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

The transcriptome-wide association study (TWAS) has discovered novel gene-trait associations that offer essential biological interpretations. TWAS integrates reference panel transcriptome data and genome-wide association study (GWAS) data. In standard TWAS methods, a prediction model for gene expression is built utilizing the transcriptome data, which is employed to impute the gene expression in the GWAS data. The complex trait in GWAS is regressed on the predicted expression to identify gene-trait associations. Such a two-step approach ignores the uncertainty of the imputed expression and can lead to reduced inference accuracy. We develop a unified Bayesian method for TWAS to avoid a two-step approach, modeling the two datasets simultaneously. We consider the horseshoe prior, an established choice of global-local shrinkage prior, to model the relationship between gene expression and local SNPs and the spike and slab prior while testing for an association between the genetic component of expression and the trait. We extend our approach to conducting a multi-ancestry TWAS, focusing on discovering genes that affect the trait in all ancestries. Using extensive simulations, we show that our approach performs better than existing methods concerning the accuracy of effect size estimation and classifying non-null genes. We perform a single-ancestry and multi-ancestry TWAS for body mass index (BMI), integrating the Geuvadis transcriptome and UK Biobank GWAS data. Our approach identified 111 genes associated with BMI for the single ancestry TWAS of Europeans, which were enriched in the Interferon gamma signaling pathway at a nominal level. For a multi-ancestry TWAS of Africans and Europeans, the method discovered 19 genes that affect BMI in both ancestries, which were nominally enriched in the Molybdenum cofactor biosynthesis pathway.

Keywords: Gene-trait associations, Gene expression, Horseshoe prior, Spike and slab prior, Markov chain Monte Carlo

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 traits [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 relations with the trait. We must study intermediate molecular phenotypes and genetic variants together to understand better the biological mechanism through which these variants affect the trait. Gene expression is a crucial molecular phenotype that impacts a complex trait. Regulatory genetic variants, such as expression quantitative trait loci (eQTLs), govern the gene expression level [2] along with other factors. Frequent overlap between GWAS signals and eQTLs underscores the importance of combining GWAS and gene expression studies to reveal the mechanistic pathway from GWAS risk SNPs to the trait. Transcriptome-wide association study (TWAS) is a powerful approach to detecting novel gene-trait associations that address this objective.

TWAS integrates a reference transcriptome and GWAS data to infer the association between a gene and a trait. A reference transcriptome data contains the tissue or cell-type specific gene expression levels and genotypes of the gene's cis/local SNPs, i.e., nearby SNPs, for a set of individuals. GWAS data consists of the trait values and the genotypes for genome-wide SNPs, including the gene's local SNPs. However, the gene expressions are unavailable in the GWAS data. 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, hence a prediction model for the expression, are used to predict the genetically regulated component of gene expression in the GWAS data. Finally, we test for an association between the complex trait and the predicted genetically regulated component of expression (GReX) in the GWAS data. TWAS enjoys a few advantages over GWAS. First, TWAS is a gene-level test with less multiple-testing burden. Second, TWAS bridges the functional gap in GWAS by incorporating gene expression as a potential causal mediator from SNPs to outcome. However, the better performance of TWAS compared to GWAS in discovering novel genetic loci associated with a trait crucially depends on the local heritability of the gene expression, i.e., the proportion of variance of expression explained by the local SNPs. TWAS is expected to perform better than GWAS when the local heritability is higher [3].

Most TWAS methods [4, 5] employ the two-step strategy. For example, a penalized regression, such as Lasso [6], is implemented in the transcriptome data to obtain the prediction model of expression; a linear or logistic regression is applied to regress the outcome on the predicted GReX in the GWAS data. Such an approach disregards the uncertainty of the imputed GReX in the second step, which emerges due to the

variance of the estimated effect size vector of local SNPs in the transcriptome data. Lack of tuning of the noise in the predicted expression can reduce the overall accuracy of TWAS, e.g., inflation of type I or II error rates [7] or less precise estimation of the gene's effect size. A popular TWAS method, PrediXcan [8], employs the Elastic net penalized regression [9] to model the relationship between the expression and local SNPs in the transcriptome data. A recent method, BGW-TWAS [10], employed the Bayesian variable selection regression based on the spike and slab prior to model the eQTL effects. However, it also follows a two-step strategy without accounting for the uncertainty of the predicted expression. Another method, TIGAR [11], which employed the Bayesian regression based on the Dirichlet process in the first step, endures due to the same limitation. These approaches finally sorted to obtain the p-values of gene-trait association, even though BGW-TWAS and TIGAR employed Bayesian regressions in the first stage of the two-step TWAS. Instead of a two-step strategy, an integrated statistical framework that jointly models the two datasets simultaneously can overcome this limitation.

We develop a unified Bayesian approach to conduct TWAS, jointly modeling the transcriptome and GWAS data. Bayesian regularization methods consider various shrinkage priors. Global local shrinkage priors are an important class of such priors that shrink the noisy small effect sizes but leave the tail of the effect size distribution unreduced. To leverage the advantage of such efficient regularization approaches, we employ the horseshoe prior [12], a prominent choice of global-local shrinkage prior while building the relationship between the gene expression and local SNPs in the transcriptome data. We utilize the spike and slab prior [13] while testing for an association between the outcome and GReX in the GWAS data under the joint framework. The integrated 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 TWAS data and performs a comprehensive Bayesian inference on the gene-trait association. We obtain the Bayes factor and the related posterior probability for testing the association, which are natural measures in a Bayesian setup. The effect size of GReX on the outcome is simultaneously estimated based on the joint framework.

The double-exponential prior corresponding to Lasso [6] and the horseshoe prior (HSP) both shrink the small effects. But the latter refrains from shrinking the sizeable effects [14]. The double-exponential prior reduces the large effect sizes by a non-vanishing amount. 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 gene expression. Because eQTL discovery is not the goal here. Hence, we prioritize a continuous shrinkage by the HSP to robustly model the various types of sparsity present in the transcriptome data. The HSP should yield reliable expression prediction models, which is crucial for TWAS. We implement the classical spike and slab prior (SSP) with a point mass at zero to test for a gene-trait association explicitly [13].

We extended our model to incorporate multiple ancestries to enhance the genetic diversity in the TWAS analyses. The approach combines a pair of reference transcriptome and GWAS data from each ancestry to identify genes associated with the outcome in all ancestries. A multi-ancestry fine-mapping method, MA-FOCUS [15], leverages heterogeneous cross-ancestry TWAS data to prioritize the causal genes in a genetic

region associated with the trait. However, it adopted the traditional two-step TWAS procedure and did not calibrate the uncertainty of the imputed expressions. Another recent approach, METRO [16], proposed a unified approach to multi-ancestry TWAS while considering a joint likelihood-based framework to model the transcriptome and GWAS data together. The method considered independent normal distributions for the regression coefficients in the linear model of the gene expression treating local SNPs' genotypes as covariates in the transcriptome data, which is analogous to a penalty function considered in ridge regression. The ridge regression applies the same amount of shrinkage to all the local SNPs' effect sizes. On the contrary, the HSP underlying our approach differentially shrinks the effect sizes in a selective manner and is a more effective approach to regularization. METRO was developed to discover genes associated with the trait in at least one of the ancestries. We focus on a slightly different and biologically compelling question: which genes are associated with the outcome in all ancestries considered in the analysis? We develop the method following the assumption in MA-FOCUS that the genes shared across ancestries have the same effect on the trait. Likewise, in our single-ancestry model, we combine the HSP and the SSP in a single setup to model the multi-ancestry TWAS data. We refer to our method as glossTWAS (**g**lobal-**lo**cal **s**hrinkage and **s**pike and **s**lab prior-based **TWAS**).

We perform extensive simulation studies to compare glossTWAS with PrediXcan for a single-ancestry TWAS. We compare the methods concerning the accuracy of the effect size estimation and testing whether a gene is associated with the trait or not. Using simulations, we contrast glossTWAS with METRO for a multi-ancestry TWAS regarding the accuracy of estimation and testing. We apply the methods to conduct a single-ancestry and multi-ancestry TWAS for body mass index (BMI) while integrating the Geuvadis transcriptome and UK Biobank GWAS data. We perform the single-ancestry TWAS for Europeans and the multi-ancestry TWAS while combining the European and African ancestries. Since glossTWAS analyzes one gene at a time, we propose strategies to adjust for the multiple comparisons implemented for many genes by calibrating the hyperparameters of the priors.

2 Materials and methods

2.1 Single-ancestry TWAS

We first describe the method for a single ancestry. Consider a gene and its expression in the tissue or cell type relevant to the trait. Suppose the reference transcriptome data comprises n individuals. Let $\mathbf{x}_{\text{ref}}$ denote the vector of gene expression levels for the n individuals. We regard m local SNPs located around the gene's transcription start and end sites. Let $\mathbf{G}_{\text{ref}}$ denote $n \times m$ 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:

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

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 $m \times 1$ 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 Y denote a continuous trait/outcome and $\mathbf{y}$ represent the $N \times 1$ vector of trait values for the N individuals. Let $\mathbf{G}$ stand for the $N \times m$ genotype matrix for m SNPs in GWAS data, where these 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 consider another linear model:

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

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 models based on the transcriptome (Equation 1) and GWAS (Equation 2) data. These two datasets are assumed to be ancestry-matched. $\mathbf{G}\boldsymbol{\alpha}$ in the Equation 2 represents the true genetically regulated component of expression (GReX) for the GWAS individuals, and β denotes the effect of true GReX on the outcome. 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 $\boldsymbol{\epsilon}$ (Equations 1 and 2) are independently distributed. The above two linear models are connected by the common parameter vector $\boldsymbol{\alpha}$. We use the horseshoe prior for $\boldsymbol{\alpha}$ which has the following hierarchical structure:

$$\begin{aligned} \alpha_j | \lambda_j, \tau, \sigma_e^2 &\sim \mathrm{N}(0, \sigma_e^2 \tau^2 \lambda_j^2), \quad j = 1(1)m \\ \lambda_j &\sim \mathrm{C}^+(0,1) \\ \tau &\sim \mathrm{C}^+(0,1) \end{aligned}$$

Here, $\alpha_j, j = 1, \ldots, m$, are independently distributed as normal. $\lambda_j, j = 1, \ldots, m$, are also independently distributed as $\mathrm{C}^+(0,1)$, which 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 $j = 1, \ldots, m$, λ_j is the local shrinkage parameter specific to the effect of the jth local SNP (α_j), and τ is the global shrinkage parameter. While τ applies an overall shrinkage on the effects toward zero, the λ_js 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 for β (Equation 2), the effect of the GReX on the trait, as:

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

$\delta_{\{0\}}(\beta)$ 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: $q \sim \text{Beta}(b_1, b_2)$. 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 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, 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:

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

Here, $\alpha_j, j = 1, \ldots, m$, are independently distributed in the prior. The same assumption is made for $\lambda_j, j = 1, \ldots, m$. The priors for q and the variance parameters are independent. We implement the Markov chain Monte Carlo (MCMC) algorithm to generate the posterior sample of the model parameters and perform a comprehensive Bayesian inference. We devise a Metropolis within Gibbs sampler to obtain the posterior sample. We provide a brief outline of the MCMC algorithm in Algorithm 1. The detailed derivations of the full conditional distributions are presented in the supplemental material.

2.2 Estimation and testing

In glossTWAS, β denotes the effect of GReX on the trait (Equation 2). We consider the posterior mean of β using the MCMC sample as the estimate of β and calculate the posterior standard deviation (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 and posterior probability of null association. Let D denote the complete data, including the gene expression and outcome in the transcriptome and GWAS data, respectively. 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 lower the value of PPNA, 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:

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

where $\left(\boldsymbol{\alpha}^{(t)}, q^{(t)}, \sigma_d^{2^{(t)}}, \sigma_\epsilon^{2^{(t)}}\right)$ denotes the tth posterior draw corresponding to parameters $(\boldsymbol{\alpha}, q, \sigma_d^2, \sigma_\epsilon^2)$ from the MCMC and M is the number of MCMC iterations. The final approximation of the PPNA in Equation 5 corresponds to its Rao-Blackwellised estimate [17]. The term $P\left[\beta = 0|\boldsymbol{\alpha}^{(t)}, q^{(t)}, \sigma_d^{2^{(t)}}, \sigma_\epsilon^{2^{(t)}}, D\right]$ evaluated at the tth step of MCMC is the full-conditional posterior probability that β takes value 0.

2.3 Hyperparamters

We discuss the choice of the hyperparameters 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 Bonferroni-type approach to adjust for multiple comparisons under this Bayesian setup while testing many genes in the real data application later.

Next, we discuss the choices of the hyperparameters for the inverse gamma priors of the variance terms. Following previous literature [18], 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 flat prior distributes the total mass over a wide range of parameter values. We choose the hyperparameters $d_1 = 1.003 \approx 1$ and $d_2 = 0.18$ for the prior of σ_d^2. This choice provides 90% prior probability that $\sigma_d^2 \in (0, 1.7)$. It is determined by a strategy that is based on our real data analyses

and existing literature, which will be discussed later in the real data application section. We note that a moderate sample size of the transcriptome data and a large sample size of the GWAS data reduce the impact of the specific choices of the hyperparameters on the Bayesian inference.

2.4 Multi-ancestry TWAS

In this section, we extend glossTWAS to a multi-ancestry setup. To enhance genetic diversity, we integrate expression and GWAS data from multiple ancestries. We assume that a pair of transcriptome and GWAS data are available for each of the 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. Let $\mathbf{G}_k$ stand for $N_k \times m$ local SNPs' genotype matrix in the GWAS data for k^{th} ancestry, $k = 1, \cdots, K$. 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:

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

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

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

We make the key assumption here that the effect of the GReX on the trait is the same across ancestries. β denotes the common effect of GReX on the outcome. The assumption of the same effect size is along the lines of a fixed-effects meta-analysis. A gene identified by this model is associated with the trait in all the ancestries. Since a single gene explains a tiny proportion of the variability of a complex trait, we assume a common error variance σ^2_ϵ across the ancestries. $\mathbf{G}_k\boldsymbol{\alpha}_k$ in the Equation 7 represents the true GReX in the kth ancestry. Therefore, we allow for the flexibility that local SNPs 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:

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

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 local SNPs' effect sizes over the populations. We also regard the same global shrinkage parameter τ over the ancestries. The same $\lambda_j, j = 1, \ldots, m$, and τ authorize 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 ancestries. 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:

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

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

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

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

2.5 Estimation and testing

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

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

where $(\boldsymbol{\alpha}_1^{(t)},\cdots,\boldsymbol{\alpha}_K^{(t)},q^{(t)},\sigma_d^{2^{(t)}},\sigma_\epsilon^{2^{(t)}})$ is the tth posterior draw from M MCMC iterations. The last line of the Equation 9 follows from the Rao-Blackwellization technique, and $P\left[\beta=0|\boldsymbol{\alpha}_1^{(t)},\cdots,\boldsymbol{\alpha}_K^{(t)},q^{(t)},\sigma_d^{2^{(t)}},\sigma_\epsilon^{2^{(t)}},D\right]$ denotes the posterior probability of β taking value 0 in the full conditional distribution evaluated at the tth step of MCMC.

2.6 MCMC algorithm for glossTWAS in single and multi-ancestry TWAS

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

Algorithm 1 Metropolis within Gibbs sampler 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 $\mathrm{N}_m\left(\boldsymbol{\Sigma}^{-1}\left(\frac{\mathbf{G}_{\mathrm{ref}}^{\mathrm{T}}\mathbf{x}_{\mathrm{ref}}}{\sigma_e^2}+\frac{\beta\mathbf{G}^{\mathrm{T}}\mathbf{y}}{\sigma_\epsilon^2}\right),\boldsymbol{\Sigma}^{-1}\right)$ distribution where $\boldsymbol{\Sigma}=\frac{\mathbf{G}_{\mathrm{ref}}^{\mathrm{T}}\mathbf{G}_{\mathrm{ref}}}{\sigma_e^2}+\frac{\beta^2\mathbf{G}^{\mathrm{T}}\mathbf{G}}{\sigma_\epsilon^2}+\frac{\mathbf{D}_\lambda^{-1}}{\sigma_e^2\tau^2}$.

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

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

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

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

9: The full conditional distribution of q is a mixture of two beta distributions. The probability corresponding to one such component $\mathrm{Beta}(b_1,b_2+1)$ 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 $U\sim\mathrm{Uniform}(0,1)$ distribution. **if** $U<w$, **then** draw a sample from $\mathrm{Beta}(b_1,b_2+1)$ distribution; **else** simulate from $\mathrm{Beta}(b_1+1,b_2)$ 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 $U\sim\mathrm{Uniform}(0,1)$ distribution. **if** $U<\frac{w_1}{w_1+w_2}$, **then** draw a sample from $\mathrm{IG}(d_1,d_2)$ distribution; **else** draw from $\mathrm{IG}\left(d_1+\frac{1}{2},\ d_2+\frac{\beta^2}{2\sigma_\epsilon^2}\right)$ 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 $w_1=\frac{(1-q)\delta_{\{0\}}(\beta)\Gamma(\frac{N}{2}+\epsilon_1)}{\left(\epsilon_2+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)^{\frac{N}{2}+\epsilon_1}}$ and $w_2=\frac{q\Gamma(\frac{N+1}{2}+\epsilon_1)}{\sqrt{2\pi}\sigma_d\left(\epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)^{\frac{N+1}{2}+\epsilon_1}}$ respectively. To update σ_ϵ^2, first simulate $U\sim\mathrm{Uniform}(0,1)$ distribution. **if** $U<\frac{w_1}{w_1+w_2}$, **then** simulate from $\mathrm{IG}\left(\frac{N}{2}+\epsilon_1,\ \epsilon_2+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)$ distribution; **else** generate a random sample from $\mathrm{IG}\left(\frac{N+1}{2}+\epsilon_1,\ \epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y}-\beta\mathbf{G}\boldsymbol{\alpha})\right)$ distribution.

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

For multi-ancestry TWAS, 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 $\boldsymbol{\alpha}_k|\boldsymbol{\lambda},\tau,\sigma_{e_k}^2\sim\mathrm{N}_m(\mathbf{0},\sigma_{e_k}^2\tau^2\mathbf{D}_\lambda)$. Let us denote $N_c=\sum_{k=1}^K N_k$, N_k being the GWAS sample size in k^{th} ancestry. The steps for the algorithm are as follows.

Algorithm 2 Metropolis-within Gibbs sampler 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 $\mathrm{N}_m\left(\boldsymbol{\Sigma}_k^{-1}\left(\frac{\mathbf{G}_{k,\mathrm{ref}}^{\mathrm{T}}\mathbf{x}_{k,\mathrm{ref}}}{\sigma_{e_k}^2} + \frac{\beta\mathbf{G}_k^{\mathrm{T}}\mathbf{y}_k}{\sigma_\epsilon^2}\right), \boldsymbol{\Sigma}_k^{-1}\right)$ distribution for each $k = 1(1)K$, where $\boldsymbol{\Sigma}_k = \frac{\mathbf{G}_{k,\mathrm{ref}}^{\mathrm{T}}\mathbf{G}_{k,\mathrm{ref}}}{\sigma_{e_k}^2} + \frac{\beta^2\mathbf{G}_k^{\mathrm{T}}\mathbf{G}_k}{\sigma_\epsilon^2} + \frac{\mathbf{D}_\lambda^{-1}}{\sigma_{e_k}^2\tau^2}$.

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

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

9: The full conditional distribution of q is a mixture of two beta distributions. The posterior probability corresponding to one such component $\mathrm{Beta}(b_1, b_2+1)$ 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 $U \sim \mathrm{Uniform}(0,1)$ distribution. **if** $U < w$, **then** draw a sample from $\mathrm{Beta}(b_1, b_2+1)$ distribution; **else** simulate from $\mathrm{Beta}(b_1+1, b_2)$ 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 $U \sim \mathrm{Uniform}(0,1)$ distribution. **if** $U < \frac{w_1}{w_1+w_2}$, **then** draw a sample from $\mathrm{IG}(d_1, d_2)$ distribution; **else** draw from $\mathrm{IG}\left(d_1 + \frac{1}{2},\ d_2 + \frac{\beta^2}{2\sigma_\epsilon^2}\right)$ 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 $w_1 = \frac{(1-q)\delta_{\{0\}}(\beta)\Gamma(\frac{N_c}{2}+\epsilon_1)}{\left(\epsilon_2+\frac{1}{2}\sum_{k=1}^K(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)^{\frac{N_c}{2}+\epsilon_1}}$ and $w_2 = \frac{q\ \Gamma(\frac{N_c+1}{2}+\epsilon_1)}{\sqrt{2\pi}\sigma_d\left(\epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}\sum_{k=1}^K(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)^{\frac{N_c+1}{2}+\epsilon_1}}$ respectively. First, draw a random sample from $\mathrm{U} \sim \mathrm{Uniform}(0,1)$ distribution. We update σ_ϵ^2 as follows. **if** $U < \frac{w_1}{w_1+w_2}$, **then** simulate from $\mathrm{IG}\left(\frac{N_c}{2}+\epsilon_1,\ \epsilon_2+\frac{1}{2}\sum_{k=1}^K(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)$ distribution; **else** generate a random sample from $\mathrm{IG}\left(\frac{N_c+1}{2}+\epsilon_1,\ \epsilon_2+\frac{\beta^2}{2\sigma_d^2}+\frac{1}{2}\sum_{k=1}^K(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)^{\mathrm{T}}(\mathbf{y}_k-\beta\mathbf{G}_k\boldsymbol{\alpha}_k)\right)$ distribution.

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

3 Simulation study

We evaluate the performance of our method, glossTWAS, through extensive simulation studies and compare it with selected existing methods. We regard the accuracy of effect size estimation and correctly identifying null and non-null genes as the two main aspects of a method's reliability. For single-ancestry TWAS, we compare glossTWAS with the well-established PrediXcan method [8]. For multi-ancestry TWAS, we contrast

glossTWAS with METRO [16]. Although METRO does not explicitly model GWAS data for multiple ancestries, comparisons between the two methods will demonstrate the usefulness of glossTWAS in jointly analyzing transcriptome and GWAS datasets from multiple populations. We consider two populations for multi-ancestry TWAS and combine the GWAS datasets to create a single group of GWAS individuals while implementing METRO.

3.1 Simulation setup

We consider the genetic region neighboring the DGCR8 gene located on chromosome 22. For single-ancestry TWAS, we chose 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 reside 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 i^{th} individual in the transcriptome data, we generate gene expression using the linear model $x_{i,\text{ref}} = \mathbf{g}_{i,\text{ref}}^{\text{T}}\boldsymbol{\alpha} + e_i, i = 1(1)n$. Here, $x_{i,\text{ref}}$ is the expression level, and $\mathbf{g}_{i,\text{ref}}$ is the vector of genotype values at the local SNPs for i^{th} individual. $\boldsymbol{\alpha}$ denotes 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 $\text{N}(0, 1-h_e^2)$, where h_e^2 denotes the heritability of 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 eQTLs affecting the expression. The non-zero elements of $\boldsymbol{\alpha}$ are generated from $\text{N}(0, \frac{h_e^2}{m_c})$. We assign zero value to the remaining components of $\boldsymbol{\alpha}$.

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 in the transcriptome data. Let $\mathbf{g}_i$ denote the genotype vector for the ith individual in the GWAS data, $i = 1(1)N$. Assuming that the effect of local SNPs on expression remains the same in both datasets, the true GReX for the ith GWAS individual is $\mathbf{g}_i^{\text{T}}\boldsymbol{\alpha}$. We simulate the trait values following the linear model: $y_i = \beta(\mathbf{g}_i^{\text{T}}\boldsymbol{\alpha}) + \epsilon_i$, where y_i is the value of the trait for i^{th} individual and β denotes the effect of true GReX on the trait. We simulate the random error term ϵ_i from $\text{N}(0, 1)$. We assume $\text{Var}(y_i) = \text{Var}(\epsilon_i) = 1$ because one gene is likely to explain a negligible proportion of the trait's variance. We choose different non-zero values of β to create various non-null simulation scenarios. We standardize the vectors of gene expression, GWAS trait, and the columns of genotype matrices in both transcriptome and GWAS panels.

For multi-ancestry TWAS, we follow the simulation setup described in Li et al. [16] who introduced

METRO. We consider two genetic ancestries. We select the Japanese and British populations from the 1000 genome data, which serve as the representative populations for East Asian Ancestry (EAS) and European Ancestry (EUR), respectively. We employ HAPGEN to simulate genotype data for the 200 local SNPs (MAF > 0.05) for the same DGCR8 gene for each ancestry. We implement the steps described above for a single-ancestry setup to simulate gene expression in separate transcriptome data for the two populations. We use the following linear model: $x_{ik,\text{ref}} = \mathbf{g}_{ik,\text{ref}}^{\text{T}}\boldsymbol{\alpha}_k + e_{ik}, \quad i = 1(1)n_k,\ k = 1(1)2$. The notations $x_{ik,\text{ref}}$ and $\mathbf{g}_{ik,\text{ref}}^{\text{T}}$ denote the gene expression level and genotype vector for i^{th} individual in the transcriptome data for k^{th} ancestry. $\boldsymbol{\alpha}_k$ is the vector of local SNPs' effects on expression, and e_{ik} is the random noise specific to k^{th} ancestry. We assume $\text{Var}(x_{ik,\text{ref}}) = 1$ and simulate e_{ik} from $\text{N}(0, 1 - h_e^2)$ where h_e^2 is the expression heritability. We consider 10% of the local SNPs to have a non-null effect on expression in each ancestry. The elements of $\boldsymbol{\alpha}_k$ that correspond to the non-null SNPs are simulated from $\text{N}(0, \frac{h_e^2}{m_c})$ where m_c stands for the number of causal SNPs. We regard a fixed proportion, f_0, of causal SNPs to be common between the two ancestries, leading to eQTL sharing or overlap. We simulate the complex trait for N_k GWAS individuals in k^{th} ancestry employing the linear model: $y_{ik} = \beta(\mathbf{g}_{ik}^{\text{T}}\boldsymbol{\alpha}_k) + \epsilon_{ik}, \quad i = 1(1)N_k$. The term $\mathbf{g}_{ik}^{\text{T}}\boldsymbol{\alpha}_k$ stands for the true GReX for i^{th} individual in k^{th} ancestry and β is the common effect of the GReX on the trait. We simulate ϵ_{ik} from $\text{N}(0, 1)$ since a single gene explains a negligible proportion of the trait's variance. We vary f_0 and β to create different simulation scenarios.

For single-ancestry TWAS, we choose the transcriptome data sample size as $n = 500$ and the GWAS data sample size as $N = 15000$. We consider both positive and negative values for the GReX effect size ($\beta = 0.1, 0.2, 0.3, 0.4, 0.5$ and $\beta = -0.1, -0.2, -0.3, -0.4, -0.5$). We fix the gene expression heritability, h_e^2, at 20%. We also separately vary n and N to create different simulation scenarios. We vary n from 500 to 1500 and N from 15000 to 30000 while keeping the other parameters fixed. In multi-ancestry TWAS, the sample sizes of the transcriptome data in EAS and EUR ancestries are set to be $n_1 = n_2 = 500$. The corresponding GWAS data sample sizes are chosen as $N_1 = N_2 = 10000$. We make two specific choices of f_0 as 0% and 50%. The choices for β remain the same as the single-ancestry TWAS setup. We choose $h_e^2 = 20\%$, and the expression heritability does not differ across ancestries. We estimate the various posterior summaries for glossTWAS based on a posterior sample of size 1000 obtained from 7000 MCMC iterations with a burn-in period of 2000 and a thinning period of 5. We compare glossTWAS with PrediXcan and METRO based on a subset of effective replications in a given simulation scenario. Specifically, we compute the p-value for testing $H_0 : \beta = 0$ for PrediXcan (METRO) in each replication. We include only those replications for which the p-values satisfy the Benjamini-Hochberg (BH) [20] false discovery rate (FDR) control with an FDR level of 5%. The effect size estimation is of interest only for those genes which initially show evidence of association. We consider a more powerful FDR correction here to consider as many replications as possible. We consider 300 such effective replications.

To assess the accuracy in correctly identifying the non-null genes, we compare glossTWAS with PrediXcan and METRO for single-ancestry and multi-ancestry TWAS, respectively. Following the same procedure

described in Coram et al. [21], we measure the accuracy 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, whereas TPR denotes the proportion of non-null genes found associated with the trait. Here we consider a set of 200 genes selected at random from chromosome 9. We assume 5% of these genes affect the trait. The effect sizes for the non-null genes are simulated from $\mathrm{N}(0, h_g^2)$, where h_g^2 denotes the trait heritability due to the GReX of the non-null genes. Using a method, we test each of the 200 genes for an association with the trait. We obtain the p-values from PrediXcan and METRO for single and multi-ancestry TWAS, respectively, and Bayes factor and PPNA from glossTWAS. The p-values are arranged in an increasing order of magnitude. For glossTWAS, the Bayes factors are arranged in a decreasing order. Then, we reject the null hypotheses for the correspondingly ordered genes sequentially based on varying choices of significance thresholds of a measure of association, e.g., Bayes factor. For such a sequence of large number of thresholds, we enumerate the actual number of correct and false rejections. For a method, the FPR and TPR values are calculated for each threshold and plotted to create a partial receiver operating characteristic (ROC) curve. To compare the two methods, we plot the TPRs for the same choice of the FPR for the methods. Note that the FPR values range in $(\frac{1}{200}, \frac{2}{200}, \ldots, 1)$. The area under the curve (AUC) is estimated for each method. A method with a higher AUC is preferred. We implement this strategy because the p-value and Bayes factor are not directly comparable. We consider four different combinations of (h_e^2, h_g^2) as $(10\%, 10\%), (10\%, 20\%), (20\%, 10\%), (20\%, 20\%)$ to create different simulation scenarios for the partial ROC curves.

3.2 Results

3.2.1 Single ancestry TWAS

We first discuss a summary of results obtained by our approach, glossTWAS, while estimating the effect of GReX on the outcome (β) in a single-ancestry TWAS. We find that the estimates based on the posterior mean and median of β are similar for varying choices of the effect size. For example, when $\beta = 0.1$, the posterior mean and median are both 0.09 with posterior standard deviation (SD) 0.03 (Table 1). In this case, the 95% central posterior interval (CPI) is $(0.04, 0.15)$. The method performs robustly for both positive and negative values of β. For $\beta = -0.1$, the posterior mean and median are -0.09 and -0.08 with the CPI as $(-0.15, -0.04)$. The width of the CPI and the posterior SD increase marginally as the absolute value of β increases (Table 1). The empirical average of $\log_{10}(\mathrm{BF})$ values across the simulation iterations increases steadily with the absolute value of β. For example, starting with a negative value of $\log_{10}(\mathrm{BF})$ as -1.1 for $\beta = 0$, we obtain the values as $3.9, 21.7, 52.6, 94.4$ for $\beta = 0.1, 0.2, 0.3, 0.4$, respectively (Table 1).

Next, we compare glossTWAS with PrediXcan in terms of the estimation accuracy of the GReX effect size for single-ancestry TWAS. We evaluate the estimates' relative bias and mean squared errors (MSE). The relative bias of β estimate is defined as $\frac{\hat{\beta}-\beta}{\beta}$, where $\hat{\beta}$ is the estimate of β. For a small non-null effect,

Table 1: Summary of GReX effect sizes on outcome estimated by glossTWAS for single-ancestry TWAS.

β	mean	SD	median	95% CPI	$\log_{10}$(BF)
0.5	0.49	0.06	0.48	(0.39, 0.60)	145.1
0.4	0.40	0.05	0.39	(0.31, 0.49)	94.4
0.3	0.30	0.04	0.30	(0.23, 0.38)	52.6
0.2	0.21	0.03	0.20	(0.14, 0.27)	21.7
0.1	0.09	0.03	0.09	(0.04, 0.15)	3.9
0	0	0.01	0	$(-1.1 \times 10^{-3}, 4.5 \times 10^{-6})$	-1.1
-0.1	-0.09	0.03	-0.08	(-0.15, -0.04)	3.6
-0.2	-0.21	0.04	-0.20	(-0.27, -0.14)	20.3
-0.3	-0.30	0.04	-0.30	(-0.38, -0.23)	49.2
-0.4	-0.40	0.05	-0.39	(-0.49, -0.31)	94.9
-0.5	-0.50	0.06	-0.49	(-0.61, -0.39)	145.6

Note: The abbreviations SD, CPI, and BF represent the standard deviation, central posterior interval, and Bayes factor, respectively. The transcriptome and GWAS data sample sizes are $n = 500$ and $N = 15000$, respectively. The local heritability of gene expression is fixed at $h_e^2 = 20\%$. The results are obtained based on 300 effective replications in each simulation scenario.

$\beta = 0.1$, glossTWAS marginally underestimates it with a mean estimate of 0.09 (Table S1). PrediXcan slightly overestimates it with a mean estimate of 0.11. The mean relative bias is -9% (SD = 39%) for the former and 11% (SD = 30%) for the latter (Table S1). For a small effect, limited downward relative bias is reasonable. As β increases to $0.2, \ldots, 0.5$, glossTWAS provides nearly unbiased estimates with tiny relative biases (Table S1). For example, the mean estimates are the same as the true values when $\beta = 0.3, 0.4$. PrediXcan overestimates these βs with higher means of relative biases and higher SDs as well (Table S1). When $\beta = 0.4$, the mean relative bias is reduced to almost zero (-0.4%) for glossTWAS with SD 10%, whereas for PrediXcan, it is 6.5% with SD 23%. The relative bias of glossTWAS diminishes faster than PrediXcan as true β increases from 0.1 to 0.5 (Figure 1A). Regarding the MSE, glossTWAS produced lower values than PrediXcan except for $\beta = 0.1$ (Table S1). When $\beta = 0.1$, glossTWAS and PrediXcan produced a slightly different MSEs of 0.003 and 0.002, respectively. For $\beta = 0.5$, glossTWAS yielded an MSE of 0.006, and PrediXcan produced an estimate of 0.015 (Table S1). Contrary to relative bias, the MSEs increase with true β. The MSE of glossTWAS increases at a lower rate than PrediXcan (Figure 1B). We observe that the SD of β estimates obtained by glossTWAS are slightly larger than PrediXcan (Table S1). For the negative choices of βs, we observe a similar comparative performance of the two approaches (Table S1 and Figure S2). For the null effect of GReX, i.e., when $\beta = 0$, glossTWAS produces substantially lower bias ($\hat{\beta} - \beta$) and MSE than PrediXcan (Figure S1). Overall, glossTWAS estimates the effect size of the GReX on the trait more precisely than PrediXcan.

Next, we vary the sample sizes of the reference transcriptome (n) and GWAS (N) data and assess its impact on the estimation accuracy. Here, we fix $\beta = 0.2$. First, we vary n, keeping $N = 15000$ unchanged. GlossTWAS provides accurate estimation, whereas, PrediXcan consistently overestimates the true β for

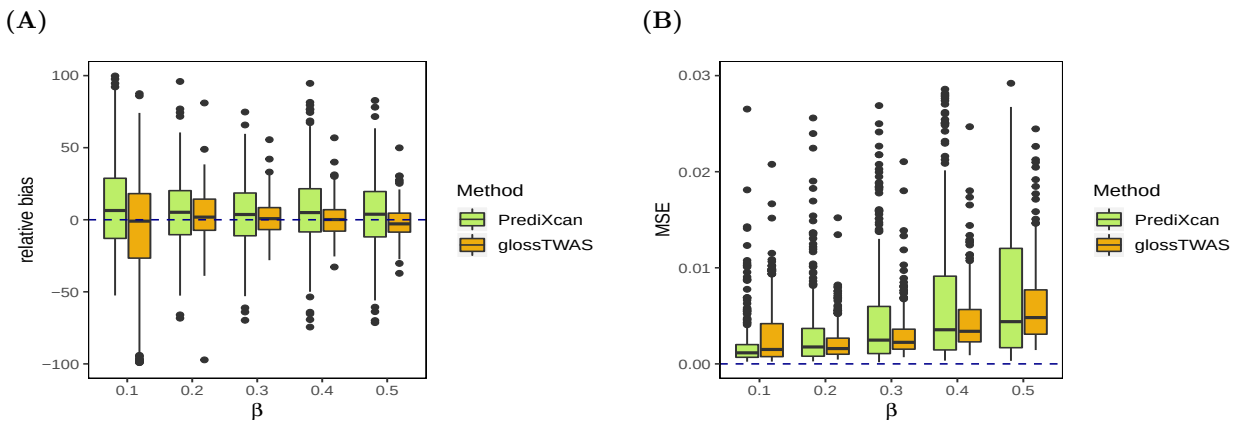

Figure 1: Comparison of TWAS effect size estimation accuracy by PrediXcan and glossTWAS for single-ancestry TWAS. Two measures are considered: (**A**) relative bias and (**B**) MSE. The transcriptome and GWAS data sample sizes are $n = 500$ and $N = 15000$, respectively. The local heritability of gene expression is fixed at $h_e^2 = 20\%$.

various choices of n (Table S2). For $n = 700, 1100$ and 1300, the mean estimates are 0.2 and 0.22 for glossTWAS and PrediXcan, respectively (Table S2). For all choices except $n = 900$, the mean estimates for glossTWAS coincide with the true value. As expected, glossTWAS consistently produced lower mean relative biases in the range of -1% to 3% than PrediXcan with $8 - 11\%$ mean relative bias (Figure S3A, Table S2). The SDs of the relative biases are also smaller for glossTWAS. Due to better estimation accuracy, glossTWAS consistently led to smaller MSEs than PrediXcan as n increased from 500 to 1500 (Figure S3B). We then vary the GWAS data sample size N, keeping n fixed. GlossTWAS again estimates β more accurately than PrediXcan (Table S3). The relative bias and MSE are lower for glossTWAS (Figure S3C and S3D). For example, for $N = 30000$, the mean relative bias and the corresponding SD for glossTWAS are 2.2% and 12%; the respective estimates for PrediXcan are 6.4% and 23% (Table S3).

Next, we compare the methods regarding the accuracy of correctly classifying the null and non-null genes. In single-ancestry TWAS, the partial ROC curve for glossTWAS dominates the partial ROC curve for PrediXcan continually (Figure 2), i.e., for a given level of FPR, glossTWAS provides a better TPR. Consequently, glossTWAS yields a higher AUC than PrediXcan for various choices of GWAS sample size (N), expression heritability (h_e^2), and trait-level heritability (h_g^2) (Figure 2). For $N = 15000$, the AUC increases in a range of $3 - 12\%$. For example, the AUC for glossTWAS is 0.77, whereas the AUC for PrediXcan is 0.65 when $h_e^2 = 10\%$ and $h_g^2 = 20\%$ (Figure 2C). For $N = 30000$, glossTWAS leads to an AUC increase of $3 - 11\%$. For example, when $h_e^2 = 10\%$ and $h_g^2 = 10\%$, the AUC for glossTWAS is 0.74 compared to the AUC as 0.63 for PrediXcan (Figure 2E). We attribute the better performance of glossTWAS to its joint modeling of the GWAS and transcriptome data under the unified Bayesian framework. Increasing the GWAS sample size (N) leads to better accuracy for both methods (Figure 2E-2H). For a given N, both approaches produce higher AUCs when the expression and/or trait heritabilities increase.

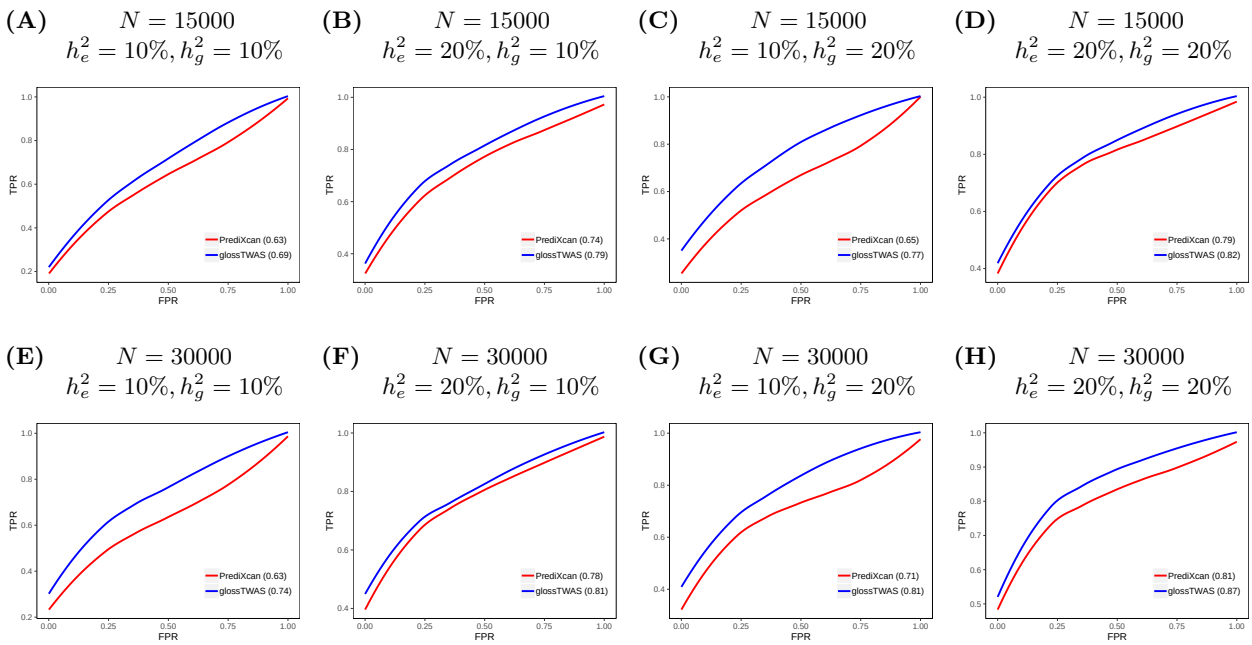

Figure 2: Partial receiver operating characteristic (ROC) curves for PrediXcan and glossTWAS to contrast their accuracy in correctly classifying the null and non-null genes. The choices of GWAS sample size (N), the local heritability of gene expression (h_e^2), and trait heritability (h_g^2) are varied. The transcriptome data sample size is fixed at $n = 500$. Each method's true positive rate (TPR) and false positive rate (FPR) values are calculated based on 50 replications in a simulation scenario. The area under the curve (AUC) is provided in the first bracket corresponding to each method's name in the plots.

3.2.2 Multi-ancestry TWAS

We first discuss a brief summary of the estimated GReX effect sizes (β) for our approach when half of the eQTLs are shared between the two ancestries, i.e., $f_0 = 50\%$. The true β is the same between the ancestries. GlossTWAS accurately estimates the positive and negative GReX effect sizes. For $\beta = 0.5$ and -0.5, the mean estimates for glossTWAS are 0.49 and -0.49, respectively (Table 2). The SDs increase marginally as we increase the magnitude of β. For $\beta = 0.1, 0.2, 0.3$, the SDs for glossTWAS are 0.03; whereas for $\beta = 0.4, 0.5$, it is 0.04 (Table 2). The posterior mean and median estimates for β are similar. The width of 95% CPI increases with an increase in the magnitude of β due to higher SDs. For example, when $\beta = 0.1$, the 95% CPI is $(0.05, 0.14)$. For $\beta = 0.5$, the 95% CPI, $(0.41, 0.57)$, has a longer width (Table 2). GlossTWAS accurately identifies the null effect of GReX on the trait, $\beta = 0$, with a mean $\log_{10}(\text{BF}) = -1.1$. For the non-null choices of β, $\log_{10}(\text{BF})$ increases with the magnitude of β. For $\beta = -0.3, -0.4, -0.5$, the values of $\log_{10}(\text{BF})$ are $68.3, 130.6, 195.7$, respectively (Table 2). When eQTLs do not overlap between ancestries with $f_0 = 0\%$, glossTWAS provides a similar posterior summary (Table S4) for different choices of β, as observed when $f_0 = 50\%$ (Table 2).

GlossTWAS estimates the GReX effect sizes more accurately than METRO. The former provides accurate estimates for different choices of β, whereas the latter overestimates it. We first discuss the results when there is 50% eQTL overlap between the ancestries, i.e., $f_0 = 50\%$. For example, when $\beta = 0.1$, the mean estimate

Table 2: Summary of the estimate of GReX effect sizes obtained by glossTWAS in multi-ancestry TWAS. Here, half of the eQTLs are shared between the two ancestries, i.e., $f_0 = 50\%$.

β	mean	SD	median	95% CPI	$\log_{10}$(BF)
0.5	0.49	0.04	0.49	(0.41, 0.57)	186.9
0.4	0.40	0.04	0.39	(0.33, 0.47)	123.8
0.3	0.30	0.03	0.30	(0.25, 0.36)	70.8
0.2	0.20	0.03	0.20	(0.16, 0.25)	27.2
0.1	0.10	0.03	0.10	(0.05, 0.14)	5.4
0	0	0.003	0	$(-9.7 \times 10^{-5}, 1.4 \times 10^{-4})$	-1.1
-0.1	-0.10	0.02	-0.10	(-0.14, -0.06)	5.8
-0.2	-0.20	0.03	-0.20	(-0.25, -0.15)	27.6
-0.3	-0.30	0.03	-0.30	(-0.36, -0.24)	68.3
-0.4	-0.40	0.04	-0.39	(-0.46, -0.33)	130.6
-0.5	-0.49	0.04	-0.49	(-0.57, -0.41)	195.7

Note: The abbreviations SD, CPI, and BF represent the standard deviation, central posterior interval, and Bayes factor, respectively. The transcriptome data sample sizes are $n_1 = n_2 = 500$, and GWAS data sample sizes are $N_1 = N_2 = 10000$ for the two ancestries. The local heritability of gene expression is fixed at $h_e^2 = 20\%$ in each ancestry. The results are obtained based on 300 effective replications in a simulation scenario.

is 0.1 for glossTWAS, and 0.13 for METRO (Table S6). Similarly, for higher choices of $\beta = 0.2, 0.3, 0.4$, the mean estimates for glossTWAS coincide with the true values (Table S6). Overall, the mean relative biases are low for glossTWAS varying in -2% to 2% (Figure 3C, Table S6). METRO invariably overestimates the effect sizes with higher mean relative biases ranging in $23 - 29\%$ with larger SDs than glossTWAS as well (Figure 3C, Table S6). For example, when $\beta = 0.5$, METRO overestimates it as 0.63 compared to 0.49 by glossTWAS. The upward bias increases for METRO as the true β increases. We observe a similar pattern for both positive and negative choices of β (Table S6). The SDs of the estimated effect sizes are also smaller for glossTWAS. We observe a very similar comparative performance of the approaches when there is no overlap in the ancestry-specific eQTLs, i.e., $f_0 = 0\%$ (Figure 3A, Table S5). Due to low biases and SDs, the MSEs for glossTWAS are substantially smaller than METRO (Figure 3B, Figure 3D). For example, when $\beta = 0.5$ with $f_0 = 50\%$, the MSE estimates are 0.003 and 0.023 for glossTWAS and METRO, respectively (Table S6). We observe that the relative bias for glossTWAS decreases rapidly as we increase β from 0.1 to 0.5 (Figure 3A, Figure 3C). The SD of the relative biases reduces faster than METRO (Table S5 and Table S6). It also happens for negative βs as their magnitudes increase. Finally, glossTWAS accurately estimates the null effects (Figure S4). Thus, our method provides robust estimation accuracy across different simulation scenarios.

Concerning the correct classification of null and non-null genes, glossTWAS performs overall better than METRO. In various simulation scenarios, where the proportion of eQTLs shared between ancestries (f_0), expression heritability (h_e^2), and trait heritability (h_g^2) vary, the partial ROC curves for glossTWAS consistently dominate those for METRO (Figure 4). For most simulation scenarios, glossTWAS yields an increase in AUC of $2 - 5\%$. In two settings, the methods led to the same AUC (Figure 4E, 4G) where the

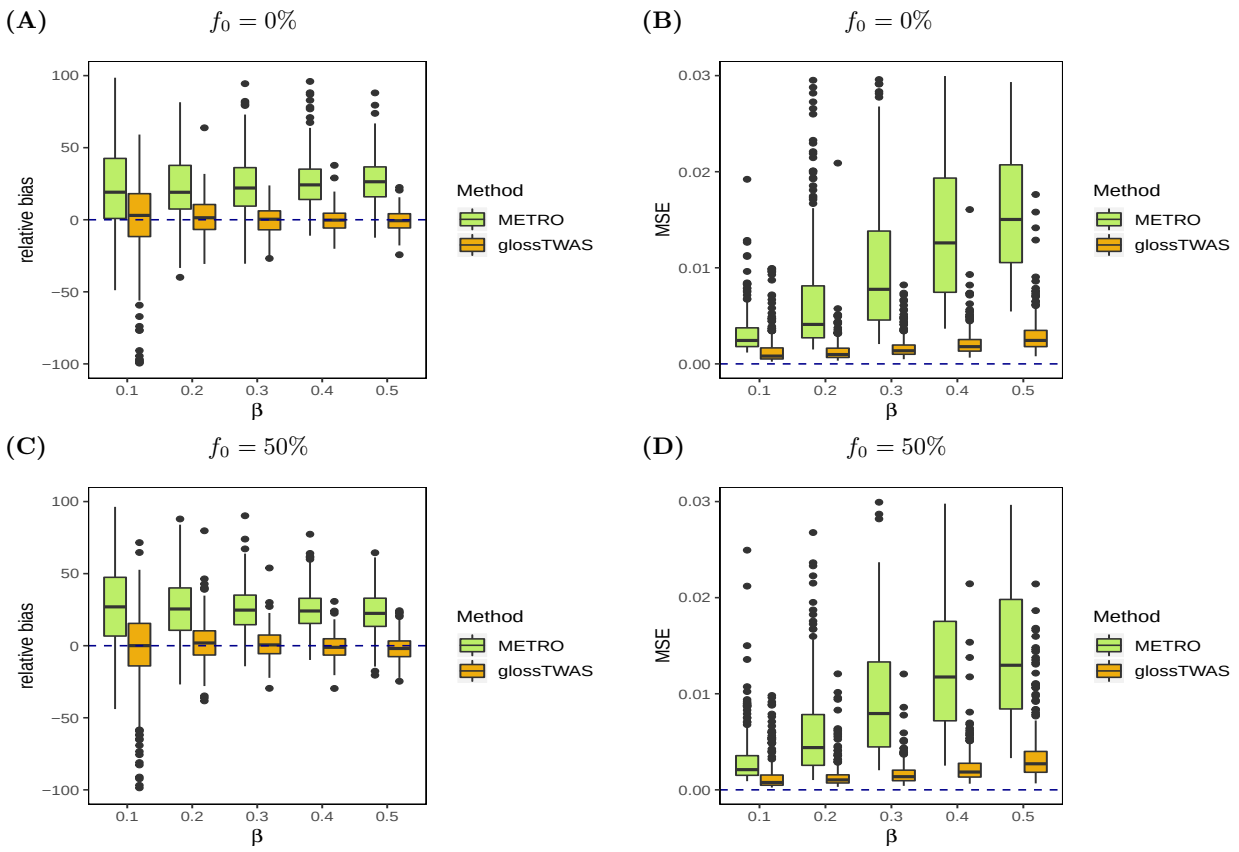

Figure 3: Comparison of TWAS effect size estimation accuracy by METRO and glossTWAS for multi-ancestry TWAS. Two measures are considered: relative bias and MSE. Both the scenarios of no sharing of eQTLs, $f_0 = 0\%$ (Figure **A**-**B**), and sharing of half of the eQTLs, $f_0 = 50\%$ (Figure **C**-**D**), between the two ancestries are considered. The transcriptome data sample sizes are $n_1 = n_2 = 500$, and GWAS data sample sizes are $N_1 = N_2 = 10000$ for the two ancestries. The local heritability of gene expression is fixed at $h_e^2 = 20\%$ in each ancestry.

ROC curves overlap. In general, the AUC increases for each method as we increase the expression heritability and/or trait heritability while keeping the other parameters fixed. For example when $f_0 = 50\%, h_e^2 = 20\%$, and $h_g^2 = 10\%$, the AUCs are 0.81 and 0.77 for glossTWAS and METRO, respectively (Figure 4F). As we increase h_g^2 from 10% to 20%, keeping f_0 and h_e^2 fixed, the AUCs increase to 0.87 and 0.84 for glossTWAS and METRO, respectively (Figure 4H). Overall, glossTWAS classifies the null and non-null genes more accurately than METRO.

3.3 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. We obtained the reference transcriptome data from the Geuvadis expression study [22] and the GWAS data from the UK Biobank (UKB). RNA sequencing was performed for lymphoblastoid cell lines in the Geuvadis data comprising 462

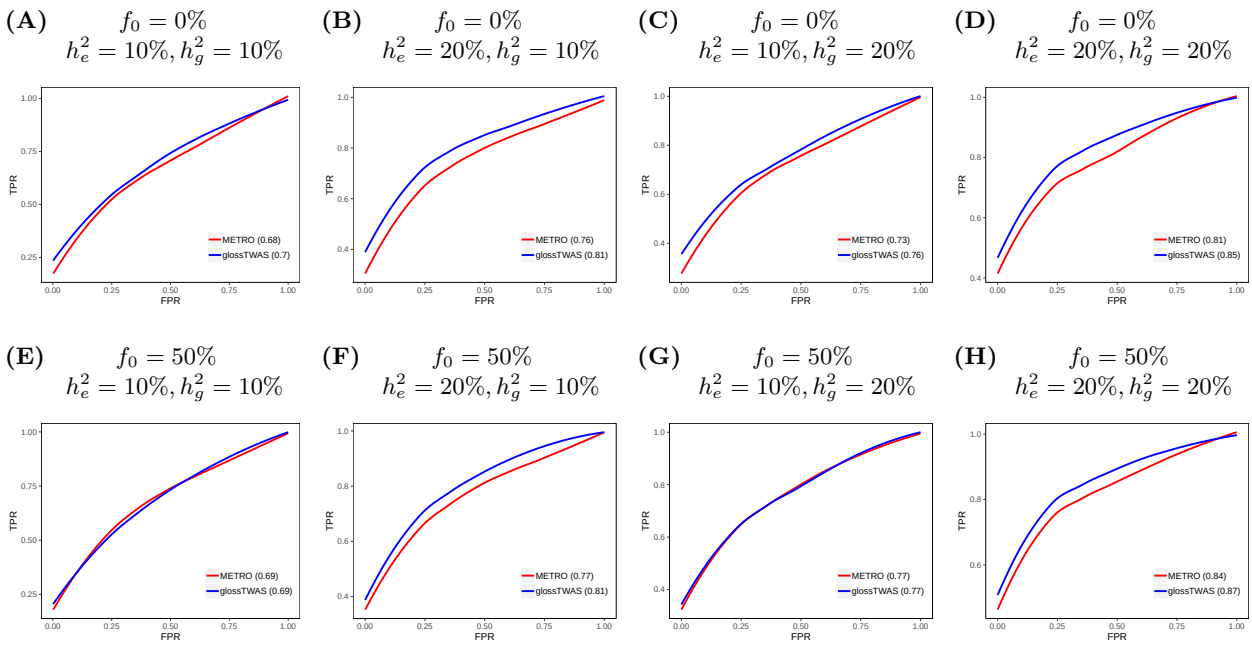

Figure 4: Partial receiver operating characteristic (ROC) curves for METRO and glossTWAS to contrast their accuracy in correctly classifying a null and non-null gene. The choices of the proportion of eQTLs shared between the two ancestries, f_0, local heritability of gene expression (h_e^2), and trait heritability (h_g^2), are varied. The transcriptome data sample sizes are $n_1 = n_2 = 500$, and GWAS data sample sizes are $N_1 = N_2 = 10000$ for the two ancestries. Each method's true positive rate (TPR) and false positive rate (FPR) values are calculated based on 50 replications in a simulation scenario. The area under the curve (AUC) is provided in the first bracket corresponding to each method's name in the plots.

individuals from the 1000 Genomes Project. The mRNA expression levels were measured for these individuals who belong 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 removed 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 [23] to impute the missing expression values for a gene. We annotated the protein-coding genes from the GENCODE (release v44) and held on to a final set of $12,517$ genes, common with the Geuvadis 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 retain good-quality SNPs. We implemented the following QC steps for the genotype data of EUR and

AFR ancestries separately. 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 removed SNPs with a genotype missingness rate of more than 10%.

We obtained the GWAS genotype and trait data for BMI in the UKB. BMI is measured in the units of Kg/m^2 in the UK Biobank. We regarded two populations, White British (WB) and Black British (BB), to constitute GWAS data for EUR and AFR ancestries. We removed a portion of GWAS individuals in either ancestry who did not meet certain criteria. Specifically, we removed those individuals whose self-reported sex did not match the genetically determined sex, who were not involved in the computation of genetic PCs available in the UKB cloud computing platform, and who had a genetic relatedness to any other participants. We finally had $287,801$ unrelated individuals from the WB population and $6,426$ from the BB population. Using multiple linear regression, BMI was adjusted for the following covariates: age, sex, and top 10 genetic PCs separately for WB and BB 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 Geuvadis 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 genes on the autosomal chromosomes only. We standardized the complex trait, gene expression, and the columns of genotype matrices both in transcriptome and GWAS data. 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. Thus, we integrate the EUR Geuvadis data with WB UKB data to conduct the single-ancestry TWAS; we combine the EUR and AFR Geuvadis data with WB and BB UKB data for the multi-ancestry TWAS.

We administered the GCTA method [25] to identify the locally heritable genes, that is, the genes for which the local SNPs significantly explain the variability of the expression in Geuvadis data. We compute the p-value for evaluating the null hypothesis for each gene, implying that the proportion of variance of gene expression explained by the local SNPs is zero. GCTA implements the restricted maximum likelihood (REML) approach to estimate this proportion and performs the likelihood ratio test (LRT) to obtain the p-value. For the EUR ancestry, we implemented the Benjamini-Hochberg (BH) false discovery rate (FDR) controlling procedure [20] based on the GCTA p-values to detect the significantly heritable genes. Using an FDR level 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 GCTA p-value < 0.05. The main reason behind identifying such a small number of heritable genes is the low sample size (89) of the transcriptome data in the AFR ancestry. At the nominal level, we found 932 heritable genes in AFR ancestry, whereas only 50 genes by controlling the FDR < 0.05. To analyze more genes in the multi-ancestry TWAS, we considered the nominal significance level to identify the heritable genes in both ancestries. We restricted the downstream analyses only to these locally heritable genes.

To adjust for testing multiple hypotheses, we applied the Bonferroni correction to the p-values obtained

by PrediXcan and METRO to control the family-wise error rate (FWER) at 5% for all the 12,517 genes analyzed. To perform Bonferroni style multiplicity adjustment in our Bayesian framework, we adopted the approach introduced by Westfall et al. [26]. Since the null hypothesis is defined as no association between the gene and the trait, we calibrate the hyperparameters in the prior distribution of q, the prior probability that a gene is associated with the trait. For r genes, the ith null hypothesis H_{0i} states that the ith gene has no effect on the trait, $i = 1(1)r$. Let π_0 denote the probability that all of the null hypotheses are true, i.e., no gene is associated with the trait; $\pi_0 = P[H_{0i} \text{ is true } \forall i = 1, \ldots, r]$. Assuming that the genes are independent apriori, we tune the prior probability that H_{0i} is true as $\pi_0^{1/r}$, which is a Bonferroni-type calibration. We note that this is a conservative adjustment similar to the Bonferroni correction in the classical setting. The posterior probability of null association (PPNA) for a gene is now compared with a nominal threshold, e.g., 0.05, to identify the genes significantly associated with the outcome at a genome-wide scale. Thus, the calibration of hyperparameters in glossTWAS induces an approximately Bonferroni-fashion adjustment in the Bayesian setting [26]. For a fixed choice of π_0, $\pi_0^{1/r}$ gets close to one for a large r, hence the adjustment for the multiple comparisons. We need to choose a sensible value of π_0, the probability that no gene is associated with the trait. Recent genetic studies have discovered numerous genes relevant for a complex trait. Thus, it is less plausible that no gene is associated with the outcome. Hence, we chose π_0 to be 0.1 in our analysis, which induces the prior knowledge that at least one gene is associated with the outcome with high probability of 0.9 and permits the alternative scenario at the same time with a small yet substantial probability of 0.1. Since we have 12517 protein-coding genes, the choice of $r = 12,517$ and $\pi_0 = 0.1$ leads to $\pi_0^{1/r} = 0.9998$. The prior probability of a gene being null is 0.9998. A gene affects the outcome with a tiny probability in the prior; the penalty for the multiple comparisons is induced in this way. To identify the significantly associated genes, we employed the threshold of PPNA as 0.05 for each gene. We obtained the corresponding threshold of $\log_{10}(\text{BF})$. We used this choice both for single and multi-ancestry TWAS.

We choose the hyperparameters of the prior distribution for q and σ_d^2 before applying glossTWAS on real data. We have $q \sim \text{Beta}(b_1, b_2)$ and $\sigma_d^2 \sim \text{IG}(d_1, d_2)$. According to the Bonferroni-style adjustment above, we set the prior probability that a gene is null as 0.9998, that is, $\frac{b_2}{b_1+b_2} = 0.9998$. We fix $b_1 = 1$, and we obtain $b_2 = 5435$ by solving this equation. To determine the hyperparameters for the prior of σ_d^2, we adopt the method of moments approach considered by Majumdar et al. [27]. Specifically, we equate $m_2^{'} = 1 + \frac{b_1}{b_1+b_2}\frac{d_2}{d_1-1}$, where $m_2^{'}$ is the second-order sample raw moment of the Z (Wald) statistic for PrediXcan across all genes evaluated for the trait. The right side of the equality is $E(Z^2)$ under a spike and slab prior assigned to a standardized effect size (multiplied by the square root of the GWAS sample size) of the gene on the trait [27]. We apply the PrediXcan to the real data for BMI in our single ancestry TWAS and calculate $m_2^{'}$ as 3.77. Along with this equation, we further restrict the hyperparameters d_1 and d_2 such that the mode of the $\text{IG}(d_1, d_2)$ distribution coincides with the mode of a weakly informative $\text{IG}(0.1, 0.1)$ distribution. The mode of $\text{IG}(0.1, 0.1)$ is 0.09. These provide two equations to solve to find d_1 and d_2. We obtain $d_1 = 1.000012$ and $d_2 = 0.18$. Besides the method of moments approach applied to real data, we also considered the mode

of the weakly informative IG prior because it has been used in other recent studies [18]. We emphasized determining a robust choice of these hyperparameters, which can be employed in TWAS for other traits as well. This prior distribution provides the following prior probability: $P[0 < \sigma_d^2 < 1.7] = 0.9$. We used the same strategy to choose the hyperparameters in the simulation studies, where we considered $P(H_0) = 0.95$ for one gene tested; it led to a similar choice as $d_1 = 1.003, d_2 = 0.18$.

Out of 1491 heritable genes in EUR ancestry, PrediXcan could not construct the GReX prediction model for 24 genes. Even though these genes were found locally heritable by GCTA, PrediXcan estimated the effect sizes of all the local SNPs to be zero. On the other hand, glossTWAS analyzed all the 1491 genes. A possible reason is that the horseshoe prior in glossTWAS shrinks larger effect sizes minimally, hence modeling the relationship between expression and local SNPs more effectively. GlossTWAS identified 111 significant genes for BMI (Table S8). PrediXcan found 40 genes to be significantly associated with BMI (Table S7). Both methods detected a common set of 32 genes (Table S9). GlossTWAS identified 79 additional genes. We note that the Bonferroni adjustment for multiple comparisons in the frequentist framework of PrediXcan and the Bayesian framework of glossTWAS may not be entirely comparable. GlossTWAS identified the RBM6 gene (Table 3) on chromosome 3 to be most significant as it has the highest $\log_{10}(\text{BF}) = 38.4$ and the lowest $\text{PPNA} = 2.3 \times 10^{-35}$. The effect size was estimated to be 0.09, with a posterior SD of 0.02. PrediXcan found the UBA7 gene on chromosome 3 to have the strongest association signal with a p-value of 5×10^{-33} (Table S7). The GReX effect size estimate by PrediXcan was 0.21 with SD 0.02. The UBA7 gene was also identified by glossTWAS as the second strongest signal with $\log_{10}(\text{BF}) = 34.3$ and $\text{PPNA} = 2.6 \times 10^{-31}$ (Table 3). The corresponding estimate of GReX effect size is 0.17 with posterior SD 0.08 (Table 3). Among the genes detected by glossTWAS, 56 genes are positively associated, and 55 genes show a negative association.

Regarding the effect size estimation, in addition to identifying genes with higher effect sizes, glossTWAS also detected some genes with weaker effects. Specifically, for 78 genes, the GReX effect size magnitude is greater than 0.1, while for the remaining 33 genes, it is lower than 0.1 (Table S8). PrediXcan detected 35 genes out of 40 genes with an effect size estimate smaller than 0.1 (Table S7). For the genes found by both approaches, the effect size estimates vary to some extent (Table S9). For example, whereas for the gene CTSW on chromosome 11, we observe a minor change in effect size estimation between PrediXcan (Effect size = 0.07) and glossTWAS (Effect size = 0.08), a noticeable difference in the GReX estimation is observed for NR1H3 gene on chromosome 11. The GReX effect size estimates for the NR1H3 gene by PrediXcan and glossTWAS are -0.04 and -0.18 respectively. Based on our simulation studies, the effect size estimates by glossTWAS should be more accurate.

Out of 111 genes detected by glossTWAS, 59 genes have already been reported to be associated with BMI in the NHGRI-EBI GWAS catalog. The remaining genes are not communicated to be associated with BMI, hence, we regard these genes as novel (Table S12). However, many of the novel genes are associated with other phenotypes, some with phenotypes related to body measurements, e.g., height, weight, and waist-hip index. For example, the gene SMIM20 on chromosome 4 was found to be associated with body height [28] (Table

S12). While PrediXcan did not show any association for this gene (p-value = 0.17), glossTWAS flagged a significant association with $\log_{10}(\text{BF}) = 20.6$ and PPNA $= 1.3 \times 10^{-17}$ (Table S8). Another gene, CBX1, on chromosome 17, was previously reported to be associated with body fat rate [29]. PrediXcan missed this gene with a nominally significant p-value of 0.04. GlossTWAS showed an association with $\log_{10}(\text{BF}) = 11.7$ and PPNA $= 1.1 \times 10^{-8}$ (Table S8).

We performed pathway enrichment analysis for the list of genes identified by glossTWAS using the Reactome method [30, 31]. We found suggestive nominally significant (p-value < 0.01) enrichment for Interferon gamma signaling pathway (p-value $= 7.1 \times 10^{-4}$), RUNX1 regulates transcription of genes involved in BCR signaling pathway (p-value = 0.002), Nuclear Receptor transcription pathway (p-value = 0.006), and Nuclear Receptor transcription pathway (p-value = 0.009). However, the enrichment p-value for none of these pathways could pass the FDR correction. We experimented with a larger list of 210 genes obtained from glossTWAS based on a $\log_{10}(\text{BF}) \geq 2$. This gene list was enriched in the Interferon gamma signaling pathway with FDR-adjusted p-value = 0.04. Interferon gamma signaling has been communicated to play an important role in obesity and insulin resistance [32, 33].

Table 3: Results for 20 most significantly associated genes for BMI identified by glossTWAS in single-ancestry TWAS for Europeans.

Chr	Gene	Effect size	SD	PPNA	$\log_{10}$(BF)
3	RBM6	0.09	0.02	2.3×10^{-35}	38.4
3	UBA7	0.17	0.08	2.6×10^{-31}	34.3
4	MANBA	0.55	0.36	7×10^{-31}	33.9
4	GUF1	-0.67	0.5	8.5×10^{-26}	28.8
12	NCKAP5L	-0.13	0.05	5.1×10^{-22}	25
3	KLHDC8B	0.5	0.25	9.2×10^{-22}	24.8
11	THYN1	0.53	0.42	2.3×10^{-21}	24.4
16	MAPK3	0.12	0.05	4.1×10^{-21}	24.1
3	WDR6	0.44	0.24	2×10^{-20}	23.4
16	KCTD13	-0.09	0.02	2.9×10^{-20}	23.3
16	TBX6	-0.08	0.02	4.4×10^{-20}	23.1
11	ACAD8	0.31	0.42	5×10^{-20}	23
16	SGF29	0.27	0.15	6.9×10^{-20}	22.9
16	DDX19B	-0.26	0.16	7.3×10^{-20}	22.9
16	APOBR	-0.1	0.04	1.3×10^{-19}	22.6
16	SH2B1	-0.14	0.07	5.7×10^{-19}	22
3	NCKIPSD	0.33	0.19	1.3×10^{-18}	21.6
11	ACP2	-0.24	0.13	2.8×10^{-18}	21.3
16	NFATC2IP	-0.09	0.03	3.1×10^{-18}	21.2
16	CD19	-0.16	0.08	3.6×10^{-18}	21.2

Note: The abbreviations SD, PPNA, and BF represent the standard deviation, posterior probability of null association, and Bayes factor, respectively. The results are obtained based on 7000 MCMC iterations with a burn-in sample size of 2000 and a thinning period of 5.

In the multi-ancestry TWAS, glossTWAS found 19 genes significantly associated with BMI in both EUR and AFR ancestries (Table 4). While detecting genes associated with BMI in at least one of the ancestries,

METRO discovered 26 genes with a non-zero effect in both ancestries (Table S10). Thus, 26 or fewer genes with an association in both ancestries were spotted by METRO since it does not explicitly test for an association in both populations. We focussed on such genes detected by METRO because our primary goal is discovering genes associated with the outcome in both ancestries. Both methods detected a subset of eight common genes (Table S11). For example, the TBX6 gene on chromosome 16, reported to be associated with BMI by an earlier study [34], was identified by both METRO (p-value $= 7 \times 10^{-24}$) and glossTWAS ($\log_{10}(\text{BF}) = 22.8$ and PPNA $= 9.1 \times 10^{-20}$). However, METRO was developed to find the susceptibility genes in at least one of the ancestries. The method spotted a total of 64 genes, which contain the above 26 genes. The GReX effect size estimated by METRO is zero for 29 genes in EUR ancestry and for another 9 genes in AFR ancestry (Table S10). Among the 19 genes detected by glossTWAS, nine genes are positively associated with BMI. GlossTWAS obtained the strongest signal of association for the TAOK2 gene, located on chromosome 16, with $\log_{10}(\text{BF}) = 23.9$ and PPNA $= 7.5 \times 10^{-21}$ (Table 4). The effect of GReX for TAOK2 on BMI was estimated at -0.16 with posterior SD 0.07 (Table 4). METRO also found the TAOK2 gene to be the most significant with p-value $= 5.6 \times 10^{-24}$ and GReX effect size estimate -0.16 (Table S10).

Again, we performed pathway enrichment analysis employing the Reactome method for the set of 19 genes identified by glossTWAS for the multi-ancestry TWAS. The Molybdenum cofactor biosynthesis pathway appeared to be nominally enriched with a p-value 8×10^{-4}. The FDR-adjusted enrichment p-value, 0.068, marginally missed the threshold of 0.05. This pathway is not reported to be directly relevant for BMI or related traits. Instead, it participates in biological processes underlying human enzymes. Along with known genes for BMI, our method also identified novel genes (Table S13) in the multi-ancestry TWAS that are not reported in the NHGRI-EBI GWAS catalog. A few of these genes were communicated as being associated with phenotypes related to body measurements. For example, Kichaev et al. [35] found the MOCS1 gene on chromosome 6 associated with body height. While METRO did not produce evidence of association for this gene in both EUR and AFR ancestries, but only in EUR, glossTWAS confirms the association for both ancestries with $\log_{10}(\text{BF}) = 20.2$ and PPNA $= 3.7 \times 10^{-17}$. The GReX effect size estimate by glossTWAS is -0.11 with posterior SD 0.04 (Table 4). We note that employing a PPNA threshold of 0.05 led to a corresponding threshold of $\log_{10}(\text{BF})$ as 5 both for single and multi-ancestry TWAS.

4 Discussion

We have proposed a novel approach to conduct TWAS that circumvents a crucial methodological limitation inherent in a standard two-step TWAS approach. Traditional TWAS methods ignore the uncertainty of predicted expressions in the second stage of the analysis, which arises due to the variance of estimated local SNPs' effect sizes in the reference transcriptome data. Our proposed method, glossTWAS, overcomes this limitation by jointly modeling the transcriptome and GWAS data in a unified Bayesian framework. Instead of utilizing two unrelated models for transcriptomic and GWAS data, we implement a hierarchical

Table 4: Results for significantly associated genes for BMI identified by glossTWAS in the multi-ancestry TWAS for European and African ancestries.

Chr	Gene	Effect size	SD	PPNA	$\log_{10}$(BF)
1	ZCCHC17	0.12	0.09	3.5×10^{-2}	5.2
1	NAV1	-0.45	0.24	3.5×10^{-18}	21.2
3	DHFR2	-0.17	0.11	3.9×10^{-9}	12.1
5	RHOBTB3	0.14	0.08	1.6×10^{-5}	8.5
6	MOCS1	-0.11	0.04	3.7×10^{-17}	20.2
6	ZNF192P1	0.10	0.03	10×10^{-8}	10.7
6	ZSCAN26	0.09	0.04	2.3×10^{-4}	7.4
6	RNF5	0.04	0.01	2.3×10^{-7}	10.4
10	HPS1	0.33	0.2	3.4×10^{-7}	10.2
10	FBXL15	-0.16	0.11	4.3×10^{-7}	10.1
10	DPYSL4	-0.02	0.01	3.9×10^{-2}	5.1
10	TRIM8	-0.28	0.24	1×10^{-6}	9.7
11	ACP2	-0.29	0.25	3.6×10^{-18}	21.2
11	OVOL1	-0.33	0.37	1.1×10^{-13}	16.7
12	CFAP54	0.13	0.06	4.2×10^{-4}	7.1
12	HCAR3	0.42	0.39	1.5×10^{-10}	13.6
16	TBX6	-0.07	0.02	9.1×10^{-20}	22.8
16	TAOK2	-0.16	0.07	7.5×10^{-21}	23.9
17	CBX1	0.19	0.11	6.9×10^{-10}	12.9

Note: The abbreviations SD, PPNA, and BF represent the standard deviation, posterior probability of null association, and Bayes factor, respectively. The results are obtained based on 7000 MCMC iterations with a burn-in sample size of 2000 and a thinning period of 5.

structure where the unknown local SNPs' effect sizes connect the seemingly separate models. We perform a comprehensive Bayesian inference based on MCMC. The simultaneous updation of model parameters in the MCMC steps, including the local SNPs' effect sizes and the corresponding variances, accommodates the uncertainty of the predicted expression in the analysis. To model the sparse effects of local SNPs or eQTLs on gene expression in transcriptome data, we adopt a Bayesian regularization approach based on a popular global-local shrinkage prior, the horseshoe prior (HSP). Lesser shrinkage on the larger eQTL effects by the HSP is more desirable than the Laplace prior, corresponding to Lasso, which penalizes potent eQTL effects substantially. This property of HSP is expected to improve the overall performance of glossTWAS through improved GReX prediction. Concerning inference accuracy, it has been demonstrated that the HSP outperforms the double exponential prior, the Bayesian dual of Lasso [14]. A major reason for the HSP's better performance is its probability spike at zero and heavy tails, which help it model the sparsity and large effects robustly.

Using extensive simulations, we demonstrate that glossTWAS yields good estimation accuracy of GReX effect sizes. The method estimates the effect sizes more precisely than PrediXcan for a single-ancestry TWAS. The accurate GReX estimates would improve the polygenic transcriptomic risk score (PTRS) calculation for disease risk or quantitative trait prediction based on the TWAS framework [36]. While classifying the non-null genes, glossTWAS performed efficiently. In most simulation settings, the partial ROC curves for

glossTWAS dominate that for PrediXcan with a higher area under the curve (AUC). This implies that, given a false positive rate (FPR) choice, the true positive rate (TPR) is higher for glossTWAS than PrediXcan. We employed the partial ROC curve to compare the two methods since one produces p-values and the other produces Bayes factors to assess the strength of associations that are not directly comparable.

European populations largely over-represent the genetic studies of complex traits. Genetically diverse studies help elucidate complex traits' etiology better. Recent studies are emphasizing the need to increase the genetic diversity of the populations included in TWAS [15, 16]. We developed a multi-ancestry TWAS approach based on the integrated Bayesian framework. Instead of discovering genes associated with the outcome in at least one of the ancestries, we emphasize finding genes relevant for all ancestries, which is biologically intriguing. Local SNPs can affect gene expression differentially across ancestries due to the difference in allele frequency and LD pattern, as well as the heterogeneity in eQTL effects. At the same time, some similarity is expected due to the sharing of eQTL effects between the ancestries. We accommodate both possibilities in the glossTWAS model by choosing independently distributed HSPs for the ancestry-specific effects of local SNPs but with the same SNP-specific local shrinkage parameters and the global shrinkage parameter across ancestries. Like in a single-ancestry setup, our multi-ancestry method models the transcriptome and GWAS data in a unified Bayesian framework assuming that the datasets across ancestries are non-overlapping and independent. Following a fixed-effects framework, we assume that the GReX effect size for a gene is the same across ancestries. Effect sizes of a single gene are typically small for a complex trait. Following Lu et al. [15], we consider the same effect size to develop a parsimonious model and boost the power of detecting a gene associated with the outcome in all ancestries. Shi et al. [37] estimated an average trans-ethnic genetic correlation of 0.85 for East Asians and Europeans across various complex traits. Even though the correlation is not perfect, it indicates a strong concordance between population-specific genetic effects. Moreover, it is a variant-level correlation; gene-level effects can be more similar across ancestries.

Using simulations, we show that glossTWAS accurately estimates the GReX effect size for a multi-ancestry TWAS. The estimation by glossTWAS is more accurate than METRO [16] with lower relative bias and MSE in the simulation setting of equal effects in two ancestries. GlossTWAS performed robustly well for various extents of eQTL sharing between ancestries. The method classified the non-null genes with better accuracy than METRO. The AUC under the partial ROC curve obtained by glossTWAS is consistently higher. We note that METRO does not explicitly model separate GWAS data from the ancestries, contrary to glossTWAS. To implement METRO, we combined the two ancestry-specific GWAS data to create a single pool of GWAS individuals. The authors of the METRO paper [16] acknowledged that their method requires to be extended to the setting of ancestry-specific non-overlapping GWAS data. Despite this major difference between METRO and our setting, we mainly compared glossTWAS with this approach to validate our findings. We emphasize that ancestry-specific distinct GWAS studies are commonly conducted across global populations. Meta-analyses are widely performed to combine these genetic studies. Hence, our multi-ancestry TWAS setup is relevant and frequent in practice.

GlossTWAS performed well in identifying numerous gene-BMI associations in the single and multi-ancestry TWAS based on Geuvadis transcriptome and UK Biobank GWAS data. The set of genes identified in single-ancestry TWAS was nominally enriched in the Interferon gamma signaling pathway, which is known to regulate obesity. GlossTWAS also detected most of the genes identified by PrediXcan. PrediXcan could not construct the GReX prediction model for some genes found locally heritable by GCTA. However, glossTWAS successfully analyzed these genes. This indicates the effectiveness of the HSP in modeling the relationship between gene expression and local SNPs. In the multi-ancestry TWAS, glossTWAS identified 5% of the 369 genes to be associated with BMI in both ancestries, which is a significant proportion of these genes, found locally heritable in both the EUR and AFR populations. A recent study [27] found BMI to be 7% polygenic using the brain cerebellum tissue type for European ancestry. Even though we considered blood tissue type, the finding of a lower 5% significantly associated genes for BMI in both European and African ancestries does not contradict the previous finding [27]. We acknowledge that a small number of locally heritable genes could be analyzed in the multi-ancestry TWAS, which is mainly due to the smaller sample size of the AFR ancestry in the Geuvadis data. When analyzing many genes in the real data application, we carefully improvised a multiple comparison strategy by calibrating the relevant hyperparameters in the Bayesian model. The tuning of the hyperparameters uses the total number of genes in a Bonferroni fashion. This procedure led to a sensible choice of the threshold for PPNA, analogously a $\log_{10}(\text{BF})$ threshold of 5, which has been used in practice before [38]. We carefully determined the hyperparameters in the model based on our real data application and existing literature. Hence, the strategy has an empirical Bayes flavor. We observed that the results remain similar despite some changes in these hyperparameters, mainly due to moderate sample sizes of reference transcriptome data and large sample sizes of GWAS data.

We discuss a few limitations of glossTWAS and future directions of work. Our method performs the gene-trait association without adjusting for the horizontal pleiotropic effects of the local SNPs on the outcome. For some genes, the local SNPs can affect the trait without modulating the gene expression due to varying LD patterns in some segments of the genome and the polygenic nature of the trait [39, 40]. In future work, we plan to extend the model of glossTWAS to account for horizontal pleiotropy. We have developed the method based on individual-level data, which can be challenging to access for various traits. We plan to extend the method for summary-level GWAS association data. The method is currently developed to analyze continuous traits. We shall extend the method for case-control binary outcomes. We also plan to extend the work to jointly model multiple tissue or cell types since multiple tissue or cell types can be relevant for a complex trait.

Overall, glossTWAS is a novel unified Bayesian approach to performing TWAS. The method is based on Bayesian priors with attractive properties. It is an efficient and powerful approach for a single and multi-ancestry TWAS. The latter aims to discover genes associated with the outcome in all ancestries, which is biologically interesting. The method can be applied to find genes associated with various complex traits and comprehend their genomic basis better.

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.

Acknowledgment

This research used the UK Biobank resource under application 77327.

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] C. Cao, B. Ding, Q. Li, D. Kwok, J. Wu, and Q. Long, "Power analysis of transcriptome-wide association study: Implications for practical protocol choice," *PLoS genetics*, vol. 17, no. 2, p. e1009405, 2021.

[4] 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.

[5] J. Mai, M. Lu, Q. Gao, J. Zeng, and J. Xiao, "Transcriptome-wide association studies: recent advances in methods, applications and available databases," *Communications Biology*, vol. 6, no. 1, p. 899, 2023.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

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

[18] 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.

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

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[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] 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.

[27] A. Majumdar and B. Pasaniuc, "A bayesian method for estimating gene-level polygenicity under the framework of transcriptome-wide association study," *Statistics in Medicine*, vol. 42, no. 26, pp. 4867–4885, 2023.

[28] L. Yengo, S. Vedantam, E. Marouli, J. Sidorenko, E. Bartell, S. Sakaue, M. Graff, A. U. Eliasen, Y. Jiang, S. Raghavan *et al.*, "A saturated map of common genetic variants associated with human height," *Nature*, vol. 610, no. 7933, pp. 704–712, 2022.

[29] 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.

[30] A. Fabregat, K. Sidiropoulos, G. Viteri, O. Forner, P. Marin-Garcia, V. Arnau, P. D'Eustachio, L. Stein, and H. Hermjakob, "Reactome pathway analysis: a high-performance in-memory approach," *BMC bioinformatics*, vol. 18, pp. 1–9, 2017.

[31] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May *et al.*, "The reactome pathway knowledgebase," *Nucleic acids research*, vol. 46, no. D1, pp. D649–D655, 2018.

[32] D. Bradley, A. J. Smith, A. Blaszczak, D. Shantaram, S. M. Bergin, A. Jalilvand, V. Wright, K. L. Wyne, R. S. Dewal, L. A. Baer *et al.*, "Interferon gamma mediates the reduction of adipose tissue regulatory t cells in human obesity," *Nature communications*, vol. 13, no. 1, p. 5606, 2022.

[33] F. C. McGillicuddy, E. H. Chiquoine, C. C. Hinkle, R. J. Kim, R. Shah, H. M. Roche, E. M. Smyth, and M. P. Reilly, "Interferon γ attenuates insulin signaling, lipid storage, and differentiation in human adipocytes via activation of the jak/stat pathway," *Journal of Biological Chemistry*, vol. 284, no. 46, pp. 31 936–31 944, 2009.

[34] S.-H. Wang, M.-H. Su, C.-Y. Chen, Y.-F. Lin, Y.-C. A. Feng, P.-C. Hsiao, Y.-J. Pan, and C.-S. Wu, "Causality of abdominal obesity on cognition: a trans-ethnic mendelian randomization study," *International Journal of Obesity*, vol. 46, no. 8, pp. 1487–1492, 2022.

[35] 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.

[36] Y. Liang, M. Pividori, A. Manichaikul, A. A. Palmer, N. J. Cox, H. E. Wheeler, and H. K. Im, "Polygenic transcriptome risk scores (ptrs) can improve portability of polygenic risk scores across ancestries," *Genome biology*, vol. 23, pp. 1–18, 2022.

[37] 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.

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

[39] M. Verbanck, C.-Y. Chen, B. Neale, and R. Do, "Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases," *Nature genetics*, vol. 50, no. 5, pp. 693–698, 2018.

[40] M. Verbanck, D. Jordan, and R. Do, "The landscape of pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases," in *EUROPEAN JOURNAL OF HUMAN GENETICS*, vol. 27. NATURE PUBLISHING GROUP MACMILLAN BUILDING, 4 CRINAN ST, LONDON N1 9XW, ENGLAND, 2019, pp. 1115–1115.

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 distributions for glossTWAS in single-ancestry TWAS setup

The hierarchical structure of glossTWAS in the single-ancestry setup is given by

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

As discussed in the main text, the terms in Equation 1 carry out their usual meaning. Let us denote $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_m)^{\text{T}}$, where α_j is the effect of jth cis-SNP on the gene expression levels for $j = 1(1)m$. We specify the SNP-specific shrinkage penalties in the form of a vector as $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_m)^{\text{T}}$. 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

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

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. The full conditional distribution of $\boldsymbol{\alpha}$ is given as follows

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

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

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

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

As for $j=1(1)m$, λ_js are independently distributed, we update $\boldsymbol{\lambda}$ in an MCMC step by sequentially updating its components. For each $j=1(1)m$, we have

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

Note that, the full conditional posterior distribution of λ_j is not a well-known distribution from which we can draw random samples directly. In such a scenario, we use the Metropolis-Hastings algorithm to update the parameter λ_j in the MCMC which we have discussed in section 2.1.

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 generate a random sample from $f(\tau|\cdot)$ by the Metropolis-Hastings algorithm, and the steps of the algorithm are discussed in section 2.2.

1.4 Full conditional distribution of σ_e^2

The full conditional distribution of σ_e^2 is given by

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

The full conditional distribution of σ_e^2 is given by

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

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 $\text{N}(\widetilde{\mu}, \widetilde{\sigma}^2)$. The weight corresponding to $\delta_{\{0\}}(\beta)$ turns out to be

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

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

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

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 $B(s,t) = \frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}$ denote the beta function, where $\Gamma(\cdot)$ indicating the gamma function. The full conditional distribution of q is given as

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

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

$$q|\cdot \sim \begin{cases} \text{Beta}(b_1, b_2+1), & \text{with probability} \quad w = \frac{\delta_{\{0\}}(\beta)B(b_1,b_2+1)}{\delta_0(\beta)B(b_1,b_2+1)+\phi(\beta;0,\sigma_d^2\sigma_\epsilon^2)B(b_1+1,b_2)} \\ \text{Beta}(b_1+1, b_2), & \text{with probability} \quad 1 - w \end{cases} \tag{4}$$

1.7 Full conditional distribution of σ_d^2

The full conditional distribution of σ_d^2 is given by

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

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

$$
\sigma_d^2|\cdot\sim\begin{cases}\mathrm{IG}(d_1,d_2), & \text{with probability } \frac{w_1}{w_1+w_2}\\ \mathrm{IG}\left(d_1+\frac{1}{2},\ d_2+\frac{\beta^2}{2\sigma_\epsilon^2}\right), & \text{with probability } \frac{w_2}{w_1+w_2}\end{cases}\tag{5}
$$

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}}\ \mathrm{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}}\ \mathrm{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

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

2 Metropolis-Hastings algorithm for shrinkage parameters

We make inferences on the model parameters by implementing the Gibbs sampling algorithm. However, we cannot draw samples directly from the full conditional distributions of shrinkage parameters $\boldsymbol{\lambda}$ and τ as these are not well-known distributions. 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_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 2) is given as follows

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

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

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

Note that transformation $Z_j = \lambda_j^2$ is one-to-one as $\lambda_j > 0$. Hence, updating the parameter λ_j is equivalent to updating Z_j in an MCMC step. Let us denote the probability density function (pdf) of $Z_j|\cdot$ by $g(\cdot)$. We choose the candidate proposal distribution as IG $\left(1, \frac{\alpha_j^2}{2\sigma_e^2\tau^2}\right)$ and denote the corresponding pdf by $h(\cdot)$. This choice of candidate proposal pdf approximates the target pdf for large values of Z_j. 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_j in the MCMC.
Let ${z_j}^{(l-1)}$ is the generated sample from the target distribution $Z_j|\cdot$ 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 $v^{(l)}$ from the IG $\left(1, \frac{\alpha_j^2}{2\sigma_e^2\tau^2}\right)$ distribution.
- Draw a random number $U \sim \text{Uniform}(0,1)$.
- Set

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

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

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

At the lth step, we either draw a new random sample $v^{(l)}$ from the target distribution with probability $r({z_j}^{(l-1)}, v^{(l)})$ or retain the sample ${z_j}^{(l-1)}$ generated at the $(l-1)^{\text{th}}$ step with probability $1 - r({z_j}^{(l-1)}, v^{(l)})$. Then $\sqrt{{z_j}^{(l)}}$ is the random sample drawn from the full conditional posterior distribution of $\lambda_j|\cdot$ 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 choose the candidate proposal distribution as $\mathrm{IG}\left(\frac{m+1}{2}, \frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2}\right)$ and denote the corresponding pdf by $h(\cdot)$. We implement the independent Metropolis-Hastings algorithm to update Z in an MCMC step.
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 $v^{(l)}$ from the $\mathrm{IG}\left(\frac{m+1}{2}, \frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}}{2\sigma_e^2}\right)$ distribution.
- Draw a random number $U \sim \mathrm{Uniform}(0,1)$.
- Set

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

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

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

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 distributions for glossTWAS in multi-ancestry TWAS setup

The hierarchical structure of glossTWAS in the multi-ancestry TWAS setting is given by

$$\begin{aligned}
\mathbf{x}_{k,\text{ref}}|\boldsymbol{\alpha}_k, \sigma^2_{e_k} &\sim \text{N}_{n_k}(\mathbf{G}_{k,\text{ref}}\boldsymbol{\alpha}_k, \sigma^2_{e_k}\text{I}_{n_k}), \quad k = 1(1)K \\
\mathbf{y}_k|\beta, \boldsymbol{\alpha}_k, \sigma^2_\epsilon &\sim \text{N}_{N_k}(\beta\mathbf{G}_k\boldsymbol{\alpha}_k, \sigma^2_\epsilon\text{I}_{N_k}) \\
\alpha_{kj}|\lambda_j, \tau, \sigma^2_{e_k} &\sim \text{N}(0, \sigma^2_{e_k}\tau^2\lambda^2_j), \quad j = 1(1)m,\ k = 1(1)K \\
\lambda_j &\sim \text{C}^+(0, 1) \\
\tau &\sim \text{C}^+(0, 1) \\
\beta|q, \sigma^2_d, \sigma^2_\epsilon &\sim (1-q)\delta_{\{0\}}(\beta) + q\text{N}(0, \sigma^2_d\sigma^2_\epsilon) \\
q &\sim \text{Beta}(b_1, b_2) \\
\sigma^2_d &\sim \text{IG}(d_1, d_2) \\
\sigma^2_{e_k} &\sim \text{IG}(e_1, e_2) \\
\sigma^2_\epsilon &\sim \text{IG}(\epsilon_1, \epsilon_2)
\end{aligned} \tag{6}$$

The notations in the Equation 6 have their usual meaning as described in the main text. For each $k = 1(1)K$, let us denote $\boldsymbol{\alpha}_k = (\alpha_{k1}, \alpha_{k2}, \cdots, \alpha_{km})^{\text{T}}$. For $j = 1(1)m$, α_{kj} denotes the effect of jth cis-SNP on the gene expression levels in the kth ancestry. We specify the SNP-specific shrinkage parameters across all ancestries in the form of a vector as $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_m)^{\text{T}}$. 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

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

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

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

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

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

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

For each $j = 1(1)m$, as λ_js are independently distributed in the prior it is convenient to update the components of $\boldsymbol{\lambda}$ sequentially in an MCMC step. The conditional distribution of λ_j is given as follows

$$\begin{aligned} f(\lambda_j|\cdot) &\propto \left\{\prod_{k=1}^{K} f(\alpha_{kj}|\lambda_j, \tau, \sigma_e^2)\right\} f(\lambda_j) \\ &\propto \left[\prod_{k=1}^{K}\left\{\frac{1}{\sigma_{e_k}\lambda_j\tau}\ \exp\left[-\frac{\alpha_{kj}^2}{2\sigma_{e_k}^2\tau^2\lambda_j^2}\right]\right\}\right]\left\{\frac{1}{1+\lambda_j^2}\right\} \\ &\propto \frac{1}{\lambda_j^K(1+\lambda_j^2)}\ \exp\left[-\frac{1}{2\tau^2\lambda_j^2}\sum_{k=1}^{K}\frac{\alpha_{kj}^2}{\sigma_{e_k}^2}\right] \end{aligned} \tag{7}$$

As the full conditional posterior distribution of λ_j is not conditionally conjugate, we employ the Metropolis-Hastings algorithm to draw a random sample. The brief steps of the algorithm are mentioned in the section 4.1.

3.3 Full conditional distribution of τ

The full conditional distribution of τ is given as

$$\begin{aligned} f(\tau|\cdot) &\propto \left\{f(\boldsymbol{\alpha}_k|\boldsymbol{\lambda}, \tau, \sigma_{e_k}^2)\right\} f(\tau) \\ &\propto \left\{\prod_{k=1}^{K}\frac{1}{\tau^m}\ \exp\left[-\frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}_k}{2\sigma_{e_k}^2\tau^2}\right]\right\}\left\{\frac{1}{1+\tau^2}\right\} \\ &\propto \frac{1}{\tau^{mK}(1+\tau^2)}\ \exp\left[-\frac{1}{2\tau^2}\sum_{k=1}^{K}\frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{D}_\lambda^{-1}\boldsymbol{\alpha}_k}{\sigma_{e_k}^2}\right] \end{aligned} \tag{8}$$

We use the Metropolis-Hastings algorithm to draw a random sample from the conditional posterior distribution of τ and the steps of the algorithm are discussed in the section 4.2.

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

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

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

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

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

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 $\mathrm{N}(\widetilde{\mu},\widetilde{\sigma}^2)$. Let us denote the weight attached to the point mass function $\delta_{\{0\}}(\beta)$ by

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

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

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

3.6 Full conditional distribution of σ_ϵ^2

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

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

where

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

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

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

The full conditional posterior distributions of q and σ_d^2 remain the same as we have pointed out in Equation 4 and Equation 5.

4 Metropolis-Hastings algorithm for shrinkage parameters

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

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

We 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 7) 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 $Z_j|\cdot$ by $g(\cdot)$. We 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

$$r({z_j}^{(l-1)}, v^{(l)}) = \min\left\{1, \frac{g(v^{(l)})h({z_j}^{(l-1)})}{g({z_j}^{(l-1)})h(v^{(l)})}\right\} = \min\left\{1, \frac{1+{z_j}^{(l-1)}}{{z_j}^{(l-1)}}\frac{v^{(l)}}{1+v^{(l)}}\right\}, \tag{9}$$

where ${z_j}^{(l-1)}$ is the generated sample from the target distribution at the $(l-1)^{\text{th}}$ step and $v^{(l)}$ 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 8) 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

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

We 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 9 and follow the similar steps as discussed in section 2.2.

Supplemental Figures

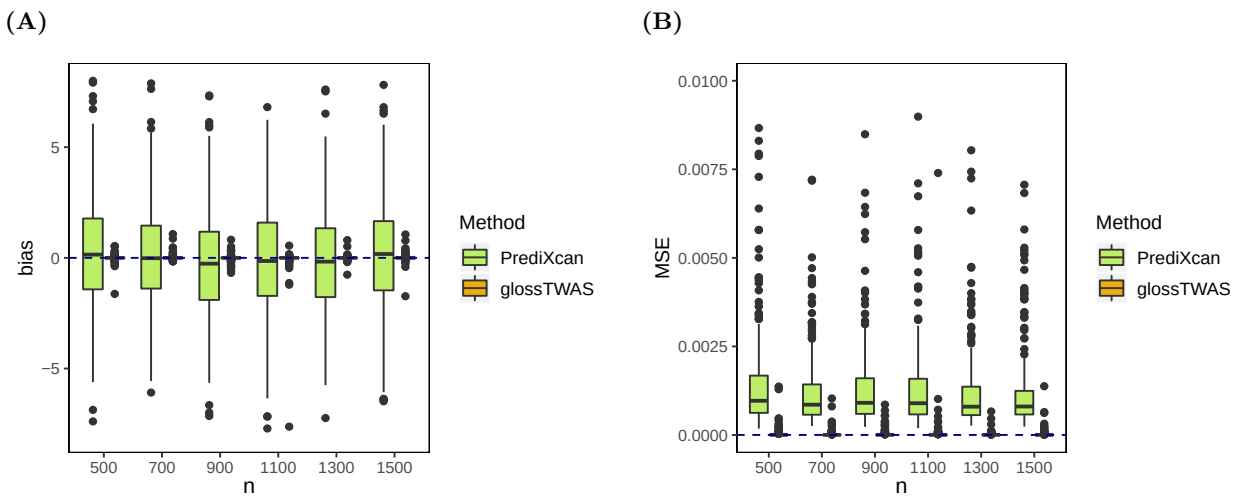

Figure S1: Comparison of estimation accuracy by PrediXcan and glossTWAS in single-ancestry TWAS when the GReX has no effect on the trait, i.e., $\beta = 0$. The transcriptome data sample size (n) is varied from 500 to 1500. Two measures are considered: (**A**) bias and (**B**) MSE. The GWAS sample size is $N = 15000$. The local heritability of gene expression is fixed at $h_e^2 = 20\%$.

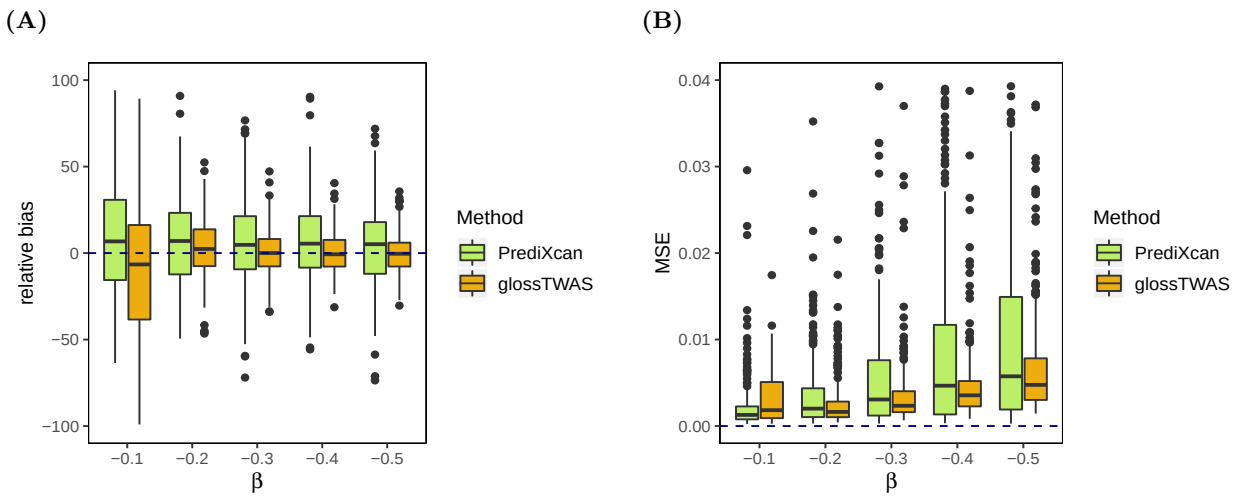

Figure S2: Comparison between PrediXcan and glossTWAS in estimating the negative GReX effect sizes for single-ancestry TWAS. Two measures are considered: (**A**) relative bias and (**B**) MSE. The transcriptome and GWAS data sample sizes are $n = 500$ and $N = 15000$, respectively. The local heritability of gene expression is fixed at $h_e^2 = 20\%$.

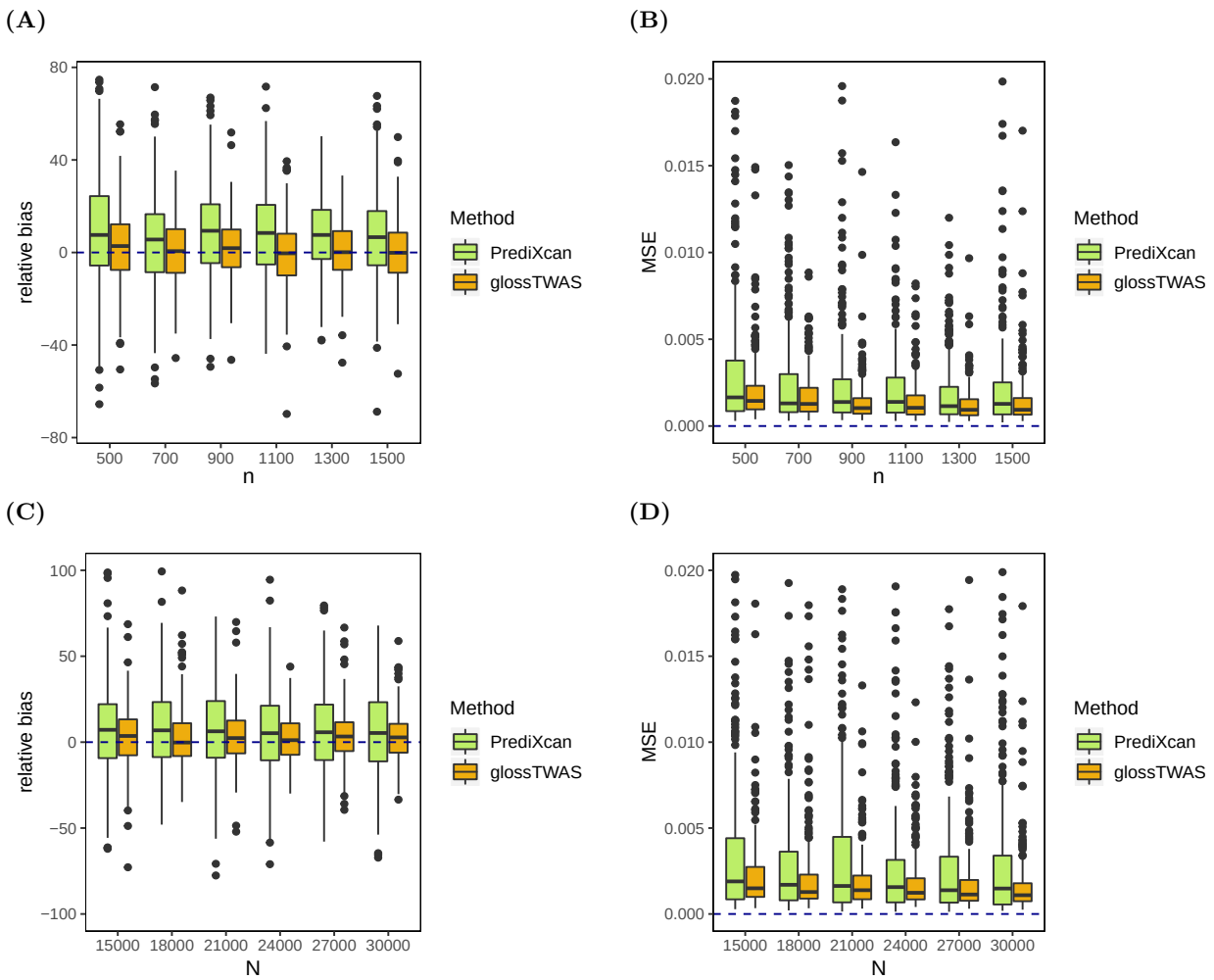

Figure S3: Comparison of GReX effect size estimation accuracy between PrediXcan and glossTWAS in single-ancestry TWAS. The true GReX effect size is considered as $\beta = 0.2$. Two measures of relative bias and MSE are considered for varying transcriptome data (n) sample sizes (Figure **A-B**) and varying GWAS (N) sample sizes (Figure **C-D**). The GWAS sample size is $N = 15000$ in Figure **A-B** and the transcriptome data sample size is $n = 500$ in (Figure **C-D**). The local heritability of gene expression is fixed at $h_e^2 = 20\%$.

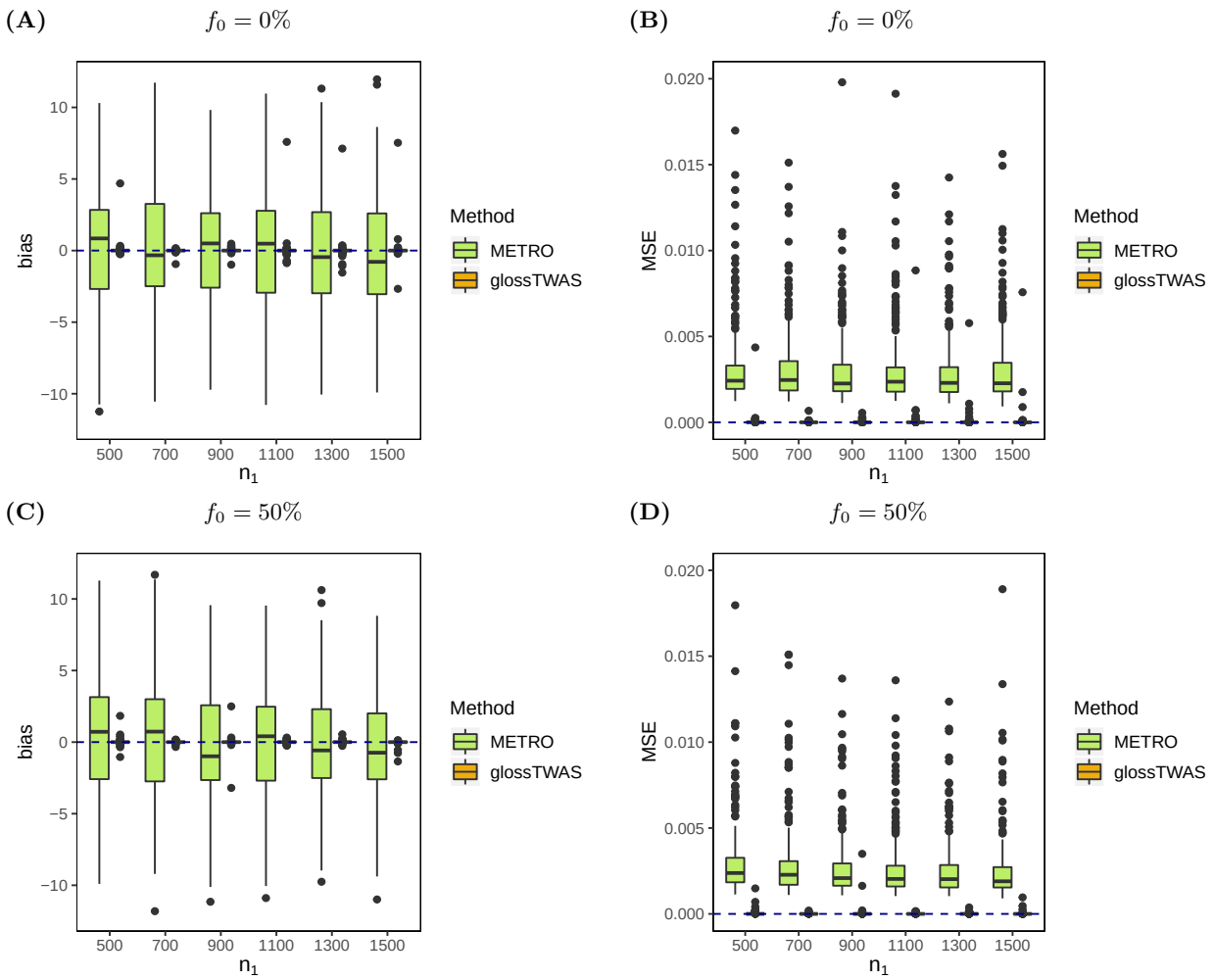

Figure S4: Comparison of estimation accuracy by METRO and glossTWAS in multi-ancestry TWAS when the GReX has no effect on the trait, i.e., $\beta = 0$. The transcriptome data sample size for East Asian ancestry (n_1) is varied from 500 to 1500. Two measures are considered: bias and MSE. Both the scenarios of no sharing of eQTLs, $f_0 = 0\%$ (Figure **A**-**B**), and sharing of half of the eQTLs, $f_0 = 50\%$ (Figure **C**-**D**), between the two ancestries are considered. The transcriptome data sample size for European ancestry is $n_2 = 500$. The GWAS sample sizes are $N_1 = N_2 = 10000$ for the two ancestries. The local heritability of gene expression is fixed at $h_e^2 = 20\%$ in each ancestry.

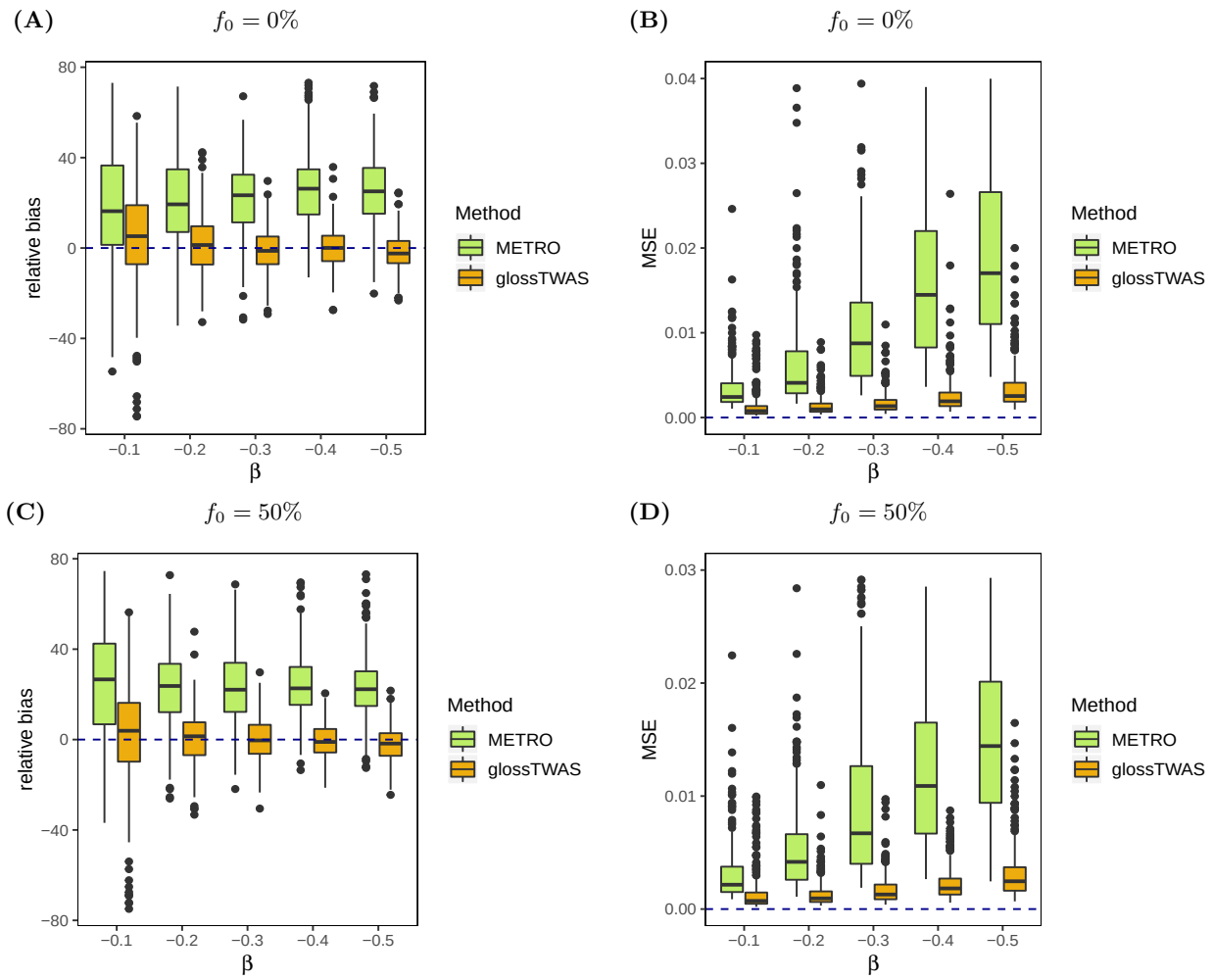

Figure S5: Comparison between METRO and glossTWAS in estimating the negative GReX effect sizes for multi-ancestry TWAS. Two measures are considered: relative bias and MSE. Both the scenarios of no sharing of eQTLs, $f_0 = 0\%$ (Figure **A**-**B**), and sharing of half of the eQTLs, $f_0 = 50\%$ (Figure **C**-**D**), between the two ancestries are considered. The transcriptome data sample sizes are $n_1 = n_2 = 500$, and GWAS data sample sizes are $N_1 = N_2 = 10000$ for the two ancestries. The local heritability of gene expression is fixed at $h_e^2 = 20\%$ in each ancestry.

Supplemental Tables

Table S1: Comparison between glossTWAS and PrediXcan based on the summary of different GReX effect sizes in single-ancestry TWAS

	glossTWAS					PrediXcan				
	GReX effect size			Relative bias		GReX effect size			Relative bias	
β	Estimate	SD	MSE	mean	SD	Estimate	SD	MSE	mean	SD
0.5	0.49	0.06	0.006	-2.2	10	0.52	0.03	0.015	3.5	23
0.4	0.40	0.05	0.004	-0.4	10	0.43	0.03	0.010	6.5	23
0.3	0.30	0.04	0.003	0.5	11	0.31	0.03	0.005	3.9	21
0.2	0.21	0.03	0.002	2.6	15	0.21	0.03	0.003	6.2	22
0.1	0.09	0.03	0.003	-9.0	39	0.11	0.03	0.002	11.0	30
-0.1	-0.09	0.03	0.003	-14.0	41	-0.11	0.03	0.002	10.4	31
-0.2	-0.21	0.03	0.002	2.8	15	-0.21	0.03	0.003	7.4	24
-0.3	-0.30	0.04	0.003	-0.2	12	-0.32	0.03	0.006	6.2	23
-0.4	-0.40	0.05	0.004	-0.6	11	-0.42	0.03	0.009	5.6	22
-0.5	-0.50	0.06	0.007	-0.9	10	-0.52	0.03	0.015	4.5	23

Note: The notations β and SD denote the GReX effect size and standard deviation respectively. The values in the column Relative bias for each method are represented in percentages. The transcriptome data and GWAS sample sizes are $n = 500$ and $N = 15000$. The local heritability of gene expression is fixed at $h_e^2 = 20\%$. The results are obtained based on 300 effective replications in each simulation scenario.

Table S2: Comparison between glossTWAS and PrediXcan for estimating the GReX effect size $\beta = 0.2$ for varying transcriptome data sample sizes (n).

	glossTWAS					PrediXcan				
	GReX effect size			Relative bias		GReX effect size			Relative bias	
n	Estimate	SD	MSE	mean	SD	Estimate	SD	MSE	mean	SD
500	0.20	0.03	0.002	1.4	14	0.21	0.03	0.003	7.0	22
700	0.20	0.03	0.002	1.2	13	0.22	0.03	0.003	8.4	22
900	0.21	0.03	0.002	2.7	14	0.22	0.02	0.003	10.9	20
1100	0.20	0.03	0.001	0.1	11	0.22	0.02	0.002	7.6	16
1300	0.20	0.03	0.001	0.8	11	0.22	0.02	0.002	8.7	15
1500	0.20	0.03	0.001	-0.6	12	0.21	0.02	0.002	6.7	17

Note: The notations n and SD denote the transcriptome data sample size and standard deviation respectively. The values in the column Relative bias for each method are represented in percentages. The true value of the GReX effect size is $\beta = 0.2$ and the GWAS sample size is $N = 15000$. The local heritability of gene expression is fixed at $h_e^2 = 20\%$. The results are obtained based on 300 effective replications in each simulation scenario.

Table S3: Comparison between glossTWAS and PrediXcan for estimating the GReX effect size $\beta = 0.2$ for varying GWAS sample sizes (N).

	glossTWAS					PrediXcan				
	GReX effect size			Relative bias		GReX effect size			Relative bias	
N	Estimate	SD	MSE	mean	SD	Estimate	SD	MSE	mean	SD
15000	0.21	0.03	0.002	2.9	14	0.21	0.03	0.004	7.4	26
18000	0.20	0.03	0.002	1.5	14	0.22	0.02	0.003	8.7	23
21000	0.20	0.03	0.002	2.5	14	0.22	0.02	0.003	8.3	24
24000	0.20	0.03	0.002	1.8	13	0.21	0.02	0.003	6.1	22
27000	0.20	0.03	0.002	2.4	13	0.21	0.02	0.002	6.2	22
30000	0.20	0.03	0.002	2.2	12	0.21	0.02	0.003	6.4	23

Note: The notations N and SD denote the GWAS sample size and standard deviation respectively. The values in the column Relative bias for each method are represented in percentages. The true value of the GReX effect size is $\beta = 0.2$ and the transcriptome data sample size is $n = 500$. The local heritability of gene expression is fixed at $h_e^2 = 20\%$. The results are obtained based on 300 effective replications in each simulation scenario.

Table S4: Summary of the estimate of GReX effect sizes obtained by glossTWAS in multi-ancestry TWAS. Here, eQTLs are not shared between the two ancestries, i.e., $f_0 = 0\%$.

β	mean	SD	median	95% CPI	$\log_{10}$(BF)
0.5	0.49	0.04	0.49	(0.42, 0.57)	197.5
0.4	0.40	0.04	0.40	(0.33, 0.47)	125.6
0.3	0.30	0.03	0.30	(0.24, 0.36)	68.0
0.2	0.20	0.03	0.20	(0.16, 0.25)	27.4
0.1	0.10	0.03	0.10	(0.05, 0.15)	5.4
0	0	0.003	0	$(-1 \times 10^{-4}, 2.5 \times 10^{-4})$	-1.1
-0.1	-0.10	0.02	-0.10	(-0.15, -0.06)	5.8
-0.2	-0.20	0.03	-0.20	(-0.25, -0.15)	27.1
-0.3	-0.30	0.03	-0.29	(-0.35, -0.24)	68.1
-0.4	-0.40	0.04	-0.40	(-0.47, -0.33)	126.4
-0.5	-0.49	0.04	-0.49	(-0.57, -0.41)	196.4

Note: The abbreviations SD, CPI, and BF represent the standard deviation, central posterior interval, and Bayes factor, respectively. The transcriptome data sample sizes are $n_1 = n_2 = 500$, and GWAS data sample sizes are $N_1 = N_2 = 10000$ for the two ancestries. The local heritability of gene expression is fixed at $h_e^2 = 20\%$ in each ancestry. The summaries are obtained based on 300 effective replications in a simulation scenario.

Table S5: Comparison between glossTWAS and METRO based on the summary of different GReX effect sizes in multi-ancestry TWAS. Here, eQTLs are not shared between the two ancestries, i.e., $f_0 = 0\%$.

	glossTWAS					METRO				
	GReX effect size			Relative bias		GReX effect size			Relative bias	
β	Estimate	SD	MSE	mean	SD	Estimate	SD	MSE	mean	SD
0.5	0.49	0.04	0.003	-1.3	6	0.63	0.08	0.031	26.5	14
0.4	0.40	0.04	0.002	-0.4	7	0.50	0.07	0.020	26.0	15
0.3	0.30	0.03	0.002	-0.4	8	0.37	0.06	0.011	22.8	17
0.2	0.20	0.03	0.001	1.8	11	0.24	0.05	0.006	22.1	21
0.1	0.10	0.02	0.001	1.4	23	0.12	0.04	0.003	21.8	27
-0.1	-0.10	0.02	0.001	3.4	22	-0.12	0.04	0.003	23.4	29
-0.2	-0.20	0.03	0.001	1.4	12	-0.24	0.05	0.006	21.5	20
-0.3	-0.30	0.03	0.002	-1.1	8	-0.37	0.06	0.010	22.1	15
-0.4	-0.40	0.04	0.002	0.1	8	-0.50	0.07	0.019	25.5	14
-0.5	-0.49	0.04	0.003	-2.2	7	-0.63	0.08	0.028	25.6	15

Note: The abbreviation SD denotes the standard deviation. The values in the column Relative bias for each method are represented in percentages. The transcriptome data sample sizes are $n_1 = n_2 = 500$, and GWAS data sample sizes are $N_1 = N_2 = 10000$ for the two ancestries. The local heritability of gene expression is fixed at $h_e^2 = 20\%$ in each ancestry. The results are obtained based on 300 effective replications in a simulation scenario.

Table S6: Comparison between glossTWAS and METRO based on the summary of different GReX effect sizes in multi-ancestry TWAS. Here, half of the eQTLs are shared between the two ancestries, i.e., $f_0 = 50\%$.

	glossTWAS					METRO				
	GReX effect size			Relative bias		GReX effect size			Relative bias	
β	Estimate	SD	MSE	mean	SD	Estimate	SD	MSE	mean	SD
0.5	0.49	0.04	0.003	-2.3	7	0.62	0.07	0.023	23.6	14
0.4	0.40	0.04	0.002	-0.9	7	0.50	0.06	0.015	24.1	13
0.3	0.30	0.03	0.002	0.8	8	0.37	0.05	0.010	24.4	14
0.2	0.20	0.03	0.001	2.2	12	0.25	0.04	0.006	25.5	20
0.1	0.10	0.02	0.001	-1.4	24	0.13	0.04	0.003	27.2	28
-0.1	-0.10	0.02	0.001	0.9	24	-0.13	0.04	0.003	29.0	26
-0.2	-0.20	0.03	0.001	0.8	11	-0.25	0.04	0.005	23.6	17
-0.3	-0.30	0.03	0.002	0.1	9	-0.37	0.05	0.010	23.9	15
-0.4	-0.40	0.03	0.002	-0.9	7	-0.49	0.06	0.015	23.6	13
-0.5	-0.49	0.04	0.003	-2.4	7	-0.62	0.07	0.022	23.0	13

Note: The abbreviation SD denotes the standard deviation. The values in the column Relative bias for each method are represented in percentages. The transcriptome data sample sizes are $n_1 = n_2 = 500$, and GWAS data sample sizes are $N_1 = N_2 = 10000$ for the two ancestries. The local heritability of gene expression is fixed at $h_e^2 = 20\%$ in each ancestry. The results are obtained based on 300 effective replications in a simulation scenario.

Table S7: Associated genes for BMI identified by PrediXcan in the single-ancestry TWAS for European ancestry.

Chr	Gene	Effect size	SD	p-value
1	PHC2	-0.04	0.01	2.5×10^{-7}
1	WARS2	-0.02	0.01	4.2×10^{-7}
1	SLC35E2B	-0.02	0.004	3.1×10^{-6}
3	UBA7	0.21	0.02	5×10^{-33}
3	RBM6	0.07	0.01	3.6×10^{-28}
3	DHFR2	-0.08	0.01	4.3×10^{-13}
3	PCYT1A	-0.04	0.01	4.8×10^{-7}
4	GRK4	-0.03	0.01	4.3×10^{-7}
5	POC5	0.03	0.01	3.2×10^{-9}
6	HLA-B	-0.03	0.01	2.1×10^{-7}
6	RNF5	0.02	0.01	2.6×10^{-7}
6	MICB	0.02	0.01	9.9×10^{-7}
6	SKIV2L	0.03	0.01	2.5×10^{-6}
8	BLK	0.04	0.01	2.9×10^{-11}
8	FAM167A	-0.02	0.004	2.4×10^{-6}
10	DPYSL4	-0.02	0.004	2.9×10^{-8}
10	FBXL15	-0.06	0.01	5.5×10^{-7}
11	CTSW	0.07	0.01	1.6×10^{-15}
11	SNX32	-0.02	0.01	8×10^{-8}
11	FIBP	-0.06	0.01	5.6×10^{-7}
11	NR1H3	-0.04	0.01	6×10^{-7}
11	SIRT3	-0.08	0.02	2.3×10^{-6}
12	NCKAP5L	-0.07	0.01	1.5×10^{-10}
12	RILPL2	0.07	0.01	2.6×10^{-9}
12	MORN3	-0.04	0.01	1.2×10^{-7}
12	RPS26	0.01	0.003	3.2×10^{-6}
16	KCTD13	-0.13	0.01	5.6×10^{-27}
16	NFATC2IP	-0.08	0.01	3.4×10^{-22}
16	SH2B1	-0.14	0.02	1.6×10^{-20}
16	MAPK3	0.07	0.01	2.8×10^{-17}
16	TBX6	-0.05	0.01	6.5×10^{-14}
16	APOBR	-0.04	0.01	1.9×10^{-10}
16	CD19	-0.05	0.01	4.4×10^{-10}
16	DDX19B	-0.10	0.02	1.7×10^{-8}
16	KAT8	-0.09	0.02	3.3×10^{-8}
16	KNOP1	0.05	0.01	1.5×10^{-7}
17	TTC19	-0.02	0.004	4.1×10^{-7}
17	GIT1	-0.10	0.02	7.1×10^{-7}
19	IL12RB1	0.06	0.01	6×10^{-11}
20	ENTPD6	-0.09	0.01	2.3×10^{-9}

Table S8: Associated genes for BMI identified by glossTWAS in the single-ancestry TWAS for European ancestry.

Chr	Gene	Effect size	SD	PPNA	$\log_{10}$(BF)
1	PABPC4	0.05	0.02	7.9×10^{-3}	5.8
1	PTBP2	-0.6	0.44	1×10^{-13}	16.7
1	PHC2	-0.05	0.01	9.9×10^{-7}	9.7
1	DNAJB4	0.15	0.11	2.7×10^{-10}	13.3
1	AMIGO1	-0.3	0.2	6.5×10^{-11}	13.9
1	ANKRD65	-0.34	0.33	2.7×10^{-8}	11.3
1	FPGT	0.29	0.21	5.5×10^{-9}	12
2	CRIPT	0.22	0.2	1.6×10^{-12}	15.5
2	ATRAID	-0.27	0.2	4.9×10^{-4}	7
2	PIGF	0.21	0.18	1.6×10^{-12}	15.5
2	FBXO36	0.33	0.23	5.8×10^{-8}	11
2	FAM228B	0.16	0.1	4.4×10^{-5}	8.1
3	RBM6	0.09	0.02	2.3×10^{-35}	38.4
3	MAP4	0.21	0.21	8.2×10^{-6}	8.8
3	IFT57	0.26	0.2	9.1×10^{-3}	5.8
3	ABHD14B	0.16	0.09	1.7×10^{-13}	16.5
3	CGGBP1	-0.53	0.33	7.6×10^{-5}	7.9
3	ZNF589	-0.35	0.42	2.7×10^{-17}	20.3
3	ZNF654	0.28	0.27	1.9×10^{-5}	8.5
3	WDR6	0.44	0.24	2×10^{-20}	23.4
3	DHFR2	-0.12	0.06	9.9×10^{-10}	12.7
3	UBA7	0.17	0.08	2.6×10^{-31}	34.3
3	KLHDC8B	0.5	0.25	9.2×10^{-22}	24.8
3	ZNF501	0.11	0.06	4.4×10^{-4}	7.1
3	CD47	-0.09	0.05	1.6×10^{-7}	10.5
3	DZIP3	-0.05	0.02	2×10^{-2}	5.4
3	NCKIPSD	0.33	0.19	1.3×10^{-18}	21.6
4	MANBA	0.55	0.36	7×10^{-31}	33.9
4	GRK4	-0.04	0.01	1.2×10^{-2}	5.6
4	GUF1	-0.67	0.5	8.5×10^{-26}	28.8
4	SMIM20	-0.34	0.23	1.3×10^{-17}	20.6
5	RNF14	0.55	0.36	2.2×10^{-9}	12.4
5	POC5	0.06	0.01	1.7×10^{-13}	16.5
5	EFNA5	-0.1	0.04	1.7×10^{-6}	9.5
6	HLA-B	-0.05	0.01	5.8×10^{-9}	12
6	MOCS1	-0.21	0.19	5.6×10^{-16}	19
6	ZNF192P1	0.27	0.22	1.8×10^{-10}	13.5
6	ZSCAN26	0.18	0.12	3.8×10^{-8}	11.2
6	HLA-DPA1	-0.35	0.26	7×10^{-14}	16.9
6	RNF5	0.05	0.01	3.8×10^{-8}	11.2
6	SKIV2L	0.1	0.03	3.3×10^{-16}	19.2
6	MICB	0.05	0.01	6.5×10^{-8}	10.9
7	PPIA	0.33	0.2	1.2×10^{-10}	13.7
8	BLK	0.04	0.01	2.6×10^{-7}	10.3
8	CTSB	-0.09	0.03	2.3×10^{-8}	11.4
8	INTS8	-0.06	0.02	2.1×10^{-2}	5.4
10	HPS1	0.39	0.27	5.2×10^{-8}	11
10	FBXL15	-0.14	0.08	2.6×10^{-8}	11.3
10	ACTR1A	-0.27	0.24	7.4×10^{-8}	10.9

Continuation of Table S8					
Chr	Gene	Effect size	SD	PPNA	$\log_{10}$(BF)
10	MFSD13A	0.33	0.23	2.3×10^{-2}	5.4
10	NRBF2	0.07	0.02	4.2×10^{-6}	9.1
10	DPYSL4	-0.02	0.01	4×10^{-2}	5.1
10	VDAC2	0.11	0.06	1.9×10^{-2}	5.4
10	TRIM8	-0.09	0.05	8.4×10^{-6}	8.8
10	BNIP3	-0.19	0.12	1.7×10^{-7}	10.5
10	ADO	-0.28	0.25	3.8×10^{-6}	9.2
10	EEF1AKMT2	0.16	0.13	1.3×10^{-5}	8.6
11	NR1H3	-0.18	0.09	5.5×10^{-18}	21
11	MADD	-0.24	0.13	4.6×10^{-18}	21.1
11	STK33	-0.07	0.04	5.8×10^{-5}	8
11	LRP4	-0.39	0.35	3.6×10^{-16}	19.2
11	ACP2	-0.24	0.13	2.8×10^{-18}	21.3
11	ARFGAP2	-0.25	0.14	6.3×10^{-18}	20.9
11	ACAD8	0.31	0.42	5.1×10^{-20}	23
11	THYN1	0.53	0.42	2.3×10^{-21}	24.4
11	FIBP	-0.13	0.09	1.9×10^{-13}	16.4
11	CTSW	0.08	0.02	1.8×10^{-13}	16.5
11	SNX32	-0.04	0.01	5×10^{-9}	12
11	OVOL1	-0.34	0.28	10×10^{-14}	16.7
11	CCS	-0.07	0.03	2.6×10^{-2}	5.3
11	EIF1AD	-0.34	0.28	3.9×10^{-14}	17.1
12	SLC25A3	0.17	0.19	2.9×10^{-7}	10.3
12	CDK2AP1	-0.16	0.14	5.6×10^{-6}	9
12	NT5DC3	0.15	0.11	9.3×10^{-10}	12.8
12	MORN3	-0.08	0.03	1.1×10^{-13}	16.7
12	RILPL2	0.08	0.04	1.1×10^{-8}	11.7
12	CFAP251	-0.39	0.26	6.5×10^{-14}	16.9
12	NCKAP5L	-0.13	0.05	5.1×10^{-22}	25
12	TCTN2	0.2	0.15	7.8×10^{-5}	7.8
12	HCAR2	0.17	0.12	1.1×10^{-8}	11.7
14	PPP1R13B	0.15	0.15	5.3×10^{-3}	6
14	BAG5	0.14	0.09	2.6×10^{-4}	7.3
15	NEO1	0.06	0.02	1.4×10^{-4}	7.6
16	MAPK3	0.12	0.05	4.1×10^{-21}	24.1
16	CLUAP1	-0.19	0.15	1.1×10^{-5}	8.7
16	STX4	-0.35	0.36	5.5×10^{-9}	12
16	KAT8	-0.08	0.03	1.3×10^{-9}	12.6
16	KNOP1	0.1	0.05	2.4×10^{-3}	6.4
16	TBX6	-0.08	0.02	4.4×10^{-20}	23.1
16	DDX19B	-0.26	0.16	7.3×10^{-20}	22.9
16	KCTD13	-0.09	0.02	3×10^{-20}	23.3
16	SGF29	0.27	0.15	6.9×10^{-20}	22.9
16	NFATC2IP	-0.09	0.03	3.1×10^{-18}	21.2
16	CD19	-0.16	0.08	3.6×10^{-18}	21.2
16	SH2B1	-0.14	0.07	5.7×10^{-19}	22
16	APOBR	-0.1	0.04	1.3×10^{-19}	22.6
17	RABEP1	-0.22	0.19	1.9×10^{-2}	5.5
17	CBX1	0.28	0.22	1.1×10^{-8}	11.7
17	SKAP1	0.12	0.06	2.5×10^{-13}	16.3
17	LRRC46	-0.27	0.21	4.8×10^{-6}	9

Continuation of Table S8					
Chr	Gene	Effect size	SD	PPNA	$\log_{10}$(BF)
17	UBE2Z	0.17	0.16	6.5×10^{-12}	14.9
17	ZNF232	0.12	0.08	4.5×10^{-5}	8.1
17	C17orf107	-0.07	0.03	1.3×10^{-3}	6.6
17	CCDC200	0.14	0.1	5.3×10^{-3}	6
19	TCF3	0.3	0.27	4.5×10^{-5}	8.1
19	ANKRD24	-0.27	0.27	3.2×10^{-7}	10.2
19	IL12RB1	0.08	0.03	1.5×10^{-8}	11.6
19	KPTN	0.53	0.53	1.2×10^{-6}	9.7
19	SIX5	0.57	0.45	5.6×10^{-16}	19
19	ADAT3	0.35	0.35	6.5×10^{-10}	12.9
20	ENTPD6	-0.07	0.03	2.5×10^{-5}	8.3

Note: The abbreviations SD, PPNA, and BF represent the standard deviation, posterior probability of null association, and Bayes factor, respectively. The results are obtained based on 7000 MCMC iterations with a burn-in sample size of 2000 and a thinning period of 5.

Table S9: Common associated genes for BMI identified by glossTWAS and PrediXcan in European ancestry for single-ancestry TWAS.

		glossTWAS				PrediXcan		
Chr	Gene	Effect size	SD	PPNA	$\log_{10}$(BF)	Effect size	SD	p-value
1	PHC2	-0.05	0.01	9.9×10^{-7}	9.7	-0.04	0.01	2.5×10^{-7}
3	RBM6	0.09	0.02	2.3×10^{-35}	38.4	0.07	0.01	3.6×10^{-28}
3	DHFR2	-0.12	0.06	9.9×10^{-10}	12.7	-0.08	0.01	4.3×10^{-13}
3	UBA7	0.17	0.08	2.6×10^{-31}	34.3	0.21	0.02	5×10^{-33}
4	GRK4	-0.04	0.01	1.2×10^{-2}	5.6	-0.03	0.01	4.3×10^{-7}
5	POC5	0.06	0.01	1.7×10^{-13}	16.5	0.03	0.01	3.2×10^{-9}
6	HLA-B	-0.05	0.01	5.8×10^{-9}	12	-0.03	0.01	2.1×10^{-7}
6	RNF5	0.05	0.01	3.8×10^{-8}	11.2	0.02	0.01	2.6×10^{-7}
6	SKIV2L	0.1	0.03	3.3×10^{-16}	19.2	0.03	0.01	2.5×10^{-6}
6	MICB	0.05	0.01	6.5×10^{-8}	10.9	0.02	0.01	9.9×10^{-7}
8	BLK	0.04	0.01	2.6×10^{-7}	10.3	0.04	0.01	2.9×10^{-11}
10	FBXL15	-0.14	0.08	2.6×10^{-8}	11.3	-0.06	0.01	5.5×10^{-7}
10	DPYSL4	-0.02	0.01	4×10^{-2}	5.1	-0.02	0.004	2.9×10^{-8}
11	NR1H3	-0.18	0.08	5.5×10^{-18}	21	-0.04	0.01	6×10^{-7}
11	FIBP	-0.13	0.09	1.9×10^{-13}	16.4	-0.06	0.01	5.6×10^{-7}
11	CTSW	0.08	0.02	1.8×10^{-13}	16.5	0.07	0.01	1.6×10^{-15}
11	SNX32	-0.04	0.01	5×10^{-9}	12	-0.02	0.01	8×10^{-8}
12	MORN3	-0.08	0.03	1.1×10^{-13}	16.7	-0.04	0.01	1.2×10^{-7}
12	RILPL2	0.08	0.03	1.1×10^{-8}	11.7	0.07	0.01	2.6×10^{-9}
12	NCKAP5L	-0.13	0.05	5.1×10^{-22}	25	-0.07	0.01	1.5×10^{-10}
16	MAPK3	0.12	0.05	4.1×10^{-21}	24.1	0.07	0.01	2.8×10^{-17}
16	KAT8	-0.08	0.03	1.3×10^{-9}	12.6	-0.09	0.02	3.3×10^{-8}
16	KNOP1	0.1	0.05	2.4×10^{-3}	6.4	0.05	0.01	1.5×10^{-7}
16	TBX6	-0.08	0.02	4.4×10^{-20}	23.1	-0.05	0.01	6.5×10^{-14}
16	DDX19B	-0.26	0.16	7.3×10^{-20}	22.9	-0.1	0.02	1.7×10^{-8}
16	KCTD13	-0.09	0.02	3×10^{-20}	23.3	-0.13	0.01	5.6×10^{-27}
16	NFATC2IP	-0.09	0.03	3.1×10^{-18}	21.2	-0.08	0.01	3.4×10^{-22}
16	CD19	-0.16	0.08	3.6×10^{-18}	21.2	-0.05	0.01	4.4×10^{-10}
16	SH2B1	-0.14	0.07	5.7×10^{-19}	22	-0.14	0.02	1.6×10^{-20}
16	APOBR	-0.1	0.04	1.3×10^{-19}	22.6	-0.04	0.001	1.9×10^{-10}
19	IL12RB1	0.08	0.03	1.5×10^{-8}	11.6	0.06	0.01	6×10^{-11}
20	ENTPD6	-0.07	0.03	2.5×10^{-5}	8.3	-0.09	0.01	2.3×10^{-9}

Note: The abbreviations SD, PPNA, and BF represent the standard deviation, posterior probability of null association, and Bayes factor, respectively. For glossTWAS, the results are obtained based on 7000 MCMC iterations with a burn-in sample size of 2000 and a thinning period of 5.

Table S10: Associated genes for BMI identified by METRO in at least one of European and African ancestry in multi-ancestry TWAS.

Chr	Gene	Effect size	EUR Effect size	AFR Effect size	p-value
1	TRIM62	-0.18	-0.117	-0.065	2×10^{-8}
1	CRYZ	0.06	0	0.065	3.4×10^{-9}
1	ZCCHC17	0.16	0.029	0.128	1.1×10^{-6}
1	NAV1	-0.06	0	-0.056	2.1×10^{-16}
1	LRRIQ3	-0.06	0	-0.060	6×10^{-10}
1	GSTM4	-0.22	-0.011	-0.207	8×10^{-8}
2	SP140	-0.07	0	-0.075	5×10^{-7}
2	KHK	0.27	0.005	0.269	2.2×10^{-7}
2	TRMT61B	-0.07	0	-0.073	1.2×10^{-6}
3	PCCB	0.09	0.001	0.092	1.1×10^{-9}
3	PCYT1A	-0.04	-0.033	-0.004	1.6×10^{-6}
3	PIGX	-0.08	-0.072	-0.004	2.2×10^{-7}
3	ULK4	-0.09	-0.008	-0.078	2×10^{-10}
3	DHFR2	-0.09	-0.093	0	1.2×10^{-12}
3	CD47	0.01	0	0.005	3.4×10^{-10}
5	NUDT12	0.47	0.465	0	2.8×10^{-6}
5	RHOBTB3	0.12	0.051	0.065	2×10^{-6}
5	ERAP1	0.15	0.007	0.144	6.8×10^{-7}
6	MOCS1	-0.10	-0.099	0	1.6×10^{-16}
6	TCF19	-0.05	0	-0.049	3.9×10^{-10}
6	HLA-C	-0.05	0	-0.046	8.7×10^{-11}
6	HLA-DRB5	-0.04	0	-0.045	2.9×10^{-12}
6	HLA-DRB1	-0.10	0	-0.101	2×10^{-11}
6	ZSCAN26	0.14	0.101	0.042	6.1×10^{-10}
6	BTN3A2	0.04	0	0.037	3.4×10^{-11}
6	RNF5	0.21	0.019	0.186	5.9×10^{-16}
6	HLA-DQA1	-0.03	0	-0.032	1.3×10^{-8}
6	HLA-DQB1	-0.03	0	-0.030	3.3×10^{-8}
6	CCHCR1	-0.07	0	-0.067	1.5×10^{-8}
7	FAM220A	0.08	0.032	0.051	4.3×10^{-7}
8	FAM167A	-0.03	0	-0.027	5.3×10^{-14}
8	FUT10	0.02	0	0.023	2.5×10^{-6}
10	HPS1	-0.05	0	-0.053	7.5×10^{-9}
10	FBXL15	-0.17	-0.174	0	3.3×10^{-11}
10	DPYSL4	-0.02	-0.024	0	3.5×10^{-7}
10	SFXN2	0.03	0	0.035	5.9×10^{-10}
10	TRIM8	0.06	0	0.061	1.5×10^{-11}
11	ACP2	-0.12	-0.005	-0.118	2.6×10^{-20}
11	ARL14EP	-0.04	0	-0.043	3.5×10^{-6}
11	LRRC56	-0.04	-0.007	-0.033	4×10^{-8}
11	OVOL1	0.09	0	0.093	4.2×10^{-14}
11	PPP1R14B	0.03	0	0.032	3.5×10^{-6}
11	TRAPPC4	-0.05	0	-0.046	5.8×10^{-7}
12	CACNA2D4	0.05	0.053	0	1.6×10^{-6}
12	CFAP54	-0.05	0	-0.046	2×10^{-7}
12	RPS26	0.09	0.010	0.081	1.1×10^{-7}
12	HCAR3	0.15	0	0.155	4.8×10^{-14}
14	PSEN1	-0.06	-0.009	-0.054	1.5×10^{-8}
15	ADPGK	0.05	0	0.048	2.1×10^{-6}

Continuation of Table S10					
Chr	Gene	Effect size	EUR Effect size	AFR Effect size	p-value
15	DIS3L	0.05	0	0.054	1.5×10^{-9}
15	MINAR1	-0.10	-0.103	0	9.4×10^{-8}
15	SCAMP5	-0.55	-0.552	0	1.8×10^{-6}
16	TBX6	-0.06	-0.014	-0.048	7×10^{-24}
16	TAOK2	-0.16	-0.117	-0.040	5.6×10^{-24}
16	NMRAL1	0.13	0.007	0.127	2.9×10^{-6}
17	CBX1	-0.14	-0.005	-0.137	5.2×10^{-11}
17	FTSJ3	0.05	0.001	0.050	4.4×10^{-7}
17	LRRC46	-0.07	0	-0.065	4.9×10^{-7}
17	NARF	-0.06	-0.008	-0.050	8.5×10^{-8}
17	NPLOC4	0.14	0.137	0	8.7×10^{-8}
17	AARSD1	0.09	0.048	0.037	4.7×10^{-8}
19	PEPD	-0.22	-0.020	-0.201	2.2×10^{-8}
19	ZNF554	0.11	0.017	0.097	1.2×10^{-7}
20	PRPF6	0.03	0	0.027	2.2×10^{-6}

Note: The columns EUR Effect size and AFR Effect size denote the effect of the gene on BMI in European and African ancestries respectively.

Table S11: Common associated genes for BMI identified by glossTWAS and METRO in the multi-ancestry TWAS for European and African ancestries.

		glossTWAS				METRO	
Chr	Gene	Effect size	SD	PPNA	$\log_{10}$(BF)	Effect size	p-value
1	ZCCHC17	0.12	0.09	3.5×10^{-2}	5.2	0.16	1.1×10^{-6}
5	RHOBTB3	0.14	0.08	1.6×10^{-5}	8.5	0.12	2×10^{-6}
6	ZSCAN26	0.09	0.04	2.3×10^{-4}	7.4	0.14	6.1×10^{-10}
6	RNF5	0.04	0.01	2.3×10^{-7}	10.4	0.21	5.9×10^{-16}
11	ACP2	-0.29	0.25	3.6×10^{-18}	21.2	-0.12	2.6×10^{-20}
16	TBX6	-0.07	0.02	9.1×10^{-20}	22.8	-0.06	7×10^{-24}
16	TAOK2	-0.17	0.07	7.5×10^{-21}	23.9	-0.16	5.6×10^{-24}
17	CBX1	0.19	0.11	6.9×10^{-10}	12.9	-0.14	5.2×10^{-11}

Note: The abbreviations SD, PPNA, and BF represent the standard deviation, posterior probability of null association, and Bayes factor, respectively. For glossTWAS, the results are obtained based on 7000 MCMC iterations with a burn-in sample size of 2000 and a thinning period of 5.

Table S12: Novel associated genes for BMI identified by glossTWAS in the single-ancestry TWAS for European ancestry.

Chr	Gene	Trait
1	ANKRD65	-
2	ATRAID	-
2	PIGF	Height, Gait measurement
3	ABHD14B	Height, Aminoacylase-1 levels
3	ZNF589	Height, Weight
3	ZNF654	Type 2 diabetes, Mean corpuscular volume
3	WDR6	Eosinophil counts
3	UBA7	Insomnia, Type 2 diabetes
3	ZNF501	-
4	MANBA	Psoriasis, Fasting glucose
4	GRK4	Body fat percentage, Hip circumference adjusted for BMI
4	GUF1	Insomnia
4	SMIM20	Height, Hip index, Waist-hip ratio
5	RNF14	Superior temporal gyrus volume measurement
6	MOCS1	Height, LDL cholesterol levels
6	ZNF192P1	-
6	ZSCAN26	Anxiety, Stress-related disorder
6	HLA-DPA1	Hip circumference adjusted for BMI, Bone mineral density mean
6	RNF5	Fat body mass, Waist circumference
6	SKIV2L	-
7	PPIA	Height, Bone mineral density mean
8	INTS8	Height, LDL cholesterol levels
10	HPS1	Calcium levels, Metabolite levels (N6-methyllysine)
10	FBXL15	-
10	MFSD13A	Height, Waist circumference adjusted for body mass index
10	NRBF2	Schizophrenia, Systolic blood pressure
10	VDAC2	Pneumonia
10	BNIP3	Height, Systolic blood pressure
10	ADO	Hip circumference adjusted for BMI, Whole body fat-free mass
11	NR1H3	HDL cholesterol, Triglyceride levels
11	ARFGAP2	Major depressive disorder
11	THYN1	neutrophil measurement
11	FIBP	Hip circumference adjusted for BMI, Inflammatory bowel disease
11	SNX32	Waist-hip index, Waist circumference adjusted for body mass index
11	CCS	Depression
11	EIF1AD	Height, Waist circumference adjusted for body mass index
12	SLC25A3	Hip circumference adjusted for BMI, Serum levels of protein C1QTNF5
12	NT5DC3	Central corneal thickness, Attention deficit hyperactivity disorder
12	MORN3	Height, Waist-hip ratio
12	TCTN2	Height, Waist-hip index
14	BAG5	Depression, Insomnia
16	KNOP1	Height, Insomnia
16	DDX19B	Height, Systolic blood pressure
17	CBX1	Height, Body fat rate
17	UBE2Z	Height, Weight
17	C17orf107	Alzheimer's disease, Serum alkaline phosphatase levels
17	CCDC200	Heel bone mineral density, Triglyceride levels
19	TCF3	Height, Triglyceride levels
19	ANKRD24	Height, Cancer

Continuation of Table S12		
Chr	Gene	Trait
19	KPTN	gait measurement
19	SIX5	Primary open angle glaucoma
19	ADAT3	-

Note: The genes in the table are not associated with BMI but are reported to be associated with other traits in the NHGRI-EBI GWAS catalog. A few of them are mentioned here.

Table S13: Novel associated genes for BMI identified by glossTWAS in the multi-ancestry TWAS for European and African ancestries.

Chr	Gene	Trait
1	ZCCHC17	Schizophrenia, Color vision defects
5	RHOBTB3	Height, Metabolite levels
6	MOCS1	Height, LDL cholesterol levels
6	ZNF192P1	-
6	ZSCAN26	Anxiety, Stress-related disorder
6	RNF5	Fat body mass, Waist circumference
10	HPS1	Calcium levels, Metabolite levels (N6-methyllysine)
10	FBXL15	-
12	CFAP54	Height, Bone density
17	CBX1	Height, Body fat rate

Note: The genes in the table are not associated with BMI but are reported to be associated with other traits in the NHGRI-EBI GWAS catalog. A few of them are mentioned here.