Multivariate adaptive shrinkage improves cross-population transcriptome prediction for transcriptome-wide association studies in underrepresented populations.

Daniel S. Araujo[1], Chris Nguyen[2], Xiaowei Hu[3], Anna V. Mikhaylova[4], Chris Gignoux[5], Kristin Ardlie[6], Kent D. Taylor[7], Peter Durda[8], Yongmei Liu[9], George Papanicolaou[10], Michael H. Cho[11], Stephen S. Rich[3], Jerome I. Rotter[7], NHLBI TOPMed Consortium, Hae Kyung Im[12], Ani Manichaikul[3], Heather E. Wheeler[1,2,*]

[1]Program in Bioinformatics, Loyola University Chicago, Chicago, IL, 60660, USA
[2]Department of Biology, Loyola University Chicago, Chicago, IL, 60660, USA
[3]Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, 22908, USA
[4]Department of Biostatistics, University of Washington, Seattle, WA, 98195, USA
[5]Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, UC Denver Anschutz Medical Campus, Aurora, CO, 80045, USA
[6]Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA
[7]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, 90502, USA
[8]Laboratory for Clinical Biochemistry Research, University of Vermont, Colchester, VT, 05446, USA
[9]Department of Medicine, Duke University School of Medicine, Durham, NC, 27710, USA
[10]Epidemiology Branch, Division of Cardiovascular Sciences, National Heart, Lung and Blood Institute, Bethesda, MD, 20892, USA
[11]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, 02115, USA
[12]Section of Genetic Medicine, The University of Chicago, Chicago, IL, 60637, USA
[*]Corresponding author: hwheeler1@luc.edu

Keywords: genetics, genomics, human genetics, transcriptome-wide association studies

Abstract:

Transcriptome prediction models built with data from European-descent individuals are less accurate when applied to different populations because of differences in linkage disequilibrium patterns and allele frequencies. We hypothesized methods that leverage shared regulatory effects across different conditions, in this case, across different populations may improve cross-population transcriptome prediction. To test this hypothesis, we made transcriptome prediction models for use in transcriptome-wide association studies (TWAS) using different methods (Elastic Net, Joint-Tissue Imputation (JTI), Matrix eQTL, Multivariate Adaptive Shrinkage in R (MASHR), and Transcriptome-Integrated Genetic Association Resource (TIGAR)) and tested their out-of-sample transcriptome prediction accuracy in population-matched and cross-population scenarios. Additionally, to evaluate model applicability in TWAS, we integrated publicly available multi-ethnic genome-wide association study (GWAS) summary statistics from the Population Architecture using Genomics and Epidemiology Study (PAGE) and Pan-UK Biobank with our developed transcriptome prediction models. In regard to transcriptome prediction accuracy, MASHR models performed better or the same as other methods in both population-matched and cross-population transcriptome predictions. Furthermore, in multi-ethnic TWAS, MASHR models yielded more discoveries that replicate in both PAGE and PanUKBB across all methods analyzed, including loci previously mapped in GWAS and new loci previously not found in GWAS. Overall, our study demonstrates the importance of using methods that benefit

from different populations' effect size estimates in order to improve TWAS for multi-ethnic or underrepresented populations.

1. INTRODUCTION

Through genome-wide association studies (GWAS), many associations between single nucleotide polymorphisms (SNPs) and diverse phenotypes have been uncovered[1]. However, most GWAS to date have been conducted on individuals of European descent, even though they make up less than one fifth of the total global population[2,3]. Ancestry diversity in human genetic studies is important because as linkage disequilibrium and allele frequencies differ among populations, associations found within European ancestry individuals may not reflect associations for individuals of other ancestries and vice versa[3]. Some efforts to increase ancestry diversity in human genetics studies include the NHLBI Trans-Omics for Precision Medicine (TOPMed) consortium[4], the Population Architecture using Genomics and Epidemiology (PAGE) study[5], the Human Heredity and Health in Africa (H3Africa) initiative[6], and the Pan-ancestry genetic analysis of the UK Biobank (PanUKBB[7]).

Alongside GWAS, transcriptome-wide association studies (TWAS) test predicted gene expression levels for association with complex traits of interest, identifying gene-trait associated pairs[8]. Different TWAS methods, such as PrediXcan and FUSION, work by estimating gene expression through genotype data using transcriptomic prediction models built on expression quantitative trait loci (eQTL) data[9,10]. Similarly to GWAS, TWAS are also negatively affected by ancestry underrepresentation, as gene expression prediction models for use in TWAS are often trained in European descent datasets, which reduces the power of studies conducted with individuals of other ancestries[11,12]. Still, we expect the underlying biological mechanisms of complex traits to be shared across human populations[11,13], and thus prediction methods that account for allelic heterogeneity and better estimate effect sizes can improve the discovery rate and interpretation of TWAS across populations.

Here, we used genomic and transcriptomic data from the Multi-Ethnic Study of Atherosclerosis (MESA)[14] multi-omics pilot study of TOPMed to build TWAS prediction models (Figure 1). Using five different methods to estimate effect sizes, Elastic-Net[15,16], Joint-Tissue Imputation (JTI)[17], Matrix eQTL[18], multivariate adaptive shrinkage (MASHR)[19], and Transcriptome-Integrated Genetic Association Resource (TIGAR)[20], we built population-specific transcriptomic prediction models for four MESA-defined populations – African American, Chinese, European, and Hispanic/Latino – across three blood cell types and evaluated their prediction performance in the Geuvadis[21] cohort using PrediXcan[9]. From there, we used S-PrediXcan[22] to apply our models to GWAS summary statistics of complex traits from the multi-ethnic PAGE[5] study and PanUKBB[7]. We hypothesized that MASHR and JTI were most likely to improve transcriptome prediction and increase the number of TWAS hits in comparison to the other methods, as they both leverage similar effect size estimates across different conditions - in this case, different populations - to adjust effect sizes. In agreement to that, our results indicated that in cross-population predictions, MASHR models have a higher transcriptome prediction accuracy than other methods. Furthermore, in our TWAS, MASHR models discovered the highest number of associated gene-trait pairs across all population models. These findings illustrate that leveraging genetic diversity and effect size estimates across populations can help improve current transcriptome prediction models, which may increase discovery and replication in association studies in underrepresented populations or multi-ethnic cohorts.

2. METHODS

a. Training dataset

To build our transcriptome prediction models, we used data from the Multi-Ethnic Study of Atherosclerosis (MESA)[14] multi-omics pilot study of the NHLBI Trans-Omics for Precision Medicine (TOPMed) consortium. This data set includes genotypes derived from whole genome sequencing and transcripts per million (TPM) values derived from RNA-Seq for individuals of four different populations – African American (AFA), Chinese (CHN), European (EUR), and Hispanic/Latino (HIS) – for three different blood cell types: peripheral blood mononuclear cells (PBMC, ALL n = 1287, AFA n = 334, CHN n = 104, EUR n = 528, HIS n= 321), CD16+ monocytes (Mono, ALL n = 395, AFA n = 75, EUR n = 221, HIS n = 99), and CD4+ T-cells (T cells, ALL n = 397, AFA n = 75, EUR n = 224, HIS n = 98).

b. Genotype and RNA-Seq QC

We performed QC on each MESA tissue-population pair separately. For the genotype data[4] (Freeze 8, phs001416.v2.p1), we excluded INDELs, multi-allelic SNPs, and ambiguous-strand SNPs (A/T, T/A, C/G, G/C), and removed the remaining variants with MAF < 0.01 and HWE < 1×10^{-6} using PLINK[23] v1.9. For chromosome X, filtering by HWE was only applied in variants found within the pseudoautosomal regions based on GRCh38 positions. Furthermore, for the non-pseudoautosomal region of X, male dosages were assigned either 0 or 2. After QC, the average numbers of non-ambiguous SNPs remaining per population across all cell types were: AFA = 15.7M; CHN = 8.4M; EUR = 9.7M; HIS = 13.2M.

For the RNA-Seq data, we also performed QC separately by tissue-population. First, we removed genes with average TPM values < 0.1. For some individuals, RNA expression levels were measured at two different time points (Exam 1 and Exam 5); thus, after log-transforming each measurement and adjusting for age and sex as covariates using linear regression and extracting the residuals, we took the mean of the two time points (or the single adjusted log-transformed value, if expression levels were only measured once), performed rank-based inverse normal transformation, and adjusted for the first 10 genotype and 10 expression PCs. To estimate principal components, we used PC-AiR[24] with kinship threshold of ~0.022, which corresponds to 4th degree relatives. No individuals were removed. For each tissue, we removed genes absent in at least one population. After QC, we had 17,585 genes in PBMC, 14,503 in Mono, and 16,647 in T cells. We used GENCODE[25] annotation v38 to annotate gene types (e.g. protein-coding, lncRNA, etc.) and gene transcription start and end sites.

c. Gene expression cis-heritability estimation

We estimated gene expression heritability (h^2) using cis-SNPs within the 1Mb region upstream of the transcription start site and 1Mb region downstream of the transcription end site. Using the genotype data filtered only by HWE P-value < 1×10^{-6}, for each tissue-population pair, we first performed LD-pruning with a 500 variants count window, a 50 variants count step, and a 0.2 r^2 threshold using PLINK[23] v1.9. Then, for each gene, we extracted cis-SNPs and excluded SNPs with MAF < 0.01. Finally, to assess cis-SNP expression heritability, we estimated the genetic relationship matrix and h^2 using GCTA-GREML[26] with the "--reml-no-constrain" option. We considered a gene heritable if it had a positive h^2 estimate (h^2 - 2*S.E. > 0.01 and p-value < 0.05) in at least one MESA population. In total, 9,206 genes were heritable in PBMC, 3,804 in Mono, and 4,053 in T cells. We only built transcriptome prediction models for these heritable genes across all populations in their respective cell types.

d. Transcriptome prediction models

With the aforementioned genotype and gene expression data, we built transcriptome prediction models for each MESA tissue-population pair, and for each gene we considered cis-SNPs as defined in the previous section. Additionally, we only considered SNPs present in the GWAS summary statistics of the Population Architecture using Genomics and Epidemiology (PAGE) study[5] to build our prediction models to make sure that there would be a high overlap between SNPs in the transcriptome models and SNPs in the GWAS summary statistics. After merging with PAGE SNPs, the average numbers of SNPs left in our dataset were: AFA = 12.8M; CHN = 6.2M; EUR = 7.4M; HIS = 10.5M.

We built our population-based models using five different approaches. The first was elastic-net (EN) regression using the *glmnet* package in R[15,16], with mixing parameter $\alpha = 0.5$. We considered EN as our baseline model, as it has been previously used to make transcriptome prediction models for the TOPMed MESA data[27].

The second method implemented was mash (Multivariate Adaptive Shrinkage)[19] in R (MASHR). Unlike EN, MASHR does not estimate weights by itself; rather, it takes z-score (or weight and standard error) matrices as input and adjusts them based on correlation patterns present in the data in an empirical Bayes algorithm, allowing for both shared and condition-specific effects. By doing so, MASHR increases power and effect size estimation accuracy[19]. Originally, MASHR applicability was demonstrated by leveraging effect size estimates across different tissues[19], however, herein we sought to assess its potential to leverage effect sizes across populations. We ran MASHR for each gene at a time, using cis-SNPs weights (effect sizes) estimated by Matrix eQTL[18] and MESA populations as different conditions (Figure 2A). Then, we split MASHR-adjusted weights according to their respective populations, and selected the top SNP (lowest local false sign rate) per gene to determine which SNPs would end up in the final models (Figure 2B). Local false sign rate is similar to false discovery rate, but it is more rigorous as it also takes into account the direction of effect[19]. Thus, by selecting one top SNP per population, the maximum number of SNPs per gene in the final model is 4, which corresponds to the number of populations in our study. If two or more populations had the same variant as top SNP, it was only included once. To make population-based models, we used population-specific effect sizes, taken from the corresponding MASHR output matrices.

The third method was based on the unadjusted effect sizes estimated by Matrix eQTL[18] using the linear regression model. We used the same approach taken to build the MASHR models, including the SNP with the lowest p-value from each population, but the key difference is that we made the models using the unadjusted effect sizes.

The fourth method we used was Transcriptome-Integrated Genetic Association Resource (TIGAR), which trains transcriptome imputation models using either EN or nonparametric Bayesian Dirichlet Process Regression (DPR)[20]. As we already used EN to make a set of transcriptome prediction models, we opted to make DPR-based models. We used TIGAR's default parameters to train our models, such as using the Variational Bayesian algorithm and outputting fixed effect sizes. However, by default, TIGAR performs 5-fold cross validation (CV) during training, and only outputs results if the final average CV R^2 is equal or greater than 0.005; thus, since we did not implement CV for any of the aforementioned methods and instead tested performance in an independent sample, we opted to skip this step of TIGAR's pipeline and generate outputs for all genes. Most gene models generated by TIGAR had

hundreds of SNPs with near-zero effect sizes. To reduce memory requirements for storage of these models, we removed SNPs with effect sizes smaller than 1×10^{-4}.

The fifth and last method we implemented was Joint-Tissue Imputation (JTI)[17]. JTI was designed to leverage similarity in gene expression and DNase 1 hypersensitive sites across different tissues to possibly improve prediction performance. Thus, similarly to MASHR, we sought to assess whether the method could be adapted to use populations instead of tissues. To assess gene expression similarity between MESA populations, we computed transcriptome-wide pairwise correlations between populations using the median TPM value per gene. Additionally, we did not have population DNase 1 hypersensitivity site data, so we set column five to 1 in our input files. By default, JTI performs 5-fold CV and only produces outputs for genes with average CV R greater than 0.1. Thus, similarly to TIGAR, we removed this filtering step of the pipeline to generate output for all genes regardless of CV performance.

To perform TWAS using GWAS summary statistics data, it is necessary to have information about the correlation between the SNPs used to predict gene expression levels[22]. Thus, for all our transcriptome prediction models previously mentioned, we computed pairwise covariances for the SNPs within each TOPMed MESA population model using the respective population dosage data. All model files are freely available for anyone to use (see Data Availability section).

e. Assessing transcriptome prediction performance

To evaluate the gene expression prediction performance of all our transcriptome prediction models, we used DNA and lymphoblastoid cell lines RNA-Seq data from 449 individuals in the Geuvadis[21] study. Individuals within the testing dataset belong to five different populations (Utah residents with Northern and Western European ancestry (CEU), n = 91; Finnish in Finland (FIN), n = 92; British in England and Scotland (GBR), n = 86; Toscani in Italy (TSI), n = 91; Yoruba in Ibadan, Nigeria (YRI), n = 89), which we analyzed both separately and together (ALL). Similarly to our training dataset, we performed rank-based inverse normal transformation on the gene expression levels, and adjusted for the first 10 genotype and 10 expression PCs, using the residuals as observed expression levels. With the Geuvadis genotype data and our transcriptome prediction models, we used PrediXcan[9] to estimate gene expression levels. PrediXcan is a two-step TWAS method, in which the first step is to estimate genetically regulated expression levels (GReX). Thus, to assess transcriptome prediction performance, we compared GReX to the adjusted, measured expression levels using Spearman correlation.

f. Assessing performance in transcriptome-wide association studies

To test the applicability of our transcriptome prediction models in multi-ethnic association studies, we applied S-PrediXcan[22] to GWAS summary statistics from the Population Architecture using Genomics and Epidemiology (PAGE) study[5]. The PAGE study consists of 28 different phenotypes tested for association with variants within a multi-ethnic, non-European cohort of 49,839 individuals (Hispanic/Latino (n=22,216), African American (n=17,299), Asian (n=4,680), Native Hawaiian (n=3,940), Native American (n=652) or Other (n=1,052)). Since we tested multiple phenotypes and transcriptome prediction models in our TWAS, we used a conservative approach and considered genes as significantly associated with a phenotype if the association p-value was less than the standard Bonferroni corrected GWAS significance threshold of 5×10^{-8}.

To replicate the associations found in PAGE, we also applied S-PrediXcan[19] to PanUKBB[7] GWAS summary statistics (N=441,331; European (n=420,531), Central/South Asian (n=8,876), African (n=6,636), East Asian (n=2,709), Middle Eastern (n=1,599) or Admixed American (n=980)). For similarity purposes, we selected summary statistics of phenotypes that overlap with the ones tested in PAGE (Table S1). As previously described, a gene-trait pair association was considered significant if its p-value was less than the Bonferroni corrected GWAS significance threshold of 5×10^{-8}. Furthermore, we deemed significant gene-trait pair associations as replicated if they were detected by the same MESA tissue-population model and had the same direction of effect in PAGE and PanUKBB. To assess if the gene-trait association pairs reported in our study are novel or not, we compared them to studies found in the GWAS Catalog[1] (All associations v1.0.2 file downloaded on 11/9/2022).

3. RESULTS

a. Increased sample sizes improve gene expression cis-heritability estimation

With the goal of improving transcriptome prediction in diverse populations, we first determined which gene expression traits were heritable and thus amenable to genetic prediction, using genome-wide genotype and RNA-Seq data from three blood cell types (PBMCs, monocytes, T cells) in TOPMed MESA. We estimated cis-heritability (h^2) using data from four different populations (African American - AFA, Chinese - CHN, European - EUR, and Hispanic/Latino - HIS). Variation in h^2 estimation between populations is expected due to differences in allele frequencies and LD patterns; however, we show that larger population sample sizes yield more significant (p-value < 0.05) h^2 estimates (Figure 3). Using the PBMC dataset as an example, with the EUR dataset (n = 528), we assessed h^2 for 10,228 genes, however, we estimated h^2 for 8,765 genes using the AFA dataset (n = 334) (Figure 3A). Moreover, we see a great impact on the CHN population, which has the smallest sample size. For that population, we managed to estimate h^2 for only 3,448 genes. The same pattern repeats when analyzing only the heritable genes (h^2 lower bound > 0.01). In EUR, 6,902 genes were deemed heritable, whereas in AFA and CHN the amount of heritable genes is 5,537 and 1,367, respectively (Figure 3B). Thus, larger sample sizes are needed to better pinpoint h^2 estimates, especially in non-European populations. In total, analyzing the union across all populations' results, we detected 9,206 heritable genes in PBMCs, 3,804 in monocytes, and 4,053 in T Cells.

b. MASHR models improve cross-population transcriptome prediction performance

To improve TWAS power for discovery and replication across all populations, we sought to improve cross-population transcriptome prediction accuracy. For this, we used data from four different populations and built gene expression prediction models using five different methods (Elastic Net (EN), Transcriptome-Integrated Genetic Association Resource (TIGAR), Matrix eQTL, multivariate adaptive shrinkage in R (MASHR), and Joint-Tissue Imputation (JTI)). We chose EN as a baseline approach for comparison in our analysis, as it has been previously shown to have better performance than other common machine learning methods such as random forest, K-nearest neighbor, and support vector regression[28]. Furthermore, we trained gene expression prediction models by applying TIGAR's nonparametric Bayesian Dirichlet Process Regression pipeline[20]. Using Matrix eQTL, we estimated univariate effect sizes for each cis-SNP-gene relationship and we developed an algorithm to include top SNPs from each population, but population-estimated effect sizes in each population's model (Figure 2). Matrix eQTL effect sizes are the input for MASHR, which we hypothesized might better estimate cross-population effect sizes, due to its flexibility in allowing both shared and population-specific effects[19,29]. Similarly, JTI was designed to leverage correlation across different tissues

to improve gene expression prediction[17]; thus, we also adapted its pipeline to perform cross-population leveraging. By filtering our models to include only genes with positive h^2 (h^2 lower bound > 0.01) in at least one population, we saw that among all methods used, we obtained more gene models in MatrixeQTL and MASHR (Figure 4A). The difference is especially greater in the CHN population model.

To evaluate model performance at population-matched and cross-population transcriptome predictions, we used data from the Geuvadis study, which comprises individuals of West African or European descent. We defined "population-matched predictions" as the scenarios in which the transcriptome model MESA training data and Geuvadis test data have the closest genetic distance with available data, and we defined "cross-population predictions" as any other pairs (Figure S1). Overall, across all Geuvadis populations, the methods tested show distinct performances (Figure S2). This result, however, may be influenced by the fact that different transcriptome models have a different number of genes in them (Figure 4A). Thus, we sought to compare performances considering the intersection of genes with expression predicted by all methods. Focusing on Geuvadis GBR and YRI populations, which have similar sample sizes and are of distinct continental ancestries, we observed that MASHR models significantly outperform the other methods in cross-population transcriptome predictions, as seen in the AFA-GBR and EUR-YRI MESA-Geuvadis populations pairs (Figure 4B, Table S2). The only exception is in AFA-GBR, in which MASHR and MatrixeQTL have similar performances. Additionally, in population-matched scenarios (AFA-YRI and EUR-GBR), prediction performance does not significantly differ between MASHR, MatrixeQTL, and EN. All three aforementioned methods significantly outperform JTI and TIGAR in population-matched predictions (Table S2). Moreover, we also performed pairwise comparisons between all methods using all Geuvadis populations, taking into account the intersection of genes with expression predicted in each case. Overall, across all MESA transcriptome models and Geuvadis populations, MASHR models either performed better or the same as other methods in both population-matched or cross-population transcriptome prediction scenarios (Table S3).

c. Leveraging effect sizes across different populations improves discovery rate in multi-ethnic TWAS

In order to investigate the applicability of the models we built in multi-ethnic TWAS, we used S-PrediXcan with GWAS summary statistics of complex traits from PAGE and PanUKBB. We show that across all tissue-population models, MASHR identified the highest number of gene-trait pair associations (208) that replicated in both PAGE and PanUKBB ($P < 5 \times 10^{-8}$), followed by Matrix eQTL (173), JTI (131), EN (94), and TIGAR (91) (Table S3). When analyzing the total number of discoveries separately for each population, MASHR had the highest number of gene-trait pairs in most population models (Figure 5A). The only exception is with HIS models, in which both MASHR and MatrixeQTL had the same number of discoveries. The discovery rate improvement by MASHR is exceptionally high in CHN models, as it had almost twice the number of discoveries as the second-highest method (27 by MASHR vs. 14 by MatrixeQTL).

Additionally, when comparing gene-trait pairs, we saw that most MASHR hits were shared between population models, whereas other methods have higher population-specific discoveries (Figure 5B). Most MatrixeQTL hits were also shared by many population models, but not to the same degree as MASHR. Altogether, these findings indicate that MASHR models show high consistency and also suggest that TWAS results are not as affected by the MASHR population model used as compared to other methods.

To contextualize our models' findings, we investigated whether the discovered gene-trait pairs had been previously reported in any studies in the GWAS Catalog (https://www.ebi.ac.uk/gwas/home). We saw that across 105 distinct gene-trait pairs associations found (totaling 697 across all models), 38 (36.19%) have not been reported in the GWAS Catalog, and therefore may be novel associations that require further investigation (Table S4). Out of those potential new biological associations, most of them (13) were discovered with MASHR AFA models (Table S4). Furthermore, out of the 67 distinct known GWAS catalog associations discovered, MASHR models identified most of them (Table S3). For instance, MASHR EUR models found 34 known associations, followed by MASHR AFA with 33, and MatrixeQTL EUR with 32 (Figure S4).

4. DISCUSSION

In this work, we sought to build population-based transcriptome prediction models for TWAS using data from the TOPMed MESA cohort using five distinct approaches. We saw that although the AFA and HIS populations' datasets contained the highest numbers of SNPs after quality control, EUR yielded the highest number of gene expression traits with significant heritability estimates across all tissues analyzed. This is most likely due to the higher sample size in EUR in comparison to AFA and HIS, as larger sample sizes provide higher statistical power to detect eQTLs with smaller effects[30]. Furthermore, we saw that the number of genes in each population transcriptome model is not the same across all methods tested. Some transcriptome prediction models, such as the ones built using EN or JTI, only contain genes for which the SNPs effect sizes converged during training, which is not a limiting factor for MASHR, MatrixeQTL, and TIGAR. One of the factors that impacts the number of genes for which SNPs effect sizes converge during training is sample size, which explains the lower number of genes in the EN and JTI CHN model in comparison to other population models. Furthermore, although sample size does not impact the number of gene models trained for TIGAR in the same degree as EN and JTI, it influences SNP effect size estimation[31]. Thus, when we removed SNPs with near-zero effects, there was a drop in the number of genes in the final population transcriptome models for TIGAR. Test data sample size has also been shown to positively correlate with gene expression prediction accuracy[32].

In addition to sample size, gene expression prediction accuracy is known to be greater when the training and testing datasets have similar ancestries[11,12,32,33]; however, non-European ancestries are vastly underrepresented in human genetics studies[2,3], which compromises the ability to build accurate TWAS models for them. Thus, using data from the Geuvadis cohort, we evaluated the transcriptome prediction performance of our models and found that MASHR models either significantly outperformed all other methods tested, or had similar performance. Previous studies have shown that by borrowing information across different conditions, such as tissues[19] or cell types[34], MASHR identifies shared- or condition-specific eQTLs, which can enhance causal gene identification[29], as well as improve effect size estimation accuracy[19]. Similarly, by leveraging effect size estimates across multiple populations, MASHR improved cross-population transcriptome prediction without compromising population-matched prediction accuracy. Interestingly, another method we tested, JTI, was also originally designed to leverage similarity in gene expression and DNase 1 hypersensitive sites across tissues in order to improve transcriptome prediction accuracy[17]. However, our results showed that it performed worse than MASHR and the same as EN in cross-population transcriptome prediction. This suggests that distinct cross-condition leveraging frameworks may have different performances when applied across populations. One possible reason for differences in performance is that JTI

uses EN weighted by condition similarity to estimate effect sizes and select SNPs to be included in the final models, whereas for MASHR, our pipeline selects one SNP per condition. Since more SNPs with less significant effect sizes were included in our EN and JTI models, greater uncertainty in effect sizes likely led to lower transcriptome prediction accuracy compared to MASHR. Furthermore, among the methods evaluated, TIGAR had the lowest prediction performance. Originally, TIGAR was benchmarked against EN, and showed better transcriptome prediction accuracy; however, unlike in our analysis, their analysis included only genes whose expression heritability was equal or lower than 0.2[20].

Discovery and replication of TWAS associations are also related to the ancestries of the transcriptome prediction model training dataset and ancestries of the TWAS sample dataset[11]. Thus, we assessed the applicability of our models in TWAS using S-PrediXcan on PAGE and PanUKBB GWAS summary statistics and found that across all tissues and populations, MASHR models yielded the highest number of total gene-trait pairs associations, with MASHR AFA reporting the highest number. In this manner, it seems that although MASHR improved gene expression prediction accuracy for all populations analyzed, using transcriptome prediction models that match the ancestries of the GWAS dataset still yields the highest number of TWAS discoveries, which is in agreement with many previous studies[11,35–38]. Our results also showed that although JTI transcriptome prediction was not as accurate as baseline EN, JTI models had more TWAS discoveries than EN. This exemplifies how integrating data from different genetic ancestries may improve TWAS.

By investigating which associations had been previously reported in the GWAS Catalog, we saw that most new discoveries were found by MASHR models. Some of these possible new discoveries are unique to MASHR models and have been corroborated previously, such as *YJEFN3* (also known as *AIBP2*) and triglycerides, whose low expression in zebrafish increases cellular unesterified cholesterol levels[39], consistent with our S-PrediXcan effect size directions (PAGE effect size = -0.52, p-value = 6.1×10^{-16}; PanUKBB effect size = -0.86, p-value = 7.1×10^{-86}). Additionally, we also saw that MASHR models showed higher consistency across the different population transcriptome prediction models, which means that TWAS results are not as affected by the population model used as other methods.

One limitation of our TWAS is that we used transcriptome prediction models trained in PBMCs, monocytes and T cells, and those tissues might not be the most appropriate for some phenotypes in PAGE or PanUKBB. Additionally, because of the smaller sample sizes for some populations in our training dataset, h^2 and eQTL effect sizes estimates have large standard errors, which may affect the ability of MASHR to adjust effect sizes across different conditions based on correlation patterns present in the data. Regardless of that, our results mainly demonstrate that we can implement cross-population effect size leveraging using a method first applied to do cross-tissue effect size leveraging - and improve cross-population transcriptome prediction accuracy in doing so. Thus, increasing sample size for underrepresented populations will improve current MASHR TWAS models' performances, as well as increase genetic diversity in the data. Another TWAS method, METRO, which implements a likelihood-based inference framework to incorporate transcriptome prediction models built on datasets of two different genetic ancestries, has also shown enhanced TWAS power[40]. METRO jointly models gene expression and the phenotype of interest[40], and thus was not directly comparable to the five methods we tested here, which all separate the transcriptome prediction step from the association test. Given that this traditional two-stage TWAS procedure ignores uncertainty in the expression prediction, the joint approach of METRO across more than two populations is an area of future TWAS methods research. Furthermore, while our study focused on transcriptome prediction, MASHR could also be adapted to possibly improve cross-population polygenic risk

scores (PRS). Indeed, other methods like PRS-CSx jointly model complex traits effects across populations in order to improve PRS[41]. MASHR is most useful when population effects are shared, as demonstrated by the more consistent S-PrediXcan results, but population-specific effects are also relevant. For instance, a study in a large African American and Latino cohort discovered eQTLs only present at appreciable allele frequencies in African ancestry populations[38]. Moreover, since our MASHR models focus on the top SNPs, we might not be including enough eQTLs in the models, especially for those genes whose expression is genetically regulated by multiple eQTLs with small effects.

In conclusion, our results demonstrate the importance and the benefits of increasing ancestry diversity in the field of human genetics, especially regarding association studies. As shown, sample size is valuable for assessing gene expression heritability and for accurately estimating eQTL effect sizes, and thus some populations are negatively affected due to the lack of data. However, by making transcriptome prediction models that leverage effect size estimates across different populations using multivariate adaptive shrinkage, we were able to increase gene expression prediction performance for scenarios in which the training data and test data have distant ("cross-population") genetic distances with available data. Additionally, when applied to multi-ethnic TWAS, the aforementioned models yielded more discoveries across all methods analyzed, even detecting well-known associations that were not detected by other methods. Thus, in order to further improve TWAS in multi-ethnic or underrepresented populations and possibly reduce health care disparities, it is necessary to use methods that consider shared- and population-specific effect sizes, as well as increase available data of underrepresented populations.

5. ACKNOWLEDGEMENTS

This work is supported by the NIH National Human Genome Research Institute Academic Research Enhancement Award R15 HG009569 (HEW). Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). WGS for "NHLBI TOPMed: Multi-Ethnic Study of Atherosclerosis (MESA)" (phs001416.v1.p1) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-13S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1), and TOPMed MESA Multi-Omics (HHSN2682015000031/HSN26800004). The MESA projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for the Multi-Ethnic Study of Atherosclerosis (MESA) projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1TR001881, DK063491, and R01HL105756. The MESA Epigenomics and Transcriptomics Studies were funded by National Institutes of Health grants 1R01HL101250, 1RF1AG054474, R01HL126477, R01DK101921, and R01HL135009. The authors thank the other investigators, the staff, and the participants of the

MESA study for their valuable contributions. A full list of participating MESA investigators and institutes can be found at http://www.mesa-nhlbi.org.

6. DATA AVAILABILITY

All scripts used for analyses, including a pipeline to derive new MASHR models, are available at https://github.com/danielsarj/TOPMed_MESA_crosspop_portability. MESA populations prediction models and raw S-PrediXcan TWAS output files are available at https://doi.org/10.5281/zenodo.7551844. TOPMed MESA data are under controlled access in dbGaP at https://www.ncbi.nlm.nih.gov/gap/ through study accession phs001416.v2.p1. Geuvadis expression data is at Array Express (E-GEUV-1) and genotype data is at http://www.internationalgenome.org/. PAGE GWAS summary statistics are available in the GWAS Catalog at https://www.ebi.ac.uk/gwas/publications/31217584. PanUKBB GWAS summary statistics are available at https://pan.ukbb.broadinstitute.org/phenotypes/index.html.

7. DECLARATION OF INTERESTS

All authors declare that they have no conflicts of interest.

8. REFERENCES

1. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Research *47*, D1005–D1012. 10.1093/nar/gky1120.

2. Morales, J., Welter, D., Bowler, E.H., Cerezo, M., Harris, L.W., McMahon, A.C., Hall, P., Junkins, H.A., Milano, A., Hastings, E., et al. (2018). A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. Genome Biol *19*, 21. 10.1186/s13059-018-1396-2.

3. Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet *51*, 584–591. 10.1038/s41588-019-0379-x.

4. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature *590*, 290–299. 10.1038/s41586-021-03205-y.

5. Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature *570*, 514–518. 10.1038/s41586-019-1310-4.

6. The H3Africa Consortium, Matovu, E., Bucheton, B., Chisi, J., Enyaru, J., Hertz-Fowler, C., Koffi, M., Macleod, A., Mumba, D., Sidibe, I., et al. (2014). Enabling the genomic revolution in Africa. Science *344*, 1346–1348. 10.1126/science.1251546.

7. Pan UKBB Team (2022). Pan UKBB. https://pan.ukbb.broadinstitute.org/.

8. Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A.N., Knowles, D.A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., et al. (2019). Opportunities and

challenges for transcriptome-wide association studies. Nat Genet *51*, 592–599. 10.1038/s41588-019-0385-z.

9. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nature Genetics *47*, 1091–1098. 10.1038/ng.3367.

10. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet *48*, 245–252. 10.1038/ng.3506.

11. Geoffroy, E., Gregga, I., and Wheeler, H.E. (2020). Population-Matched Transcriptome Prediction Increases TWAS Discovery and Replication Rate. iScience *23*, 101850. 10.1016/j.isci.2020.101850.

12. Keys, K.L., Mak, A.C.Y., White, M.J., Eckalbar, W.L., Dahl, A.W., Mefford, J., Mikhaylova, A.V., Contreras, M.G., Elhawary, J.R., Eng, C., et al. (2020). On the cross-population generalizability of gene expression prediction models. PLOS Genetics *16*, e1008927. 10.1371/journal.pgen.1008927.

13. Hou, K., Ding, Y., Xu, Z., Wu, Y., Bhattacharya, A., Mester, R., Belbin, G.M., Buyske, S., Conti, D.V., Darst, B.F., et al. (2023). Causal effects on complex traits are similar for common variants across segments of different continental ancestries within admixed individuals. Nat Genet *55*, 549–558. 10.1038/s41588-023-01338-6.

14. Bild, D.E., Bluemke, D.A., Burke, G.L., Detrano, R., Diez Roux, A.V., Folsom, A.R., Greenland, P., JacobsJr., D.R., Kronmal, R., Liu, K., et al. (2002). Multi-Ethnic Study of Atherosclerosis: Objectives and Design. Am J Epidemiol *156*, 871–881. 10.1093/aje/kwf113.

15. Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) *67*, 301–320. 10.1111/j.1467-9868.2005.00503.x.

16. Friedman, J.H., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software *33*, 1–22. 10.18637/jss.v033.i01.

17. Zhou, D., Jiang, Y., Zhong, X., Cox, N.J., Liu, C., and Gamazon, E.R. (2020). A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. Nat Genet *52*, 1239–1246. 10.1038/s41588-020-0706-2.

18. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358. 10.1093/bioinformatics/bts163.

19. Urbut, S.M., Wang, G., Carbonetto, P., and Stephens, M. (2019). Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. Nat Genet *51*, 187–195. 10.1038/s41588-018-0268-8.

20. Nagpal, S., Meng, X., Epstein, M.P., Tsoi, L.C., Patrick, M., Gibson, G., Jager, P.L.D., Bennett, D.A., Wingo, A.P., Wingo, T.S., et al. (2019). TIGAR: An Improved Bayesian Tool for Transcriptomic Data Imputation Enhances Gene Mapping of Complex Traits. The American Journal of Human Genetics *105*, 258–266. 10.1016/j.ajhg.2019.05.018.

21. Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A.C., Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506–511. 10.1038/nature12531.

22. Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L., et al. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun *9*, 1825. 10.1038/s41467-018-03621-1.

23. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. 10.1086/519795.

24. Conomos, M.P., Miller, M.B., and Thornton, T.A. (2015). Robust Inference of Population Structure for Ancestry Prediction and Correction of Stratification in the Presence of Relatedness. Genetic Epidemiology *39*, 276–293. 10.1002/gepi.21896.

25. Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J.E., Mudge, J.M., Sisu, C., Wright, J.C., Armstrong, J., Barnes, I., et al. (2021). GENCODE 2021. Nucleic Acids Research *49*, D916–D923. 10.1093/nar/gkaa1087.

26. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat Genet *42*, 565–569. 10.1038/ng.608.

27. Mogil, L.S., Andaleon, A., Badalamenti, A., Dickinson, S.P., Guo, X., Rotter, J.I., Johnson, W.C., Im, H.K., Liu, Y., and Wheeler, H.E. (2018). Genetic architecture of gene expression traits across diverse populations. PLOS Genetics *14*, e1007586. 10.1371/journal.pgen.1007586.

28. Okoro, P.C., Schubert, R., Guo, X., Johnson, W.C., Rotter, J.I., Hoeschele, I., Liu, Y., Im, H.K., Luke, A., Dugas, L.R., et al. (2021). Transcriptome prediction performance across machine learning models and diverse ancestries. Human Genetics and Genomics Advances *2*, 100019. 10.1016/j.xhgg.2020.100019.

29. Barbeira, A.N., Melia, O.J., Liang, Y., Bonazzola, R., Wang, G., Wheeler, H.E., Aguet, F., Ardlie, K.G., Wen, X., and Im, H.K. (2020). Fine-mapping and QTL tissue-sharing information improves the reliability of causal gene identification. Genetic Epidemiology *44*, 854–867. 10.1002/gepi.22346.

30. Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi, P., Park, Y., Parsana, P., Segrè, A.V., et al. (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213. 10.1038/nature24277.

31. Parrish, R.L., Gibson, G.C., Epstein, M.P., and Yang, J. (2022). TIGAR-V2: Efficient TWAS tool with nonparametric Bayesian eQTL weights of 49 tissue types from GTEx V8. Human Genetics and Genomics Advances *3*, 100068. 10.1016/j.xhgg.2021.100068.

32. Fryett, J.J., Morris, A.P., and Cordell, H.J. (2020). Investigation of prediction accuracy and the impact of sample size, ancestry, and tissue in transcriptome-wide association studies. Genet Epidemiol *44*, 425–441. 10.1002/gepi.22290.

33. Mikhaylova, A.V., and Thornton, T.A. (2019). Accuracy of Gene Expression Prediction From Genotype Data With PrediXcan Varies Across and Within Continental Populations. Front. Genet. *10*, 261. https://doi.org/10.3389/fgene.2019.00261.

34. Sheng, X., Guan, Y., Ma, Z., Wu, J., Liu, H., Qiu, C., Vitale, S., Miao, Z., Seasock, M.J., Palmer, M., et al. (2021). Mapping the genetic architecture of human traits to cell types in the kidney identifies mechanisms of disease and potential treatments. Nat Genet *53*, 1322–1333. 10.1038/s41588-021-00909-9.

35. Schubert, R., Geoffroy, E., Gregga, I., Mulford, A.J., Aguet, F., Ardlie, K., Gerszten, R., Clish, C., Berg, D.V.D., Taylor, K.D., et al. (2022). Protein prediction for trait mapping in diverse populations. PLOS ONE *17*, e0264341. 10.1371/journal.pone.0264341.

36. Bhattacharya, A., García-Closas, M., Olshan, A.F., Perou, C.M., Troester, M.A., and Love, M.I. (2020). A framework for transcriptome-wide association studies in breast cancer in diverse study populations. Genome Biol *21*, 42. 10.1186/s13059-020-1942-6.

37. Bhattacharya, A., Hirbo, J.B., Zhou, D., Zhou, W., Zheng, J., Kanai, M., the Global Biobank Meta-analysis Initiative, Pasaniuc, B., Gamazon, E.R., and Cox, N.J. (2021). Best practices for multi-ancestry, meta-analytic transcriptome-wide association studies: lessons from the Global Biobank Meta-analysis Initiative (Genetic and Genomic Medicine) 10.1101/2021.11.24.21266825.

38. Kachuri, L., Mak, A.C.Y., Hu, D., Eng, C., Huntsman, S., Elhawary, J.R., Gupta, N., Gabriel, S., Xiao, S., Keys, K.L., et al. (2021). Gene expression in African Americans and Latinos reveals ancestry-specific patterns of genetic architecture (Genetics) 10.1101/2021.08.19.456901.

39. Fang, L., Choi, S.-H., Baek, J.S., Liu, C., Almazan, F., Ulrich, F., Wiesner, P., Taleb, A., Deer, E., Pattison, J., et al. (2013). Control of angiogenesis by AIBP-mediated cholesterol efflux. Nature *498*, 118–122. 10.1038/nature12166.

40. Li, Z., Zhao, W., Shang, L., Mosley, T.H., Kardia, S.L.R., Smith, J.A., and Zhou, X. (2022). METRO: Multi-ancestry transcriptome-wide association studies for powerful gene-trait association detection. The American Journal of Human Genetics *109*, 783–801. 10.1016/j.ajhg.2022.03.003.

41. Ruan, Y., Lin, Y.-F., Feng, Y.-C.A., Chen, C.-Y., Lam, M., Guo, Z., Stanley Global Asia Initiatives, Ahn, Y.M., Akiyama, K., Arai, M., et al. (2022). Improving polygenic prediction in ancestrally diverse populations. Nat Genet *54*, 573–580. 10.1038/s41588-022-01054-7.

Figure Legends & Supplementary Data

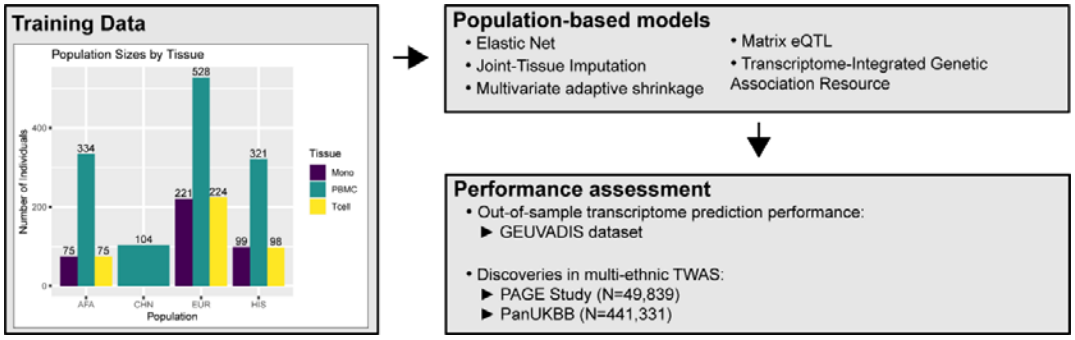


Figure 1: Overall study methodology. Using TOPMed MESA as a training dataset, we built population-based transcriptome prediction models using five different methods (Elastic Net (EN), Joint-Tissue Imputation (JTI), Multivariate adaptive shrinkage (MASHR), Matrix eQTL, and Transcriptome-Integrated Genetic Association Resource (TIGAR)). With these transcriptome models, we evaluated their out-of-sample transcriptome prediction accuracy using the GEUVADIS dataset. Additionally, we assessed their applicability in multi-ethnic TWAS using GWAS summary statistics from the PAGE Study and PanUKBB. AFA = African American, CHN = Chinese, EUR = European, HIS = Hispanic/Latino.

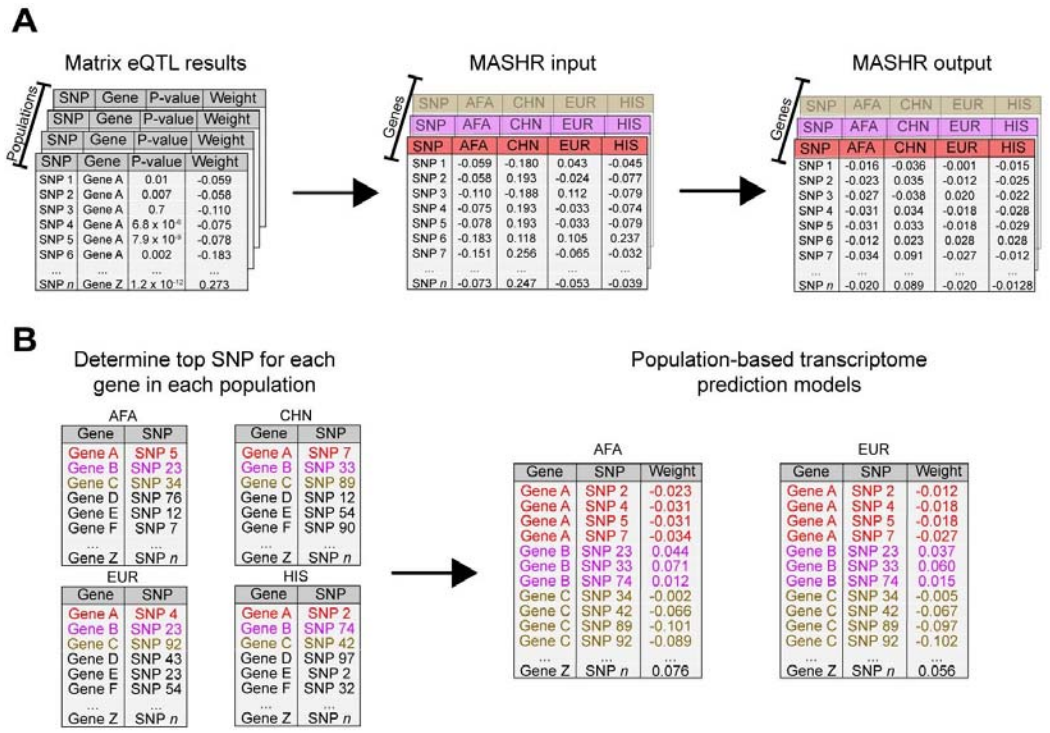


Figure 2: Design of the methodology implemented to make MASHR models. (A) Using effect sizes estimated using Matrix eQTL within each population dataset, we combined them across genes, with the different populations as conditions, to use as input for MASHR. The output matrixes contain adjusted effect sizes. (B) For each population, we selected the top SNP (lowest local false sign rate) per gene. Then, we concatenated the Gene-top SNP pairs across populations to determine which SNPs would end up in the final models. Lastly, to make our population-based transcriptome prediction models, we used population-specific effect sizes, taken from the corresponding MASHR output matrices. AFA = African American, CHN = Chinese, EUR = European, HIS = Hispanic/Latino.

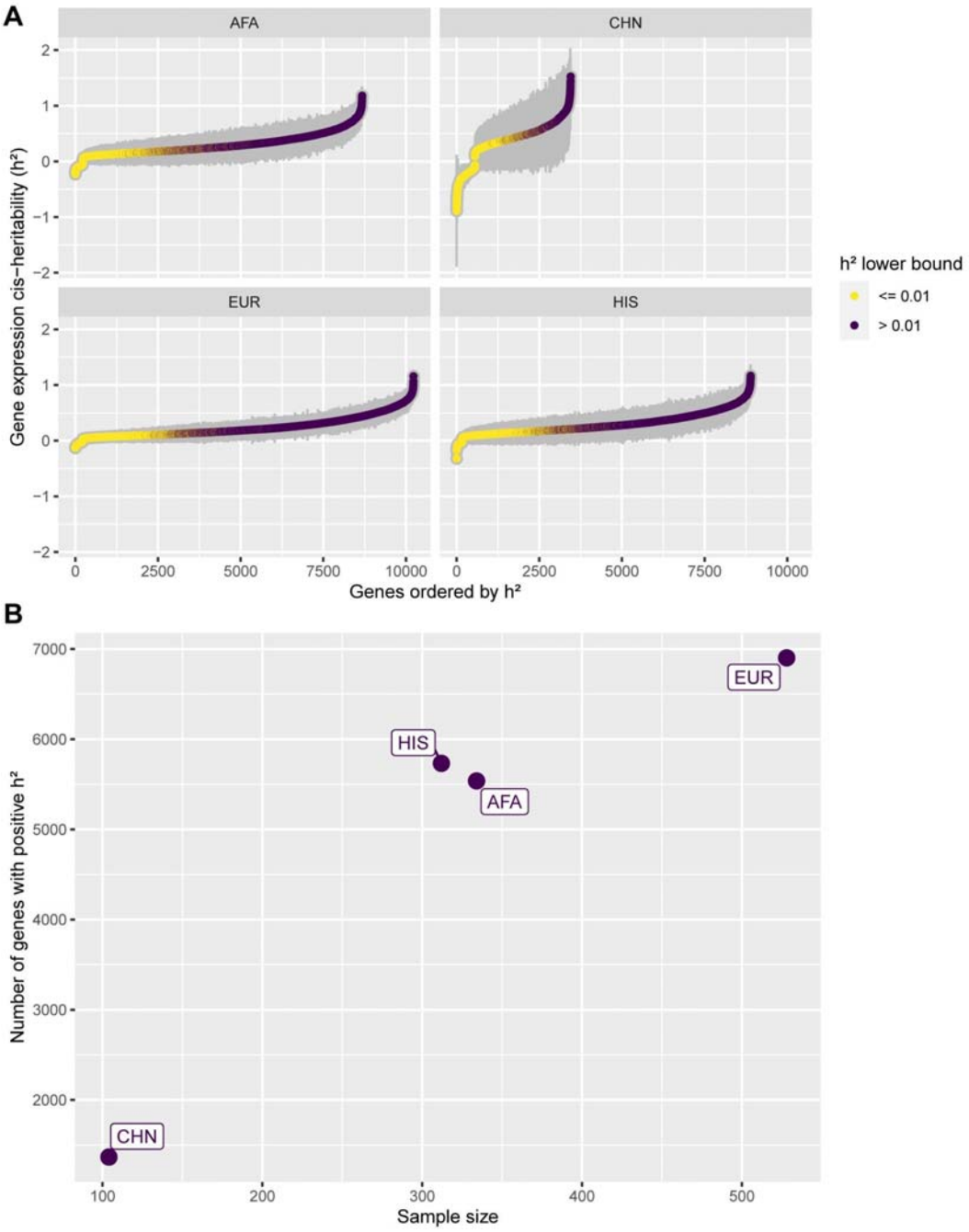


Figure 3: PBMC gene expression cis-heritability estimates across MESA populations. (A) Gene expression cis-heritability (h^2) estimated for different genes across different MESA population datasets in PBMC. Only genes with significant estimated h^2 (p-value < 0.05) are shown. Gray bars represent the standard errors (2*S.E.). Genes are ordered on the x-axis in ascending h^2 order, and colored according to the h^2 lower bound (h^2 - 2*S.E.). (B) Number of significant heritable genes (p-value < 0.05 and h^2 lower bound > 0.01) within each PBMC population dataset, by sample size. AFA = African American, CHN = Chinese, EUR = European, HIS = Hispanic/Latino.

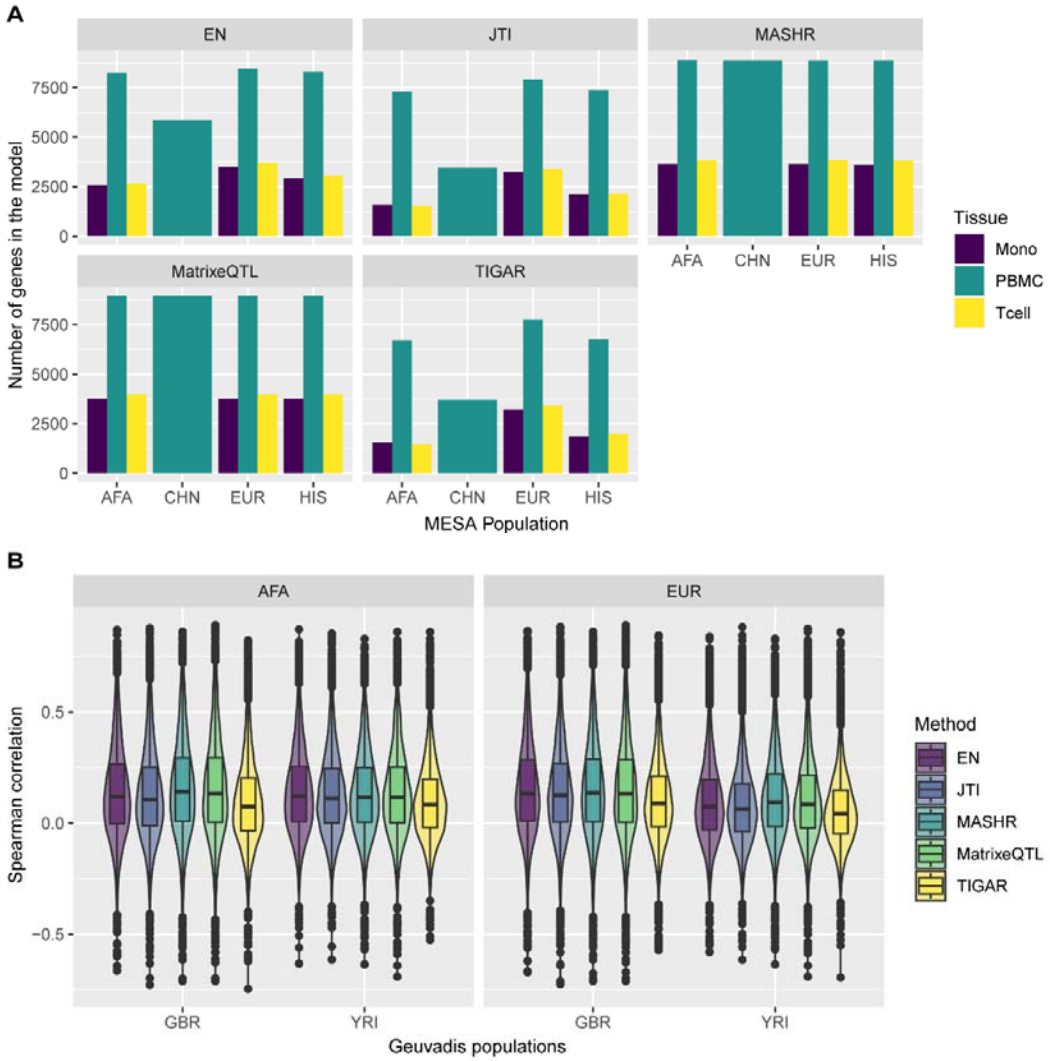


Figure 4: Comparison of MESA population transcriptome prediction models. (A) The number of genes in each MESA population model, by method and tissue. (B) Prediction performance (Spearman's rho) of EN, JTI, MASHR, MatrixeQTL, and TIGAR PBMC MESA population models in Geuvadis GBR and YRI populations. Only the intersection of genes with expression predicted by all methods for each MESA-Geuvadis population pair are shown. MASHR performed better than or the same as all other methods (see Table S2 for all pairwise comparisons).

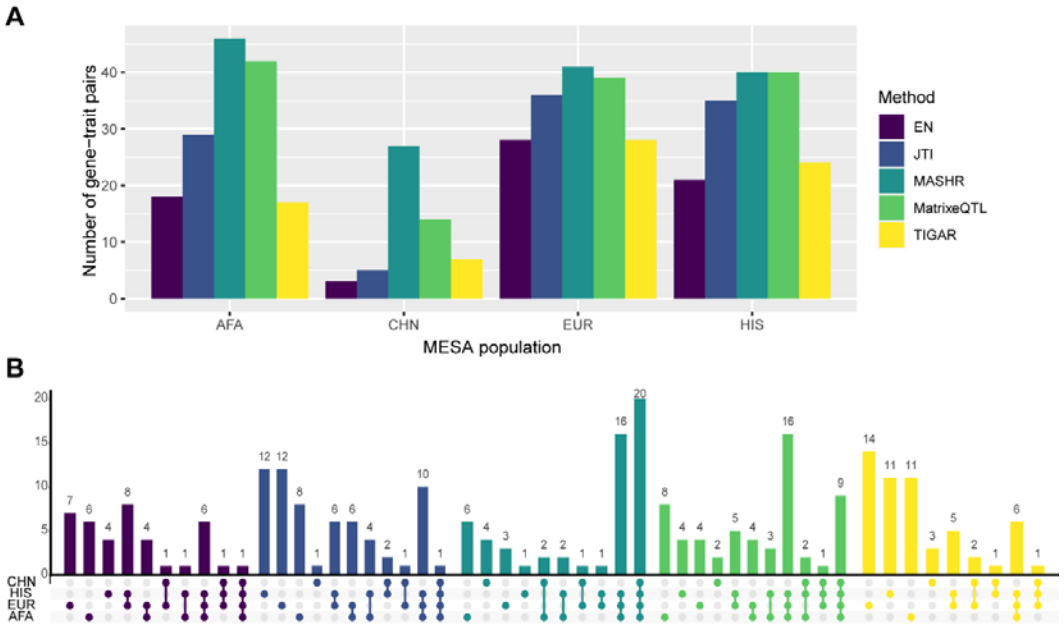


Figure 5: Number of significant S-PrediXcan gene-trait pairs in PAGE and PanUKBB GWAS summary statistics. (A) Total number of significant gene-trait pairs discovered by each MESA population model (considering the union of the three tissues), by method. (B) Number of significant gene-trait pairs discovered with individual or multiple MESA populations colored by method (considering the union of the three tissues). Population set intersections are indicated on the x-axis in color.

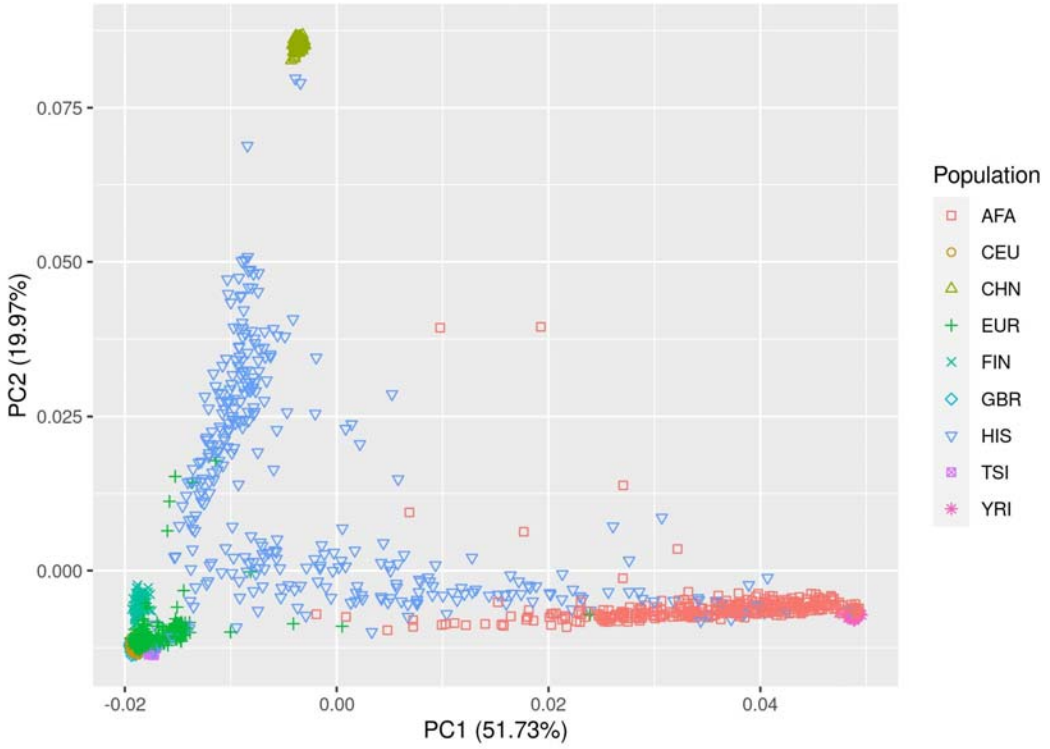


Figure S1: Genotype principal component analysis. Plot of the first two principal components of TOPMed MESA populations with Geuvadis populations. AFA = African American (TOPMed), CEU = Utah residents with Northern and Western European ancestry

(Geuvadis), CHN = Chinese (TOPMed), EUR = European (TOPMed), FIN = Finnish in Finland (Geuvadis), GBR = British in England and Scotland (Geuvadis), HIS = Hispanic/Latino (TOPMed), TSI = Toscani in Italy (Geuvadis), YRI = Yoruba in Ibadan, Nigeria (Geuvadis).

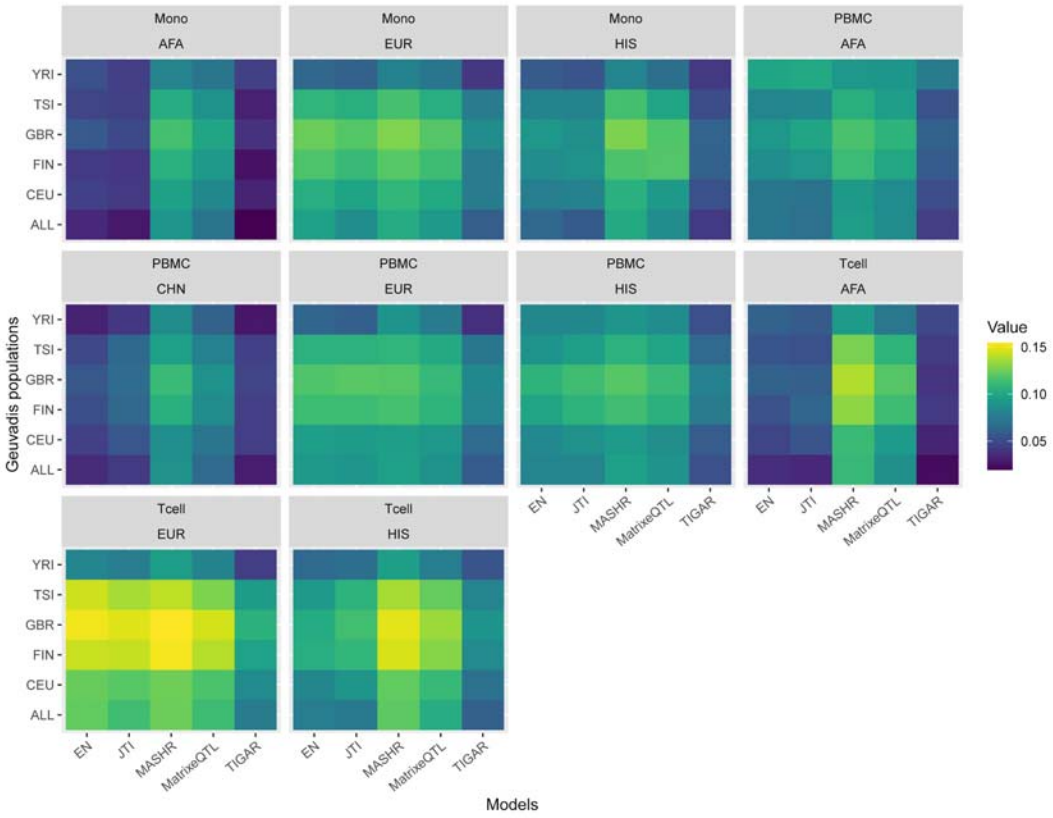


Figure S2: Overall prediction performance of MESA population models in Geuvadis. Prediction performance (median Spearman's rho) of EN, JTI, MASHR, MatrixeQTL, and TIGAR MESA population models in all Geuvadis populations.

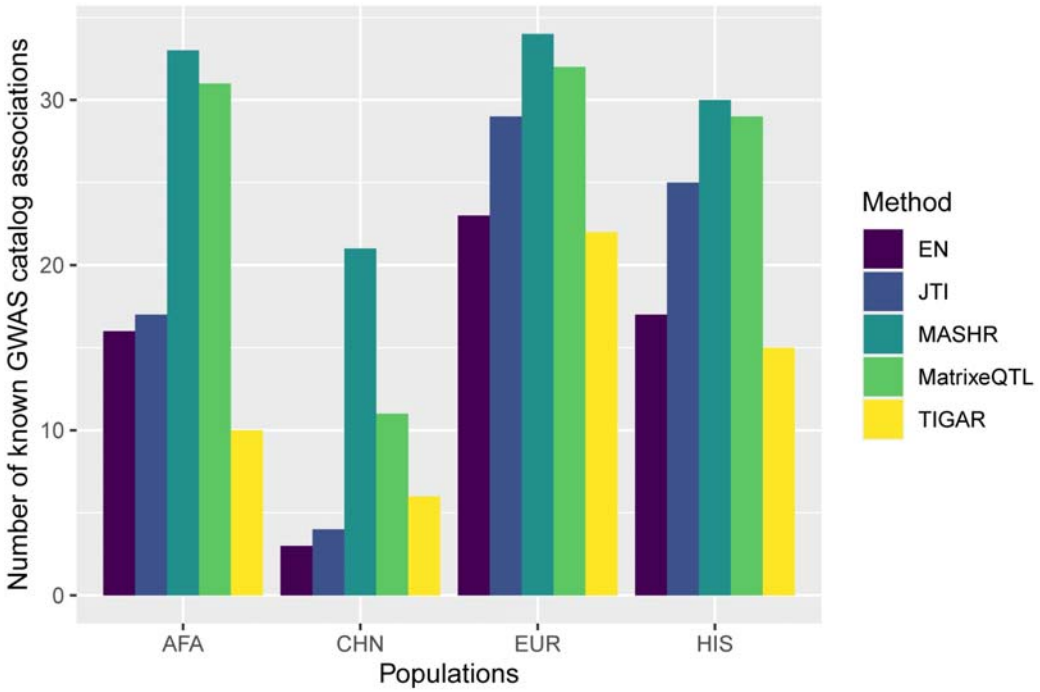


Figure S3: Number of significant S-PrediXcan gene-trait pairs in PAGE and PanUKBB GWAS summary statistics that have been reported in the GWAS catalog. Total number of significant gene-trait pairs discovered by each MESA population model (considering the union of the three tissues), by method.

Table S1: PAGE and PanUKBB summary statistics used in this study.

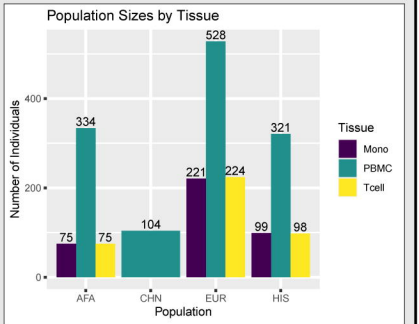
Table S2: Performance comparisons of PBMC AFA and EUR MESA transcriptome prediction models in the GBR and YRI Geuvadis populations between all methods.

Table S3: Pairwise comparisons of the performance of EN, JTI, MASHR, MatrixeQTL, and TIGAR MESA transcriptome prediction models in all Geuvadis populations.

Table S4: Compiled S-PrediXcan gene-trait pair discoveries, significant in PAGE and PanUKBB GWAS summary statistics with the same direction of effect.

Table S5: List of NHLBI TOPMed consortium members.

Training Data

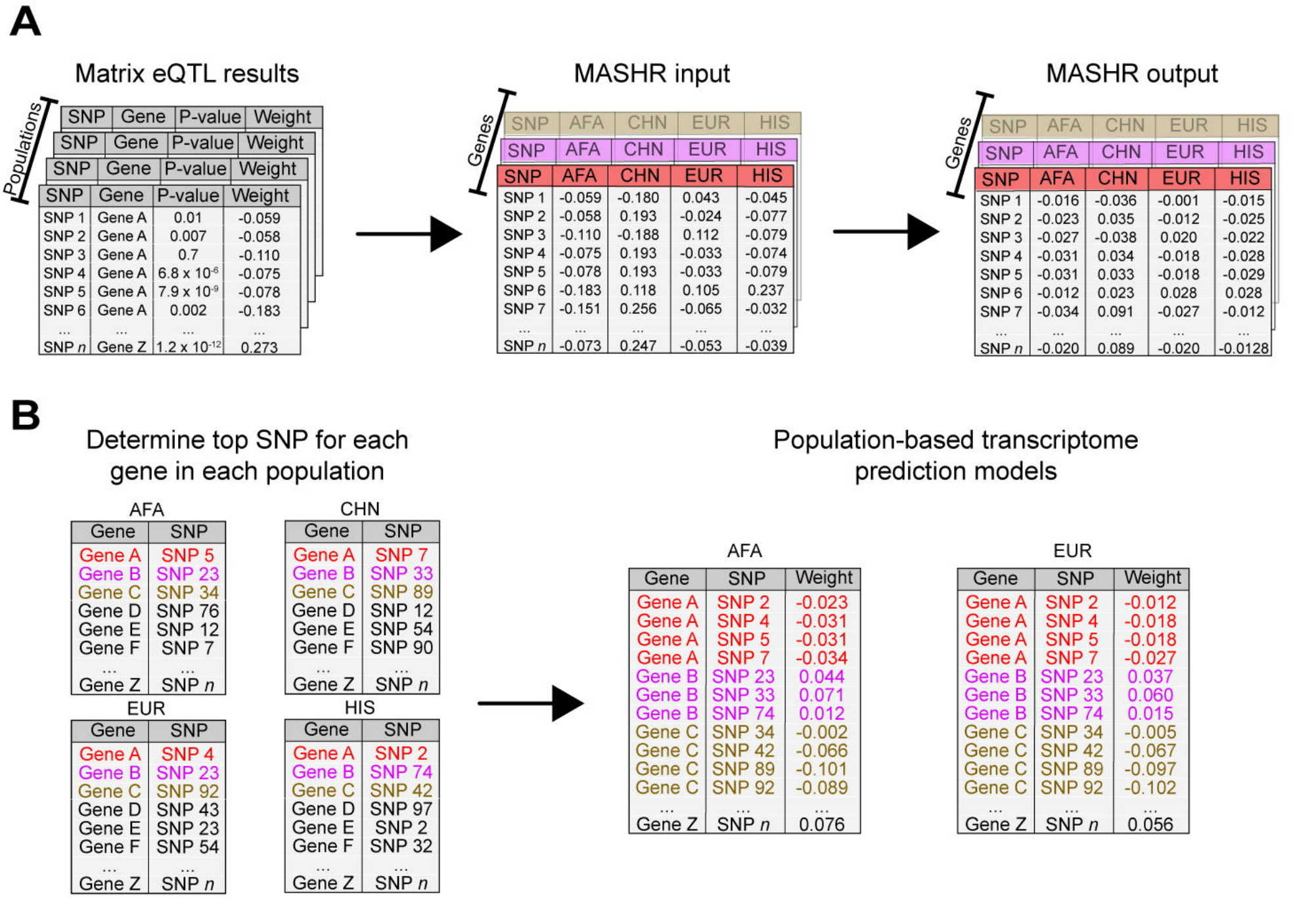


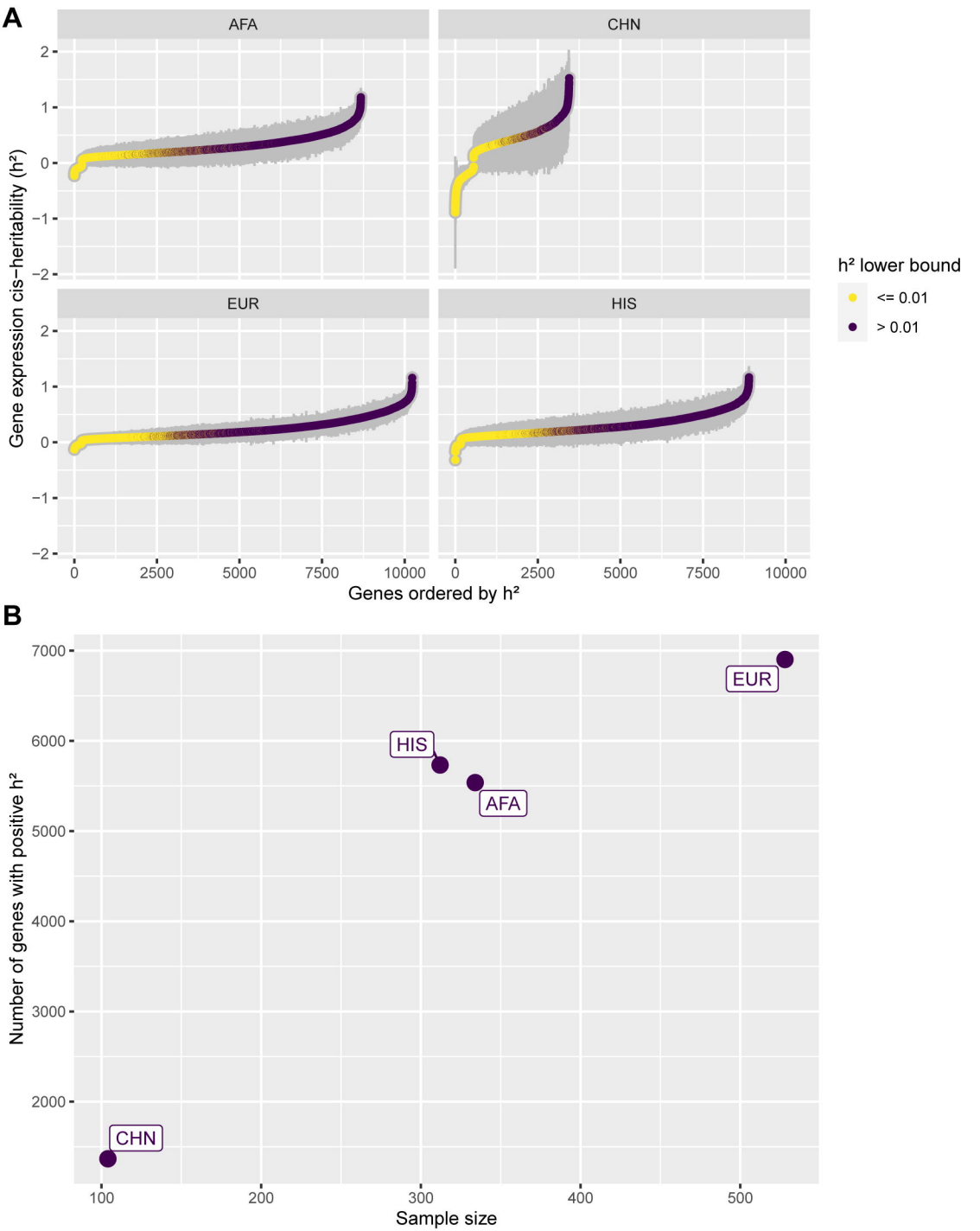
Population-based models

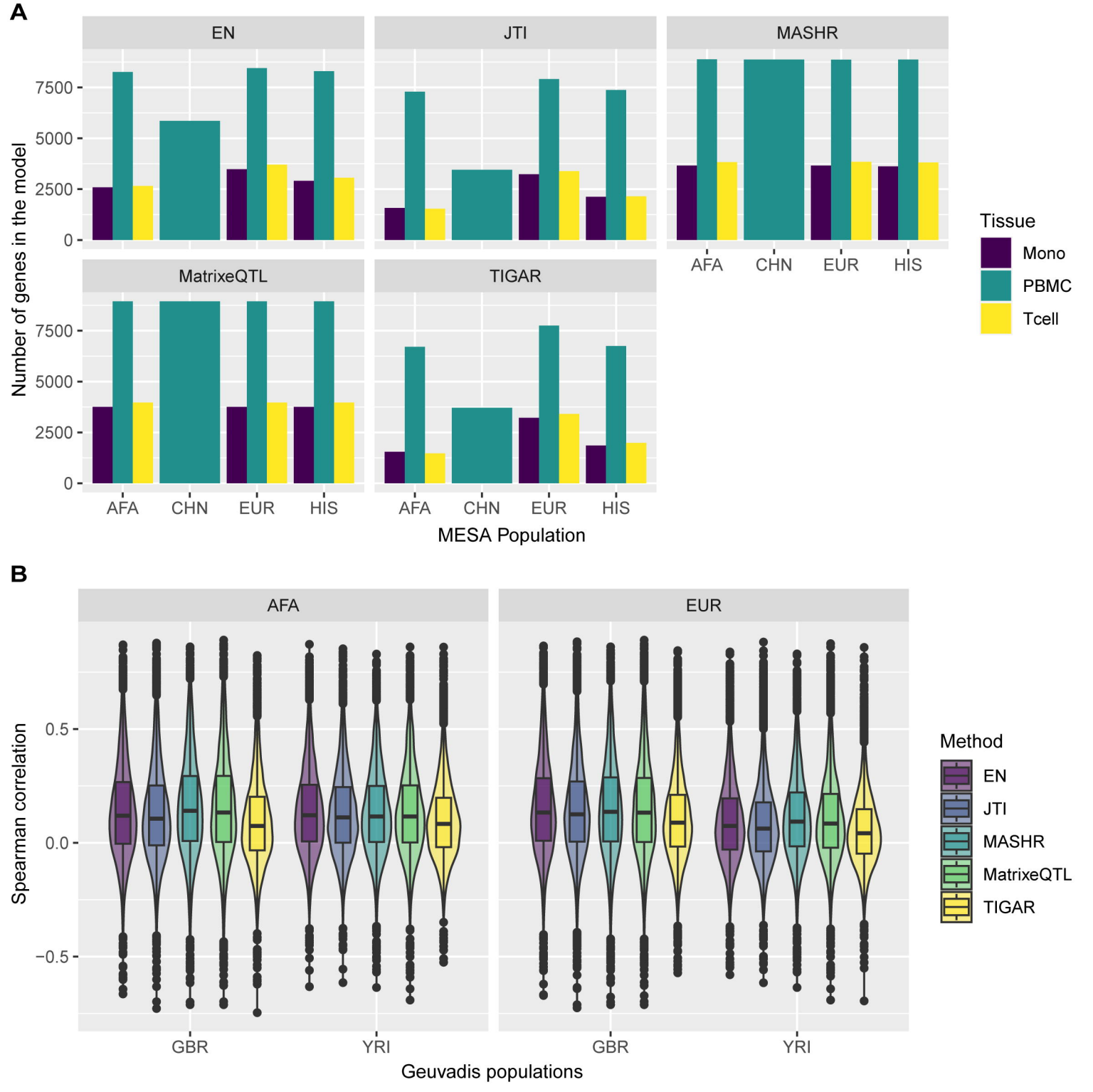
- Elastic Net
- Joint-Tissue Imputation
- Multivariate adaptive shrinkage
- Matrix eQTL
- Transcriptome-Integrated Genetic Association Resource

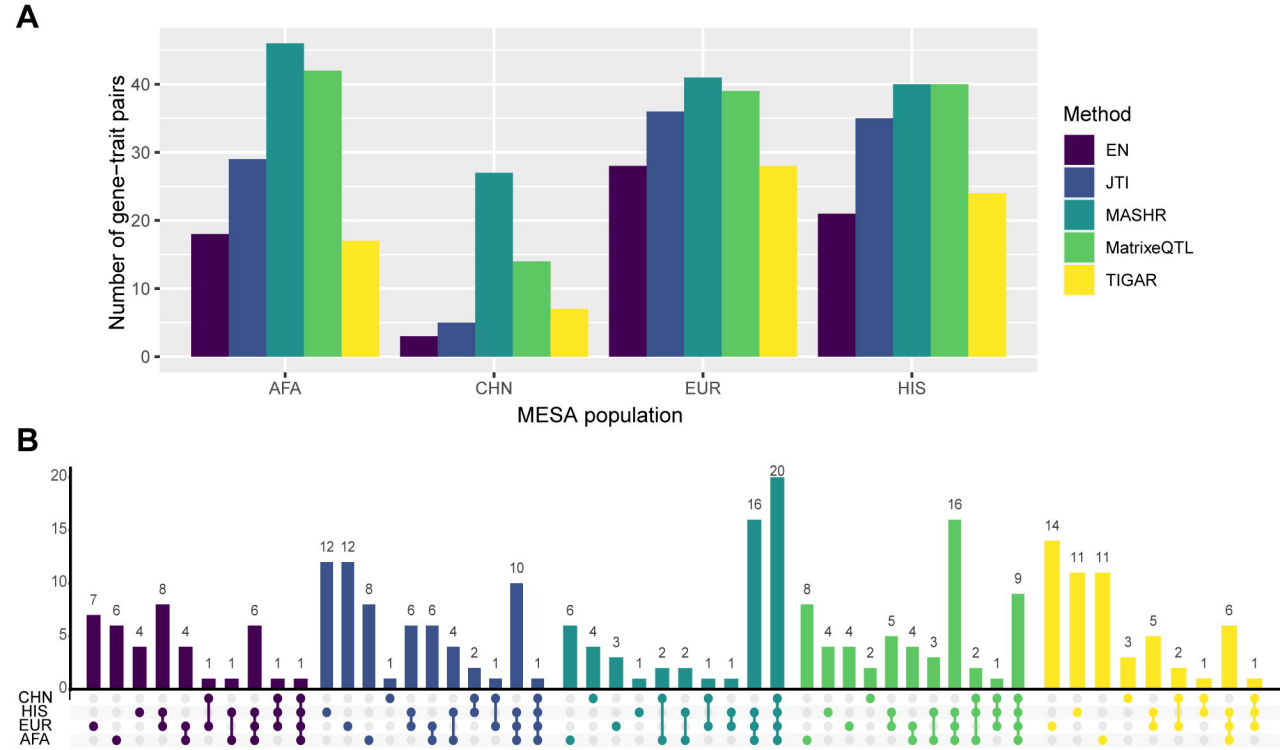
Performance assessment

- Out-of-sample transcriptome prediction performance:
  - ► GEUVADIS dataset
- Discoveries in multi-ethnic TWAS:
  - ► PAGE Study (N=49,839)
  - ► PanUKBB (N=441,331)


A
Matrix eQTL results
Populations
SNP Gene P-value Weight
SNP 1 Gene A 0.01 -0.059
SNP 2 Gene A 0.007 -0.058
SNP 3 Gene A 0.7 -0.110
SNP 4 Gene A 6.8 x 10-6 -0.075
SNP 5 Gene A 7.9 x 10-9 -0.078
SNP 6 Gene A 0.002 -0.183
SNP n Gene Z 1.2 x 10-12 0.273
MASHR input
Genes
SNP AFA CHN EUR HIS
SNP 1 -0.059 -0.180 0.043 -0.045
SNP 2 -0.058 0.193 -0.024 -0.077
SNP 3 -0.110 -0.188 0.112 -0.079
SNP 4 -0.075 0.193 -0.033 -0.074
SNP 5 -0.078 0.193 -0.033 -0.079
SNP 6 -0.183 0.118 0.105 0.237
SNP 7 -0.151 0.256 -0.065 -0.032
SNP n -0.073 0.247 -0.053 -0.039
MASHR output
Genes
SNP AFA CHN EUR HIS
SNP 1 -0.016 -0.036 -0.001 -0.015
SNP 2 -0.023 0.035 -0.012 -0.025
SNP 3 -0.027 -0.038 0.020 -0.022
SNP 4 -0.031 0.034 -0.018 -0.028
SNP 5 -0.031 0.033 -0.018 -0.029
SNP 6 -0.012 0.023 0.028 0.028
SNP 7 -0.034 0.091 -0.027 -0.012
SNP n -0.020 0.089 -0.020 -0.0128
B
Determine top SNP for each gene in each population
AFA
Gene SNP
Gene A SNP 5
Gene B SNP 23
Gene C SNP 34
Gene D SNP 76
Gene E SNP 12
Gene F SNP 7
Gene Z SNP n
CHN
Gene SNP
Gene A SNP 7
Gene B SNP 33
Gene C SNP 89
Gene D SNP 12
Gene E SNP 54
Gene F SNP 90
Gene Z SNP n
EUR
Gene SNP
Gene A SNP 4
Gene B SNP 23
Gene C SNP 92
Gene D SNP 43
Gene E SNP 23
Gene F SNP 54
Gene Z SNP n
HIS
Gene SNP
Gene A SNP 2
Gene B SNP 74
Gene C SNP 42
Gene D SNP 97
Gene E SNP 2
Gene F SNP 32
Gene Z SNP n
Population-based transcriptome prediction models
AFA
Gene SNP Weight
Gene A SNP 2 -0.023
Gene A SNP 4 -0.031
Gene A SNP 5 -0.031
Gene A SNP 7 -0.034
Gene B SNP 23 0.044
Gene B SNP 33 0.071
Gene B SNP 74 0.012
Gene C SNP 34 -0.002
Gene C SNP 42 -0.066
Gene C SNP 89 -0.101
Gene C SNP 92 -0.089
Gene Z SNP n 0.076
EUR
Gene SNP Weight
Gene A SNP 2 -0.012
Gene A SNP 4 -0.018
Gene A SNP 5 -0.018
Gene A SNP 7 -0.027
Gene B SNP 23 0.037
Gene B SNP 33 0.060
Gene B SNP 74 0.015
Gene C SNP 34 -0.005
Gene C SNP 42 -0.067
Gene C SNP 89 -0.097
Gene C SNP 92 -0.102
Gene Z SNP n 0.056


A
AFA
CHN
EUR
HIS
Gene expression cis−heritability (h²)
Genes ordered by h²
h² lower bound
<= 0.01
> 0.01
B
Number of genes with positive h²
Sample size
EUR
HIS
AFA
CHN


A
EN
JTI
MASHR
MatrixeQTL
TIGAR
Number of genes in the model
0
2500
5000
7500
AFA
CHN
EUR
HIS
MESA Population
Tissue
Mono
PBMC
Tcell
B
AFA
EUR
Spearman correlation
0.5
0.0
−0.5
GBR
YRI
Geuvadis populations
Method
EN
JTI
MASHR
MatrixeQTL
TIGAR


A
Number of gene−trait pairs
0
10
20
30
40
AFA
CHN
EUR
HIS
MESA population
Method
EN
JTI
MASHR
MatrixeQTL
TIGAR
B
0
5
10
15
20
7
6
4
8
4
1
1
6
1
1
12
12
8
1
6
6
4
2
1
10
1
6
4
3
1
2
2
1
1
16
20
8
4
4
2
5
4
3
16
2
1
9
14
11
11
3
5
2
1
6
1
CHN
HIS
EUR
AFA