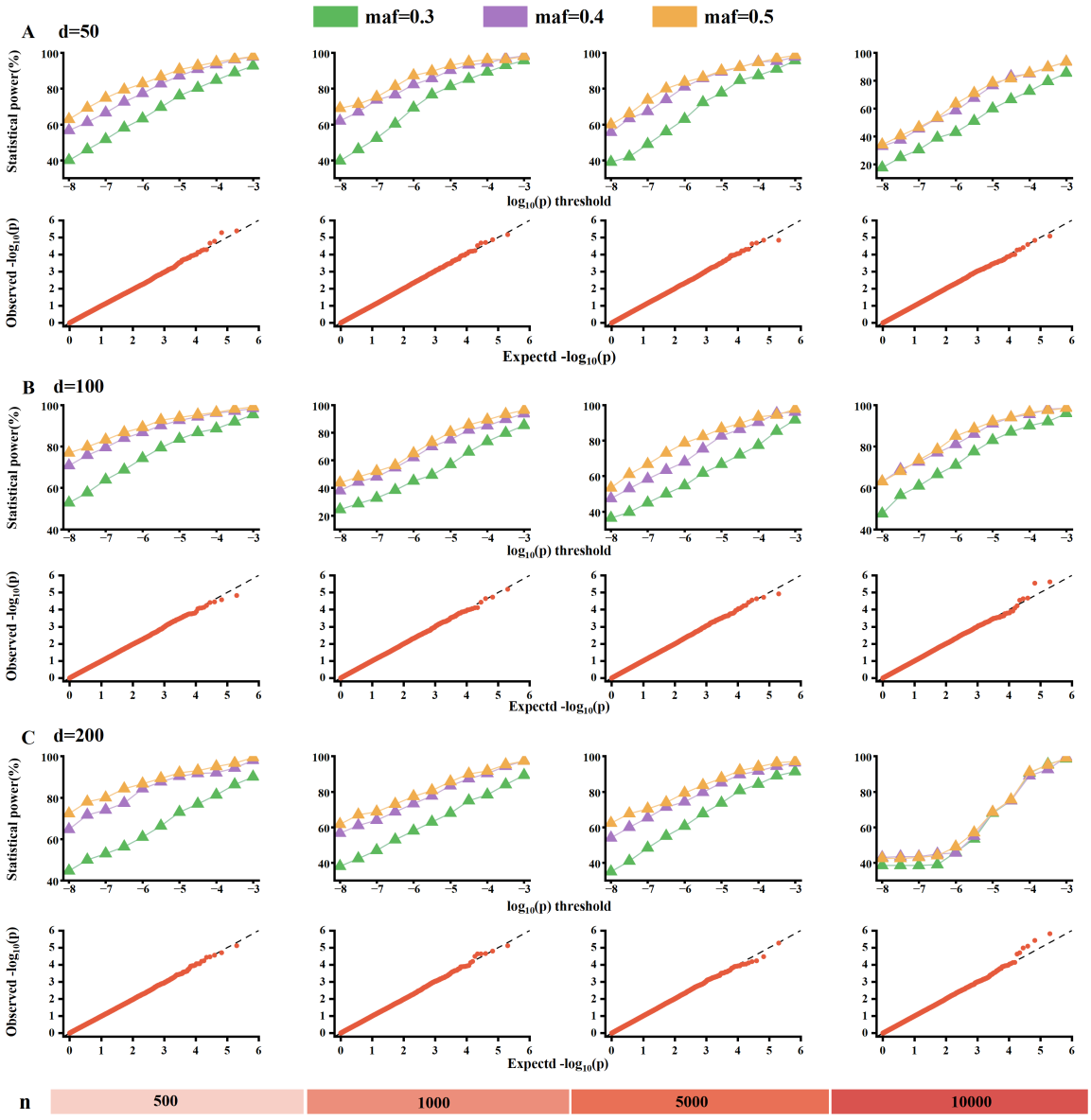


Figure 3: Statistical power and false positive rate (FPR) of RmvLMM in simulation analyses with a large number of phenotypes ($d = 50, 100, 200$). All simulation were performed with sparse-signal phenotypes, where the proportion of phenotypes with zero effects corresponded to $\lfloor 0.618d \rfloor$. $\lfloor \cdot \rfloor$ denotes the floor function. We considered three different minor allele frequencies ($maf \in \{0.3, 0.4, 0.5\}$) and small-to-moderate sample sizes ($n = 500, 1\,000, 5\,000, 10\,000$).

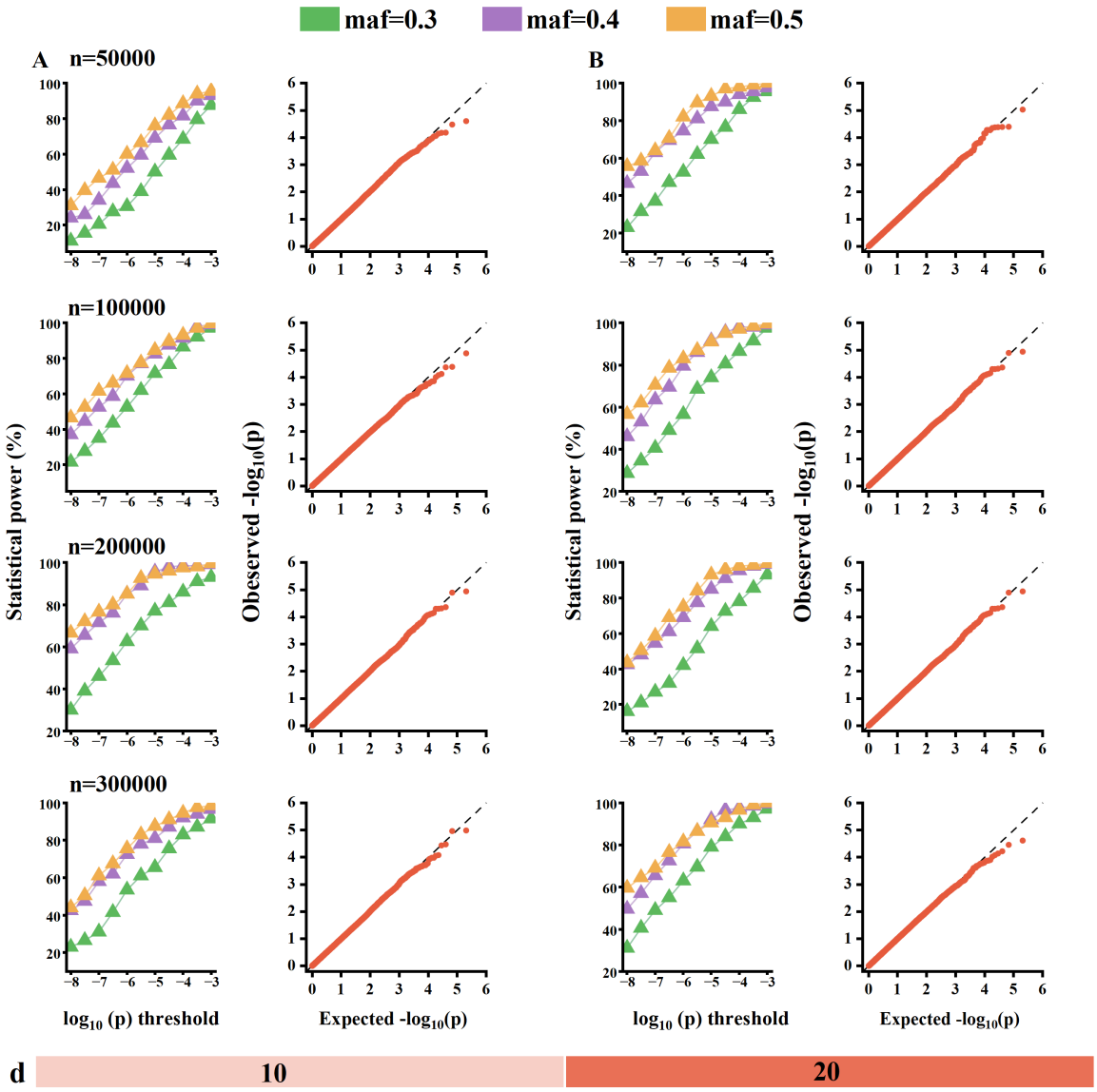


Figure 4: Statistical power and false positive rate (FPR) of RmvLMM in simulation analyses with large sample sizes (n = 50,000, 100,000, 200,000, 300,000). Simulation were performed with $d = 10$ dense-signal phenotypes and $d = 20$ sparse-signal phenotypes. We considered three different minor allele frequencies ($maf \in \{0.3, 0.4, 0.5\}$).

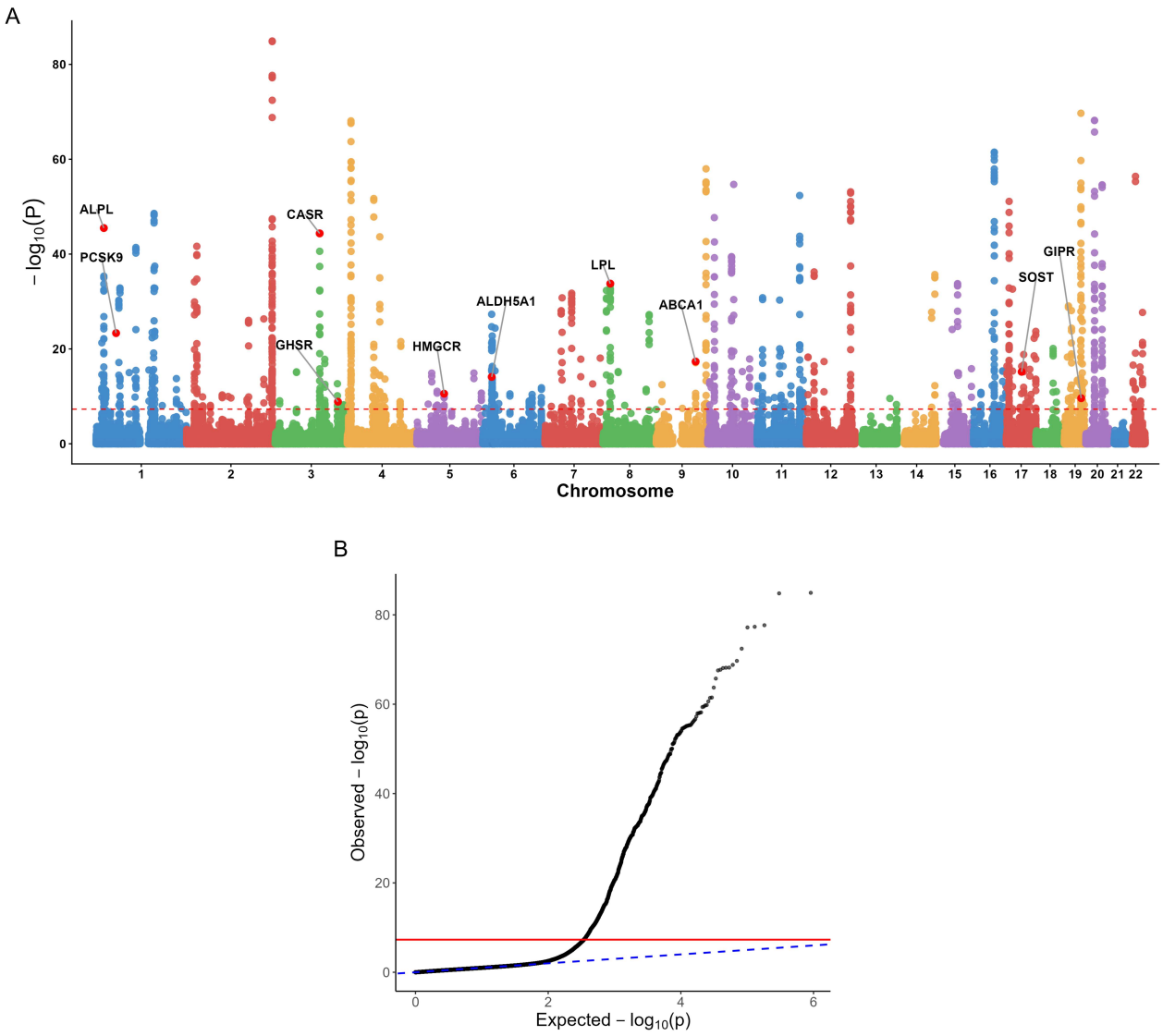


Figure 5: (A) Manhattan plot of the multi-trait GWAS for 26 blood biochemical phenotypes using RmvLMM (323,839 individuals; 454,296 SNPs). 10 genes annotated in the plot have been validated as drug target genes, as identified via searches of the DrugBank and TTD databases. The horizontal red dashed line indicates the genome-wide significance threshold of $p < 5 \times 10^{-8}$. Autosomes are numbered 1–22; X and Y sex chromosomes were not included. (B) Q–Q plot of the multi-trait GWAS for 26 blood biochemical phenotypes using RmvLMM (323,839 individuals; 454,296 SNPs). The blue dashed line represents the 45° reference line, and the horizontal red solid line indicates the genome-wide significance threshold of $p < 5 \times 10^{-8}$.

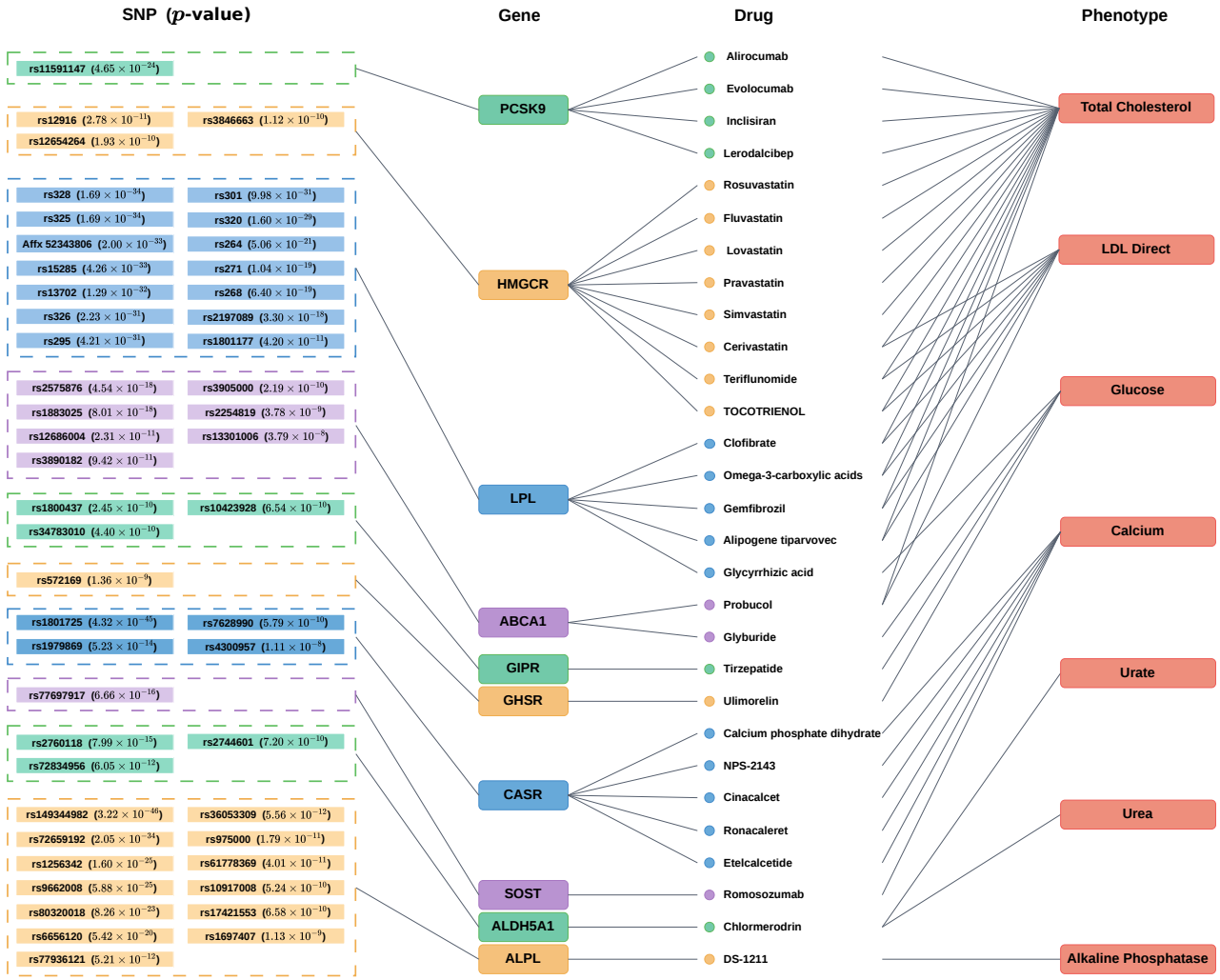


Figure 6: The multi-layer association network illustrates the relationships between GWAS-identified SNPs, pleiotropic genes, therapeutic drugs, and blood biochemistry phenotypes. The diagram illustrates 10 pleiotropic genes (central column) identified by RmvLMM multi-trait GWAS, with their associated genome-wide significant SNPs (left column, colored by gene locus and labeled with discovery p-values), 29 known therapeutic drugs targeting these genes (middle-right column), and the corresponding blood biochemistry phenotypes (right column). Lines denote direct associations: SNPs map to their host genes, genes link to their approved or investigational drugs, and drugs connect to the blood traits they influence.

References

[1] Southam, L. & Zeggini, E. Twenty years of genome-wide association studies. *Nature* **641**, 47–49 (2025).

[2] Abdellaoui, A., Yengo, L., Verweij, K. J. & Visscher, P. M. 15 years of GWAS discovery: realizing the promise. *American Journal of Human Genetics* **110**, 179–194 (2023).

[3] Yu, J. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**, 203–208 (2006).

[4] Kang, H. M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723 (2008).

[5] Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* **42**, 348–354 (2010).

[6] Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* **42**, 355–360 (2010).

[7] Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nature Methods* **8**, 833–835 (2011).

[8] Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44**, 825–830 (2012).

[9] Listgarten, J. *et al.* Improved linear mixed models for genome-wide association studies. *Nature Methods* **9**, 525–526 (2012).

[10] Svishcheva, G. R., Axenovich, T. I., Belonogova, N. M., van Duijn, C. M. & Aulchenko, Y. S. Rapid variance components-based method for whole-genome association analysis. *Nature Genetics* **44**, 1166–1170 (2012).

[11] Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**, 821–824 (2012).

[12] Chen, H. *et al.* Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *The American Journal of Human Genetics* **98**, 653–666 (2016).

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

[14] Loh, P. R. *et al.* Efficient bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics* **47**, 284–290 (2015).

[15] Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nature Genetics* **50**, 906–908 (2018).

[16] Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics* **50**, 1335–1341 (2018).

[17] Jiang, L., Zheng, Z., Fang, H. & Yang, J. A generalized linear mixed model association tool for biobank-scale data. *Nature Genetics* **53**, 1616–1621 (2021).

[18] Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics* **53**, 1097–1103 (2021).

[19] Loya, H., Kalantzis, G., Cooper, F. & Palamara, P. F. A scalable variational inference approach for increased mixed-model association power. *Nature Genetics* **57**, 461–468 (2025).

[20] Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics* **14**, 483–495 (2013).

[21] Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics* **51**, 1339–1348 (2019).

[22] Mackay, T. F. C. & Anholt, R. R. H. Pleiotropy, epistasis and the genetic architecture of quantitative traits. *Nature Reviews Genetics* **25**, 639–657 (2024).

[23] Sivakumaran, S. *et al.* Abundant pleiotropy in human complex diseases and traits. *American Journal of Human Genetics* **89**, 607–618 (2011).

[24] Cotsapas, C. *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genetics* **7**, e1002254 (2011).

[25] Li, X. *et al.* A statistical framework for multi-trait rare variant analysis in large-scale whole-genome sequencing studies. *Nature Computational Science* **5**, 125–143 (2025).

[26] Korte, A. *et al.* A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics* **44**, 1066–1071 (2012).

[27] Aschard, H. *et al.* Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *The American Journal of Human Genetics* **94**, 662–676 (2014).

[28] Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods* **11**, 407–409 (2014).

[29] Zhang, W. *et al.* GATE: an efficient procedure in study of pleiotropic genetic associations. *BMC Genomics* **18**, 552 (2017).

[30] Furlotte, N. A. & Eskin, E. Efficient multiple-trait association and estimation of genetic correlation using the matrix-variate linear mixed model. *Genetics* **200**, 59–68 (2015).

[31] Zaitlen, N. *et al.* Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genetics* **9**, e1003520 (2013).

[32] Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale data. *Nature Genetics* **51**, 1749–1755 (2019).

[33] de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of gwas data. *PLoS Computational Biology* **11**, e1004219 (2015).

[34] Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research* **44**, W90–W97 (2016).

[35] Schwartz, G. G. *et al.* Alirocumab and cardiovascular outcomes after acute coronary syndrome. *New England Journal of Medicine* **379**, 2097–2107 (2018).

[36] Sabatine, M. S. *et al.* Efficacy and safety of evolocumab in reducing lipids and cardiovascular events. *New England Journal of Medicine* **372**, 1500–1509 (2015).

[37] Abifadel, M. *et al.* Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nature Genetics* **34**, 154–156 (2003).

[38] Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).

[39] Collins, R. *et al.* Interpretation of the evidence for the efficacy and safety of statin therapy. *The Lancet* **388**, 2532–2561 (2016).

[40] Astle, W. & Balding, D. J. Population structure and cryptic relatedness in genetic association studies. *Statistical Science* **24**, 451–471 (2009).

[41] Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).

[42] Bush, W. S., Oetjens, M. T. & Crawford, D. C. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nature Reviews Genetics* **17**, 129–145 (2016).

[43] Yan, J., Zhang, X., Wang, S., Wang, J. & Li, Q. The sequence kernel association test for the proportional odds model. *Bioinformatics* **41**, btaf304 (2025).

Data availability

The individual-level genotype and phenotype data are available through formal application to the UK Biobank (https://www.ukbiobank.ac.uk/).

Code availability

RmvLMM is available at https://github.com/amss-stat/RmvLMM.

Acknowledgements

This research was supported by National Natural Science Foundation of China (Grant No. 12325110, 12401344, 12288201). This study has been conducted using UK Biobank resource under Application Number 97563.

Author contributions

Q.L. conceived and supervised the study. Q.L., H.G., and Q.Z. designed the experiment. Q.Z. developed the software tools. X.Z. performed the simulations and data analyses under the assistance and guidance from Q.L., H.G., and Q.Z.. Y.Q. provided the real datasets and participated in real data analysis. H.G. and Q.Z. wrote the manuscript with the participation of all authors. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available in the submitted version of this paper.

Correspondence and requests for materials should be addressed to Qizhai Li.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RmvLMMsupplementary20260516.pdf