A unified Bayesian approach to transcriptome-wide association study

Arnab Kumar Khan[1], Tanushree Haldar[2], and Arunabha Majumdar[3*]

[1]Human Genetics Unit, Indian Statistical Institute Kolkata, Kolkata, West Bengal 700103, India
[2]Institute for Human Genetics, University of California San Francisco, San Francisco, California 94143, USA
[3]Department of Mathematics, Indian Institute of Technology Hyderabad, Kandi, Telangana 502285, India

*Correspondence: arun.majum@math.iith.ac.in

Abstract

Transcriptome-wide association study (TWAS) has shed light on molecular mechanisms by examining the roles of genes in complex disease etiology. TWAS facilitates gene expression mapping studies based on a reference panel of transcriptomic data to build a prediction model to identify expression quantitative loci (eQTLs) affecting gene expressions. These eQTLs leverage the construction of genetically regulated gene expression (GReX) in the GWAS data and a test between imputed GReX and the trait indicates gene-trait association. Such a two-step approach ignores the uncertainty of the predicted expression and can lead to reduced inference accuracy, e.g., inflated type-I error in TWAS. To circumvent a two-step approach, we develop a unified Bayesian method for TWAS, combining the two datasets simultaneously. We consider the horseshoe prior in the transcriptome data while modeling the relationship between the gene expression and local SNPs and the spike and slab prior while testing for an association between the GReX and the trait. We extend our approach to conducting a multi-ancestry TWAS, focusing on discovering genes that affect the trait in all ancestries. We have shown through simulation that our method gives better estimation accuracy for GReX effect size than other methods. In real data, applying our method to the GEUVADIS expression study and the GWAS data from the UK Biobank revealed several novel genes associated with the trait body mass index (BMI).

1 Introduction

High throughput genotyping motivated the simple yet powerful design of genome-wide association studies (GWAS), which have revolutionized the mapping of novel genetic variants modulating complex phenotypes [1]. However, the genetic variants identified by GWAS often fall in the non-coding region of the genome, making it challenging to establish their functional relationship with the phenotype. To better understand the biological mechanism by which these variants affect the phenotype, we need to consider intermediate molecular phenotypes together with genotypes of genetic variants. Gene expression crucially participates in complex phenotype variation. Regulatory variants, e.g., expression quantitative trait loci (eQTLs) [2], control the gene expression levels. The co-localization of eQTLs with GWAS signals underscores the importance of integrating GWAS and gene expression data to explain the functional role of GWAS associations in phenotype biology.

TWAS combines two data sets to test for the association between a gene and phenotype. In the transcriptome data, the gene expression levels and genotypes of cis/local SNPs for the gene are available. On the other hand, we have data on the genotypes of local SNPs and the phenotype in the GWAS data. However, the gene expressions are unavailable mainly due to limited resources. The sample size of the transcriptome data

is usually substantially smaller than the GWAS data. We train a regression model in the transcriptome data, considering gene expression as the response variable and the genotypes of local SNPs as the predictors. The estimated local SNPs' effect sizes from the transcriptome data are used to impute the genetically regulated component of gene expression (GReX) in the GWAS data. Finally, we test for an association between the outcome phenotype and predicted GReX in the GWAS data. TWAS offers a few advantages over GWAS. First, TWAS is a gene-level test with less of a multiple-testing burden. Second, TWAS bridges the functional gap in GWAS by incorporating gene expression as a potential causal mediator from SNPs to phenotype.

Most currently available TWAS methods [3] employ the above two-step strategy. For example, a penalized regression, such as Lasso [4] or Elastic net [5], is used in the first step to fit the prediction model of expression, and a linear or logistic regression is used in the second step to regress the outcome phenotype on the GReX. These approaches ignore the uncertainty of the predicted expression in the second step, which emerges due to the variance of the estimated local SNPs' effect sizes in the transcriptome data. Unadjustment of the noise in the imputed expression can reduce the overall accuracy of TWAS, e.g., inflation of type 1 error rate in a frequentist framework (Leeuw et al. [6]). Instead of a two-step strategy, A unified statistical framework that jointly models the two datasets simultaneously can resolve this limitation. One such method is METRO (Li et al. [7]), which proposed a unified approach to multi-ancestry TWAS. However, the method implemented the ridge regression while modeling the relationship between the expression and the local SNPs. The ridge regression applies the same shrinkage to the regression coefficients corresponding to all the local SNPs. However, more attractive penalized regressions, such as Lasso, Adaptive Lasso [8], Elastic net, etc., shrink the effect sizes differentially to prioritize the most important predictors. Bayesian regularization methods consider various shrinkage priors for the same purpose. Global local shrinkage priors are one such important class of priors. To leverage the advantage of such efficient regularization methods, we consider the horseshoe prior (Carvalho et al. [9]), a popular choice of global-local shrinkage priors, while building the relationship between the gene expression and local SNPs in the transcriptome data.

In this article, we develop a unified Bayesian framework for TWAS, combining the two datasets simultaneously. We consider the horseshoe prior in the transcriptome data while modeling the relationship between the gene expression and local SNPs and the spike and slab prior (SSP) [10] while testing for an association between the GReX and the phenotype. The joint Bayesian approach passes around the impromptu genre of the two-step approaches. It avails the benefits of well-founded priors to model the sparsity in the transcriptome data and provides a comprehensive Bayesian inference on the gene-phenotype association. In TWAS, we do not need an explicit variable selection when modeling the eQTL effects in the transcriptome data. Hence, we consider the horseshoe prior (HSP) for its robustness in modeling various sparsity in the data. Contrary to Lasso, the HSP shrinks the small effects more toward zero than the large effects. Thus, the HSP should lead to a reliable expression prediction model that is crucial for TWAS. We implement the classical SSP to test for a gene-trait association explicitly.

A recent method, BGW-TWAS [11], used the Bayesian variable selection regression based on the spike and slab prior to modeling the eQTL effects. However, the method also follows a two-step strategy without accounting for the uncertainty of the predicted expression. In principle, another method TIGAR [12], which employed the Bayesian regression based on the Dirichlet process in the first step, suffers due to the same limitation. These methods finally sorted to obtain the p-values of gene-phenotype association. However, a Bayesian approach is well-suited to develop a unified framework to perform a Bayesian inference in this scenario. This is the primary motivation behind our approach, which combines HSP and the classical SSP in a single set-up to model the TWAS data. We obtain the Bayes factor and the related posterior probabilities for testing the gene-trait association. Since our approach analyzes one gene at a time, similarly to standard TWAS, we propose strategies to adjust for the multiple comparisons by calibrating the hyper-parameters of the priors. We refer to our Bayesian method as bayesTWAS.

2 Materials and methods

Single-ancestry TWAS

We first develop our method for a single ancestry. We consider a gene and its expression in the tissue or cell type relevant to the trait. Suppose, the reference panel transcriptome data comprises n_r individuals. Let $\mathbf{x}_{\text{ref}}$ denote the vector of gene expression levels for the n_r individuals. We regard m local SNPs located

around the transcription start and end sites of the gene. Let $\mathbf{G}_{\text{ref}}$ denote $n_r \times m$ genotype matrix for the local SNPs for the individuals in the transcriptome data. We consider the following linear model regressing the expression on the local SNPs' genotypes:

$$\mathbf{x}_{\text{ref}} = \mathbf{G}_{\text{ref}}\boldsymbol{\alpha} + \mathbf{e}, \quad \text{where } \mathbf{e} \sim N_{n_r}(\mathbf{0}, \sigma_e^2 \mathrm{I}_{n_r}). \tag{1}$$

We do not include an intercept term since both vectors of gene expressions and genotypes for each local SNP are standardized. Here, $\boldsymbol{\alpha}$ is a $m \times 1$ vector of local SNPs' effects on the gene expression levels. Each component of the random error vector $\mathbf{e}$ independently follows a normal distribution with mean 0 and common variance σ_e^2.

Next, we consider separate GWAS data that comprises n individuals. Let $\mathbf{y}$ denote the $n \times 1$ vector of trait values for the n individuals in the GWAS data. Let $\mathbf{G}$ stand for the $n \times m$ genotype matrix for the m SNPs in the GWAS sample, where the m SNPs are the same local SNPs considered in the transcriptome data. To model the relation between the trait and the gene expression for the GWAS individuals, we regard another linear model as:

$$\mathbf{y} = \beta(\mathbf{G}\boldsymbol{\alpha}) + \boldsymbol{\epsilon}, \quad \text{where } \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma_\epsilon^2 \mathrm{I}_n). \tag{2}$$

We do not include an intercept term considering $\mathbf{y}$ to be mean-centered. The effect of local SNPs on the expression, $\boldsymbol{\alpha}$, remains the same between the transcriptome (Equation 1) and GWAS data (Equation 2). The term $\mathbf{G}\boldsymbol{\alpha}$ in the Equation 2 represents the true GReX for the GWAS individuals, and β denotes the effect of GReX on the trait. σ_ϵ^2 is the common error variance for each component in the residual vector $\boldsymbol{\epsilon}$. Since the individuals are non-overlapping between the transcriptome and GWAS data, $\mathbf{e}$ and $\boldsymbol{\epsilon}$ are independently distributed. The above two linear models in the Equations 1 and 2 are connected by the common parameter vector $\boldsymbol{\alpha}$. We use the horseshoe prior for $\boldsymbol{\alpha}$, which has attractive statistical properties. The double-exponential prior underlying Lasso and the horseshoe prior shrink the small effects. But the latter refrains from shrinking the sizeable effects [13]. The prior corresponding to Lasso reduces the large effect sizes by a non-vanishing amount. Another appealing property of the horseshoe prior is that the estimated predictive density converges to the true density very fast [9]. In the context of TWAS, we do not require explicit variable selection as done by Lasso or two-group spike and slab prior while modeling the effects of local SNPs on the expression in the transcriptome data. Because eQTL discovery is not the main goal here. We expect very small effects due to many local SNPs and significant effects due to a few. Hence, we prioritize a continuous shrinkage by the horseshoe prior to model the sparsity of the data in a robust manner. Thus,

$$\begin{aligned} \alpha_i | \lambda_i, \tau, \sigma_e^2 &\sim N(0, \sigma_e^2 \tau^2 \lambda_i^2), \quad i = 1(1)m \\ \lambda_i &\sim C^+(0,1) \\ \tau &\sim C^+(0,1) \end{aligned}$$

Here, $\alpha_i, i = 1, \ldots, m$ are independently distributed, and $\lambda_i, i = 1, \ldots, m$ are also independently distributed. $C^+(0,1)$ is the standard half-Cauchy distribution with its support on the positive real numbers. Two distinct kinds of shrinkage parameters are considered to regulate the overall sparsity. For $i = 1, \ldots, m$, λ_i is the local shrinkage parameter specific to the effect of the ith local SNP, whereas τ is the global shrinkage parameter. While τ applies an overall shrinkage of the effects toward zero, the λ_is help the local SNPs with sizeable effect sizes escape the overall shrinkage.

To evaluate the impact of the GReX on the outcome, we employ the Dirac spike and slab prior (cite) for β (Equation 2), the effect of the GReX on the trait as:

$$\beta | q, \sigma_d^2, \sigma_\epsilon^2 \sim (1-q)\delta_{\{0\}}(\beta) + qN(0, \sigma_d^2 \sigma_\epsilon^2) \tag{3}$$

$\delta_{\{0\}}(\beta)$ is the point mass function that takes the value 1 if $\beta = 0$ and 0 when $\beta \neq 0$. The spike component constitutes the null effect, whereas the slab component represents the non-null effect of the GReX on the trait. q is the prior probability that β is non-null. We consider a Beta prior for q as: $q \sim Beta(b_1, b_2)$. Here, $\sigma_d^2 \sigma_\epsilon^2$ denotes the slab variance for the GReX effect size β. The slab variance is proportional to the trait variance σ_ϵ^2 (Equation 2) with σ_d^2 as the ratio between them. Thus, σ_d^2 is a crucial component of the prior variability of β. In the prior, we consider inverse gamma (IG) distributions for the variances, σ_e^2

(Equation 1), σ_ϵ^2, and σ_d^2. Collating the layers of the hierarchical structure (Equations 1 and 2), we model the transcriptome and GWAS data together in a unified Bayesian framework as:

$$
\begin{aligned}
\mathbf{x}_{\text{ref}}|\boldsymbol{\alpha}, \sigma_e^2 &\sim N_{n_r}(\mathbf{G}_{\text{ref}}\boldsymbol{\alpha}, \sigma_e^2 \mathrm{I}_{n_r}) \\
\mathbf{y}|\beta, \boldsymbol{\alpha}, \sigma_\epsilon^2 &\sim N_n(\beta \mathbf{G}\boldsymbol{\alpha}, \sigma_\epsilon^2 \mathrm{I}_n) \\
\alpha_i|\lambda_i, \tau, \sigma_e^2 &\sim N(0, \sigma_e^2 \tau^2 \lambda_i^2), \quad i = 1(1)m \\
\lambda_i &\sim C^+(0,1) \\
\tau &\sim C^+(0,1) \\
\beta|q, \sigma_d^2, \sigma_\epsilon^2 &\sim (1-q)\delta_{\{0\}}(\beta) + qN(0, \sigma_d^2 \sigma_\epsilon^2) \\
q &\sim Beta(b_1, b_2) \\
\sigma_d^2 &\sim IG(d_1, d_2) \\
\sigma_e^2 &\sim IG(e_1, e_2) \\
\sigma_\epsilon^2 &\sim IG(\epsilon_1, \epsilon_2)
\end{aligned}
\tag{4}
$$

Here, $\alpha_i, i = 1, \ldots, m$, are independently distributed in the prior. The same assumption is made for $\lambda_i, i = 1, \ldots, m$.

We implement the MCMC algorithm (cite the algorithm with reference number) to generate the posterior sample of the model parameters and perform a comprehensive Bayesian inference. We devise a Metropolis within the Gibbs sampler to obtain the posterior sample. We provide the MCMC algorithm in Algorithm 1. The detailed derivations of the full conditional distributions are presented in the supplemental material.

Estimation and testing

In our TWAS approach, β denotes the effect of GReX on the outcome (Equation 2). We consider the posterior mean of β using the MCMC sample as the estimate of β and calculate the posterior SD based on the sample to assess the uncertainty involved in the estimate. We regard the following null hypothesis $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$. We design a Bayesian testing strategy based on the Bayes factor. Let D denote the complete data, including the gene expression and outcome in the transcriptome and GWAS data. The posterior odds of H_1 versus H_0 is $\frac{P[H_1|D]}{P[H_0|D]}$ and the prior odds is $\frac{P[H_1]}{P[H_0]}$. The Bayes factor for testing H_1 versus H_0 is given by:

$$
\text{BF} = \frac{\text{Posterior odds}}{\text{Prior odds}} = \frac{P[H_1|D]}{P[H_0|D]} \frac{P[H_0]}{P[H_1]} = \frac{P[\beta \neq 0|D]}{P[\beta = 0|D]} \frac{P[\beta = 0]}{P[\beta \neq 0]}
$$

$P[H_0|D] = P[\beta = 0|D]$ stands for the posterior probability of null association (PPNA). The higher the value of the Bayes Factor, the stronger the evidence against H_0. The expression of prior odds of H_1 versus H_0 is $\frac{b_2}{b_1}$. For the computation of posterior odds, we estimate the PPNA as follows:

$$
\begin{aligned}
P[H_0|D] &= P[\beta = 0|D] \\
&= \int_{\sigma_\epsilon^2} \int_{\sigma_d^2} \int_q \int_{\boldsymbol{\alpha}} P[\beta = 0, \boldsymbol{\alpha}, q, \sigma_d^2, \sigma_\epsilon^2|D] \; d\boldsymbol{\alpha} \; dq \; d\sigma_d^2 \; d\sigma_\epsilon^2 \\
&= \int_{\sigma_\epsilon^2} \int_{\sigma_d^2} \int_q \int_{\boldsymbol{\alpha}} P[\beta = 0|\boldsymbol{\alpha}, q, \sigma_d^2, \sigma_\epsilon^2, D] f(\boldsymbol{\alpha}, q, \sigma_d^2, \sigma_\epsilon^2|D) \; d\boldsymbol{\alpha} \; dq \; d\sigma_d^2 \; d\sigma_\epsilon^2 \\
&= E_{\boldsymbol{\alpha}, q, \sigma_d^2, \sigma_\epsilon^2|D} \left[P[\beta = 0|\boldsymbol{\alpha}, q, \sigma_d^2, \sigma_\epsilon^2, D] \right] \\
&\approx \frac{1}{N} \sum_{i=1}^{N} P\left[\beta = 0|\boldsymbol{\alpha}^{(i)}, q^{(i)}, \sigma_d^{2(i)}, \sigma_\epsilon^{2(i)}, D\right],
\end{aligned}
\tag{5}
$$

where $\left(\boldsymbol{\alpha}^{(i)}, q^{(i)}, \sigma_d^{2^{(i)}}, \sigma_\epsilon^{2^{(i)}}\right)$ denotes the ith posterior draw corresponding to parameters $(\boldsymbol{\alpha}, q, \sigma_d^2, \sigma_\epsilon^2)$ from MCMC and N is the number of MCMC iterations. The last line in Equation 5 corresponds to the Rao-Blackwellised estimate (Casella and Robert [14]) for the PPNA. The term $P\left[\beta = 0 | \boldsymbol{\alpha}^{(i)}, q^{(i)}, \sigma_d^{2^{(i)}}, \sigma_\epsilon^{2^{(i)}}, D\right]$ evaluated at the ith step of MCMC denotes the posterior probability that β takes value 0 in the full conditional distribution and the detailed expression is presented in the supplemental material.

Hyperparamters

We discuss the choice of the hyper-parameters while testing a single gene for association. The prior distribution of q is $Beta(b_1, b_2)$ with the shape parameters b_1 and b_2. We note that $E(q) = \frac{b_1}{b_1+b_2} = P(H_1)$, the prior probability that the gene is associated. Since only a small proportion of genes are expected to be associated at a genome-wide level, we regard $P(H_1) = 0.05$. For simplicity, we fix $b_1 = 1$. We choose b_2 such that $P(H_1) = 0.05$. For this choice, $P(0 < q < 0.2)$ turns out to be 0.99. We will discuss a strategy to adjust for multiple comparisons while testing many genes in the real data application later.

Next, we discuss the choices of the hyper-parameters for the inverse-gamma priors of the variance terms. Following previous literature [15], we consider weakly informative priors for σ_e^2 and σ_ϵ^2. We choose $e_1 = e_2 = \epsilon_1 = \epsilon_2 = 0.1$. This is a flat prior that distributes the total mass over a wide range of parameter values. We choose the hyperparameters $d_1 = 1$ and $d_2 = 0.18$ for the prior of σ_d^2. This choice provides 95% prior probability that $\sigma_d^2 \in (0, 3.6)$. This choice is guided by our real data application, which will be discussed later. We note that a moderate sample size of the transcriptome data and a large sample size of the GWAS data alleviate the impact of the specific choices of the hyperparameters on the Bayesian inference.

Multi-ancestry TWAS

In this section, we extend our TWAS approach to a multi-ancestry set-up. To enhance the genetic diversity in TWAS analysis, we leverage expression and GWAS data from multiple ancestries. We assume that a pair of transcriptome and GWAS data are available for each of K ancestries. Suppose we consider a single gene and its expression in the same tissue or cell type of relevance in all the ancestries. We denote $\mathbf{x}_{k,\text{ref}}$ as $n_k \times 1$ vector of gene expression levels collected from n_k individuals in the kth ancestry, $k = 1, \cdots, K$. Let $\mathbf{G}_{k,\text{ref}}$ denote $n_k \times m$ local SNPs' genotype matrix in the kth ancestry, where m is the number of local SNPs. Suppose $\mathbf{y}_k$ is the vector of trait values for N_k individuals in the GWAS data for k^{th} ancestry, $k = 1, \cdots, K$. Let $\mathbf{G}_k$ stand for $N_k \times m$ local SNPs' genotype matrix in the GWAS data for k^{th} ancestry. The local SNPs are the same in the transcriptome and GWAS data. We consider the ancestry-specific linear model for the transcriptome data as follows:

$$\mathbf{x}_{k,\text{ref}} = \mathbf{G}_{k,\text{ref}}\boldsymbol{\alpha}_k + \mathbf{e}_k, \quad \text{where } \mathbf{e}_k \sim N_{n_k}(\mathbf{0}, \sigma_{e_k}^2 \mathrm{I}_{n_k}), \quad k = 1, \ldots, K \tag{6}$$

Here, $\boldsymbol{\alpha}_k$ represents the ancestry-specific effect sizes of the local SNPs on the gene expression, and $\sigma_{e_k}^2$ is the error variance for each component of the residual vector $\mathbf{e}_k$. We include different $\boldsymbol{\alpha}_k$s since eQTLs can vary across populations with possibly distinct effect sizes. Since the heritability of the gene expression due to the local SNPs can differ among the ancestries, we regard ancestry-specific error variance term $\sigma_{e_k}^2$. The gene-outcome relationship is modelled for the kth GWAS data as:

$$\mathbf{y}_k = \beta(\mathbf{G}_k\boldsymbol{\alpha}_k) + \boldsymbol{\epsilon}_k, \quad \text{where } \boldsymbol{\epsilon}_k \sim N_{N_k}(\mathbf{0}, \sigma_\epsilon^2 \mathrm{I}_{N_k}). \tag{7}$$

We make the key assumption here that the effect of the GReX on the trait is the same across ancestries. Recent literature has demonstrated that genetic effect sizes for a complex trait are strongly correlated over populations [16]. Lu et al. [17] considered the same effect sizes of the genes. Hence, we consider the same effect size of the GReX on the trait across ancestries to better identify the gene-trait association. The assumption is along the lines of a fixed-effects meta-analysis. If this model discovers a gene associated with the trait, the association holds for all the ancestries. Since a single gene explains a tiny proportion of the variability of a complex trait, we assume a common error variance term σ_ϵ^2 across the ancestries. $\mathbf{G}_k\boldsymbol{\alpha}_k$ in the Equation 7 represents the true GReX in the kth ancestry. Therefore, we include the flexibility that local

SNPs (or eQTLs) can impact the gene expression differentially across populations, whereas the population-specific GReX has the same effect on the outcome. However, we carefully choose the parameters of the horseshoe prior to persuade an overall sharing of the eQTLs and their effect sizes across ancestries as follows:

$$\begin{aligned}
\alpha_{kj}|\lambda_j, \tau, \sigma^2_{e_k} &\sim N(0, \sigma^2_{e_k}\tau^2\lambda_j^2), \quad j = 1(1)m,\ k = 1(1)K. \\
\lambda_j &\sim C^+(0,1) \\
\tau &\sim C^+(0,1)
\end{aligned}$$

For $j = 1, \ldots, m$, α_{kj} is the effect of the jth local SNP on the gene expression in the kth ancestry. We introduce a common local SNP-specific shrinkage parameter λ_j irrespective of the ancestry. This implies a similar shrinkage of the effect sizes of the local SNPs over the populations. The same λ_j authorizes the possible sharing of eQTLs and the correlation between their effect sizes across populations. We note that the sharing is induced weakly because we assumed independence of the effect size distributions across the populations. Similarly, as for single ancestry, we employed the Dirac spike and slab prior (Equation 3) to evaluate the effect of the gene on the trait.

$$\beta|q, \sigma_d^2, \sigma_\epsilon^2 \sim (1-q)\delta_{\{0\}}(\beta) + qN(0, \sigma_d^2\sigma_\epsilon^2)$$

A complete hierarchical structure of the multi-ancestry TWAS model is given by:

$$\begin{aligned}
\mathbf{x}_{k,\text{ref}}|\boldsymbol{\alpha}_k, \sigma^2_{e_k} &\sim N_{n_k}(\mathbf{G}_{k,\text{ref}}\boldsymbol{\alpha}_k, \sigma^2_{e_k}\mathrm{I}_{n_k}), \quad k = 1(1)K \\
\mathbf{y}_k|\beta, \boldsymbol{\alpha}_k, \sigma^2_\epsilon &\sim N_{N_k}(\beta\mathbf{G}_k\boldsymbol{\alpha}_k, \sigma^2_\epsilon\mathrm{I}_{N_k}) \\
\alpha_{kj}|\lambda_j, \tau, \sigma^2_{e_k} &\sim N(0, \sigma^2_{e_k}\tau^2\lambda_j^2), \quad k = 1(1)K,\ j = 1(1)m \\
\lambda_j &\sim C^+(0,1) \\
\tau &\sim C^+(0,1) \\
\beta|q, \sigma_d^2, \sigma_\epsilon^2 &\sim (1-q)\delta_{\{0\}}(\beta) + qN(0, \sigma_d^2\sigma_\epsilon^2) \\
q &\sim Beta(b_1, b_2) \\
\sigma_d^2 &\sim IG(d_1, d_2) \\
\sigma^2_{e_k} &\sim IG(e_1, e_2) \\
\sigma_\epsilon^2 &\sim IG(\epsilon_1, \epsilon_2)
\end{aligned} \tag{8}$$

Similarly, as for the single ancestry, we perform a comprehensive Bayesian inference by implementing the MCMC based on a Metropolis within the Gibbs sampler outlined in Algorithm 2. We present the derivations of the full conditional posterior distributions in the supplemental materials. We use the same choices of hyper-parameters considered for the single-ancestry model. Thus, our approach proposes a unified Bayesian framework to model the transcriptome and GWAS data for performing TWAS in a synchronized manner.

Estimation and testing

The estimation procedure remains the same for the multi-ancestry model, i.e., we use the posterior mean of β and the posterior variance to assess the uncertainty. Even though the prior odds remain unchanged from the single-ancestry model, the explicit expression of the posterior odds, in particular, $P[H_0|D]$, changes as follows.

$$
\begin{aligned}
&P[H_0|D] \\
&= \int_{\sigma_\epsilon^2}\int_{\sigma_d^2}\int_{q}\int_{\boldsymbol{\alpha}_1}\cdots\int_{\boldsymbol{\alpha}_K} P[\beta=0,\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_K,q,\sigma_d^2,\sigma_\epsilon^2|D]\ d\boldsymbol{\alpha}_K\cdots d\boldsymbol{\alpha}_1\ dq\ d\sigma_d^2\ d\sigma_\epsilon^2 \\
&= \int_{\sigma_\epsilon^2}\int_{\sigma_d^2}\int_{q}\int_{\boldsymbol{\alpha}_1}\cdots\int_{\boldsymbol{\alpha}_K} P[\beta=0|\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_K,q,\sigma_d^2,\sigma_\epsilon^2,D]f(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_K,q,\sigma_d^2,\sigma_\epsilon^2|D)d\boldsymbol{\alpha}_K\cdots d\boldsymbol{\alpha}_1\ dq\ d\sigma_d^2\ d\sigma_\epsilon^2 \\
&= E_{\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_K,q,\sigma_d^2,\sigma_\epsilon^2|D}\left[P[\beta=0|\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_K,q,\sigma_d^2,\sigma_\epsilon^2,D]\right] \\
&\approx \frac{1}{N}\sum_{i=1}^{N} P\left[\beta=0|\boldsymbol{\alpha}_1^{(i)},\cdots,\boldsymbol{\alpha}_K^{(i)},q^{(i)},\sigma_d^{2^{(i)}},\sigma_\epsilon^{2^{(i)}},D\right],
\end{aligned}
\tag{9}
$$

where $(\boldsymbol{\alpha}_1^{(i)},\cdots,\boldsymbol{\alpha}_k^{(i)},q^{(i)},\sigma_d^{2^{(i)}},\sigma_\epsilon^{2^{(i)}})$ is the ith posterior draw from N MCMC iterations. The last line of the Equation 9 follows from the Rao-Blackwellization technique and the explicit expression for that is provided in supplemental material.

Gibbs sampling algorithm for single-ancestry TWAS

In this section, we present a brief outline of the Metropolis within the Gibbs sampling algorithm for our TWAS models. First, we present the algorithm for the single-ancestry TWAS model. The notations used in the algorithm carry their usual meaning as discussed in section 2. Let us define $\mathbf{D}_\lambda = diag(\lambda_1^2,\lambda_2^2,\cdots,\lambda_m^2)$, where $\mathbf{D}_\lambda$ is a $m\times m$ diagonal matrix. Then we have $\boldsymbol{\alpha}|\boldsymbol{\lambda},\tau,\sigma_e^2 \sim N_m(\mathbf{0},\sigma_e^2\tau^2\mathbf{D}_\lambda)$. The brief steps of the algorithm are presented as follows.

Algorithm 1 Metropolis within the Gibbs sampling algorithm for single-ancestry TWAS

1: Start

2: Choose initial values for parameters σ_e^2, $\boldsymbol{\lambda}$, τ, β, q, σ_d^2, σ_ϵ^2.

3: loop:

4: Update $\boldsymbol{\alpha}$ by drawing a random vector from $N_m\left(\boldsymbol{\Sigma}^{-1}\left(\frac{\mathbf{G}_{\text{ref}}^{\text{T}}\mathbf{x}_{\text{ref}}}{\sigma_e^2}+\frac{\beta\mathbf{G}^{\text{T}}\mathbf{y}}{\sigma_\epsilon^2}\right),\boldsymbol{\Sigma}^{-1}\right)$ distribution where $\boldsymbol{\Sigma}=\frac{\mathbf{G}_{\text{ref}}^{\text{T}}\mathbf{G}_{\text{ref}}}{\sigma_e^2}+\frac{\beta^2\mathbf{G}^{\text{T}}\mathbf{G}}{\sigma_\epsilon^2}+\frac{\mathbf{D}_\lambda^{-1}}{\sigma_e^2\tau^2}$.

5: The conditional posterior distribution of λ_i is given by $f(\lambda_i|\cdot)\propto\frac{1}{\lambda_i(1+\lambda_i^2)}\ \exp\left[-\frac{\alpha_i^2}{2\sigma_e^2\tau^2\lambda_i^2}\right],\quad i=1(1)m$. We update each λ_i by the Metropolis-Hastings algorithm outlined in the supplemental material.

6: The full conditional distribution of τ is given by $f(\tau|\cdot)\propto\frac{1}{\tau^m(1+\tau^2)}\exp\left[-\frac{\boldsymbol{\alpha}^{\text{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2\tau^2}\right]$. We update τ by the Metropolis-Hastings algorithm discussed in the supplemental material.

7: Update σ_e^2 by drawing a sample from $IG\left(e_1+\frac{n_r}{2}+\frac{m}{2},\ e_2+\frac{\boldsymbol{\alpha}^{\text{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\tau^2}+\frac{1}{2}(\mathbf{x}_{\text{ref}}-\mathbf{G}_{\text{ref}}\boldsymbol{\alpha})^{\text{T}}(\mathbf{x}_{\text{ref}}-\mathbf{G}_{\text{ref}}\boldsymbol{\alpha})\right)$.

8: To update β first compute the probability for the point mass function $\delta_{\{0\}}(\beta)$ in the full conditional distribution of β as $w=\frac{1-q}{1-q+\frac{q\widetilde{\sigma}}{\sigma_d\sigma_\epsilon}\exp\left(\frac{\widetilde{\mu}^2}{2\widetilde{\sigma}^2}\right)}$, where $\widetilde{\mu}=\frac{\boldsymbol{\alpha}^{\text{T}}\mathbf{G}^{\text{T}}\mathbf{y}}{\boldsymbol{\alpha}^{\text{T}}\mathbf{G}^{\text{T}}\mathbf{G}\boldsymbol{\alpha}+\frac{1}{\sigma_d^2}}$ and $\widetilde{\sigma}^2=\frac{\sigma_\epsilon^2}{\boldsymbol{\alpha}^{\text{T}}\mathbf{G}^{\text{T}}\mathbf{G}\boldsymbol{\alpha}+\frac{1}{\sigma_d^2}}$. Generate $U\sim Uniform(0,1)$ distribution. **if** U<w, **then** assign $\beta=0$; **else** simulate β from $N(\widetilde{\mu},\widetilde{\sigma}^2)$ distribution.

9: The full conditional distribution of q is a mixture of two beta distributions. The probability corresponding to one such component $Beta(b_1,b_2+1)$ is $w=\frac{\delta_{\{0\}}(\beta)B(b_1,b_2+1)}{\delta_0(\beta)B(b_1,b_2+1)+\phi(\beta;0,\sigma_d^2\sigma_\epsilon^2)B(b_1+1,b_2)}$, where $\phi(x;\mu,\sigma^2)$ is the value of pdf of normal distribution at point x with mean μ, variance σ^2 and $B(\cdot,\cdot)$ denotes the beta function. We update q as follows. Simulate $U\sim Uniform(0,1)$ distribution. **if** U<w, **then** draw a sample from $Beta(b_1,b_2+1)$ distribution; **else** simulate from $Beta(b_1+1,b_2)$ distribution.

10: To update σ_d^2 first compute two quantities $w_1=\frac{(1-q)\delta_{\{0\}}(\beta)\Gamma(d_1)}{d_2^{d_1}}$ and $w_2=\frac{q\Gamma(d_1+\frac{1}{2})}{\sqrt{2\pi}\sigma_\epsilon\left(d_2+\frac{\beta^2}{2\sigma_\epsilon^2}\right)^{d_1+\frac{1}{2}}}$, where $\Gamma(\cdot)$ denotes the gamma function. Generate $U\sim Uniform(0,1)$ distribution. **if** U<$\frac{w_1}{w_1+w_2}$, **then** draw a sample from $IG(d_1,d_2)$ distribution; **else** draw from $IG\left(d_1+\frac{1}{2},\ d_2+\frac{\beta^2}{2\sigma_\epsilon^2}\right)$ distribution.

11: The full conditional distribution of σ_ϵ^2 is a mixture of two inverse gamma distributions and the weights corresponding to two components are $w_1=\frac{(1-q)\delta_{\{0\}}(\beta)\Gamma(\frac{n}{2}+\epsilon_1)}{\left(\epsilon_2+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\text{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)^{\frac{n}{2}+\epsilon_1}}$ and $w_2=\frac{q\Gamma(\frac{n+1}{2}+\epsilon_1)}{\sqrt{2\pi}\sigma_d\left(\epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\text{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)^{\frac{n+1}{2}+\epsilon_1}}$ respectively. To update σ_ϵ^2, first simulate $U\sim Uniform(0,1)$ distribution. **if** U<$\frac{w_1}{w_1+w_2}$, **then** simulate from $IG\left(\frac{n}{2}+\epsilon_1,\ \epsilon_2+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\text{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)$ distribution; **else** generate a random sample from $IG\left(\frac{n+1}{2}+\epsilon_1,\ \epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\text{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)$ distribution.

12: Repeat the steps (4) − (11) until all MCMC iterations are finished.

For multi-ancestry TWAS we follow the same notations used in the section 2. We regard the same notation $\mathbf{D}_\lambda=diag(\lambda_1^2,\lambda_2^2,\cdots,\lambda_m^2)$ as a $m\times m$ diagonal matrix. Then we have $\boldsymbol{\alpha}_k|\boldsymbol{\lambda},\tau,\sigma_{e_k}^2\sim N_m(\mathbf{0},\sigma_{e_k}^2\tau^2\mathbf{D}_\lambda)$. The steps for the algorithm are as follows.

Algorithm 2 Metropolis-within the Gibbs sampling algorithm for multi-ancestry TWAS

1: Start

2: Choose initial values for the parameters $\boldsymbol{\lambda}, \tau, \beta, q, \sigma_d^2, \sigma_\epsilon^2$ and $\sigma_{e_k}^2$ for each $k = 1(1)K$.

3: loop:

4: Update $\boldsymbol{\alpha}_k$ by drawing a $m \times 1$ random vector from $N_m\left(\boldsymbol{\Sigma}_k^{-1}\left(\frac{\mathbf{G}_{\mathrm{k,ref}}^{\mathrm{T}}\mathbf{x}_{\mathrm{k,ref}}}{\sigma_{e_k}^2} + \frac{\beta\mathbf{G}_k^{\mathrm{T}}\mathbf{y}_k}{\sigma_\epsilon^2}\right), \boldsymbol{\Sigma}_k^{-1}\right)$ distribution for each $k = 1(1)K$, where $\boldsymbol{\Sigma}_k = \frac{\mathbf{G}_{\mathrm{k,ref}}^{\mathrm{T}}\mathbf{G}_{\mathrm{k,ref}}}{\sigma_{e_k}^2} + \frac{\beta^2\mathbf{G}_k^{\mathrm{T}}\mathbf{G}_k}{\sigma_\epsilon^2} + \frac{\mathbf{D}_\lambda^{-1}}{\sigma_{e_k}^2\tau^2}$.

5: The full conditional distribution of λ_j is given as $f(\lambda_j|\cdot) \propto \frac{1}{\lambda_j^k(1+\lambda_j^2)}\ \exp\left[-\frac{1}{2\tau^2\lambda_j^2}\sum_{k=1}^K \frac{\alpha_{kj}^2}{\sigma_{e_k}^2}\right]$, $j = 1(1)m$. We update each λ_j by the Metropolis-Hastings algorithm discussed in the supplemental material.

6: The conditional posterior distribution of τ is given by $f(\tau|\cdot) \propto \frac{1}{\tau^{mK}(1+\tau^2)}\ \exp\left[-\frac{1}{2\tau^2}\sum_{k=1}^K \frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}_k}{\sigma_{e_k}^2}\right]$. We update τ by the Metropolis-Hastings algorithm outlined in the supplemental material.

7: Update $\sigma_{e_k}^2$ by simulating from $IG\left(e_1 + \frac{n_k}{2} + \frac{m}{2},\ e_2 + \frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}_k}{2\tau^2} + \frac{1}{2}(\mathbf{x}_{\mathrm{k,ref}} - \mathbf{G}_{\mathrm{k,ref}}\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{x}_{\mathrm{k,ref}} - \mathbf{G}_{\mathrm{k,ref}}\boldsymbol{\alpha}_k)\right)$ for each $k = 1(1)K$.

8: The posterior probability that β takes value 0 in the full conditional distribution is $w = \frac{1-q}{1-q+\frac{q\tilde{\sigma}}{\sigma_d\sigma_\epsilon}\exp\left(\frac{\tilde{\mu}^2}{2\tilde{\sigma}^2}\right)}$, where $\tilde{\mu} = \frac{\sum_{k=1}^K \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{G}_k^{\mathrm{T}}\mathbf{y}_k}{\sum_{k=1}^K \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{G}_k^{\mathrm{T}}\mathbf{G}_k\boldsymbol{\alpha}_k + \frac{1}{\sigma_d^2}}$ and $\tilde{\sigma}^2 = \frac{\sigma_\epsilon^2}{\sum_{k=1}^K \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{G}_k^{\mathrm{T}}\mathbf{G}_k\boldsymbol{\alpha}_k + \frac{1}{\sigma_d^2}}$. We update β as follows: first simulate $U \sim Uniform(0,1)$ distribution. **if** $U < w$, **then** assign $\beta = 0$; **else** draw a sample from $N(\tilde{\mu}, \tilde{\sigma}^2)$ distribution.

9: The full conditional distribution of q is a mixture of two beta distributions. The posterior probability corresponding to one such component $Beta(b_1, b_2+1)$ is $w = \frac{\delta_{\{0\}}(\beta)B(b_1,b_2+1)}{\delta_0(\beta)B(b_1,b_2+1)+\phi(\beta;0,\sigma_d^2\sigma_\epsilon^2)B(b_1+1,b_2)}$, where $\phi(x;\mu,\sigma^2)$ is the value of pdf of normal distribution with mean μ, variance σ^2 at point x and $B(\cdot,\cdot)$ denotes the beta function. We update q as follows. Simulate $U \sim Uniform(0,1)$ distribution. **if** U<w, **then** draw a sample from $Beta(b_1, b_2+1)$ distribution; **else** simulate from $Beta(b_1+1, b_2)$ distribution.

10: To update σ_d^2, first compute two quantities $w_1 = \frac{(1-q)\delta_{\{0\}}(\beta)\Gamma(d_1)}{d_2^{d_1}}$ and $w_2 = \frac{q\Gamma(d_1+\frac{1}{2})}{\sqrt{2\pi}\sigma_\epsilon\left(d_2+\frac{\beta^2}{2\sigma_\epsilon^2}\right)^{d_1+\frac{1}{2}}}$, where $\Gamma(\cdot)$ denotes the gamma function. Generate $U \sim Uniform(0,1)$ distribution. **if** U$<\frac{w_1}{w_1+w_2}$, **then** draw a sample from $IG(d_1, d_2)$ distribution; **else** draw from $IG\left(d_1 + \frac{1}{2},\ d_2 + \frac{\beta^2}{2\sigma_\epsilon^2}\right)$ distribution.

11: The full conditional distribution of σ_ϵ^2 is a mixture of two inverse gamma distributions and the weights corresponding to two components are $w_1 = \frac{(1-q)\delta_{\{0\}}(\beta)\Gamma(\frac{N}{2}+\epsilon_1)}{\left(\epsilon_2+\frac{1}{2}\sum_{k=1}^K(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)^{\frac{N}{2}+\epsilon_1}}$ and $w_2 = \frac{q\ \Gamma(\frac{N+1}{2}+\epsilon_1)}{\sqrt{2\pi}\sigma_d\left(\epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}\sum_{k=1}^K(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)^{\frac{N+1}{2}+\epsilon_1}}$ respectively. First, draw a random sample from $U \sim Uniform(0,1)$ distribution. We update σ_ϵ^2 as follows. **if** U$<\frac{w_1}{w_1+w_2}$, **then** simulate from $IG\left(\frac{N}{2}+\epsilon_1,\ \epsilon_2+\frac{1}{2}\sum_{k=1}^K(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)$ distribution; **else** generate a random sample from $IG\left(\frac{N+1}{2}+\epsilon_1,\ \epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}\sum_{k=1}^K(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)$ distribution.

12: Repeat the steps (4) − (11) until all MCMC iterations are finished.

Simulation study

We assess the performance of our method through in-depth simulation studies and compare it with selected existing methods. For the single-ancestry TWAS, we compare the performance of our method, bayesTWAS, with the well-established PrediXcan method (Gamazon et al. [18]). For multi-ancestry TWAS, we consider the METRO method for comparison. Next, we describe the procedure for simulating the data for the single-ancestry TWAS; the extension to the multi-ancestry scenario immediately follows.

Simulation setup

For the simulation of genotype data, we consider the genetic region neighboring the DGCR8 gene located on chromosome 22. We choose the Japanese population from the 1000 Genome Project which serves as a template population for simulating the genotype data. Surrounding DGCR8, we retain only those SNPs in the analysis that fall within the 500 kilobases of the upstream and downstream regions of the gene. We randomly select 200 SNPs from this set of local SNPs with minor allele frequencies (MAF) more than 0.05. To simulate genotype data for the individuals in the reference transcriptome data, We implement the HAPGEN software [19], which mimics the real LD pattern of the local SNPs in the Japanese population. HAPGEN is an established method based on meticulous theory for simulating genotype data that resembles a chosen real-world population. We run HAPGEN twice to simulate genotype data for the individuals in the transcriptome and GWAS data.

For each individual in the transcriptome data, we generate gene expression using the linear model $x_{i,\text{ref}} = \mathbf{g}_{i,\text{ref}}^{'}\boldsymbol{\alpha} + e_i, \quad i = 1(1)n_r$. $x_{i,\text{ref}}$ is the gene expression level, and $\mathbf{g}_{i,\text{ref}}^{'}$ is the vector of genotype values at the local SNPs for i^{th} individual in the transcriptome data. Here, $\boldsymbol{\alpha}$ represents the effect size vector for local SNPs, and e_i is the random residual. Under the assumption that $\text{Var}(x_{i,\text{ref}}) = 1$, we simulate e_i from $N(0, 1 - h_e^2)$; h_e^2 denotes the heritability of gene expression due to the local SNPs. We assume that 10% of local SNPs affect the gene expression. Let m_c stand for the number of non-null SNPs affecting the gene expression. The non-zero components of $\boldsymbol{\alpha}$ are generated from $N(0, \frac{h_e^2}{m_c})$. We assign zero value to the remaining components of $\boldsymbol{\alpha}$ that do not influence the gene expression.

Next, we discuss the steps to simulate the complex trait for the individuals in GWAS. The GWAS individuals do not overlap with those in the reference transcriptome data. HAPGEN generates the genotype data for the GWAS individuals for the same set of local SNPs considered. Let $\mathbf{g}_l$ denote the genotype vector for the lth individual in the GWAS panel, $l = 1(1)n$. As the effect of local SNPs on gene expression remains the same in both datasets, the true GReX for the lth individual in the GWAS panel is $\mathbf{g}_l^{'}\boldsymbol{\alpha}$. We simulate the outcome for the GWAS individuals following the linear model: $y_l = \beta(\mathbf{g}_l^{'}\boldsymbol{\alpha}) + \epsilon_l$. y_l is the value of the trait for l^{th} individual in GWAS panel and β denotes the effect of GReX on the trait. We simulate the random error term ϵ_l from $N(0, 1)$ distribution. One gene is anticipated to explain a negligible proportion of the trait's variability. Hence, we assume $\text{Var}(y_l) = \text{Var}(\epsilon_l) = 1$. We choose different values of β to create the null and various non-null simulation scenarios. We standardize gene expression, outcome vector, and the columns of genotype matrices in both reference transcriptome and GWAS panels while applying the methods.

For multi-ancestry TWAS, we follow the simulation setup described in the Metro method. We consider two genetic ancestries: East Asian Ancestry (EAS) and European Ancestry (EUR). We select the Japanese and British populations in the 1000 genome data, which serve as the representative populations for EAS and EUR ancestries, respectively. We employ HAPGEN to simulate genotype data for the 200 local SNPs (MAF > 0.05) for the DGCR8 gene as considered for the single-anestry scenario separately for each ancestry. We apply the steps described above for single-ancestry to simulate gene expression in separate transcriptome data for the two ancestries.

We regard the following linear model: $x_{lk,\text{ref}} = \mathbf{g}_{lk,\text{ref}}^{'}\boldsymbol{\alpha}_k + e_{lk}, \quad l = 1(1)n_k, \; k = 1(1)2$. The notations $x_{lk,\text{ref}}$ and $\mathbf{g}_{lk,\text{ref}}^{'}$ denote the gene expression level and genotype vector for l^{th} individual in the transcriptome data for k^{th} ancestry. $\boldsymbol{\alpha}_k$ is the vector of local SNPs' effects on expression, and e_{lk} is the random noise specific to k^{th} ancestry. We assume $\text{var}(x_{lk,\text{ref}}) = 1$ and simulate e_{lk} from $N(0, 1 - h_e^2)$ where h_e^2 is the gene-expression heritability in k^{th} ancestry. We select 10% of the local SNPs, m_c causal SNPs, to have a non-null effect on gene expression in each ancestry. The elements of $\boldsymbol{\alpha}_k$ that correspond to the causal SNPs are simulated from $N(0, \frac{h_e^2}{m_c})$. We consider a fixed proportion, f_0, of causal SNPs overlap between the two ancestries to induce eQTL sharing. We simulate the complex trait for GWAS individuals in k^{th} ancestry employing the linear model: $y_{lk} = \beta(\mathbf{g}_{lk}^{'}\boldsymbol{\alpha}_k) + \epsilon_{lk}, \quad l = 1(1)N_k$. Here, N_k denotes the GWAS sample size for k^{th} ancestry. The term $\mathbf{g}_{lk}^{'}\boldsymbol{\alpha}_k$ stands for the true GReX for l^{th} individual in k^{th} ancestry and β is the common effect of the GReX on the outcome trait. We simulate ϵ_{lk} from $N(0, 1)$ since a single gene explains a negligible proportion of the trait's variability. We vary β to create different simulation scenarios of null and non-null effects of the gene.

In the single ancestry TWAS framework, we consider reference panel sample size $n_r = 500$ and GWAS sample size $n = 5000$. We consider five different values of $\beta = 0.1, 0.2, 0.3, 0.4, 0.5$. We fix the gene-expression

heritability at 20%. In the multi-ancestry TWAS setup, we choose reference panel sample sizes $n_1 = n_2 = 500$ and GWAS sample sizes $N_1 = N_2 = 5000$ respectively, for EAS and EUR ancestries. We consider three different choices of $f_0 = 0\%, 50\%, 100\%$. We fix h_e^2 at 20% in each ancestry so that the overall heritability is controlled at the same 20% level. We evaluate the result for bayesTWAS based on 7000 MCMC iterations with a burn-in sample size of 2000. In each scenario, we compare our method with the PrediXcan and METRO methods based on an effective number of replications. Specifically, we compute the p-value for testing H_0 for PrediXcan and METRO methods in each replication. We consider those replications for which the p-value falls below the Bonferroni-corrected threshold, i.e., p-value $< \frac{0.05}{R}$, where R is a pre-fixed number of replications. In our simulation setting, we consider 50 effective replications.

We also compare the performance of our Bayesian method with the PrediXcan and METRO methods in identifying non-null genes affecting the trait from a given set of null and non-null genes. Following the same procedure described in Coram et al. [20] we measure the selection accuracy between methods by computing two quantities false positive rates (FPR) and true positive rates (TPR). The FPR is defined as the proportion of null genes declared significant by any method, whereas TPR indicates the proportion of non-null genes showing an association for the trait. In this setting, we will consider 200 genes with 5% genes having a non-null effect on the trait. The effect sizes for non-null genes are simulated by considering a fixed value of trait-level heritability (h_g^2). The term h_g^2 denotes the proportion of the trait variability that the significant genes can explain. We conduct a test for each of the 200 genes to check whether the gene has a non-null effect on the trait. We arrange the p-values obtained from the PrediXcan and METRO methods for each test in increasing order of magnitude. The corresponding posterior probability of null association for each test for our Bayesian method is also arranged in increasing order. Then, for varying choices of statistically significant thresholds, we compute FPR and TPR values for each method and plot them to obtain the receiver operating characteristic (ROC) curve. The method with the higher area under the curve (AUC) is preferred over the method with a lower AUC.

Result

We first discuss the overall summary of the results obtained by our approach, bayesTWAS, while estimating the effect size of GReX on the outcome, β, in a single-ancestry TWAS. We find that the posterior mean and median of β based on the posterior sample of β provide similar estimates of β. For example, when $\beta = 0.2$, the posterior mean and median are both 0.18 with posterior variance 0.004. In this case the 95% central posterior interval (CPI) as $(0.09, 0.29)$ (Table 1). The CPI constructed for our Bayesian method has the shortest length among all possible 95% credible intervals. The method performs robustly for both positive and negative values of β. For $\beta = -0.2$, the posterior mean and median are -0.19 and -0.18 with the CPI as $(-0.29, -0.1)$. The width of the CPIs and the posterior variances increase marginally as the absolute value of β increases (Table 1). The empirical average of log_{10}(BF) values across the simulation iterations increases steadily with the absolute value of β. For example, starting with a negative value of log_{10}(BF) as -1.37 for $\beta = 0$, we obtain the values as $0.03, 5.13, 14.69, 29.26$ for $\beta = 0.1, 0.2, 0.3, 0.4$, respectively (Table 1).

Next, we compare bayesTWAS with PrediXcan in terms of the estimation accuracy of the GReX effect size for single-ancestry TWAS. We regard the estimates' relative bias and mean squared errors (MSE) for comparison. The relative bias in the estimation of β is defined as $\frac{\hat{\beta}-\beta}{\beta}$, where $\hat{\beta}$ is an estimate of β. bayesTWAS produced a lower upward bias than PrediXcan (Figure 1A). In general, the relative bias by bayesTWAS and PrediXcan diminishes fast as true β increases (Figure 1A). Regarding the MSE, bayesTWAS produced lower values than PrediXcan. The MSE of bayesTWAS is substantially smaller for higher βs. Contrary to relative bias, the MSE estimates increase with true β. However, the MSE of PrediXcan increases much faster than bayesTWAS (Figure 1B); the latter yielded a negligible increase. Generally, both methods improve the estimation accuracy for higher values of β, e.g., $\beta = 0.4, 0.5$. Although the MSE of PrediXcan is noticeably inflated for these choices. For the negative choices of βs, we observe a similar comparative performance of the two methods. Overall, bayesTWAS estimates the effect size of the GReX on the outcome more precisely.

Next, we varied the sample sizes of the reference transcriptome (n_r) and GWAS (n) data. Here, we chose a fixed $\beta = 0.2$. First, we varied n_r, keeping $n = 5000$ fixed. Regarding bias, we find that bayesTWAS

Table 1: Posterior summary of GReX effect size for single-ancestry model

β	mean	variance	median	95% CPI	log_{10}(BF)
0.5	0.49	0.0066	0.49	$(0.37, 0.63)$	46.36
0.4	0.4	0.0048	0.39	$(0.29, 0.53)$	29.26
0.3	0.3	0.0038	0.29	$(0.2, 0.42)$	14.69
0.2	0.18	0.0039	0.18	$(0.09, 0.29)$	5.13
0.1	0.04	0.0022	0.03	$(0.07, 0.09)$	0.03
0	-0.0002	0.0002	0	$(-0.002, 0.001)$	-1.37
-0.1	-0.04	0.0018	-0.03	$(-0.1, -0.07)$	0.08
-0.2	-0.19	0.0034	-0.18	$(-0.29, -0.1)$	5.35
-0.3	-0.3	0.0042	-0.29	$(-0.41, -0.2)$	15.16
-0.4	-0.39	0.0055	-0.38	$(-0.52, -0.27)$	28.5
-0.5	-0.49	0.0056	-0.48	$(-0.63, -0.37)$	47.22

Note: The notations CPI and BF stand for the central posterior interval and Bayes factor, respectively. The reference and GWAS panel sample sizes are $n_r = 500$ and $n = 5000$. The SNP-level heritability (h_e^2) is fixed at 20%. The results are obtained based on the 500 replication of data.

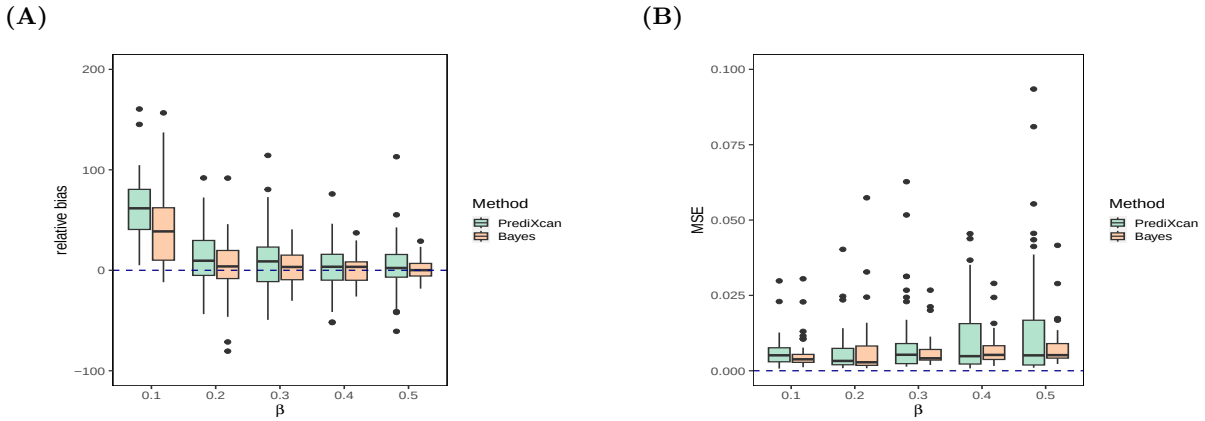

Figure 1: Comparison between PrediXcan and Bayesian method for (**A**) relative bias and (**B**) MSE for varying effect sizes of β. The reference panel sample size is $n_r = 500$ and the GWAS sample size is $n = 5000$. The heritability of the gene expression is fixed at $h_e^2 = 20\%$.

consistently produced lower relative biases than PrediXcan for all choices of n_r ranging between 500 to 1500 (Figure 2A). PrediXcan yielded substantial upward bias. However, bayesTWAS produced a mean bias of nearly zero for some choices of n_r, such as 500, 900, and 1500 (Figure 2A). Similarly, bayesTWAS steadily produced smaller MSEs than PrediXcan for different sample sizes. We then varied the GWAS sample size n keeping n_r fixed. We again find that bayesTWAS estimates β more accurately with lower relative bias and MSE than PrediXcan (Figure 2C and 2D). However, we did not observe a conclusive decrease in the bias and MSE as the sample sizes increased. It is possible that larger sample sizes could reflect this, which is computationally intensive.

Next, we compare the methods regarding the accuracy of correctly classifying the null and non-null genes. For a single ancestry TWAS, the ROC of bayesTWAS dominates the ROC of PrediXcan (Figure 3), i.e., for a given level of FPR, bayesTWAS provides a better TPR. Consequently, bayesTWAS yielded a consistently higher AUC in the range of $1 - 10\%$ than PrediXcan (Figure 3). As we increase the GWAS sample size (n) from 5000 to 10000 or increase the trait heritability (h_g^2) from 10% to 20% for a given choice of expression heritability (h_e^2), the AUC increases for both methods. We observe that a change in expression heritability affects the performance of PrediXcan more than a change in trait heritability. For example, when $h_e^2 = 20\%$ and $h_g^2 = 10\%$, the AUC for PrediXcan is 0.65 (Figure 3A). As we decrease h_e^2 to 10% and increase h_g^2 to

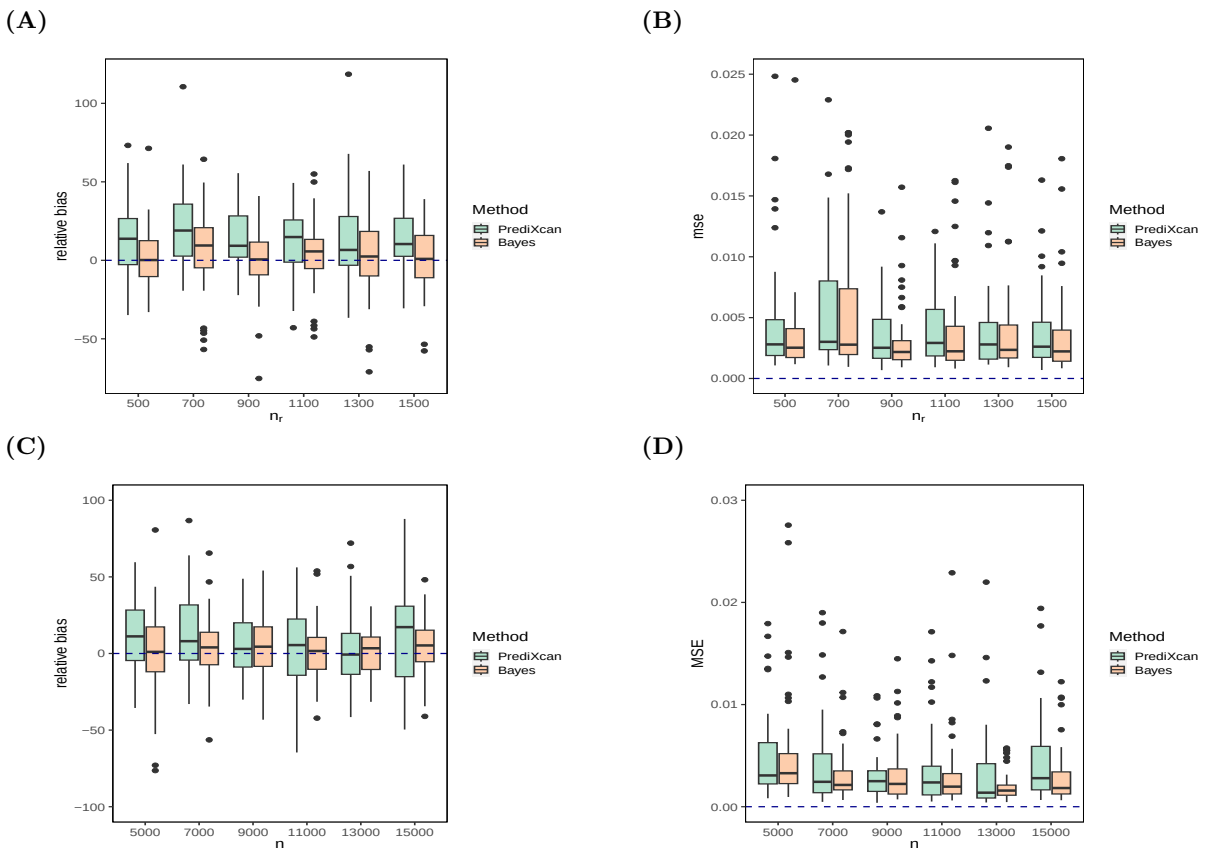

Figure 2: Comparison between PrediXcan and Bayesian method for relative bias and MSE for varying reference panel sample sizes (Figure **A**-**B**) and varying GWAS sample sizes (Figure **C**-**D**). The GWAS sample size is $n = 5000$ in Figure **A**-**B** and the reference panel sample size is $n_r = 500$ in (Figure **C**-**D**). The GReX effect size is 0.2 and gene-expression heritability is fixed at $h_e^2 = 20\%$.

20%, the AUC decreases from 0.65 to 0.59 (Figure 3B) for PrediXcan. On the other hand, the AUCs for bayesTWAS are 0.66 (Figure 3A) and 0.65 (Figure 3B) in these two scenarios, respectively. The improved performance of bayesTWAS to PrediXcan is due to the joint modeling of the GWAS and transcriptome data under the unified Bayesian framework based on the horseshoe and spike and slab prior contrary to PrediXcan which does not adjust for the uncertainty of imputed expression in the two-step approach.

For multi-ancestry TWAS, bayesTWAS estimates the effect sizes of GReX on the outcome with lower variability than single-ancestry TWAS. For example, when $\beta = 0.1$, single-ancestry TWAS delivers an estimate of 0.04 with a posterior variance of 0.0022; whereas the variance is 0.0014 for multi-ancestry TWAS (Table 2). Along with the accurate estimates, the width of the credible intervals is also smaller for varying effect sizes of β in the multi-ancestry TWAS setup due to having lower variances. In this setup, incorporating complementary information collected from different ancestries results in better estimation accuracy than a single-ancestry procedure.

For the multi-ancestry TWAS framework, our method outperforms the METRO method in terms of having lesser relative bias and MSE across all simulation scenarios. Each ancestry's reference panel and GWAS panel sample sizes are fixed for all these simulation settings at $n_1 = n_2 = 500$ and $N_1 = N_2 = 5000$ respectively. We fix the gene level heritability (h_e^2) at 20%. We observe that relative bias for our method is higher a bit for $\beta = 0.1$ and decreases rapidly as we increase β from 0.1 to 0.5 (Figure 4). Our method produces significantly lower relative biases for different choices of β across different settings than the METRO method. We also note that our method produces similar or lower MSE than the METRO method for various choices of β. The variation between the eQTLs underlying gene expressions across ancestries does not impact our Bayesian method in estimating the GReX effect sizes. Across different simulation settings, our method

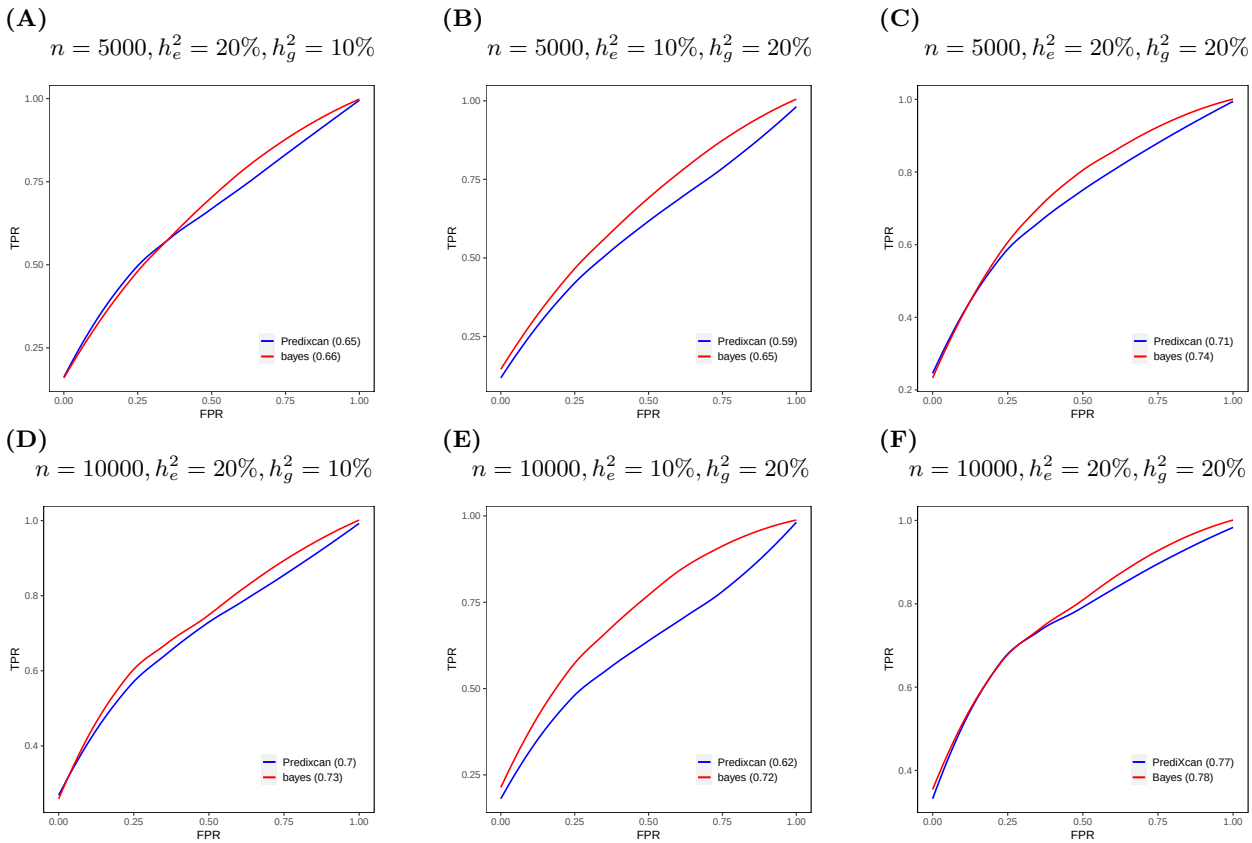

Figure 3: Partial receiver operating characteristic (ROC) curves for PrediXcan and Bayesian method. The ROC curves are obtained for reference panel sample size $n_r = 500$ and varying choices of GWAS sample size (n), gene-expression heritability (h_e^2), and trait-level heritability (h_g^2). The TPR and FPR for each method are calculated based on 50 replications. The numbers in the first bracket in the legend section of the plot denote the area under the curve (AUC) for each method.

maintains good estimation accuracy.

In the case of the multi-ancestry TWAS framework, our method gives higher AUC compared to the METRO method (Figure 5) across different simulation scenarios. For example, from Figure 5A, we observe that the AUC for our method is 0.74, while for the METRO method, it is 0.67. The AUC increases for each method as we increase the gene-level heritability (h_e^2) or trait-level heritability (h_g^2) while keeping the other parameters fixed. Given a fixed choice of gene and trait-level heritability, the METRO method's AUC decreases as we change f_0 from 50% to 0%. For example given $h_e^2 = 10\%$ and $h_g^2 = 20\%$, METRO method's AUC decreases from 0.7 (Figure 5E) to 0.65 (Figure 5B) when we change f_0 from 50% to 0%. METRO loses its efficiency in identifying the true non-null set of genes When the eQTLs do not overlap between the ancestries. On the other hand, the AUC for our method remains almost the same for fixed values of h_e^2 and h_g^2 across different choices of f_0. The sharing pattern of eQTLs across genetic ancestries does not impact our Bayesian method.

Real Data Application

We apply the methods to conduct a single-ancestry and multi-ancestry TWAS for body mass index (BMI), a commonly used measure of obesity which is a crucial risk factor for many diseases. The main goal here is to contrast the performance of the methods for real-life data. We obtained the transcriptomic data from the GEUVADIS expression study [21] and the GWAS data from the UK Biobank (UKBB). BMI was measured in the units of Kg/m^2 in the UK Biobank. RNA sequencing was performed for lymphoblastoid cell lines in the

Table 2: Posterior summary of GReX effect size for multi-ancestry model

β	mean	variance	median	95% CPI	log(BF)
0.5	0.49	0.0022	0.49	$(0.4, 0.58)$	97.57
0.4	0.4	0.0017	0.39	$(0.32, 0.48)$	60.77
0.3	0.3	0.0014	0.3	$(0.23, 0.37)$	32.88
0.2	0.19	0.0012	0.19	$(0.14, 0.26)$	12.38
0.1	0.07	0.0014	0.06	$(0.02, 0.12)$	1.4
0	-8.57×10^{-5}	2.61×10^{-5}	-7.98×10^{-5}	$(-0.0007, 0.0003)$	-1.55
-0.1	-0.07	0.0013	-0.06	$(-0.12, -0.02)$	1.5
-0.2	-0.2	0.0011	-0.2	$(-0.27, -0.14)$	12.79
-0.3	-0.29	0.0013	-0.29	$(-0.37, -0.23)$	32.24
-0.4	-0.39	0.0017	-0.39	$(-0.48, -0.32)$	62.37
-0.5	-0.49	0.0023	-0.49	$(-0.58, -0.41)$	97.34

Note: The abbreviations CPI and BF denote the central posterior interval and Bayes factor respectively. The reference panel sample sizes for two ancestries are taken as $n_1 = n_2 = 500$. The corresponding GWAS sample sizes are $N_1 = N_2 = 5000$. The results are obtained based on 500 replications.

GEUVADIS data comprising 462 individuals from the 1000 Genomes Project. The mRNA expression levels are measured for these individuals who belonged to five different populations: CEPH (CEU), Finns (FIN), British (GBR), Toscani (TSI), and Yoruba (YRI). We combine the individuals from CEU, FIN, GBR, and TSI to form a broader European (EUR) ancestry with a sample size of 373. The remaining 89 individuals in the YRI population constitute an African (AFR) ancestry. We considered the EUR ancestry to perform a single-ancestry TWAS and the EUR and AFR ancestries to conduct a multi-ancestry TWAS for BMI.

We remove those genes from the analysis for which the gene expression measurements are missing in at least 50% of the individuals. Among the genes retained, we implemented the KNNimpute algorithm [22] to impute the missing expression values for a gene. We annotated the protein-coding genes from GENCODE [23] and held on to a final set of $12,517$ genes, common with the GEUVADIS expression data. Using multiple linear regression, the expression levels of a gene were adjusted for covariates such as age, sex, and top 10 genetic principal components (PCs) separately within each ancestry EUR and AFR. The genetic PCs were included to adjust for the effect of population stratification on the expression. We then applied the rank-based inverse normal transformation [24] on the adjusted expression so that the transformed residuals satisfy the normality assumption in the transcriptome data. We follow the standard quality control (QC) procedure to clean and retain good-quality SNP genotype data for analysis. We implemented the following steps for the genotype data of EUR and AFR ancestries separately in the transcriptome data. We filtered out SNPs with MAF < 0.05. We tested for the Hardy-Weinberg equilibrium (HWE) and removed the SNPs with p-value $< 5 \times 10^{-8}$. We also remove SNPs with a genotype missingness rate of more than 10%.

We obtained the GWAS genotype data and the outcome data for BMI in the UKBB. We regarded two ancestries, White British (WB) and African-British (AB), to constitute GWAS data for EUR and AFR ancestries. We remove a portion of GWAS individuals in either ancestry who do not meet certain criteria. Specifically, we removed those individuals whose self-reported sex does not match the genetically determined sex, who were not included in the computation of genetic PCs available in the UKBB data, and who have a genetic kinship to any other study participants. After these filtering, we had $287,801$ individuals in the WB ancestry and $6,426$ in AB ancestry. Using multiple linear regression, the complex trait BMI was adjusted for the following covariates: age, sex, and top 10 genetic PCs separately for WB and AB ancestries. We then implemented the inverse normal transformation to the adjusted BMI to induce normality. We employed a similar QC procedure outlined above for the GWAS genotype data. We filtered out SNPs based on MAF < 0.05, HWE test p-value $< 5 \times 10^{-8}$, and retained a final set of $241,977$ SNPs that overlapped with the QCed SNPs in the GEUVADIS data. We restricted our analysis to the SNPs on the autosomal chromosomes only. We standardized the complex trait, gene expression, and the columns of genotype matrices both in transcriptome and GWAS. For the gene under consideration, we considered those local SNPs falling within 500 kilobases of the gene's upstream and downstream regions beginning from the transcription start and end sites.

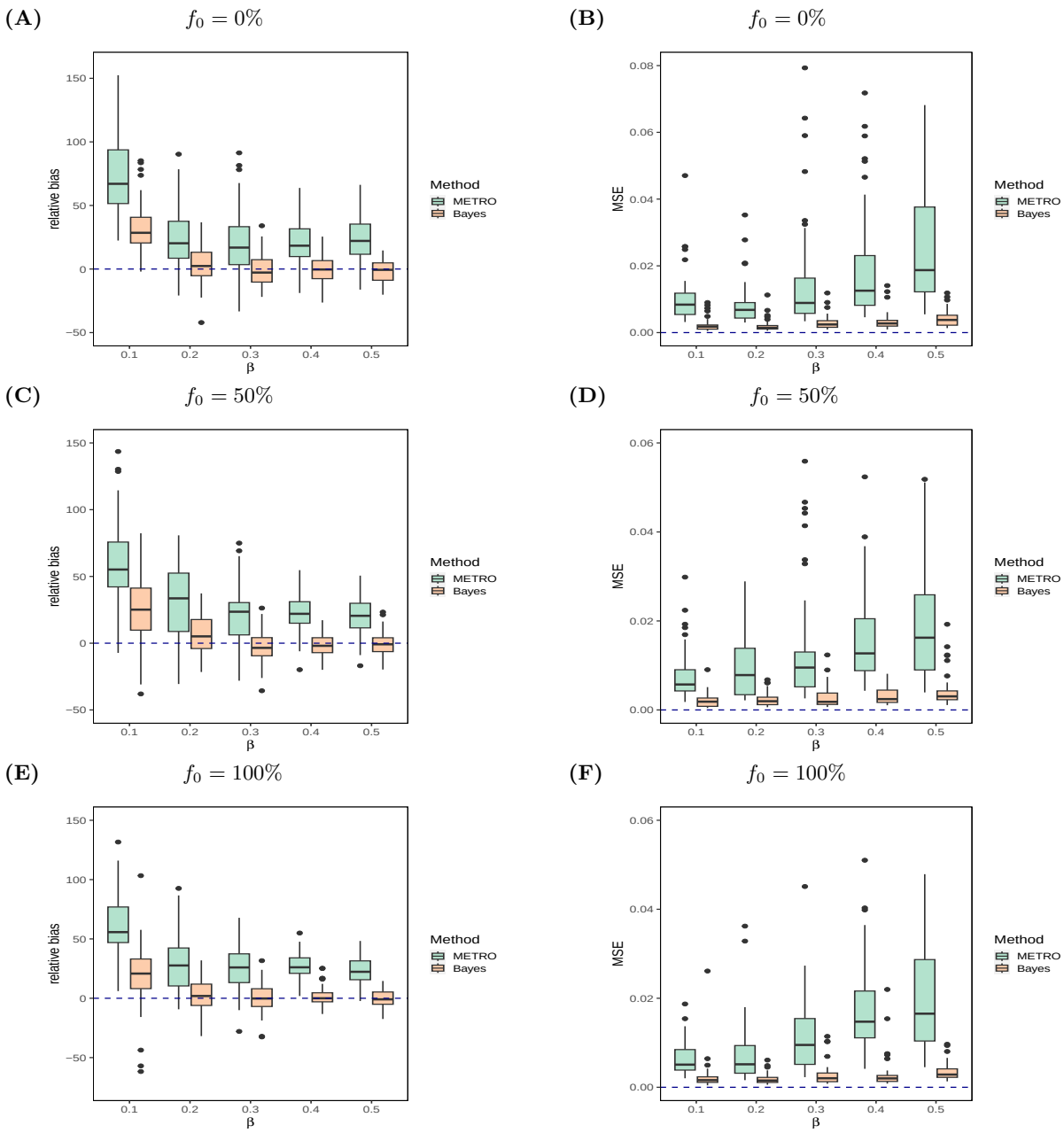

Figure 4: Comparison between METRO and Bayesian method for relative bias and MSE for different choices of $f_0 = 0\%$(Figure **A**-**B**), 50%(Figure **C**-**D**), 100%(Figure **F**-**G**). The reference panel sample sizes are $n_1 = n_2 = 500$ and the GWAS sample sizes are $N_1 = N_2 = 5000$ respectively. The gene-expression heritability is fixed at $h_e^2 = 20\%$.

We administered the GCTA method [25] to identify the locally heritable genes, i.e., the genes, the variance of the expression of which is explained significantly by the local SNPs. For each gene, we compute the p-value for the null hypothesis indicating the gene is not locally heritable. For the EUR ancestry, we implemented the Benjamini-Hochberg (BH) false discovery rate (FDR) controlling procedure [26] based on the GCTA p-values. Based on an FDR of 5%, we detected 1491 heritable genes for the EUR ancestry. For the multi-ancestry TWAS, we found 369 genes heritable in both EUR and AFR ancestries based on a nominal level of p-value < 0.05. The reason behind identifying such a small number of genes is mainly the low sample

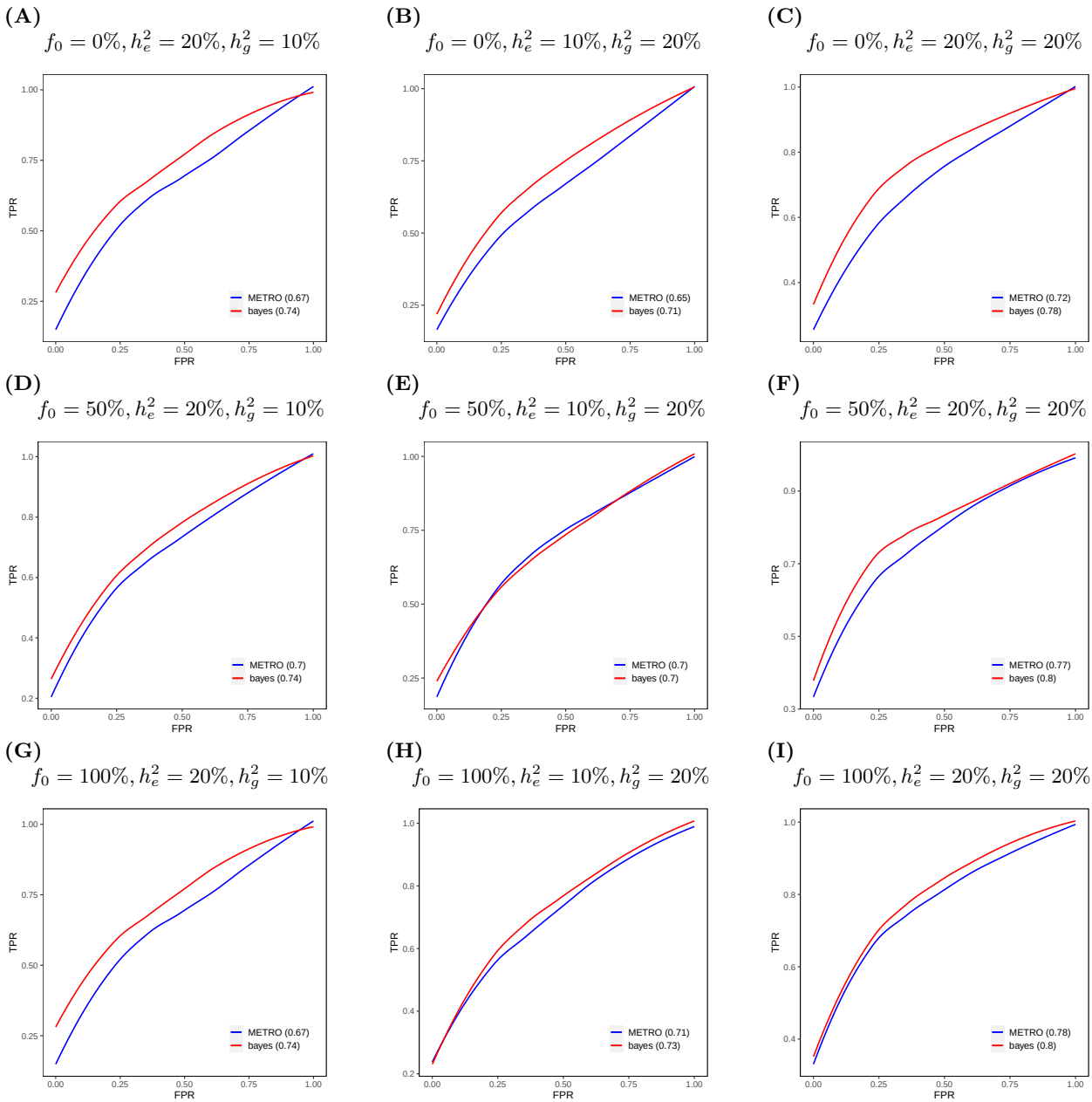

Figure 5: Partial receiver operating characteristic (ROC) curves for METRO and our Bayesian method. The ROC curves are obtained for reference panel sample sizes $n_1 = n_2 = 500$ and GWAS sample sizes $N_1 = N_2 = 5000$, and for varying choices of f_0, gene-expression heritability (h_e^2), and trait-level heritability (h_g^2). The f_0 indicates the proportion of overlapped eQTLs between two ancestries and we consider three specific choices for f_0 as $f_0 = 0\%, 50\%, 100\%$. The TPR and FPR for each method are calculated based on 50 replications. The numbers in the first bracket in the legend section of the plot denote the area under the curve (AUC) for each method.

size (89) of the transcriptome data in the AFR ancestry. At a nominal level (p-value < 0.05), we found 932 heritable genes in AFR ancestry, whereas only 50 genes while controlling for the FDR < 0.05. Hence, to analyze more genes in the multi-ancestry TWAS, we considered a list of genes that were heritable in both ancestries based on a nominal level of GCTA p-values. We restricted our downstream analysis only to these locally heritable genes.

We apply the Bonferroni correction for the PrediXcan and METRO methods to control the overall Type-I error rate at 5% level. A similar adjustment has been made for our Bayesian models while performing the association tests for multiple genes. Following the procedure suggested by Westfall et al. [27] we revise the hyperparameters in prior q, the prior probability of the gene to be non-null, in an appropriate way such that Bayesian correction is a good approximation of Bonferroni correction. We mention a brief outline of the procedure. Suppose we have r hypotheses and for $i = 1(1)r$; i^{th} hypothesis H_{0i} indicating the gene has a null effect on the trait. Let, π_0 be the probability that all the null hypotheses are true. That is, $\pi_0 = P[H_{0i}$ is true for all $i]$. Then we revise the marginal probabilities $P[H_{0i}$ is true$]$ by $\pi_0^{1/r}$ to perform multiple testing corrections in Bayesian setups. We take π_0 to be 0.1 in our analysis. We report the strength of the association test for our Bayesian model by computing the Bayes Factor (BF). We declare a gene to be significant for our Bayesian models if the $log_{10}(\text{BF}) > 5$. Although Kass and Raftery [28] point out that a cut-off 2 for $log_{10}(BF)$ will indicate decisive evidence against H_{0i}, the more stringent choice of threshold 5 is generally preferred in genetics [29] to control the number of false signals.

Out of 1491 heritable genes in EUR ancestry, PrediXcan constructs the GReX prediction model for 1467 genes. For the remaining 24 genes PrediXcan fails to construct GReX even though the genes are heritable. Among the 1467 genes, PrediXcan finds 40 genes to be significantly associated with BMI after Bonferroni correction. On the other hand, our Bayesian model performs the analysis for all 1491 genes. As our model does not require the imputation of GReX, it analyzes all heritable genes. Our model identifies 111 significant genes for BMI after Bayes-Bonferroni correction. There are 32 genes common for both methods. In addition, our method identifies 79 additional genes. PrediXcan identified the UBA7 gene (located on chromosome 3) to have a significant association (p-value 5×10^{-33}) with BMI with GReX effect size estimate 0.21 and the corresponding variance estimate 0.0003. For our Bayesian method, the RBM6 gene (Table 3) on chromosome 3 is the most significant as it has the highest BF ($log_{10}(\text{BF}) = 38.4$) and the lowest posterior probability of null association (PPNA) value of 2.3×10^{-35}. The corresponding GReX effect size and variance estimates are 0.09 and 0.0004 respectively. Our method also detects novel risk genes for BMI that were not earlier reported in the EBI GWAS catalog. For example, the gene CBX1 on chromosome 17 was previously reported to be associated with body fat rate by a study conducted by Chen et al. [30]. PrediXcan fails to identify this gene at the Bonferroni-corrected threshold (p-value 0.04). However, our method shows a significant association signal ($log_{10}(\text{BF})$ is 11.7 and PPNA is 1×10^{-8}) for this gene with corresponding estimates of GReX effect and variance as 0.28 and 0.048 respectively.

In the multi-ancestry TWAS framework, out of 369 heritable genes, 64 genes cross the Bonferroni-corrected threshold for the METRO method. On the other hand, our Bayesian method finds 19 genes to be significantly associated with BMI in both EUR and AFR ancestry. Although METRO identifies more associated genes than our method, we need to critically examine the nature of association for the METRO method. METRO was developed to find the risk genes associated with the trait in at least one ancestry. In contrast, the gene-trait association for our Bayesian method holds for all ancestries. Specifically, among 64 risk genes identified by the METRO, the GReX effect size estimate is 0 for 28 genes in EUR ancestry and for another 9 genes in AFR ancestry. Hence, METRO finds at most 27 genes that have a significant association in both EUR and AFR ancestries. Both the METRO and our Bayesian method identify 8 common genes. METRO finds the TAOK2 gene (located on chromosome 16) to be the most significant gene (p-value 5.6×10^{-24}) with GReX effect size estimate -0.16. Our method also reports the same gene to have the strongest association with BMI with a value of $log_{10}(\text{BF})$ 23.9 and the PPNA of 7.5×10^{-21} (Table 4). The corresponding estimates for the GReX effect and variance are -0.17 and 0.0052 respectively. Our method also identifies novel genes in this multi-ancestry TWAS setting that were not reported in the EBI GWAS catalog. For example, Kichaev et al. [31] found the MOCS1 gene in chromosome 6 associated with body height. While METRO finds an association for this gene with BMI only in EUR ancestry, our method confirms the association ($log_{10}(\text{BF})$ 20.2 and PPNA 3.7×10^{-17}) for both EUR and AFR ancestries. The GReX effect and variance estimates for the MOCS1 gene (Table 4) are -0.11 and 0.001 respectively.

Conclusion

In this article, we proposed a novel approach to conduct TWAS that circumvents the problem of adjusting the uncertainty of the predicted expression in the gene-trait association. Standard TWAS methods ignore the

Table 3: Top risk genes for BMI identified by our method in EUR ancestry

Chr	Gene	Estimate	Variance	$P(H_0\|\text{data})$	$log_{10}(BF)$
3	RBM6	0.09	0.0004	2.3×10^{-35}	38.4
3	UBA7	0.17	0.0058	2.6×10^{-31}	34.3
4	MANBA	0.55	0.1258	7×10^{-31}	33.9
4	GUF1	−0.67	0.2462	8.5×10^{-26}	28.8
12	NCKAP5L	−0.13	0.0027	5.1×10^{-22}	25
3	KLHDC8B	0.5	0.0634	9.2×10^{-22}	24.8
11	THYN1	0.53	0.1726	2.3×10^{-21}	24.4
16	MAPK3	0.12	0.0028	4.1×10^{-21}	24.1
3	WDR6	0.44	0.0592	2×10^{-20}	23.4
16	KCTD13	−0.09	0.0006	2.9×10^{-20}	23.3
16	TBX6	−0.08	0.0004	4.4×10^{-20}	23.1
11	ACAD8	0.31	0.179	5×10^{-20}	23
16	SGF29	0.27	0.0209	6.9×10^{-20}	22.9
16	DDX19B	−0.26	0.0256	7.3×10^{-20}	22.9
16	APOBR	−0.1	0.0013	1.3×10^{-19}	22.6
16	SH2B1	−0.14	0.0055	5.7×10^{-19}	22
3	NCKIPSD	0.33	0.0362	1.3×10^{-18}	21.6
11	ACP2	−0.24	0.0175	2.8×10^{-18}	21.3
16	NFATC2IP	−0.09	0.0007	3.1×10^{-18}	21.2
16	CD19	−0.16	0.0069	3.6×10^{-18}	21.1

Note: The abbreviations Chr and BF denote the chromosome number and Bayes factor respectively. The results are obtained based on 7000 MCMC iterations with a burn-in sample of 2000.

variability of estimated eQTLs' effect sizes while imputing the GReX. This results in an imprecise estimation of GReX effect size and inflated type-I error in TWAS. We developed a model that integrates the data from the reference and GWAS panels and provides a unified Bayesian framework to conduct the TWAS in a single step. Instead of fitting two separate models for transcriptomic and GWAS data like existing TWAS methods, we implement a hierarchical structure where cis-SNP effect sizes connect our reference and GWAS panel models. We develop a simple computational technique by implementing the MCMC to infer the model parameters through the Gibbs sampling algorithm. The prior variance of the effect size vector of cis-SNPs in our model is updated by the posterior variance in an MCMC step. Our model's simultaneous estimation and testing framework automatically incorporates the variability of cis-SNP effect sizes in the analysis.

We use the horseshoe prior in our model to separate non-null SNPs from null SNPs in the transcriptome data. Horseshoe prior has an advantage over the elastic net penalty used in PrediXcan for accurately estimating the larger eQTL effects. The elastic net penalty applies severe shrinkage on larger eQTL effect sizes whereas for the horseshoe prior large effects remain unshrunk. The non-vanishing amount of shrinkage of large eQTLS for the PrediXcan method leads to the imprecise GReX imputation. Moreover, in the presence of a small number of eQTLs with small to moderate effect sizes, PrediXcan might fail to identify any of the significant eQTLs underlying gene expressions. The real data analysis section supports that PrediXcan fails to construct GReX for 24 heritable genes. Thus the applicability of the PrediXcan method is hindered in those situations.

We demonstrate through simulation that our method maintains good estimation accuracy for different choices of GReX effect sizes. Our method estimates the GReX more precisely than PrediXcan while keeping the MSE lower. The estimation accuracy remains unaffected for different choices of reference panel and GWAS sample sizes. The increase in GWAS sample size reduces the relative bias and MSE in estimating β for our method whereas for the PrediXcan method, it does not have any impact. The accurate estimate of GReX is crucial in calculating the polygenic risk score (PRS) to determine the risk profile of an individual underlying a disease. In addition to maintaining good estimation accuracy, our method can correctly identify the true set of non-null genes. For a given level of specificity, our method maintains the same or higher level

Table 4: Risk genes for BMI identified by our method in both EUR and AFR ancestry

Chr	Gene	Estimate	Variance	$P(H_0\|\text{data})$	$log_{10}(\text{BF})$
1	ZCCHC17	0.12	0.0085	3.5×10^{-2}	5.2
1	NAV1	-0.45	0.0561	3.5×10^{-18}	21.2
3	DHFR2	-0.18	0.0118	3.9×10^{-9}	12.1
5	RHOBTB3	0.14	0.0058	1.6×10^{-5}	8.5
6	MOCS1	-0.11	0.0015	3.7×10^{-17}	20.2
6	ZNF192P1	0.1	0.0011	10×10^{-8}	10.7
6	ZSCAN26	0.09	0.0014	2.3×10^{-4}	7.4
6	RNF5	0.04	0.0001	2.3×10^{-7}	10.4
10	HPS1	0.33	0.0391	3.3×10^{-7}	10.2
10	FBXL15	-0.16	0.0131	4.2×10^{-7}	10.1
10	DPYSL4	-0.02	0.0001	3.9×10^{-2}	5.1
10	TRIM8	-0.28	0.0565	1×10^{-6}	9.7
11	ACP2	-0.29	0.0642	3.6×10^{-18}	21.2
11	OVOL1	-0.33	0.1363	1.1×10^{-13}	16.7
12	CFAP54	0.13	0.0038	4.2×10^{-4}	7.1
12	HCAR3	0.42	0.1515	1.5×10^{-10}	13.6
16	TBX6	-0.07	0.0003	9.1×10^{-20}	22.8
16	TAOK2	-0.17	0.0052	7.5×10^{-21}	23.9
17	CBX1	0.19	0.0127	6.9×10^{-10}	12.9

Note: The abbreviations Chr and BF denote the chromosome number and Bayes factor respectively. The results are obtained based on 7000 MCMC iterations with a burn-in sample of 2000.

of sensitivity than the PrediXcan for varying choices of sample sizes and heritability. The sensitivity of our method improves with the increase in either gene or trait level heritability for a given GWAS sample size. Our method is particularly beneficial compared to the PrediXcan method when the gene level heritability is lower.

We extend our single-ancestry method to incorporate multiple expression and GWAS studies to conduct TWAS effectively. Our multi-ancestry TWAS model aims to identify risk genes for traits relevant to all ancestries. Our method effectively combines complementary information from each ancestry to decipher genetic architecture. SNPs affect gene expression differently in each ancestry due to the difference in allele frequency and LD, so our model's heterogeneous parameters for cis-SNP effect sizes can capture the distinct eQTL effects. Moreover, the SNP-specific shrinkage penalty induced by horseshoe prior can powerfully locate the shared eQTLs between ancestries. The local shrinkage parameters in horseshoe prior also induce marginal correlation among SNP effects, increasing the accuracy in effect size estimation. Hence our model can handle different sparsity levels of eQTLs without making any particular assumption on genetic architecture. Likewise, single-ancestry setup, our multi-ancestry method adjusts the uncertainty of predicted expression by connecting the reference panel and GWAS models by cis-SNP effects. The only assumption in our model is that causal genes are shared across ancestries. Such an assumption is consistent with overlapped risk regions between ancestries validated by multiple GWAS studies [32]. Thus, utilizing a single parameter for the gene-trait association test can powerfully detect the genes in those ancestries with a lower GWAS sample size. Through simulation, we show that our method accurately estimates the GReX effect and keeps MSE lower than METRO. Irrespective of each ancestry's eQTL architecture, our method efficiently identifies the non-null genes underlying the trait. Our method shows higher AUC than METRO when eQTLs do not overlap between ancestries.

Our Bayesian method suffers from a few limitations. Our method performs the gene-trait association without adjusting the horizontal pleiotropic effect of SNP on the trait. SNPs can affect the trait without modulating the expression levels of a gene. Such an effect is observed in some segments of the genome block due to varying patterns of LD and the polygenic nature of a trait. Another limitation of our model is that it ignores the effect of gene or environment interaction in TWAS. The TWAS methods are traditionally

developed to perform the association test between a single gene and a trait. As the neighboring genes on the same locus share a proportion of causal eQTLs underlying gene expression, gene-gene interaction (G×G) is a potential confounding factor that needs to be controlled during the association test. Gene-environment interaction (G×E) is another important factor that confers different disease risks to individuals exposed to various environmental conditions.

Overall, we proposed a novel Bayesian approach that explicitly adjusts the uncertainty of the predicted expression by integrating the reference and GWAS panel data to conduct TWAS in a unified manner. Our method is simple to implement, computationally efficient, and has a natural extension to analyze multiple ancestries jointly. Its numerous applications in real data, such as BMI, identify many genes associated with complex traits that can better elucidate the complex disease etiology.

Data and software information

The dataset and software used in this article are available (either openly or via application) on the following websites.
Geuvadis data: https://www.internationalgenome.org/data-portal/data-collection/geuvadis
UK Biobank: https://www.ukbiobank.ac.uk/
HAPGEN: https://mathgen.stats.ox.ac.uk/genetics_software/hapgen/hapgen2.html
GCTA: https://cnsgenomics.com/software/gcta/
PLINK: https://www.cog-genomics.org/plink/

Declaration of interests

The authors have no competing interests.

References

[1] J. N. Hirschhorn and M. J. Daly, "Genome-wide association studies for common diseases and complex traits," *Nature reviews genetics*, vol. 6, no. 2, pp. 95–108, 2005.

[2] D. L. Nicolae, E. Gamazon, W. Zhang, S. Duan, M. E. Dolan, and N. J. Cox, "Trait-associated snps are more likely to be eqtls: annotation to enhance discovery from gwas," *PLoS genetics*, vol. 6, no. 4, p. e1000888, 2010.

[3] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. De Geus, D. I. Boomsma, F. A. Wright *et al.*, "Integrative approaches for large-scale transcriptome-wide association studies," *Nature genetics*, vol. 48, no. 3, pp. 245–252, 2016.

[4] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.

[5] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.

[6] C. de Leeuw, J. Werme, J. E. Savage, W. J. Peyrot, and D. Posthuma, "On the interpretation of transcriptome-wide association studies," *PLoS Genetics*, vol. 19, no. 9, p. e1010921, 2023.

[7] Z. Li, W. Zhao, L. Shang, T. H. Mosley, S. L. Kardia, J. A. Smith, and X. Zhou, "Metro: Multi-ancestry transcriptome-wide association studies for powerful gene-trait association detection," *The American Journal of Human Genetics*, vol. 109, no. 5, pp. 783–801, 2022.

[8] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American statistical association*, vol. 101, no. 476, pp. 1418–1429, 2006.

[9] C. M. Carvalho, N. G. Polson, and J. G. Scott, "The horseshoe estimator for sparse signals," *Biometrika*, vol. 97, no. 2, pp. 465–480, 2010.

[10] H. ISHWARAN and J. S. RAO, "Spike and slab variable selection: Frequentist and bayesian strategies," *The Annals of Statistics*, vol. 33, no. 2, pp. 730–773, 2005.

[11] J. M. Luningham, J. Chen, S. Tang, P. L. De Jager, D. A. Bennett, A. S. Buchman, and J. Yang, "Bayesian genome-wide twas method to leverage both cis-and trans-eqtl information through summary statistics," *The American Journal of Human Genetics*, vol. 107, no. 4, pp. 714–726, 2020.

[12] S. Nagpal, X. Meng, M. P. Epstein, L. C. Tsoi, M. Patrick, G. Gibson, P. L. De Jager, D. A. Bennett, A. P. Wingo, T. S. Wingo *et al.*, "Tigar: an improved bayesian tool for transcriptomic data imputation enhances gene mapping of complex traits," *The American Journal of Human Genetics*, vol. 105, no. 2, pp. 258–266, 2019.

[13] A. Bhadra, J. Datta, N. G. Polson, and B. Willard, "Lasso meets horseshoe," *Statistical Science*, vol. 34, no. 3, pp. 405–427, 2019.

[14] G. Casella and C. P. Robert, "Rao-blackwellisation of sampling schemes," *Biometrika*, vol. 83, no. 1, pp. 81–94, 1996.

[15] J. Yang, L. G. Fritsche, X. Zhou, and G. Abecasis, "A scalable bayesian method for integrating functional information in genome-wide association studies," *The American Journal of Human Genetics*, vol. 101, no. 3, pp. 404–416, 2017.

[16] M. M. Momin, J. Shin, S. Lee, B. Truong, B. Benyamin, and S. H. Lee, "A method for an unbiased estimate of cross-ancestry genetic correlation using individual-level data," *Nature communications*, vol. 14, no. 1, p. 722, 2023.

[17] Z. Lu, S. Gopalan, D. Yuan, D. V. Conti, B. Pasaniuc, A. Gusev, and N. Mancuso, "Multi-ancestry fine-mapping improves precision to identify causal genes in transcriptome-wide association studies," *The American Journal of Human Genetics*, vol. 109, no. 8, pp. 1388–1404, 2022.

[18] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E. Eyler, J. C. Denny, G. Consortium, D. L. Nicolae *et al.*, "A gene-based association method for mapping traits using reference transcriptome data," *Nature genetics*, vol. 47, no. 9, pp. 1091–1098, 2015.

[19] Z. Su, J. Marchini, and P. Donnelly, "Hapgen2: simulation of multiple disease snps," *Bioinformatics*, vol. 27, no. 16, pp. 2304–2305, 2011.

[20] M. A. Coram, S. I. Candille, Q. Duan, K. H. K. Chan, Y. Li, C. Kooperberg, A. P. Reiner, and H. Tang, "Leveraging multi-ethnic evidence for mapping complex traits in minority populations: an empirical bayes approach," *The American Journal of Human Genetics*, vol. 96, no. 5, pp. 740–752, 2015.

[21] T. Lappalainen, M. Sammeth, M. R. Friedländer, P. A. 't Hoen, J. Monlong, M. A. Rivas, M. Gonzalez-Porta, N. Kurbatova, T. Griebel, P. G. Ferreira *et al.*, "Transcriptome and genome sequencing uncovers functional variation in humans," *Nature*, vol. 501, no. 7468, pp. 506–511, 2013.

[22] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for dna microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001.

[23] J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D. Barrell, A. Zadissa, S. Searle *et al.*, "Gencode: the reference human genome annotation for the encode project," *Genome research*, vol. 22, no. 9, pp. 1760–1774, 2012.

[24] T. M. Beasley, S. Erickson, and D. B. Allison, "Rank-based inverse normal transformations are increasingly used, but are they merited?" *Behavior genetics*, vol. 39, pp. 580–595, 2009.

[25] J. Yang, S. H. Lee, M. E. Goddard, and P. M. Visscher, "Gcta: a tool for genome-wide complex trait analysis," *The American Journal of Human Genetics*, vol. 88, no. 1, pp. 76–82, 2011.

[26] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[27] P. H. Westfall, W. O. Johnson, and J. M. Utts, "A bayesian perspective on the bonferroni adjustment," *Biometrika*, vol. 84, no. 2, pp. 419–427, 1997.

[28] R. E. Kass and A. E. Raftery, "Bayes factors," *Journal of the american statistical association*, vol. 90, no. 430, pp. 773–795, 1995.

[29] M. Stephens and D. J. Balding, "Bayesian statistical methods for genetic association studies," *Nature Reviews Genetics*, vol. 10, no. 10, pp. 681–690, 2009.

[30] C.-Y. Chen, T.-T. Chen, Y.-C. A. Feng, M. Yu, S.-C. Lin, R. J. Longchamps, S.-H. Wang, Y.-H. Hsu, H.-I. Yang, P.-H. Kuo *et al.*, "Analysis across taiwan biobank, biobank japan, and uk biobank identifies hundreds of novel loci for 36 quantitative traits," *Cell genomics*, vol. 3, no. 12, 2023.

[31] G. Kichaev, G. Bhatia, P.-R. Loh, S. Gazal, K. Burch, M. K. Freund, A. Schoech, B. Pasaniuc, and A. L. Price, "Leveraging polygenic functional enrichment to improve gwas power," *The American Journal of Human Genetics*, vol. 104, no. 1, pp. 65–75, 2019.

[32] H. Shi, S. Gazal, M. Kanai, E. M. Koch, A. P. Schoech, K. M. Siewert, S. S. Kim, Y. Luo, T. Amariuta, H. Huang *et al.*, "Population-specific causal disease effect sizes in functionally important regions impacted by selection," *Nature communications*, vol. 12, no. 1, p. 1098, 2021.

Supplemental information: A unified Bayesian approach to transcriptome-wide association study

Arnab Kumar Khan[1], Tanushree Haldar[2], and Arunabha Majumdar[3*]

[1]Human Genetics Unit, Indian Statistical Institute Kolkata, Kolkata, West Bengal 700103, India
[2]Institute for Human Genetics, University of California San Francisco, San Francisco, California 94143, USA
[3]Department of Mathematics, Indian Institute of Technology Hyderabad, Kandi, Telangana 502285, India

1 Derivation of the full conditional posterior distribution for single ancestry model

The hierarchical structure for our single ancestry model is given by

$$\begin{aligned}
\mathbf{x}_{\text{ref}}|\boldsymbol{\alpha}, \sigma_e^2 &\sim N_{n_r}(\mathbf{G}_{\text{ref}}\boldsymbol{\alpha}, \sigma_e^2 \mathrm{I}_{n_r}) \\
\mathbf{y}|\beta, \boldsymbol{\alpha}, \sigma_\epsilon^2 &\sim N_n(\beta\mathbf{G}\boldsymbol{\alpha}, \sigma_\epsilon^2 \mathrm{I}_n) \\
\alpha_i|\lambda_i, \tau, \sigma_e^2 &\sim N(0, \sigma_e^2\tau^2\lambda_i^2), \quad i = 1(1)m \\
\lambda_i &\sim C^+(0,1) \\
\tau &\sim C^+(0,1) \\
\beta|q, \sigma_d^2, \sigma_\epsilon^2 &\sim (1-q)\delta_{\{0\}}(\beta) + qN(0, \sigma_d^2\sigma_\epsilon^2) \\
q &\sim Beta(b_1, b_2) \\
\sigma_d^2 &\sim IG(d_1, d_2) \\
\sigma_e^2 &\sim IG(e_1, e_2) \\
\sigma_\epsilon^2 &\sim IG(\epsilon_1, \epsilon_2)
\end{aligned} \tag{1}$$

As discussed in the main text, the terms in the Equation 1 carry out their usual meaning. Let us denote $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_m)^{\mathrm{T}}$, where α_i is the effect of ith cis-SNP on the gene expression levels for $i = 1(1)m$. We specify the SNP-specific shrinkage penalties in the form of a vector as $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_m)^{\mathrm{T}}$. Let us define $\mathbf{D}_\lambda = diag(\lambda_1^2, \lambda_2^2, \cdots, \lambda_m^2)$, where $\mathbf{D}_\lambda$ is a $m \times m$ diagonal matrix. Then we have

$$\boldsymbol{\alpha}|\boldsymbol{\lambda}, \tau, \sigma_e^2 \sim N_m(\mathbf{0}, \sigma_e^2\tau^2\mathbf{D}_\lambda)$$

Now, we derive the full conditional distributions for the model parameters.

1.1 Full conditional distribution of $\boldsymbol{\alpha}$

Let us use the notation $\boldsymbol{\alpha}|\cdot$ to denote the full conditional posterior distribution of $\boldsymbol{\alpha}$ and $f(\boldsymbol{\alpha}|\cdot)$ is the corresponding probability density function (pdf). We have used similar abbreviations for the other model parameters to derive the full conditional distribution. The full conditional distribution of $\boldsymbol{\alpha}$ is given as

follows

$$
\begin{aligned}
f(\boldsymbol{\alpha}|\cdot) &\propto f(\mathbf{x}_{\text{ref}}|\boldsymbol{\alpha},\sigma_e^2)f(\mathbf{y}|\beta,\boldsymbol{\alpha},\sigma_\epsilon^2)f(\boldsymbol{\alpha}|\boldsymbol{\lambda},\tau,\sigma_e^2)\\
&\propto \exp\left[-\frac{1}{2\sigma_e^2}(\mathbf{x}_{\text{ref}}-\mathbf{G}_{\text{ref}}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{x}_{\text{ref}}-\mathbf{G}_{\text{ref}}\boldsymbol{\alpha})\right]\exp\left[-\frac{1}{2\sigma_\epsilon^2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right]\exp\left[-\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2\tau^2}\right]\\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\alpha}^{\mathrm{T}}\left(\frac{\mathbf{G}_{\text{ref}}^{\mathrm{T}}\mathbf{G}_{\text{ref}}}{\sigma_e^2}+\frac{\beta^2\mathbf{G}^{\mathrm{T}}\mathbf{G}}{\sigma_\epsilon^2}+\frac{\mathbf{D}_\lambda^{-1}}{\sigma_e^2\tau^2}\right)\boldsymbol{\alpha}-2\boldsymbol{\alpha}^{\mathrm{T}}\left(\frac{\mathbf{G}_{\text{ref}}^{\mathrm{T}}\mathbf{x}_{\text{ref}}}{\sigma_e^2}+\frac{\beta\mathbf{G}^{\mathrm{T}}\mathbf{y}}{\sigma_\epsilon^2}\right)\right]\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\alpha}^{\mathrm{T}}\Sigma\boldsymbol{\alpha}-2\boldsymbol{\alpha}^{\mathrm{T}}\left(\frac{\mathbf{G}_{\text{ref}}^{\mathrm{T}}\mathbf{x}_{\text{ref}}}{\sigma_e^2}+\frac{\beta\mathbf{G}^{\mathrm{T}}\mathbf{y}}{\sigma_\epsilon^2}\right)\right]\right\}, \quad \text{where } \boldsymbol{\Sigma}=\frac{\mathbf{G}_{\text{ref}}^{\mathrm{T}}\mathbf{G}_{\text{ref}}}{\sigma_e^2}+\frac{\beta^2\mathbf{G}^{\mathrm{T}}\mathbf{G}}{\sigma_\epsilon^2}+\frac{\mathbf{D}_\lambda^{-1}}{\sigma_e^2\tau^2}
\end{aligned}
$$

Hence, the full conditional distribution of $\boldsymbol{\alpha}$ is given by

$$
\boldsymbol{\alpha}|\cdot \sim N_m\left(\boldsymbol{\Sigma}^{-1}\left(\frac{\mathbf{G}_{\text{ref}}^{\mathrm{T}}\mathbf{x}_{\text{ref}}}{\sigma_e^2}+\frac{\beta\mathbf{G}^{\mathrm{T}}\mathbf{y}}{\sigma_\epsilon^2}\right),\boldsymbol{\Sigma}^{-1}\right)
$$

1.2 Full conditional distribution of $\boldsymbol{\lambda}$

As $\boldsymbol{\lambda}$ is a m dimensional vector, it is simpler to continuously update components of $\boldsymbol{\lambda}$ in the Gibbs sampler. Instead of updating the entire $\boldsymbol{\lambda}$ vector at once, we alternately update each component λ_i in the MCMC. For each $i=1(1)m$, we have

$$
\begin{aligned}
f(\lambda_i|\cdot) &\propto f(\alpha_i|\lambda_i,\tau,\sigma_e^2)f(\lambda_i)\\
&\propto \left\{\frac{1}{\sigma_e\lambda_i\tau}\ \exp\left[-\frac{\alpha_i^2}{2\sigma_e^2\tau^2\lambda_i^2}\right]\right\}\left\{\frac{1}{1+\lambda_i^2}\right\}\\
&\propto \frac{1}{\lambda_i(1+\lambda_i^2)}\ \exp\left[-\frac{\alpha_i^2}{2\sigma_e^2\tau^2\lambda_i^2}\right]
\end{aligned}
\tag{2}
$$

Note that, the full conditional posterior distribution of λ_i is not a well-known distribution from which we can draw random samples directly. In such a scenario, we use the Metropolis-Hastings algorithm within the Gibbs sampler to update the parameter λ_i in the MCMC.

1.3 Full conditional distribution of τ

The full conditional distribution of τ is given as

$$
\begin{aligned}
f(\tau|\cdot) &\propto f(\boldsymbol{\alpha}|\boldsymbol{\lambda},\tau,\sigma_e^2)f(\tau)\\
&\propto \left\{\frac{1}{\tau^m}\exp\left[-\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2\tau^2}\right]\right\}\left\{\frac{1}{1+\tau^2}\right\}\\
&\propto \frac{1}{\tau^m(1+\tau^2)}\exp\left[-\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2\tau^2}\right]
\end{aligned}
\tag{3}
$$

As the posterior distribution of τ is not conditionally conjugate, we update the parameter τ by the Metropolis-Hastings algorithm within the Gibbs sampler.

1.4 Full conditional distribution of σ_e^2

The full conditional distribution of σ_e^2 is given by

$$
\begin{aligned}
f(\sigma_e^2|\cdot) &\propto f(\mathbf{x}_{\text{ref}}|\boldsymbol{\alpha},\sigma_e^2)f(\boldsymbol{\alpha}|\boldsymbol{\lambda},\tau,\sigma_e^2)f(\sigma_e^2)\\
&\propto \left\{(\sigma_e^2)^{-\frac{n_r}{2}}\exp\left[-\frac{1}{2\sigma_e^2}(\mathbf{x}_{\text{ref}}-\mathbf{G}_{\text{ref}}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{x}_{\text{ref}}-\mathbf{G}_{\text{ref}}\boldsymbol{\alpha})\right]\right\}\left\{(\sigma_e^2)^{-\frac{m}{2}}\exp\left[-\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2\tau^2}\right]\right\}\\
&\quad\left\{(\sigma_e^2)^{-(e_1+1)}\exp\left[-\frac{e_2}{\sigma_e^2}\right]\right\}\\
&\propto \left(\sigma_e^2\right)^{-\left(e_1+\frac{n_r}{2}+\frac{m}{2}+1\right)}\exp\left[-\frac{1}{\sigma_e^2}\left(e_2+\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\tau^2}+\frac{1}{2}(\mathbf{x}_{\text{ref}}-\mathbf{G}_{\text{ref}}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{x}_{\text{ref}}-\mathbf{G}_{\text{ref}}\boldsymbol{\alpha})\right)\right]
\end{aligned}
$$

The full conditional distribution of σ_e^2 is given by

$$\sigma_e^2|\cdot \sim IG\left(e_1 + \frac{n_r}{2} + \frac{m}{2},\ e_2 + \frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\tau^2} + \frac{1}{2}(\mathbf{x}_{\mathrm{ref}} - \mathbf{G}_{\mathrm{ref}}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{x}_{\mathrm{ref}} - \mathbf{G}_{\mathrm{ref}}\boldsymbol{\alpha})\right)$$

1.5 Full conditional distribution of β

The full conditional distribution of β is given by

$$\begin{aligned}
f(\beta|\cdot) &\propto f(\mathbf{y}|\boldsymbol{\alpha}, \beta, \sigma_\epsilon^2) f(\beta|q, \sigma_d^2, \sigma_\epsilon^2) \\
&\propto \exp\left[-\frac{1}{2\sigma_\epsilon^2}(\mathbf{y} - \beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y} - \beta\mathbf{G}\boldsymbol{\alpha})\right]\left\{(1-q)\delta_{\{0\}}(\beta) + \frac{q}{\sqrt{2\pi}\sigma_d\sigma_\epsilon}\ \exp\left[-\frac{\beta^2}{2\sigma_d^2\sigma_\epsilon^2}\right]\right\} \\
&\propto \exp\left[-\frac{1}{2\sigma_\epsilon^2}\left(\beta^2\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\boldsymbol{\alpha} - 2\beta\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{y}\right)\right]\left\{(1-q)\delta_{\{0\}}(\beta) + \frac{q}{\sqrt{2\pi}\sigma_d\sigma_\epsilon}\ \exp\left[-\frac{\beta^2}{2\sigma_d^2\sigma_\epsilon^2}\right]\right\} \\
&\propto (1-q)\delta_{\{0\}}(\beta) + \frac{q}{\sqrt{2\pi}\sigma_d\sigma_\epsilon}\ \exp\left[-\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\boldsymbol{\alpha} + \frac{1}{\sigma_d^2}}{2\sigma_\epsilon^2}\left(\beta^2 - 2\beta\ \frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{y}}{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\boldsymbol{\alpha} + \frac{1}{\sigma_d^2}}\right)\right] \\
&\propto (1-q)\delta_{\{0\}}(\beta) + \frac{q}{\sqrt{2\pi}\sigma_d\sigma_\epsilon}\ \exp\left[-\frac{1}{2\widetilde{\sigma}^2}(\beta^2 - 2\beta\widetilde{\mu})\right] \\
&\propto (1-q)\delta_{\{0\}}(\beta) + \left(\frac{q\widetilde{\sigma}}{\sigma_d\sigma_\epsilon}\exp\left[\frac{\widetilde{\mu}^2}{2\widetilde{\sigma}^2}\right]\right)\left\{\frac{1}{\sqrt{2\pi}\widetilde{\sigma}}\exp\left[-\frac{1}{2\widetilde{\sigma}^2}(\beta - \widetilde{\mu})^2\right]\right\}, \\
&\text{where}\quad \widetilde{\mu} = \frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{y}}{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\boldsymbol{\alpha} + \frac{1}{\sigma_d^2}}\quad \text{and}\quad \widetilde{\sigma}^2 = \frac{\sigma_\epsilon^2}{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\boldsymbol{\alpha} + \frac{1}{\sigma_d^2}}
\end{aligned}$$

Hence, the full conditional distribution of β is represented as a mixture of point mass function $\delta_{\{0\}}(\beta)$ and normal distribution $N(\widetilde{\mu}, \widetilde{\sigma}^2)$. The weight corresponding to $\delta_{\{0\}}(\beta)$ turns out to be

$$w = \frac{1-q}{1 - q + \frac{q\widetilde{\sigma}}{\sigma_d\sigma_\epsilon}\exp\left(\frac{\widetilde{\mu}^2}{2\widetilde{\sigma}^2}\right)} \tag{4}$$

Hence, the conditional posterior distribution of β is given as follows

$$\beta|\cdot \sim \begin{cases} \delta_{\{0\}}(\beta), & \text{with probability} \quad w = \frac{1-q}{1-q+\frac{q\tilde{\sigma}}{\sigma_d\sigma_\epsilon}\exp\left(\frac{\tilde{\mu}^2}{2\tilde{\sigma}^2}\right)} \\ N(\widetilde{\mu}, \widetilde{\sigma}^2), & \text{with probability} \quad 1 - w \end{cases}$$

1.6 Full conditional distribution of q

In this section, we derive the full conditional distribution of q. We denote $\phi(x; \mu, \sigma^2)$ as the value of the pdf of the normal distribution at the point x with mean μ and variance σ^2. Let $B(s,t) = \frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}$ denotes the beta function, where $\Gamma(\cdot)$ indicating the gamma function. The full conditional distribution of q is given as

$$\begin{aligned}
f(q|\cdot) &\propto f(\beta|q, \sigma_d^2, \sigma_\epsilon^2) f(q) \\
&\propto \left\{(1-q)\delta_{\{0\}}(\beta) + q\ \phi(\beta; 0, \sigma_d^2\sigma_\epsilon^2)\right\}\left\{q^{b_1-1}(1-q)^{b_2-1}\right\} \\
&\propto \left\{\delta_{\{0\}}(\beta)q^{b_1-1}(1-q)^{b_2} + \phi(\beta; 0, \sigma_d^2\sigma_\epsilon^2)q_1^b(1-q)^{b_2-1}\right\} \\
&\propto \left\{\delta_{\{0\}}(\beta)B(b_1, b_2+1)Beta(b_1, b_2+1) + \phi(\beta; 0, \sigma_d^2\sigma_\epsilon^2)B(b_1+1, b_2)Beta(b_1+1, b_2)\right\}
\end{aligned}$$

Hence, our full conditional distribution of q is represented as a mixture of Beta distributions as follows

$$q|\cdot \sim \begin{cases} Beta(b_1, b_2+1), & \text{with probability} \quad w = \frac{\delta_{\{0\}}(\beta)B(b_1,b_2+1)}{\delta_0(\beta)B(b_1,b_2+1)+\phi(\beta;0,\sigma_d^2\sigma_\epsilon^2)B(b_1+1,b_2)} \\ Beta(b_1+1, b_2), & \text{with probability} \quad 1 - w \end{cases} \tag{5}$$

1.7 Full conditional distribution of σ_d^2

The full conditional distribution of σ_d^2 is given by

$$
\begin{aligned}
f(\sigma_d^2|\cdot) &\propto f(\beta|q,\sigma_d^2,\sigma_\epsilon^2)f(\sigma_d^2)\\
&\propto \left\{(1-q)\delta_{\{0\}}(\beta)+\frac{q}{\sqrt{2\pi}\sigma_d\sigma_\epsilon}\ \exp\left[-\frac{\beta^2}{2\sigma_d^2\sigma_\epsilon^2}\right]\right\}\left\{(\sigma_d^2)^{-(d_1+1)}\exp\left[-\frac{d_2}{\sigma_d^2}\right]\right\}\\
&\propto \left\{(1-q)\delta_{\{0\}}(\beta)(\sigma_d^2)^{-(d_1+1)}\exp\left[-\frac{d_2}{\sigma_d^2}\right]+\frac{q}{\sqrt{2\pi}\sigma_\epsilon}(\sigma_d^2)^{-(d_1+\frac{1}{2}+1)}\exp\left[-\frac{1}{\sigma_d^2}\left(d_2+\frac{\beta^2}{2\sigma_\epsilon^2}\right)\right]\right\}\\
&\propto \frac{(1-q)\delta_{\{0\}}(\beta)\Gamma(d_1)}{d_2^{d_1}}\ IG(d_1,d_2)+\frac{q\Gamma(d_1+\frac{1}{2})}{\sqrt{2\pi}\sigma_\epsilon\left(d_2+\frac{\beta^2}{2\sigma_\epsilon^2}\right)^{d_1+\frac{1}{2}}}\ IG\left(d_1+\frac{1}{2},\ d_2+\frac{\beta^2}{2\sigma_\epsilon^2}\right)
\end{aligned}
$$

Let us define

$$
w_1=\frac{(1-q)\delta_{\{0\}}(\beta)\Gamma(d_1)}{d_2^{d_1}}\quad\text{and}\quad w_2=\frac{q\Gamma(d_1+\frac{1}{2})}{\sqrt{2\pi}\sigma_\epsilon\left(d_2+\frac{\beta^2}{2\sigma_\epsilon^2}\right)^{d_1+\frac{1}{2}}}
$$

Then, $\sigma_d^2|\cdot$ follows a mixture of inverse gamma distributions as

$$
\sigma_d^2|\cdot\sim\begin{cases}IG(d_1,d_2), & \text{with probability } \frac{w_1}{w_1+w_2}\\ IG\left(d_1+\frac{1}{2},\ d_2+\frac{\beta^2}{2\sigma_\epsilon^2}\right) & \text{with probability } \frac{w_2}{w_1+w_2}\end{cases}\tag{6}
$$

1.8 Full conditional distribution of σ_ϵ^2

The full conditional distribution of σ_ϵ^2 is given as

$$
\begin{aligned}
&f(\sigma_\epsilon^2|\cdot)\propto f(\mathbf{y}|\boldsymbol{\alpha},\beta,\sigma_\epsilon^2)f(\beta|q,\sigma_d^2,\sigma_\epsilon^2)f(\sigma_\epsilon^2)\\
&\propto\left\{\left(\sigma_\epsilon^2\right)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma_\epsilon^2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right]\right\}\left\{(1-q)\delta_{\{0\}}(\beta)+\frac{q}{\sqrt{2\pi}\sigma_d\sigma_\epsilon}\ \exp\left[-\frac{\beta^2}{2\sigma_d^2\sigma_\epsilon^2}\right]\right\}\\
&\quad\left\{(\sigma_\epsilon^2)^{-(\epsilon_1+1)}\exp\left[-\frac{\epsilon_2}{\sigma_\epsilon^2}\right]\right\}\\
&\propto\left\{(1-q)\delta_{\{0\}}(\beta)(\sigma_\epsilon^2)^{-(\frac{n}{2}+\epsilon_1+1)}\exp\left[-\frac{1}{\sigma_\epsilon^2}\left(\epsilon_2+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)\right]+\right.\\
&\quad\left.\frac{q}{\sqrt{2\pi}\sigma_d}(\sigma_\epsilon^2)^{-(\frac{n+1}{2}+\epsilon_1+1)}\exp\left[-\frac{1}{\sigma_\epsilon^2}\left(\epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)\right]\right\}\\
&\propto\left\{\frac{(1-q)\delta_{\{0\}}(\beta)\Gamma(\frac{n}{2}+\epsilon_1)}{\left(\epsilon_2+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)^{\frac{n}{2}+\epsilon_1}}\ IG\left(\frac{n}{2}+\epsilon_1,\ \epsilon_2+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)+\right.\\
&\left.\frac{q\Gamma(\frac{n+1}{2}+\epsilon_1)}{\sqrt{2\pi}\sigma_d\left(\epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)^{\frac{n+1}{2}+\epsilon_1}}\ IG\left(\frac{n+1}{2}+\epsilon_1,\ \epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)\right\}
\end{aligned}
$$

Let us define

$$
w_1=\frac{(1-q)\delta_{\{0\}}(\beta)\Gamma(\frac{n}{2}+\epsilon_1)}{\left(\epsilon_2+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)^{\frac{n}{2}+\epsilon_1}}\quad\text{and}\quad w_2=\frac{q\Gamma(\frac{n+1}{2}+\epsilon_1)}{\sqrt{2\pi}\sigma_d\left(\epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)^{\frac{n+1}{2}+\epsilon_1}}
$$

Then $\sigma_\epsilon^2|\cdot$ follows a mixture of inverse gamma distributions as

$$
\sigma_\epsilon^2|\cdot\sim\begin{cases}IG\left(\frac{n}{2}+\epsilon_1,\ \epsilon_2+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right) & \text{with probability } \frac{w_1}{w_1+w_2}\\ IG\left(\frac{n+1}{2}+\epsilon_1,\ \epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right) & \text{with probability } \frac{w_2}{w_1+w_2}\end{cases}
$$

2 Outline of Metropolis-Hastings algorithm for shrinkage parameters

We make inferences on the model parameters by implementing the Gibbs sampling algorithm. However, we cannot draw samples from these posterior distributions because the full conditional distributions of shrinkage parameters $\boldsymbol{\lambda}$ and τ are not in known forms. As a remedy, we implement the Metropolis-Hastings algorithm within the Gibbs sampler to draw random samples from these conditional posterior distributions.

2.1 Metropolis-Hastings algorithm for $\boldsymbol{\lambda}$

We shall derive the Metropolis-Hasting algorithm for λ_i, which is the ith component of the vector $\boldsymbol{\lambda}$. The derivation for the other components can be done similarly. The full conditional posterior distribution of λ_i (Equation 2) is given as follows

$$f(\lambda_i|\cdot) \propto \frac{1}{\lambda_i(1+\lambda_i^2)} \exp\left[-\frac{\alpha_i^2}{2\sigma_e^2\tau^2\lambda_i^2}\right], \quad \lambda_i > 0$$

Consider the transformation $Z_i = \lambda_i^2$. Then the conditional posterior distribution of Z_i turns out to be

$$f(Z_i|\cdot) \propto \frac{1}{Z_i(1+Z_i)} \exp\left[-\frac{\alpha_i^2}{2\sigma_e^2\tau^2 Z_i}\right], \quad Z > 0$$

Here the updation of the parameter λ_i is equivalent to update Z_i as they are related by one-to-one transformation $Z_i = \lambda_i^2$. Let us denote the probability density function (pdf) of $Z_i|\cdot$ by $g(\cdot)$. We shall choose the candidate proposal distribution as $IG\left(1, \frac{\alpha_i^2}{2\sigma_e^2\tau^2}\right)$ and denote the corresponding pdf by $h(\cdot)$. The reason for using such candidate distribution is that our proposal pdf approximates the target pdf for large values of the support of Z_i. As our proposal distribution does not depend on the generated sample from the target pdf at any step, an independent Metropolis-Hasting algorithm is applied to update Z_i in the MCMC.
Let ${z_i}^{(l-1)}$ is the generated sample from the target distribution $Z_i|\cdot$ at the $(l-1)^{\text{th}}$ step. We consider the following steps to draw a sample from the target distribution at the lth step.

- Generate a random sample $y^{(l)}$ from the $IG\left(1, \frac{\alpha_i^2}{2\sigma_e^2\tau^2}\right)$ distribution.
- Draw a random number $U \sim \text{Uniform}(0,1)$.
- Set

$${z_i}^{(l)} = \begin{cases} y^{(l)}, & \text{if } 0 < U < r({z_i}^{(l-1)}, y^{(l)}) \\ {z_i}^{(l-1)}, & \text{if } U > r({z_i}^{(l-1)}, y^{(l)}) \end{cases}$$

Then ${z_i}^{(l)}$ is the generated sample from the target distribution at the lth step. $r({z_i}^{(l-1)}, y^{(l)})$ is the acceptance ratio of the Metropolis-Hastings algorithm whose analytic form is given as follows

$$\begin{aligned} r({z_i}^{(l-1)}, y^{(l)}) &= \min\left\{1, \frac{g(y^{(l)})h({z_i}^{(l-1)})}{g({z_i}^{(l-1)})h(y^{(l)})}\right\} \\ &= \min\left\{1, \frac{\frac{1}{y^{(l)}\left(1+y^{(l)}\right)} \exp\left[-\frac{\alpha_i^2}{2\sigma_e^2\tau^2 y^{(l)}}\right] \frac{1}{\left(z^{(l-1)}\right)^2} \exp\left[-\frac{\alpha_i^2}{2\sigma_e^2\tau^2 z^{(l-1)}}\right]}{\frac{1}{z^{(l-1)}\left(1+z^{(l-1)}\right)} \exp\left[-\frac{\alpha_i^2}{2\sigma_e^2\tau^2 z^{(l-1)}}\right] \frac{1}{\left(y^{(l)}\right)^2} \exp\left[-\frac{\alpha_i^2}{2\sigma_e^2\tau^2 y^{(l)}}\right]}\right\} \\ &= \min\left\{1, \frac{1+{z_i}^{(l-1)}}{{z_i}^{(l-1)}} \frac{y^{(l)}}{1+y^{(l)}}\right\} \end{aligned}$$

At the lth step, we either draw a new random sample $y^{(l)}$ from the target distribution with probability $r({z_i}^{(l-1)}, y^{(l)})$ or retain the sample ${z_i}^{(k-1)}$ generated at the $(l-1)^{\text{th}}$ step with probability $1 - r({z_i}^{(l-1)}, y^{(l)})$. Then $\sqrt{{z_i}^{(l)}}$ is the random sample drawn from the full conditional posterior distribution of $\lambda_i|\cdot$ at the lth step.

2.2 Metropolis-Hastings algorithm for τ

The full conditional posterior distribution of τ (Equation 3) is given by

$$f(\tau|\cdot) \propto \frac{1}{\tau^m(1+\tau^2)} \exp\left[-\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2\tau^2}\right], \quad \tau > 0$$

Consider the transformation $Z = \tau^2$. Then the conditional posterior distribution of Z turns out to be

$$f(Z|\cdot) \propto \frac{1}{Z^{\frac{m+1}{2}}(1+Z)} \exp\left[-\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2 Z}\right], \quad Z > 0$$

Here the updation of the parameter τ is equivalent to update Z as they are related by one-to-one transformation $Z = \tau^2$. Let us denote the pdf of $Z|\cdot$ by $g(\cdot)$. We shall choose the candidate proposal distribution as $IG\left(\frac{m+1}{2}, \frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2}\right)$ and denote the corresponding pdf by $h(\cdot)$. As our proposal pdf approximates the target pdf, the independent Metropolis-Hastings algorithm updates Z in the MCMC.
Let $z^{(l-1)}$ is the generated sample from the target distribution $Z|\cdot$ at the $(l-1)^{\text{th}}$ step. We consider the following steps to draw a sample from the target distribution at the lth step.

- Generate a random sample $y^{(l)}$ from the $IG\left(\frac{m+1}{2}, \frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2}\right)$ distribution.
- Draw a random number $U \sim \text{Uniform}(0,1)$.
- Set

$$z^{(l)} = \begin{cases} y^{(l)}, & \text{if } 0 < U < r(z^{(l-1)}, y^{(l)}) \\ z^{(l-1)}, & \text{if } U > r(z^{(l-1)}, y^{(l)}) \end{cases}$$

Then $z^{(l)}$ is the generated sample from the target distribution at the lth step. $r(z^{(l-1)}, y^{(l)})$ is the acceptance ratio of the Metropolis-Hastings algorithm whose analytic form is given below

$$\begin{aligned} r(z^{(l-1)}, y^{(l)}) &= \min\left\{1, \frac{g(y^{(l)})h(z^{(l-1)})}{g(z^{(l-1)})h(y^{(l)})}\right\} \\ &= \min\left\{1, \frac{\frac{1}{(y^{(l)})^{\frac{m+1}{2}}(1+y^{(l)})} \exp\left[-\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2 y^{(l)}}\right] \frac{1}{(z^{(l-1)})^{\frac{m+1}{2}+1}} \exp\left[-\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2 z^{(l-1)}}\right]}{\frac{1}{(z^{(l-1)})^{\frac{m+1}{2}}(1+z^{(l-1)})} \exp\left[-\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2 z^{(l-1)}}\right] \frac{1}{(y^{(l)})^{\frac{m+1}{2}+1}} \exp\left[-\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2 y^{(l)}}\right]}\right\} \\ &= \min\left\{1, \frac{1+z^{(l-1)}}{z^{(l-1)}} \frac{y^{(l)}}{1+y^{(l)}}\right\} \end{aligned}$$

Then $\sqrt{z^{(l)}}$ is the random sample drawn from the full conditional posterior distribution of $\tau|\cdot$ at the lth step.

3 Derivation of the full conditional posterior distribution for multiple ancestry model

The hierarchical structure of our TWAS model in the multi-ancestry set-up is given as

$$
\begin{aligned}
\mathbf{x}_{\mathrm{k,ref}}|\boldsymbol{\alpha}_k, \sigma^2_{e_k} &\sim N_{n_k}(\mathbf{G}_{\mathrm{k,ref}}\boldsymbol{\alpha}_k, \sigma^2_{e_k}\mathrm{I}_{n_k}), \quad k = 1(1)K \\
\mathbf{y}_k|\beta, \boldsymbol{\alpha}_k, \sigma^2_\epsilon &\sim N_{N_k}(\beta\mathbf{G}_k\boldsymbol{\alpha}_k, \sigma^2_\epsilon\mathrm{I}_{N_k}) \\
\alpha_{kj}|\lambda_j, \tau, \sigma^2_{e_k} &\sim N(0, \sigma^2_{e_i}\tau^2\lambda^2_j), \quad k = 1(1)K,\ j = 1(1)m \\
\lambda_j &\sim C^+(0,1) \\
\tau &\sim C^+(0,1) \\
\beta|q, \sigma^2_d, \sigma^2_\epsilon &\sim (1-q)\delta_{\{0\}}(\beta) + qN(0, \sigma^2_d\sigma^2_\epsilon) \\
q &\sim Beta(b_1, b_2) \\
\sigma^2_d &\sim IG(d_1, d_2) \\
\sigma^2_{e_k} &\sim IG(e_1, e_2) \\
\sigma^2_\epsilon &\sim IG(\epsilon_1, \epsilon_2)
\end{aligned}
\tag{7}
$$

The notations in the Equation 7 have their usual meaning as described in the main text. For each $k = 1(1)K$, let us denote $\boldsymbol{\alpha}_k = (\alpha_{k1}, \alpha_{k2}, \cdots, \alpha_{km})^{\mathrm{T}}$. α_{kj} denotes the effect of jth cis-SNP on the gene expression levels in the kth ancestry for $j = 1(1)m$. We specify the SNP-specific shrinkage coefficients across all ancestries in the form of a vector as $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_m)^{\mathrm{T}}$. Let us define $\mathbf{D}_\lambda = diag(\lambda^2_1, \lambda^2_2, \cdots, \lambda^2_m)$, where $\mathbf{D}_\lambda$ is a $m \times m$ diagonal matrix. Then we have

$$\boldsymbol{\alpha}_k|\boldsymbol{\lambda}, \tau, \sigma^2_{e_k} \sim N_m(\mathbf{0}, \sigma^2_{e_i}\tau^2\mathbf{D}_\lambda)$$

In the next section, we derive the full conditional posterior distributions of the model parameters.

3.1 Full conditional distribution of $\boldsymbol{\alpha}_k$

The full conditional distribution of $\boldsymbol{\alpha}_k$ is given below

$$
\begin{aligned}
&f(\boldsymbol{\alpha}_k|\cdot) \propto f(\mathbf{x}_{\mathrm{k,ref}}|\boldsymbol{\alpha}_k, \sigma^2_{e_k})f(\mathbf{y}_k|\beta, \boldsymbol{\alpha}_k, \sigma^2_\epsilon)f(\boldsymbol{\alpha}_k|\boldsymbol{\lambda}, \tau, \sigma^2_{e_k}) \\
&\propto \exp\left[-\frac{1}{2\sigma^2_{e_k}}(\mathbf{x}_{\mathrm{k,ref}} - \mathbf{G}_{\mathrm{k,ref}}\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{x}_{\mathrm{k,ref}} - \mathbf{G}_{\mathrm{k,ref}}\boldsymbol{\alpha}_k)\right]\exp\left[-\frac{1}{2\sigma^2_\epsilon}(\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right]\exp\left[-\frac{\boldsymbol{\alpha}^{\mathrm{T}}_k\mathbf{D}^{-1}_\lambda\boldsymbol{\alpha}_k}{2\sigma^2_{e_k}\tau^2}\right] \\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\alpha}^{\mathrm{T}}_k\left(\frac{\mathbf{G}^{\mathrm{T}}_{\mathrm{k,ref}}\mathbf{G}_{\mathrm{k,ref}}}{\sigma^2_{e_k}} + \frac{\beta^2\mathbf{G}^{\mathrm{T}}_k\mathbf{G}_k}{\sigma^2_\epsilon} + \frac{\mathbf{D}^{-1}_\lambda}{\sigma^2_{e_k}\tau^2}\right)\boldsymbol{\alpha}_k - 2\boldsymbol{\alpha}^{\mathrm{T}}_k\left(\frac{\mathbf{G}^{\mathrm{T}}_{\mathrm{k,ref}}\mathbf{x}_{\mathrm{k,ref}}}{\sigma^2_{e_k}} + \frac{\beta\mathbf{G}^{\mathrm{T}}_k\mathbf{y}_k}{\sigma^2_\epsilon}\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\alpha}^{\mathrm{T}}_k\boldsymbol{\Sigma}_k\boldsymbol{\alpha}_k - 2\boldsymbol{\alpha}^{\mathrm{T}}_k\left(\frac{\mathbf{G}^{\mathrm{T}}_{\mathrm{k,ref}}\mathbf{x}_{\mathrm{k,ref}}}{\sigma^2_{e_k}} + \frac{\beta\mathbf{G}^{\mathrm{T}}_k\mathbf{y}_k}{\sigma^2_\epsilon}\right)\right]\right\}, \quad \text{where } \boldsymbol{\Sigma}_k = \frac{\mathbf{G}^{\mathrm{T}}_{\mathrm{k,ref}}\mathbf{G}_{\mathrm{k,ref}}}{\sigma^2_{e_k}} + \frac{\beta^2\mathbf{G}^{\mathrm{T}}_k\mathbf{G}_k}{\sigma^2_\epsilon} + \frac{\mathbf{D}^{-1}_\lambda}{\sigma^2_{e_k}\tau^2}
\end{aligned}
$$

Hence, the full conditional distribution of $\boldsymbol{\alpha}_k$ is given by

$$\boldsymbol{\alpha}_k|\cdot \sim N_m\left(\boldsymbol{\Sigma}^{-1}_k\left(\frac{\mathbf{G}^{\mathrm{T}}_{\mathrm{k,ref}}\mathbf{x}_{\mathrm{k,ref}}}{\sigma^2_{e_k}} + \frac{\beta\mathbf{G}^{\mathrm{T}}_k\mathbf{y}_k}{\sigma^2_\epsilon}\right), \boldsymbol{\Sigma}^{-1}_k\right), \quad k = 1(1)K$$

3.2 Full conditional distribution of $\boldsymbol{\lambda}$

As $\boldsymbol{\lambda}$ is a m dimensional vector, it is simpler to continuously update components of $\boldsymbol{\lambda}$ in the Gibbs sampler. For each $j = 1(1)m$, we have

$$\begin{aligned} f(\lambda_j|\cdot) &\propto \left\{\prod_{k=1}^{K} f(\alpha_{kj}|\lambda_j, \tau, \sigma_e^2)\right\} f(\lambda_j) \\ &\propto \left[\prod_{k=1}^{K}\left\{\frac{1}{\sigma_{e_k}\lambda_j\tau}\ \exp\left[-\frac{\alpha_{kj}^2}{2\sigma_{e_k}^2\tau^2\lambda_j^2}\right]\right\}\right]\left\{\frac{1}{1+\lambda_j^2}\right\} \\ &\propto \frac{1}{\lambda_j^K(1+\lambda_j^2)}\ \exp\left[-\frac{1}{2\tau^2\lambda_j^2}\sum_{k=1}^{K}\frac{\alpha_{kj}^2}{\sigma_{e_k}^2}\right] \end{aligned} \tag{8}$$

As the full conditional posterior distribution of λ_j is not conditionally conjugate, we use the Metropolis-Hastings algorithm to draw a random sample.

3.3 Full conditional distribution of τ

The full conditional distribution of τ is given as

$$\begin{aligned} f(\tau|\cdot) &\propto \left\{f(\boldsymbol{\alpha}_k|\boldsymbol{\lambda}, \tau, \sigma_{e_k}^2)\right\} f(\tau) \\ &\propto \left\{\prod_{k=1}^{K}\frac{1}{\tau^m}\ \exp\left[-\frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}_k}{2\sigma_{e_k}^2\tau^2}\right]\right\}\left\{\frac{1}{1+\tau^2}\right\} \\ &\propto \frac{1}{\tau^{mK}(1+\tau^2)}\ \exp\left[-\frac{1}{2\tau^2}\sum_{k=1}^{K}\frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}_k}{\sigma_{e_k}^2}\right] \end{aligned} \tag{9}$$

As the posterior distribution of τ is not conditionally conjugate, we update the parameter τ by the Metropolis-Hastings algorithm within the Gibbs sampler.

3.4 Full conditional distribution of $\sigma_{e_k}^2$

The full conditional distribution of $\sigma_{e_k}^2$ is given by

$$\begin{aligned} f(\sigma_{e_k}^2|\cdot) &\propto f(\mathbf{x}_{\mathrm{k,ref}}|\boldsymbol{\alpha}_k, \sigma_{e_k}^2) f(\boldsymbol{\alpha}_k|\boldsymbol{\lambda}, \tau, \sigma_{e_k}^2) f(\sigma_{e_k}^2) \\ &\propto \left\{(\sigma_{e_k}^2)^{-\frac{n_k}{2}}\exp\left[-\frac{1}{2\sigma_{e_k}^2}(\mathbf{x}_{\mathrm{k,ref}} - \mathbf{G}_{\mathrm{k,ref}}\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{x}_{\mathrm{k,ref}} - \mathbf{G}_{\mathrm{k,ref}}\boldsymbol{\alpha}_k)\right]\right\}\left\{(\sigma_{e_k}^2)^{-\frac{m}{2}}\exp\left[-\frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}_k}{2\sigma_{e_k}^2\tau^2}\right]\right\} \\ &\quad \left\{(\sigma_{e_k}^2)^{-(e_1+1)}\exp\left[-\frac{e_2}{\sigma_{e_k}^2}\right]\right\} \\ &\propto \left(\sigma_{e_k}^2\right)^{-\left(e_1+\frac{n_k}{2}+\frac{m}{2}+1\right)}\exp\left[-\frac{1}{\sigma_{e_k}^2}\left(e_2 + \frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}_k}{2\tau^2} + \frac{1}{2}(\mathbf{x}_{\mathrm{k,ref}} - \mathbf{G}_{\mathrm{k,ref}}\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{x}_{\mathrm{k,ref}} - \mathbf{G}_{\mathrm{k,ref}}\boldsymbol{\alpha}_k)\right)\right] \end{aligned}$$

The full conditional distribution of $\sigma_{e_k}^2$ is given by

$$\sigma_{e_k}^2|\cdot \sim IG\left(e_1 + \frac{n_k}{2} + \frac{m}{2},\ e_2 + \frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}_k}{2\tau^2} + \frac{1}{2}(\mathbf{x}_{\mathrm{k,ref}} - \mathbf{G}_{\mathrm{k,ref}}\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{x}_{\mathrm{k,ref}} - \mathbf{G}_{\mathrm{k,ref}}\boldsymbol{\alpha}_k)\right), \quad k = 1(1)K$$

3.5 Full conditional distribution of β

The full conditional distribution of β is given by

$$
\begin{aligned}
f(\beta|\cdot) &\propto \left\{\prod_{k=1}^{K} f(\mathbf{y}_k|\boldsymbol{\alpha}_k,\beta,\sigma_\epsilon^2)\right\} f(\beta|q,\sigma_d^2,\sigma_\epsilon^2) \\
&\propto \left\{\prod_{k=1}^{K} \exp\left[-\frac{1}{2\sigma_\epsilon^2}(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right]\right\}\left\{(1-q)\delta_{\{0\}}(\beta)+\frac{q}{\sqrt{2\pi}\sigma_d\sigma_\epsilon}\ \exp\left[-\frac{\beta^2}{2\sigma_d^2\sigma_\epsilon^2}\right]\right\} \\
&\propto \exp\left[-\frac{1}{2\sigma_\epsilon^2}\sum_{k=1}^{K}\left(\beta^2\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{G}_k^{\mathrm{T}}\mathbf{G}_k\boldsymbol{\alpha}_k-2\beta\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{G}_k^{\mathrm{T}}\mathbf{y}_k\right)\right]\left\{(1-q)\delta_{\{0\}}(\beta)+\frac{q}{\sqrt{2\pi}\sigma_d\sigma_\epsilon}\ \exp\left[-\frac{\beta^2}{2\sigma_d^2\sigma_\epsilon^2}\right]\right\} \\
&\propto (1-q)\delta_{\{0\}}(\beta)+\frac{q}{\sqrt{2\pi}\sigma_d\sigma_\epsilon}\ \exp\left[-\frac{\sum_{k=1}^{K}\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{G}_k^{\mathrm{T}}\mathbf{G}_k\boldsymbol{\alpha}_k+\frac{1}{\sigma_d^2}}{2\sigma_\epsilon^2}\left(\beta^2-2\beta\ \frac{\sum_{k=1}^{K}\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{G}_k^{\mathrm{T}}\mathbf{y}_k}{\sum_{k=1}^{K}\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{G}_k^{\mathrm{T}}\mathbf{G}_k\boldsymbol{\alpha}_k+\frac{1}{\sigma_d^2}}\right)\right] \\
&\propto (1-q)\delta_{\{0\}}(\beta)+\frac{q}{\sqrt{2\pi}\sigma_d\sigma_\epsilon}\ \exp\left[-\frac{1}{2\widetilde{\sigma}^2}(\beta^2-2\beta\widetilde{\mu})\right] \\
&\propto (1-q)\delta_{\{0\}}(\beta)+\left(\frac{q\widetilde{\sigma}}{\sigma_d\sigma_\epsilon}\exp\left[\frac{\widetilde{\mu}^2}{2\widetilde{\sigma}^2}\right]\right)\left\{\frac{1}{\sqrt{2\pi}\widetilde{\sigma}}\exp\left[-\frac{1}{2\widetilde{\sigma}^2}(\beta-\widetilde{\mu})^2\right]\right\}, \\
&\text{where}\quad \widetilde{\mu}=\frac{\sum_{k=1}^{K}\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{G}_k^{\mathrm{T}}\mathbf{y}_k}{\sum_{k=1}^{K}\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{G}_k^{\mathrm{T}}\mathbf{G}_k\boldsymbol{\alpha}_k+\frac{1}{\sigma_d^2}}\quad\text{and}\quad \widetilde{\sigma}^2=\frac{\sigma_\epsilon^2}{\sum_{k=1}^{K}\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{G}_k^{\mathrm{T}}\mathbf{G}_k\boldsymbol{\alpha}_k+\frac{1}{\sigma_d^2}}
\end{aligned}
$$

Hence, the full conditional distribution of β is a mixture of point mass function $\delta_{\{0\}}(\beta)$ and normal distribution $N(\widetilde{\mu},\widetilde{\sigma}^2)$. Let us denote the weight attached to the mixing component point mass function by

$$
v=\frac{1-q}{1-q+\frac{q\widetilde{\sigma}}{\sigma_d\sigma_\epsilon}\exp\left(\frac{\widetilde{\mu}^2}{2\widetilde{\sigma}^2}\right)} \tag{10}
$$

Hence, the conditional posterior distribution of β turns out to be

$$
\beta|\cdot\sim\begin{cases}\delta_{\{0\}}(\beta), & \text{with probability}\quad v\\ N(\widetilde{\mu},\widetilde{\sigma}^2), & \text{with probability}\quad 1-v\end{cases}
$$

3.6 Full conditional distribution of σ_ϵ^2

Let us denote $N = \sum_{k=1}^K N_k$. The full conditional distribution of σ_ϵ^2 is given as

$$
\begin{aligned}
f(\sigma_\epsilon^2|\cdot) &\propto \left\{\prod_{k=1}^K f(\mathbf{y}_k|\boldsymbol{\alpha}_k, \beta, \sigma_\epsilon^2)\right\} f(\beta|q, \sigma_d^2, \sigma_\epsilon^2) f(\sigma_\epsilon^2) \\
&\propto \left\{\prod_{k=1}^K (\sigma_\epsilon^2)^{-\frac{N_k}{2}} \exp\left[-\frac{1}{2\sigma_\epsilon^2}(\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right]\right\} \left\{(1-q)\delta_{\{0\}}(\beta) + \frac{q}{\sqrt{2\pi}\sigma_d\sigma_\epsilon} \exp\left[-\frac{\beta^2}{2\sigma_d^2\sigma_\epsilon^2}\right]\right\} \\
&\quad \left\{(\sigma_\epsilon^2)^{-(\epsilon_1+1)} \exp\left[-\frac{\epsilon_2}{\sigma_\epsilon^2}\right]\right\} \\
&\propto \left\{(1-q)\delta_{\{0\}}(\beta)(\sigma_\epsilon^2)^{-(\frac{N}{2}+\epsilon_1+1)} \exp\left[-\frac{1}{\sigma_\epsilon^2}\left(\epsilon_2 + \frac{1}{2}\sum_{k=1}^K (\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)\right] + \right. \\
&\quad \left. \frac{q}{\sqrt{2\pi}\sigma_d}(\sigma_\epsilon^2)^{-(\frac{N+1}{2}+\epsilon_1+1)} \exp\left[-\frac{1}{\sigma_\epsilon^2}\left(\epsilon_2 + \frac{\beta^2}{2\sigma_d^2} + \frac{1}{2}\sum_{k=1}^K (\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)\right]\right\} \\
&\propto \left\{w_1\ IG\left(\frac{N}{2} + \epsilon_1,\ \epsilon_2 + \frac{1}{2}\sum_{k=1}^K (\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right) + \right. \\
&\qquad \left. w_2\ IG\left(\frac{N+1}{2} + \epsilon_1,\ \epsilon_2 + \frac{\beta^2}{2\sigma_d^2} + \frac{1}{2}\sum_{k=1}^K (\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
w_1 &= \frac{(1-q)\delta_{\{0\}}(\beta)\Gamma(\frac{N}{2}+\epsilon_1)}{\left(\epsilon_2 + \frac{1}{2}\sum_{k=1}^K (\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)^{\frac{N}{2}+\epsilon_1}} \\
w_2 &= \frac{q\ \Gamma(\frac{N+1}{2}+\epsilon_1)}{\sqrt{2\pi}\sigma_d\left(\epsilon_2 + \frac{\beta^2}{2\sigma_d^2} + \frac{1}{2}\sum_{k=1}^K (\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)^{\frac{N+1}{2}+\epsilon_1}}
\end{aligned}
$$

Then the full conditional distribution of σ_ϵ^2 can be represented as a mixture of inverse gamma distribution as

$$
\sigma_\epsilon^2|\cdot \sim \begin{cases} IG\left(\frac{N}{2} + \epsilon_1,\ \epsilon_2 + \frac{1}{2}\sum_{k=1}^K (\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right) & \text{with probability } \frac{w_1}{w_1+w_2} \\ IG\left(\frac{N+1}{2} + \epsilon_1,\ \epsilon_2 + \frac{\beta^2}{2\sigma_d^2} + \frac{1}{2}\sum_{k=1}^K (\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k - \beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right) & \text{with probability } \frac{w_2}{w_1+w_2} \end{cases}
$$

The full conditional posterior distributions of q and σ_d^2 remain the same as we have mentioned in Equation 5 and Equation 6.

4 Metropolis-Hastings algorithm for shrinkage parameters

The shrinkage parameters of our multi-ancestry model are $\boldsymbol{\lambda}$ and τ. As the conditional posterior distributions of shrinkage parameters lack conditional conjugacy, we can not generate random samples from these distributions directly. Hence, we use the Metropolis-Hastings algorithm to generate random numbers from these conditional distributions.

4.1 Metropolis-Hastings algorithm for $\boldsymbol{\lambda}$

We shall derive the Metropolis-Hasting algorithm for λ_j, which is the jth component of the vector $\boldsymbol{\lambda}$. The derivation for the other components can be done similarly. The full conditional posterior distribution of λ_j

(Equation 8) is given as follows

$$f(\lambda_j|\cdot) \propto \frac{1}{\lambda_j^K(1+\lambda_j^2)} \exp\left[-\frac{1}{2\tau^2\lambda_j^2}\sum_{k=1}^{K}\frac{\alpha_{kj}^2}{\sigma_{e_k}^2}\right], \quad \lambda_j > 0$$

Consider the transformation $Z_j = \lambda_j^2$. Then the conditional posterior distribution of Z_j turns out to be

$$f(Z_j|\cdot) \propto \frac{1}{Z_j^{\frac{K+1}{2}}(1+Z_j)} \exp\left[-\frac{1}{2\tau^2 Z_j}\sum_{k=1}^{K}\frac{\alpha_{kj}^2}{\sigma_{e_k}^2}\right], \quad Z_j > 0$$

Here the updation of the parameter λ_j is equivalent to update Z_j as they are related by one-to-one transformation $Z_j = \lambda_j^2$. Let us denote the probability density function (pdf) of $Z_i|\cdot$ by $g(\cdot)$. We shall choose the candidate proposal distribution as $IG\left(\frac{K+1}{2}, \frac{1}{2\tau^2}\sum_{k=1}^{K}\frac{\alpha_{kj}^2}{\sigma_{e_k}^2}\right)$ and denote the corresponding pdf by $h(\cdot)$. We follow the same procedure to generate a sample from the target distribution as discussed in the section 2.1. The expression of the acceptance ratio in the Metropolis-Hastings algorithm turns out to be

$$r({z_j}^{(l-1)}, y^{(l)}) = \min\left\{1, \frac{g(y^{(l)})h({z_j}^{(l-1)})}{g({z_j}^{(l-1)})h(y^{(l)})}\right\} = \min\left\{1, \frac{1+{z_j}^{(l-1)}}{{z_j}^{(l-1)}}\frac{y^{(l)}}{1+y^{(l)}}\right\}, \tag{11}$$

where ${z_j}^{(l-1)}$ is the generated sample from the target distribution at the $(l-1)^{\text{th}}$ step and $y^{(l)}$ is the sample from the proposal distribution at the lth step.

4.2 Metropolis-Hastings algorithm for τ

The full conditional posterior distribution of τ (Equation 9) is given by

$$f(\tau|\cdot) \propto \frac{1}{\tau^{mK}(1+\tau^2)} \exp\left[-\frac{1}{2\tau^2}\sum_{k=1}^{K}\frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}_k}{\sigma_{e_k}^2}\right], \quad \tau > 0$$

Consider the transformation $Z = \tau^2$. Then the conditional posterior distribution of Z turns out to be

$$f(Z|\cdot) \propto \frac{1}{Z^{\frac{mK+1}{2}}(1+Z)} \exp\left[-\frac{1}{2Z}\sum_{k=1}^{K}\frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}_k}{\sigma_{e_k}^2}\right], \quad Z > 0$$

Here the updation of the parameter τ is equivalent to update Z as they are related by one-to-one transformation $Z = \tau^2$. We shall choose the candidate proposal distribution as $IG\left(\frac{mK+1}{2}, \frac{1}{2}\sum_{k=1}^{K}\frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}_k}{\sigma_{e_k}^2}\right)$. The expression for Metropolis-Hastings acceptance ratio coincides with the term mentioned in Equation 11 and the necessary steps of the algorithm has already been discussed in section 2.2.

Supplemental Figures

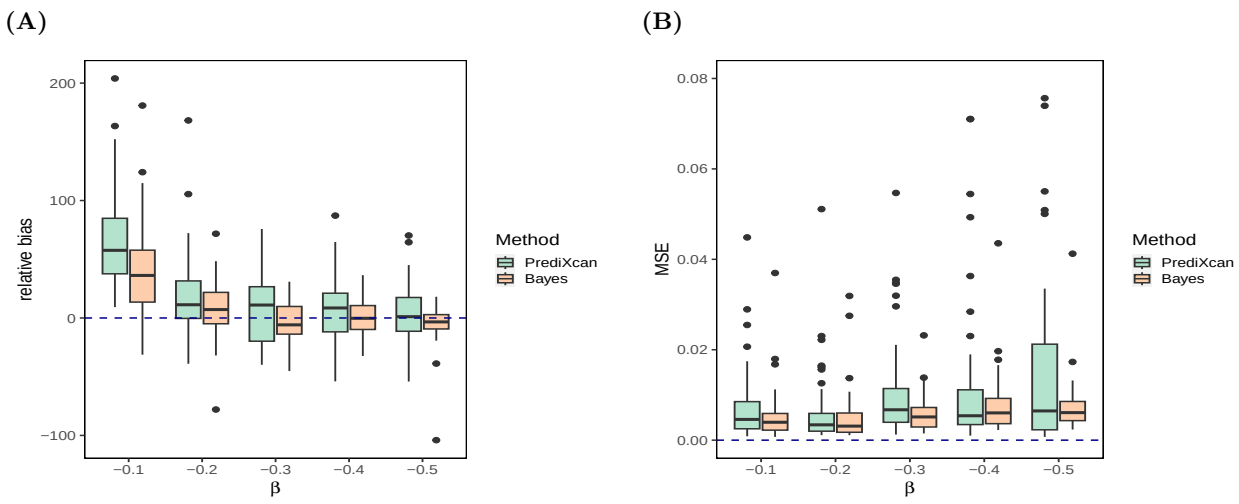

Figure S1: Comparison between PrediXcan and Bayesian method for (**A**) relative bias and (**B**) MSE for negative effect sizes of β. The reference panel sample size is $n_r = 500$ and the GWAS panel sample size is $n = 5000$. The heritability of the gene expression is fixed at $h_e^2 = 20\%$.

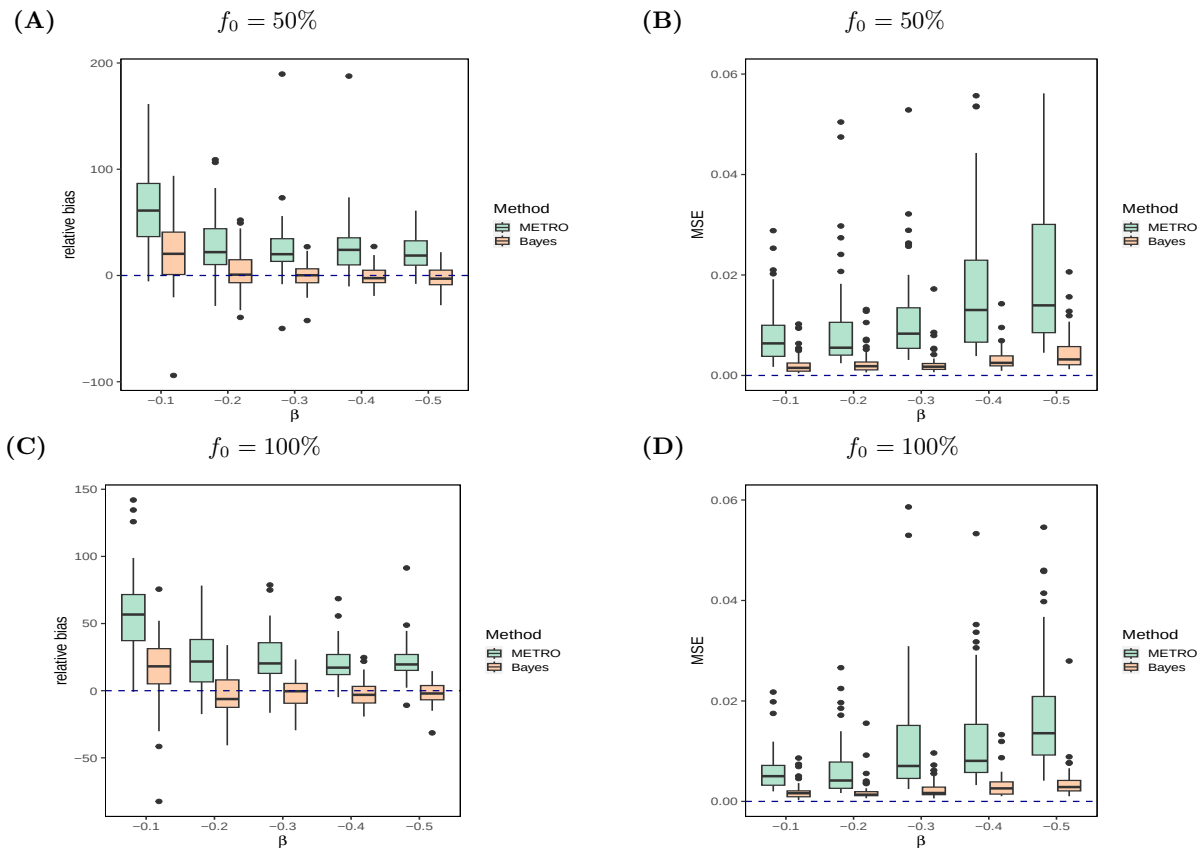

Figure S2: Comparison between METRO and Bayesian method for relative bias and MSE for different choices of $f_0 = 50\%$(Figure **A**-**B**), 100%(Figure **C**-**D**). The reference panel sample sizes are $n_1 = n_2 = 500$ and the GWAS sample sizes are $N_1 = N_2 = 5000$ respectively. The gene-expression heritability is fixed at $h_e^2 = 20\%$.