OTTERS: A powerful TWAS framework leveraging summary-level reference data

Qile Dai[1,12], Geyu Zhou[2], Hongyu Zhao[2,3], Urmo Võsa[4], Lude Franke[5,6], Alexis Battle[7], Alexander Teumer[8], Terho Lehtimäki[9], Olli Raitakari[10,11], Tõnu Esko[4], eQTLGen Consortium, Michael P. Epstein[12*], Jingjing Yang[12*]

1. Department of Biostatistics and Bioinformatics, Emory University School of Public Health, Atlanta, Georgia 30322, United States of America.
2. Program of Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06511, United States of America.
3. Department of Biostatistics, Yale School of Public Health, New Haven, CT 06520, United States of America.
4. Estonian Genome Centre, Institute of Genomics, University of Tartu, 50090 Tartu, Estonia.
5. Department of Genetics, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, Netherlands.
6. Oncode Institute, 3521 AL Utrecht, Netherlands.
7. Department of Computer Science, and Departments of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States of America
8. Institute for Community Medicine, University Medicine Greifswald, 17489 Greifswald, Germany.
9. Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center, Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33520, Finland.
10. Centre for Population Health Research, and Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, 20500 Turku, Finland.
11. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, 20521 Turku, Finland.
12. Center for Computational and Quantitative Genetics, Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia 30322, United States of America.

*Correspondence Authors: M.P.E. (mpepste@emory.edu) and J.Y. (jingjing.yang@emory.edu)

Abstract

Most existing TWAS tools require individual-level eQTL reference data and thus are not applicable to summary-level reference eQTL datasets. The development of TWAS methods that can harness summary-level reference data is valuable to enable TWAS in broader settings and enhance power due to increased reference sample size. Thus, we develop a TWAS framework called OTTERS (Omnibus Transcriptome Test using Expression Reference Summary data) that adapts multiple polygenic risk score (PRS) methods to estimate eQTL weights from summary-level eQTL reference data and conducts an omnibus TWAS. We show that OTTERS is a practical and powerful TWAS tool by both simulations and application studies.

Keywords:

Transcriptome-wide association study; Summary-level eQTL reference data; PRS method; GWAS; UK Biobank; eQTLGen; Cardiovascular disease

Transcriptome-wide association study (TWAS) is a valuable analysis strategy for identifying genes that influence complex traits and diseases through genetic regulation of gene expression[1–5]. Researchers have successfully deployed TWAS analyses to identify risk genes for complex human diseases, including Alzheimer's disease[6–8], breast cancer[9–11], ovarian cancer[12,13], and cardiovascular disease[14,15]. A typical TWAS consists of two separate stages. In Stage I, TWAS acquires individual-level genetic and expression data from relevant tissues available in a reference dataset like the Genotype-Tissue Expression (GTEx) project[16,17] or the North American Brain Expression Consortium (NABEC)[18], and trains multivariable regression models on the reference data treating gene expression as outcome and SNP genotype data (typically cis-SNPs nearby the test gene) as predictors to determine genetically regulated expression (GReX). After

Stage I that uses the GReX regression models to estimate effect sizes of SNP predictors that, in the broad sense, are expression quantitative trait loci (eQTLs), Stage II of TWAS proceeds by using these trained eQTL effect sizes to impute GReX within an independent GWAS of a complex human disease or trait. One can then test for association between the imputed GReX and phenotype, which is equivalent to a gene-based association test taking these eQTL effect sizes as corresponding test SNP weights[19–21].

For Stage I of TWAS, a variety of training tools exist for fitting GReX regression models using reference expression and genetic data, including PrediXcan[19], FUSION[20], and TIGAR[22]. While these methods all employ different techniques for model fitting, they all require individual-level reference expression and genetic data to estimate eQTL effect sizes for TWAS. Therefore, these methods cannot be applied to emerging reference summary-level eQTL results such as those generated by the eQTLGen[23] and CommonMind[24] consortia, which provide eQTL effect sizes and p-values relating individual SNPs to gene expression. The development of TWAS methods that can utilize such summary-level reference data is valuable to permit applicability of the technique to broader analysis settings. Moreover, as TWAS power increases with increasing reference sample size[25], TWAS using summary-level reference datasets can lead to enhanced performance compared to using individual-level reference datasets since the sample sizes of the former often are considerably larger than the latter. For example, the sample size of the summary-based eQTLGen reference sample is 31,684 for blood, whereas the sample size of the individual-level GTEx V6 reference is only 338 for the same tissue. Consequently, TWAS analysis leveraging the summary-based eQTLGen dataset as reference likely can provide novel insights into genetic regulation of complex human traits.

In this work, we propose a framework that can use summary-level reference data to train GReX regression models required for Stage I of TWAS analysis. Our method is motivated by a variety of published polygenic risk score (PRS) methods[26–31] that can predict phenotype in a test

dataset using summary-level SNP effect-size estimates and p-values based on single SNP tests from an independent reference GWAS. We can adapt these PRS methods for TWAS since eQTL effect sizes are essentially SNP effect sizes resulting from a reference "GWAS" of gene expression. Thus, our predicted GReX in Stage II of TWAS is analogous to the PRS constructed based on training GWAS summary statistics of single SNP-trait association. Here, we adapt four representative summary-data based PRS methods — P-value Thresholding with linkage disequilibrium (LD) clumping (*P+T*)[26], frequentist LASSO[32] regression based method *lassosum*[27], nonparametric Bayesian Dirichlet Process Regression (DPR) model[33] based method *SDPR*[29], and Bayesian multivariable regression model based method with continuous shrinkage (CS) priors *PRS-CS*[28] for TWAS analysis. We apply each of these PRS methods to first train eQTL effect sizes based on a multivariable regression model from summary-level reference eQTL data (Stage I), and subsequently use these eQTL effect sizes (i.e., eQTL weights) to impute GReX and then test GReX-trait association in an independent test GWAS (Stage II).

As we will show, the PRS method with optimal performance for TWAS depends on the underlying genetic architecture for gene expression. Since the genetic architecture of expression is unknown apriori, we maximize the performance of TWAS over different possible architectures by proposing a novel TWAS framework called OTTERS (**O**mnibus **T**ranscriptome **T**est using **E**xpression **R**eference **S**ummary data). OTTERS first constructs individual TWAS tests and p-values using eQTL weights trained by each of the PRS techniques outlined above, and then calculates an omnibus test p-value using the aggregated Cauchy association test[34] (ACAT-O) with all individual TWAS p-values (Figure 1). OTTERS is applicable to both summary-level and individual-level test GWAS data within Stage II TWAS analysis.

In subsequent sections, we first describe how to use the PRS methods on summary-level reference eQTL data in Stage I TWAS, and then describe how we can use the resulting eQTL weights to perform Stage II TWAS using OTTERS. We then evaluate the performance of

individual PRS methods and OTTERS using simulated expression and real genetic data based on patterns observed in real datasets. Interestingly, when we assume individual-level reference data are available, we observe that OTTERS outperforms the popular FUSION[20] approach across all simulation settings considered. Many of the individual PRS methods also outperform FUSION in these settings. We then apply OTTERS to blood eQTL summary-level data (n=31,684) from the eQTLGen consortium[23] and GWAS summary data of cardiovascular disease from the UK Biobank (UKBB)[35]. By comparing OTTERS results to those of FUSION[20] using individual-level GTEx reference data of whole blood tissue, we demonstrate that OTTERS using large summary-level reference datasets and multiple gene expression imputation models can successfully reveal potential risk genes missed by FUSION based on smaller individual-level reference datasets and only one model. Finally, we conclude with a discussion.

Results

Method Overview

For the standard two-stage TWAS approach, Stage I estimates a GReX imputation model using individual-level expression and genotype data available from a reference dataset, and then Stage II uses the eQTL effect sizes from Stage I to impute gene expression (GReX) in an independent GWAS and test for association between GReX and phenotype. GReX for test samples can be imputed from individual-level genotype data and eQTL effect size estimates. When individual-level GWAS data are not available, one can instead use summary-level GWAS data for TWAS by applying the TWAS Z-score statistics proposed by FUSION[20] and S-PrediXcan[36] (see details in Methods).

Since eQTL summary data are analogous to GWAS summary data where gene expression represents the phenotype, we can follow the idea from PRS methods to estimate the eQTL effect sizes based on a multivariable regression model using only marginal least squared

effect estimates and p-values (based on a single variant test) from the eQTL summary data as well as a reference LD panel from samples of the same ancestry[26–29]. Although all PRS methods are applicable to TWAS Stage I, we only consider four representative methods — *P+T*[26], *Frequentist lassosum*[27], *Nonparametric Bayesian SDPR*[29], *Bayesian PRS-CS*[28] (see details in Methods).

As shown in Figure 1, OTTERS first trains GReX imputation models per gene *g* using *P+T, lassosum, SDPR,* and *PRS-CS* methods that each infers cis-eQTL weights using cis-eQTL summary data and an external LD reference panel of the same ancestry (Stage I). Once we derive cis-eQTL weights for each training method, we can impute the respective GReX using that method and perform the respective gene-based association analysis in the test GWAS dataset. We thus derive a set of TWAS p-values for gene *g,* one per training method. We then use these TWAS p-values to create an omnibus test using the ACAT-O[34] approach that employs a Cauchy distribution for inference (see details in Supplemental Methods). We refer to the p-value derived from ACAT-O test as the OTTERS p-value. The ACAT-O[34] approach has been widely used in hypothesis testing to combine multiple testing methods for the same hypothesis[37–39], which has been shown as an effective approach to leverage different test methods to increase the power while still managing to control for type I error. Adding TWAS p-values based on additional PRS methods to the ACAT-O test can possibly improve the power further at the cost of additional computation.

Simulation Study

We used real genotype data from 1894 whole genome sequencing (WGS) samples from the Religious Orders Study and Rush Memory and Aging Project (ROS/MAP) cohort[40,41] and Mount Sinai Brain Bank (MSBB) study[42] for simulation. We divided 14,772 genes into five groups according to gene length, and randomly selected 100 genes from each group (500 genes in total). We randomly split samples into 568 training (30%) and 1326 testing samples

(70%) to mimic a relatively small sample size in the real reference panel for training gene expression imputation models.. From the real genotype data, we simulated 6 scenarios with 2 different proportions of causal cis-eQTL, $p_{causal} = (0.001, 0.01)$, as well as 3 different proportions of gene expression variance explained by causal eQTL, $h_e^2 = (0.01, 0.05, 0.1)$.

We generated gene expression of gene g ($\boldsymbol{E}_g$) using the multivariable regression model $\boldsymbol{E}_g = \boldsymbol{X_g w} + \boldsymbol{\epsilon_g}$, where $\boldsymbol{X}_g$ represents the standardized genotype matrix of the randomly selected causal eQTL of gene g, $\boldsymbol{\epsilon}_g \sim N(0, (1 - h_e^2)\boldsymbol{I})$. We generated the eQTL effect sizes $\boldsymbol{w}$ from $N(0,1)$ and then re-scaled these effects to ensure that the expression variance explained by all causal variants is h_e^2. We generated 10 replicates of gene expression per scenario. For each simulated gene expression, we then generated 10 sets of GWAS Z-scores to perform a total of 50,000 TWAS simulations. We generated the GWAS Z-scores from a multivariate normal distribution with $Z \sim MVN\left(\boldsymbol{\Sigma_g w}\sqrt{n_{gwas}\, h_p^2},\ \boldsymbol{\Sigma_g}\right)$[38], where $\boldsymbol{w}$ is the true causal eQTL effect sizes, $\boldsymbol{\Sigma_g}$ is the correlation matrix of the standardized genotype X_g from test samples, n_{gwas} is the assumed GWAS sample size, and h_p^2 denotes the amount of phenotypic variance explained by simulated GReX=$\boldsymbol{X_g w}$ (see Methods). We set $h_p^2 = 0.025$. To calibrate power, we considered n_{gwas} = (200K, 300K, 400K, 500K) for scenarios with h_e^2 = 0.01, n_{gwas} = (25K, 50K, 75K, 100K) for scenarios with $h_e^2 = 0.05$, and n_{gwas} = (10K, 20K, 30K, 40K) for scenarios with $h_e^2 = 0.1$.

In Stage I of our TWAS analysis, we applied *P+T (0.001)*, *P+T (0.05)*, *lassosum*, *SDPR*, and *PRS-CS* methods to estimate eQTL weights using eQTL summary data and the reference LD of training samples. In Stage II of the TWAS, we used the estimated eQTL weights and the simulated GWAS Z-scores to conduct a gene-based association test. In addition to gene-based association tests based on eQTL weights per training method, we further constructed the corresponding OTTERS p-values. We evaluated the performance of the training methods with

test samples, comparing test R^2 that was the squared Pearson correlation coefficient between imputed GReX and simulated gene expression. We evaluated TWAS power given by the proportion of 50,000 repeated simulations with TWAS p-value $< 2.5 \times 10^{-6}$ (genome-wide significance threshold adjusting for testing 20K independent genes). We only considered genes with GReX test $R^2 > 0.01$ in the power analysis.

As shown in Figure 2, we demonstrated that the Stage I training method with optimal test R^2 and TWAS power depended on the underlying genetic architecture of gene expression (p_{causal}) as well as gene expression heritability (h_e^2). In situations where true cis-eQTLs were sparse (p_{causal} = 0.001) and the gene expression heritability was small (h_e^2 = 0.01), *P+T (0.05)* method performed the best with the highest TWAS power among all individual methods. When gene expression heritability is low (h_e^2 = 0.01), the power of *P+T (0.001)* and *lassosum* methods were shown as the lowest for considering only genes with test $R^2 > 0.01$. When gene expression heritability increased (h_e^2 = 0.05 or 0.1) within this sparse eQTL model, *P+T (0.001)* and PRS-CS were generally the optimal methods. For a less sparse model with p_{causal} = 0.01, SDPR and PRS-CS generally performed best among the individual methods. Relative to individual methods, we found that combining the TWAS p-values based on the four PRS training methods together for analysis in our OTTERS framework obtained the highest power across all scenarios.

To evaluate the type I error of the individual PRS methods along with OTTERS, we picked one simulated replicate per gene from the scenario with $h_e^2 = 0.1$ and $p_{causal} = 0.001$, simulated 2×10^3 phenotypes from $N(0,1)$, and permuted the eQTL weights for TWAS to perform a total of 10^6 null simulations. OTTERS was shown well calibrated in the tails of the distribution as shown by quantile-quantile (Q-Q) plots of TWAS p-values in Figure S1. We also observed that OTTERS had well-controlled type I error for stringent significance levels between 10^{-4} and 2.5×10^{-6} (Table S1), which are typically utilized in TWAS. For more modest significance thresholds ($\alpha = 10^{-2}$), we noted that OTTERS had a slightly inflated type-I error rate.

This modest inflation is consistent with the findings of the original ACAT-O work, which showed that the Cauchy-distribution-based approximation that ACAT-O employs may not be accurate for larger p-values when correlation among tests is strong[34]. This suggests that modest OTTERS p-values may be interpreted with caution.

We also compared the performance of our individual PRS training methods to those of FUSION assuming individual-level reference data were available for the latter method to train GReX models. As shown in Figure 2A, we interestingly observed that our training methods yielded similar or improved test R^2 compared to FUSION in this situation, with *SDPR* and *PRS-CS* outperforming FUSION across all simulation settings. Comparing TWAS power, we found that OTTERS outperformed FUSION by a considerable margin in our simulations (Figure 2B). These simulation results suggest that, while we developed OTTERS based on PRS training methods to handle summary-level reference data, OTTERS can still improve TWAS power when individual-level reference data are available. This is likely because OTTERS accounts for multiple possible models of genetic architectures of gene expression assumed by the different PRS training methods.

GReX Imputation Accuracy in GTEx V8 Blood Samples

To evaluate the imputation accuracy of *P+T (0.001)*, *P+T (0.05)*, *lassosum*, *SDPR*, and *PRS-CS* methods in real data, we applied these training methods to summary-level eQTL reference data from the eQTLGen consortium[23] with n=31,684 blood samples, to train GReX imputation models for 16,699 genes. For test data, we downloaded the transcriptomic data of 315 blood tissue samples that are in GTEx V8 but were not part of GTEx V6 (as GTEx V6 samples contributed to the reference eQTLGen consortium summary data). For these 315 samples, we compared imputed GReX to observed expression levels. We considered trained imputation models with test $R^2 > 0.01$ as "valid" models, as suggested by previous TWAS methods[20,43]. We also compared imputation accuracy of these five training models to those using FUSION based

on a smaller individual-level training dataset (individual-level GTEx V6 reference dataset; see Methods). For such models, we compared the test R^2 for genes that had test $R^2 > 0.01$ by at least one training method.

We observed that *PRS-CS* obtained the most "valid" GReX imputation models with test R^2>0.01. Among 16,699 tested genes, *PRS-CS* obtained "valid" GReX imputation models for 10,337 genes, compared to 9,816 genes by *P+T (0.001)* (5.0% less valid genes than PRS-CS), 9,662 genes by *P+T (0.05)* (6.5% less), 8,718 genes by *lassosum* (15.7% less), 9,670 genes by *SDPR* (6.5% less), and 4,704 genes by FUSION (54.5% less) (Table 1). Among the "valid" GReX imputation models obtained by each method, the ones trained by *PRS-CS* have the highest median test R^2. The *P+T (0.001)* method obtained the second most "valid" GReX imputation models with the second largest median test R^2, as compared to *P+T (0.05)*, *lassosum*, and *SDPR* (Table 1). We note that the performance of PRS-CS method was not sensitive to the global shrinkage parameter (Figure S2).

By comparing test R^2 per "valid" GReX imputation model by *PRS-CS* versus the other methods (Figure 3), we observed that *PRS-CS* had the best overall performance for imputing GReX as it provided the most "valid" models with higher GReX imputation accuracy compared to *P+T* methods, *lassosum*, *SDPR*, and FUSION. Comparing the test R^2 among the other four training methods, we observed that these two *P+T* methods obtained similar test R^2 per "valid" model. Meanwhile, the test R^2 per valid model varied widely among the *P+T* methods, *lassosum*, and *SDPR* (Figure S3), suggesting that none of these four were optimal across all genes and their performance likely depended on the underlying unknown genetic architecture. These results are consistent with our simulation results.

TWAS of Cardiovascular Disease

Using the eQTL weights trained by *P+T (0.001), P+T (0.05), lassosum, SDPR, and PRS-CS* methods with the eQTLGen[23] reference data and reference LD from GTEx V8 WGS samples[44], we applied our OTTERS framework to the summary-level GWAS data of Cardiovascular Disease from UKBB (n=459,324, case fraction = 0.319)[35] (see Methods). We performed TWAS of cardiovascular disease for 16,678 genes. First, for each gene, we obtained TWAS p-values per individual training method (*P+T (0.001), P+T (0.05), lassosum, SDPR, and PRS-CS*). Second, we performed genomic control[45] for TWAS test statistics generated under each specific training model, by scaling all test statistics to ensure that the median test p-value equals to 0.5. Last, we only considered genes with test GReX $R^2 > 0.01$ by at least one PRS training method in additional GTEx V8 samples in the follow-up ACAT-O test. We combined the adjusted p-values across all five training models using ACAT-O to obtain our OTTERS test statistics and p-values. Genes with OTTERS p-values < 2.998×10^{-6} (Bonferroni corrected significance level) were identified as significant TWAS genes for cardiovascular risk.

In total, we identified 40 significant TWAS genes by using OTTERS. To identify independently significant TWAS genes, we calculated the R^2 (squared correlation) between the GReX predicted by *PRS-CS* for of each pair of genes. For a pair of genes with the predicted GReX $R^2 > 0.5$, we only kept the gene with the smaller TWAS p-value as the independently significant gene. OTTERS obtained 38 independently significant TWAS genes (Table 2, Figure 3B), compared to 17 independently significant genes by *P+T (0.001)*, 11 by *P+T (0.05)*, 10 by *lassosum*, 41 by *SDPR*, and 12 by *PRS-CS*. Among these 38 independent TWAS risk genes identified by OTTERS, gene *RP11-378A13.1* (OTTERS p-value = 9.78×10^{-9}) was not within 1 MB of any known GWAS risk loci with genomic-control corrected p-value < 5×10^{-8} in the UKBB summary-level GWAS data. This novel risk gene *RP11-378A13.1* was also identified to be a significant TWAS risk gene in blood tissue for systolic blood pressure, high cholesterol, and cardiovascular disease by FUSION[1].

We compared our OTTERS results with the TWAS results shown on TWAS hub (see Web Resource) obtained by FUSION using the same UKBB GWAS summary data of cardiovascular disease but using a smaller individual-level reference expression dataset from GTEx V6 (whole blood tissue, n=338). Of the 38 independent genes that OTTERS identified from TWAS with eQTLGen reference data of whole blood, FUSION only identified 8 of these genes (*CLCN6b, PSRC1, RP11-378A13.1, CAMK1D, SIDT2, MTHFSD, NTN5, OPRL1*) when using the GTEx V6 reference data of the same tissue. FUSION did identify 13 additional OTTERS genes (*NPPA, CPEB4, NT5C2, TNNT3, C11orf49, CSK, FES, MBTPS1, ACE, MRI1, HAUS8, RPL28, CTSZ*), when considering all available tissue types in GTEx V6 reference data. These genes were identified by FUSION when considering the GTEx V6 reference data of artery, thyroid, adipose visceral, and nerve tibial tissues. For example, the most significant gene *FES* (OTTERS p-value = 2.87×10^{-32}) was identified by FUSION using GTEx reference data of artery tibial, thyroid, and adipose visceral omentum tissues, and was also identified as a TWAS risk gene for high blood pressure, which is strongly related to cardiovascular disease[46].

Our OTTERS method also identified 17 novel risk genes (*LINC01093, SERPINB6, CARMIL1, ZSCAN12P1, HCG4P7, HCG4P3, HLA-S, PSPHP1, LPL, PTP4A3, SLCO3A1, RALBP1, SULT2B1, EDN3, ZBTB46, FAM3B, MX1*) that were not detected by FUSION using GTEx V6 data, where *EDN3* (Endothelin 3, a member of the endothelin family) was shown to be active in the cardiovascular system and play an important role in the maintenance of blood pressure or generation of hypertension[47].

By comparing OTTERS results with the ones obtained by individual methods (Table 2; Figure 4; Figure S4), we found that all individual methods contributed to the OTTERS results. For example, the novel risk gene *LINC01093* was only identified by *lassosum*, while genes *CPEB4, SIDT2,* and *ACE* were only detected by *PRS-CS* and *SDPR* and the novel risk gene *EDN3* was only identified by the *P+T* methods. To better understand the differences among individual

methods, we plotted the eQTL weights estimated by *P+T (0.001)*, *P+T (0.05)*, *lassosum*, *SDPR*, and *PRS-CS* for three example genes that were only detected by one or two individual methods (Figures S5-S7). For these genes, we plotted the eQTL weights produced by each method with such weights color coded with respect to $-log_{10}$(GWAS p-values) from the UKBB GWAS summary statistics and shape coded with respect to the direction of UKBB GWAS Z-score statistics. Generally, significant TWAS p-values would be obtained by methods that obtained eQTL weights with relatively large magnitude for SNPs with relatively more significant GWAS p-values.

In Figure S5, we showed the eQTL weights for gene *SIDT2*, which was a significant risk gene identified by both *PRS-CS* and *SDPR*, and had p-values < 10^{-4} by other methods. Compared to *lassosum*, *SDPR* had more significant GWAS SNPs colocalized with eQTLs having relatively large weights in the test region, and *PRS-CS* had more non-significant GWAS SNPs colocalized with eQTLs having zero weights. Compared to the P+T methods, SDPR and PRS-CS based on a multivariate regression model modeled LD among all test SNPs, and thus estimated eQTL weights leading to significant TWAS findings. In Figure S6, we provided the results of gene *EDN3*, which was only identified by *P+T* methods (p-values $\leq 9.15 \times 10^{-8}$). Compared to *P+T* methods, *SDPR* (p-value = 5.9×10^{-3}) and *PRS-CS* (p-value = 0.0158) had fewer significant GWAS SNPs colocalized with eQTLs that had relatively large weights in the test region, while *lassosum* (p-value = 8.6×10^{-6}) assigned relatively large weights to more non-significant GWAS SNPs. In Figure S7, we provided results for gene *LINC01093*, which was only identified by *lassosum*. For this gene, *SDPR* and *PRS-CS* estimated near-zero weights for most test SNPs with significant GWAS p-values in the test region. Most significant GWAS SNPs did not have eQTL test p-values < 0.001 or 0.05, and were thus filtered out by *P+T* methods. *lassosum* was the only method that produced relatively large eQTL weights that co-localized with GWAS significant SNPs.

These results were consistent with our simulation study results, demonstrating that the performance of different individual methods depended on the underlying genetic architecture. We do note that there were a handful of genes identified by an individual method that were not significant using OTTERS (Table S2). Nonetheless, the omnibus test borrows strength across all individual methods, thus generally achieves higher TWAS power and identifies the group of most robust TWAS risk genes.

By examining the Q-Q plots of TWAS p-values, we observed a moderate inflation for all methods (Figure S8). Such inflation in TWAS results is not uncommon[48–50], which could be due to similar inflation in the GWAS summary data and not distinguishing the pleiotropy and mediation effects for considered gene expression and phenotype of interest[51] (Figure S9). We also observed a notable inflation in the GWAS p-values of cardiovascular disease from UKBB (Figure S9), as we estimated the LD score regression[52] intercept to be 1.1 from the GWAS summary data.

We did not consider directly comparing to FUSION in our above TWAS analyses of cardiovascular disease since we used the summary-level reference data eQTLGen. However, to assess the performance of OTTERS and FUSION in a real study where individual-level reference data are available, we performed an additional TWAS analysis of cardiovascular disease in the UK Biobank using the GTEx V8 data of 574 whole blood samples as the reference data. We trained OTTERS Stage I using cis-eQTL summary statistics obtained from these 574 GTEx V8 whole blood samples and reference LD from GTEx V8 WGS samples, and trained FUSION models using individual-level genotype data and gene expression data of the same 574 whole blood samples.

We tested TWAS association for 19,653 genes and identified genes with TWAS p-values $< 2.53 \times 10^{-6}$ (Bonferroni corrected significance level) as significant TWAS genes. Training $R^2 >$

0.01 was used to select "valid" GReX imputation models for TWAS (Figure S10). To identify independently significant TWAS genes, we calculated the training R^2 between the GReX predicted by lassosum for of each pair of genes, since lassosum had the best training R^2 (Figure S10). For a pair of genes with the predicted GReX $R^2 > 0.5$, we only kept the gene with the smaller TWAS p-value as the independently significant gene. As a result, OTTERS obtained 34 independently significant TWAS genes, while FUSION identified 21 independently significant TWAS genes (Figure S11). A total of 14 genes were identified by both FUSION and OTTERS (Table S3).

These results demonstrate the advantages of OTTERS for using multiple PRS training methods to account for the unknown genetic architecture of gene expression, which is consistent in our simulation results. These results also showed the advantage of using eQTL summary data with a larger training sample size, as more independently significant TWAS genes were identified by using the eQTLGen summary reference data (38 vs. 34), even with a more stringent rule (test instead of training $R^2 > 0.01$) applied to select test genes with "valid" GReX imputation models.

Computational Time

The computational time per gene of different PRS methods depends on the number of test variants considered for the target gene. Thus, we calculated the computational time and memory usage for 4 groups of genes whose test variants were <2000, between 2000 and 3000, between 3000 and 4000, and >4000, respectively. Among all tested genes in our real studies, the median number of test variants per gene is 3152, and the proportion of genes in each group is 10.3%, 33.4%, 34.5%, and 21.8%, respectively. For each group, we randomly selected 10 genes on Chromosome 4 to evaluate the average computational time and memory usage per gene. We benchmarked the computational time and memory usage of each method on one Intel(R) Xeon(R) processor (2.10 GHz). The evaluation was based on 1000 MCMC iterations for SDPR and PRS-CS (default) without parallel computation (Table S4). We showed that *P+T* and *lassosum were*

computationally more efficient than *SDPR* and *PRS-CS*, whose speed were impeded by the need of MCMC iterations. Between the two Bayesian methods, SDPR implemented in C++ uses significantly less time and memory than PRS-CS implemented in Python.

Discussion

Our OTTERS framework represents an omnibus TWAS tool that can leverage summary-level expression and genotype results from a reference sample, thereby robustly expanding the use of TWAS into more settings. To this end, we adapted and evaluated five different PRS methods assuming different underlying genetic models, including the relatively simple method *P+T*[26] with two different p-value thresholds (0.001 and 0.05), the frequentist method *lassosum*[27], as well as the Bayesian methods *PRS-CS*[28] and *SDPR*[29] within our omnibus test for optimal inference. We note that additional PRS methods such as MegaPRS[30] or PUMAS[31] could also be implemented as additional OTTERS Stage I training methods. Higher TWAS power might be obtained by adding more PRS methods in OTTERS Stage I, with additional computation cost. We also note that the existing SMR-HEIDI[53] method, which uses summary-level data from GWAS and eQTL studies to test for possible causal genetic effects of a trait of interest that were mediated through gene expression, could also be used as an alternative method besides TWAS. However, the SMR method generally restricts eQTL for consideration, excluding those where the eQTL p-values larger than a certain threshold, e.g., 0.05.

In simulation studies, we demonstrated that the performance of each of these five PRS methods depended substantially on the underlying genetic architecture for gene expression, with *P+T* methods generally performing better for sparse architecture whereas the Bayesian methods performing better for denser architecture. Consequently, since genetic architecture of gene expression is unknown apriori, we believe this justifies the use of the omnibus TWAS test implemented in OTTERS for practical use as this test had near-optimal performance across all simulation scenarios considered. While we developed our methods with summary-level reference

data in mind, we note that our prediction methods and OTTERS perform well (in terms of imputation accuracy and power) relative to existing TWAS methods like FUSION when individual-level reference data are available.

In our real data application using UKBB GWAS summary-level data, we compared OTTERS TWAS results using reference eQTL summary data from eQTLGen consortium to FUSION TWAS results using a substantially smaller individual-level reference dataset from GTEx V6. OTTERS identified 13 significant TWAS risk genes that were missed by FUSION using individual-level GTEx V6 reference data of blood tissue, suggesting that the use of larger reference datasets like eQTLGen in TWAS can identify novel findings. Interestingly, the genes missed by FUSION were instead detected using individual-level GTEx reference data of other tissue types that are more directly related to cardiovascular disease. By comparing OTTERS to FUSION when the same individual-level GTEx V8 reference data of whole blood samples were used, we still observed that OTTERS identified more risk genes than FUSION, which we believe is due to the former method accounting for the unknown genetic architecture of gene expression by using multiple regression methods to train GReX imputation models. These applied results were consistent with our simulation results.

Among all individual methods, *P+T* is the most computationally efficient method. The Bayesian methods *SDPR* and *PRS-CS* require more computation time than the frequentist method *lassosum* as the former set of methods require a large number of MCMC iterations for model fit. By comparing the performance of these five methods in terms of the imputation accuracy and TWAS power in simulations and real applications, we conclude that none of these methods were optimal across different genetic architectures. We found that all methods provided distinct and considerable contributions to the final OTTERS TWAS results. These results demonstrate the benefits of OTTERS in practice, since OTTERS can combine the strength of these individual methods to achieve the optimal performance.

To enable the use of OTTERS by the public, we provide an integrated tool (see Availability of data and materials) to: (1) Train GReX imputation models (i.e., estimate eQTL weights in Stage I) using eQTL summary data by *P+T*, *lassosum*, *SDPR*, and *PRS-CS*; (2) Conduct TWAS (i.e., testing gene-trait association in Stage II) using both individual-level and summary-level GWAS data with the estimated eQTL weights; and (3) Apply ACAT-O to aggregate the TWAS p-values from individual training methods. Since the existing tools for *P+T*, *lassosum*, *SDPR*, and *PRS-CS* were originally developed for PRS calculations, we adapted and optimized them for training GReX imputation models in our OTTERS tool. For example, we integrate TABIX[54] and PLINK[55] tools in OTTERS to extract input data per target gene more efficiently. We also enable parallel computation in OTTERS for training GReX imputation models and testing gene-trait association of multiple genes.

The OTTERS framework does have its limitations. First, training GReX imputation models by all individual methods on average cost ~20 minutes for all 5 training models per gene, which might be computationally challenging for studying eQTL summary data of multiple tissue types and for ~20K genome-wide genes. Users might consider prioritize *P+T(0.001)*, *lassosum,* and *SDPR* training methods as these three provide complementary results in our studies. Second, the currently available eQTL summary statistics are mainly derived from individuals of European descent. Our OTTERS trained GReX imputations model based on these eQTL summary statistics, and the resulting imputed GReX could consequently have attenuated cross-population predictive performance[56]. This might limit the transferability of our TWAS results across populations. Third, our OTTERS cannot provide the direction of the identified gene-phenotype associations, which should be referred to the sign of the TWAS Z-score statistic per training method. Last, even though the method applies to integrate both cis- and trans- eQTL with GWAS data, the computation time and availability of summary-level trans-eQTL reference data are still the main obstacles. Our current OTTERS tool only considers cis-eQTL effects. Extension of

OTTERS to enable cross-population TWAS and incorporation of trans-eQTL effects is part of our ongoing research but out of the scope of this work.

Our novel OTTERS framework using large-scale eQTL summary data has the potential to identify more significant TWAS risk genes than standard TWAS tools that use smaller individual-level reference transcriptomic data and use only a single regression method for training GReX imputation models. This tool provides the opportunity to leverage not only available public eQTL summary data of various tissues for conducting TWAS of complex traits and diseases, but also the emerging summary-level data of other types of molecular QTL such as splicing QTLs, methylation QTLs, metabolomics QTLs, and protein QTLs. For example, OTTERS could be applied to perform proteome-wide association studies using summary-level reference data of genetic-protein relationships such as those reported by the SCALLOP consortium[57], and epigenome-wide association studies using summary-level reference data of methylation-phenotype relationships reported by Genetics of DNA Methylation Consortium (GoDMC) (see Web Resources). OTTERS would be most useful for the broad researchers who only have access to summary-level QTL reference data and summary-level GWAS data. The feasibility of integrating summary-level molecular QTL data and GWAS data makes our OTTERS tool valuable for wide application in current multi-omics studies of complex traits and diseases.

Methods

Traditional Two-Stage TWAS Analysis

Stage I of TWAS estimates a GReX imputation model using individual-level expression and genotype data available from a reference dataset. Consider the following GReX imputation model from n individuals and m SNPs (multivariable regression model assuming linear additive genetic effects) within the reference dataset:

$$\boldsymbol{E}_g = \boldsymbol{X}\boldsymbol{w} + \boldsymbol{\epsilon},\ \ \boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2 \boldsymbol{I}). \qquad \text{(Equation 1)}$$

Here, $\boldsymbol{E}_g$ is a vector representing gene expression levels of gene g, $\boldsymbol{X}$ is an $n \times m$ matrix of genotype data of SNP predictors proximal or within gene g, $\boldsymbol{w}$ is a vector of genetic effect sizes (referred to as a broad sense of eQTL effect sizes), and $\boldsymbol{\epsilon}$ is the error term. Here, we consider only cis-SNPs within 1 MB of the flanking 5' and 3' ends as genotype predictors that are coded within $\boldsymbol{X}$[19,20,22]. Once we configure the model in Equation 1, we can employ methods like PrediXcan, FUSION, and TIGAR to fit the model and obtain estimates of eQTL effect sizes ($\widehat{\boldsymbol{w}}$).

Stage II of TWAS uses the eQTL effect sizes ($\widehat{\boldsymbol{w}}$) from Stage I to impute gene expression (GReX) in an independent GWAS and then test for association between GReX and phenotype. Given individual-level GWAS data with genotype data $\boldsymbol{X}_{new}$ and eQTL effect sizes ($\widehat{\boldsymbol{w}}$) from Stage I, the GReX for $\boldsymbol{X}_{new}$ can be imputed by $\widehat{\boldsymbol{GReX}} = \boldsymbol{X}_{new}\,\widehat{\boldsymbol{w}}$. The follow-up TWAS would test the association between $\widehat{\boldsymbol{GReX}}$ and phenotype $\boldsymbol{Y}$ based on a generalized linear regression model, which is equivalent to a gene-based association test taking $\widehat{\boldsymbol{w}}$ as test SNP weights. When individual-level GWAS data are not available, one can apply FUSION and S-PrediXcan test statistics to summary-level GWAS data as follows:

$$Z_{g,FUSION} = \frac{\sum_{j=1}^{J}(\widehat{w}_j Z_j)}{\sqrt{\widehat{\boldsymbol{w}}'\boldsymbol{V}\widehat{\boldsymbol{w}}}}, \quad Z_{g,S-PrediXcan} = \frac{\sum_{j=1}^{J}(\widehat{w}_j \widehat{\sigma}_j Z_j)}{\sqrt{\widehat{\boldsymbol{w}}'\boldsymbol{V}\widehat{\boldsymbol{w}}}} \qquad \text{(Equation 2)}$$

where Z_j is the single variant *Z-score* test statistic in GWAS for the jth SNP, $j = 1, \ldots, J$, for all test SNPs that have both eQTL weights with respect to the test gene g and GWAS Z-scores; $\hat{\sigma}_j$ is the genotype standard deviation of the jth SNP; and $\boldsymbol{V}$ denotes the genotype correlation matrix in FUSION Z-score statistic and genotype covariance matrix in S-PrediXcan Z-score statistic of the test SNPs. In particular, $\hat{\sigma}_j$ and $\boldsymbol{V}$ can be approximated from a reference panel with genotype data of samples of the same ancestry such as those available from the 1000 Genomes Project[58]. If $\widehat{\boldsymbol{w}}$ are standardized effect sizes estimated assuming standardized genotype $\boldsymbol{X}$ and

gene expression $\boldsymbol{E}_g$ in Equation 1, FUSION and S-PrediXcan Z-score statistics are equivalent[13]. Otherwise, the S-PrediXcan Z-score should be applied to avoid false positive inflation.

TWAS Stage I Analysis using Summary-Level Reference Data

We now consider a variation of TWAS Stage I to estimate cis-eQTL effect sizes $\widehat{\boldsymbol{w}}$ based on a multivariable regression model (Equation 1) from summary-level reference data. We assume that the summary-level reference data provide information on the association between a single genetic variant j ($j = 1, \ldots, m$) and expression of gene g. This information generally consists of effect size estimates ($\widetilde{w}_j, j = 1, \ldots, m$) and p-values derived from the following single variant regression models:

$$\boldsymbol{E}_g = \boldsymbol{X}_j w_j + \boldsymbol{\epsilon},\ \ \boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2 \boldsymbol{I}),\ j\ =\ 1, \ldots, m. \qquad \text{(Equation 3)}$$

Here, $\boldsymbol{X}_j$ is an $n \times 1$ vector of genotype data for genetic variant j. Since eQTL summary data are analogous to GWAS summary data where gene expression represents the phenotype, we can estimate the eQTL effect sizes $\widehat{\boldsymbol{w}}$ using marginal least squared effect estimates ($\widetilde{w}_j, j = 1, \ldots, m$) and p-values from the QTL summary data as well as reference linkage disequilibrium (LD) information of the same ancestry[26–29]. Although all PRS methods apply to the TWAS Stage I framework, we only consider four representative methods as follows:

P+T: The *P+T* method selects eQTL weights by LD-clumping and P-value Thresholding[26]. Given threshold P_T for p-values and threshold R_T for LD R^2, we first exclude SNPs with marginal p-values from eQTL summary data greater than P_T or strongly correlated (LD R^2 greater than R_T) with another SNP having a more significant marginal p-value (or Z-score statistic value). For the remaining selected test SNPs, we use marginal standardized eQTL effect sizes from eQTL summary data as eQTL weights for TWAS in Stage II. We considered $R_T = 0.99$ and $P_T = (0.001, 0.05)$ in this paper and implemented the *P+T* method using PLINK 1.9[55] (see Web

Resources). We denote the *P+T* method with P_T equal to 0.001 and 0.05 as *P+T (0.001)* and *P+T (0.05)*, respectively.

Frequentist lassosum: With standardized $\boldsymbol{E_g}$ and **X**, we can show that the marginal least squared eQTL effect size estimates from the single variant regression model (Equation 3) is $\widetilde{\boldsymbol{w}} = \boldsymbol{X}^T\boldsymbol{E}_g/n$ and that the LD correlation matrix is $\boldsymbol{R} = \boldsymbol{X}^T\boldsymbol{X}/n$. That is,

$$\boldsymbol{X}^T\boldsymbol{E}_g = n\widetilde{\boldsymbol{w}} \text{ and } \boldsymbol{X}^T\boldsymbol{X} = n\boldsymbol{R}. \qquad \text{(Equation 4)}$$

By approximating $n\boldsymbol{R}$ by $n\boldsymbol{R_s}$ ($\boldsymbol{R}_s = (1-s)\boldsymbol{R_r} + s\boldsymbol{I}$ with a tuning parameter $0 < s < 1$, a reference LD correlation matrix $\boldsymbol{R_r}$ from an external panel such as one from the 1000 Genomes Project[58], and an identity matrix $\boldsymbol{I}$) in the LASSO[32] penalized loss function, the frequentist *lassosum* method[27] can tune the LASSO penalty parameter and s using a pseudovalidation approach and then solve for eQTL effect size estimates $\widehat{\boldsymbol{w}}$ by minimizing the approximated LASSO loss function requiring no individual-level data (see details in Supplemental Methods).

Bayesian SDPR: Bayesian DPR method[33] as implemented in TIGAR[22] estimates $\widehat{\boldsymbol{w}}$ for the underlying multivariable regression model in Equation 1 by assuming a normal prior $N(0, \sigma_w^2)$ for w_j and a Dirichlet process prior[59] $DP(H, \alpha)$ for σ_w^2 with base distribution H and concentration parameter α. *SDPR*[29] assumes the same DPR model but can be applied to estimate the eQTL effect sizes $\widehat{\boldsymbol{w}}$ using only eQTL summary data (see details in Supplemental Methods).

Bayesian PRS-CS: The *PRS-CS* method[28] assumes the following normal prior for w_j and non-informative scale-invariant Jeffreys prior on the residual variance ${\sigma_\epsilon}^2$ in Equation 1

$$w_j \sim N\left(0, \frac{\sigma_\epsilon^2}{n}\psi_j\right), \;\; p(\sigma_\epsilon^2) \propto {\sigma_\epsilon}^2; \;\; \psi_j \sim Gamma(a, \delta_j), \delta_j \sim Gamma(b, \phi),$$

where local shrinkage parameter ψ_j has an independent gamma-gamma prior and ϕ is a global-shrinkage parameter controlling the overall sparsity of $\boldsymbol{w}$. *PRS-CS* sets hyper parameters a = 1

and b = 1/2 to ensure the prior density of w_j to have a sharp peak around zero to shrink small effect sizes of potentially false eQTL towards zero, as well as heavy, Cauchy-like tails which asserts little influence on eQTLs with larger effects. Posterior estimates $\widehat{\boldsymbol{w}}$ will be obtained from eQTL summary data (i.e., marginal effect size estimates $\widetilde{\boldsymbol{w}}$ and p-values) and reference LD correlation matrix $\boldsymbol{R}$ by Gibbs Sampler (see details in Supplemental Methods). We set ϕ as the square of the proportion of causal variants in the simulation and as 10^{-4} per gene in the real data application.

OTTERS Framework

As shown in Figure 1, OTTERS first trains GReX imputation models per gene *g* using *P+T, lassosum, SDPR,* and *PRS-CS* methods that each infers cis-eQTLs weights using cis-eQTL summary data and an external LD reference panel of similar ancestry (Stage I). Once we derive cis-eQTLs weights for each training method, we can impute the respective GReX using that method and perform the respective gene-based association analysis in the test GWAS dataset using the formulas given in Equation 2 (Stage II). We thus derive a set of TWAS p-values for gene *g*; one p-value for each training model that we applied. We then use these TWAS p-values to create an omnibus test using the ACAT-O[34] approach that employs a Cauchy distribution for inference (see details in Supplemental Methods). We refer to the p-value derived from ACAT-O test as the OTTERS p-value.

Marginal eQTL Effect Sizes

In practice of training GReX imputation models using reference eQTL summary data, the marginal standardized eQTL effect sizes were approximated by $\widetilde{w}_j \approx Z_j/\sqrt{median(n_{g,j})}$, where Z_j denotes the corresponding eQTL Z-score statistic value by single variant test and $median(n_{g,j})$ denotes the median sample size of all cis-eQTLs for the target gene g. The median cis-eQTL sample size per gene was also taken as the sample size value required by *lassosum,*

SDPR, and *PRS-CS* methods, for robust performance. Since summary eQTL datasets (e.g., eQTLGen) were generally obtained by meta-analysis of multiple cohorts, the sample size per test SNP could vary across all cis-eQTLs of the test gene. The median cis-eQTL sample size ensures a robust performance for applying those eQTL summary data based methods.

LD Clumping

We performed LD-clumping with R_T=0.99 for all individual methods in both simulation and real studies. Using PRS-CS as an example, we also showed that LD-clumping does not affect the GReX imputation accuracy compared to no clumping in the real data testing (Figure S12).

LD Blocks for lassosum, PRS-CS, and SDPR

LD blocks were determined externally by ldetect[60] for *lassosum* and *PRS-CS*, while internally for *SDPR* which ensure that SNPs in one LD block do not have nonignorable correlation (R^2 > 0.1) with SNPs in other blocks.

Simulate GWAS Z-score

Given gene expression $\boldsymbol{E}_g$ simulated from the multivariate regression model $\boldsymbol{E}_g = \boldsymbol{X_g w} + \boldsymbol{\epsilon_g}$ with standardized genotype matrix $\boldsymbol{X_g}$ and $\boldsymbol{\epsilon}_g \sim N(0, (1-h_e^2)\boldsymbol{I}$, we assume GWAS phenotype data of n_{gwas} samples are simulated from the following linear regression model

$$\boldsymbol{Y} = h_p(\boldsymbol{X_g w}) + \boldsymbol{\epsilon_p}, \qquad \boldsymbol{\epsilon}_p \sim N(0, \boldsymbol{I}).$$

Conditioning on true genetic effect sizes, the GWAS Z-score test statistics of all test SNPs will follow a multivariate normal distribution, $MVN\left(\boldsymbol{\Sigma_g w}\sqrt{n_{gwas}h_p^2}, \boldsymbol{\Sigma_g}\right)$, where $\boldsymbol{\Sigma_g}$ is the correlation matrix of the standardized genotype X_g from test samples, and h_p^2 denotes the amount of phenotypic variance explained by simulated GReX=$\boldsymbol{X_g w}$[38]. Thus, for a given GWAS sample size, we can generate GWAS Z-score statistic values from this multivariate normal distribution.

FUSION using Individual-level Reference Data

To train GReX imputation models by FUSION with individual-level reference data, we trained Best Linear Unbiased Predictor (BLUP) model[61], Elastic-net regression[62], LASSO regression[32], and single best eQTL model as implemented in the FUSION tool (see Web Resource). Default settings were used to train GReX imputation models by FUSION in our simulation and real studies. LASSO regression was performed only for genes with positive estimated expression heritability. The eQTL weights of the best trained GReX imputation model will be used to conduct TWAS by FUSION.

GTEx V8 Dataset

GTEx V8 dataset (dbGaP phs000424.v8.p2) contains comprehensive profiling of WGS genotype data and RNA-sequencing (RNA-seq) transcriptomic data across 54 human tissue types of 838 donors. The GTEx V8 WGS genotype data of all samples were used to construct reference LD in our studies. The GTEx V6 RNA-seq data of whole blood samples were used to train GReX imputation models by FUSION, and the GTEx V8 RNA-seq data of additional whole blood samples (n=315) were used to test GReX imputation accuracy in our studies. GTEx V8 RNA-seq data of all whole blood samples (n=574) were also used as reference data for comparing the performance of OTTERS and FUSION.

eQTLGen Consortium Dataset

The eQTLGen consortium[23] dataset was generated based on meta-analysis across 37 individual cohorts (n=31,684) including GTEx V6 as a sub-cohort. eQTLGen samples consist of 25,482 blood (80.4%) and 6,202 peripheral blood mononuclear cell (19.6%) samples. We considered SNPs with minor allele frequency (MAF) > 0.01, Hardy–Weinberg P value >0.0001, call rate >0.95, genotype imputation r^2 > 0.5 and observed in at least 2 cohorts[23]. We only considered cis-eQTL (within $\pm$1MB around gene transcription start sites (TSS)) with test sample

size > 3000. As a result, we used cis-eQTL summary data of 16,699 genes from eQTLGen to train GReX imputation models for use in OTTERS in this study.

UK Biobank GWAS Data of Cardiovascular Disease

Summary-level GWAS data of Cardiovascular Disease from UKBB (n=459,324, case fraction = 0.319)[35] were generated by BOLT-LMM based on the Bayesian linear mixed model per SNP[63] with assessment centered, sex, age, and squared age as covariates. Although BOLT-LMM was derived based on a quantitative trait model, it can be applied to analyze case-control traits and has well-controlled false positive rate when the trait is sufficiently balanced with case fraction ≥ 10% and samples are of the same ancestry . The tested dichotomous cardiovascular disease phenotype includes a list of sub-phenotypes: hypertension, heart/cardiac problem, peripheral vascular disease, venous thromboembolic disease, stroke, transient ischaemic attack (tia), subdural haemorrhage/haematoma, cerebral aneurysm, high cholesterol, and other venous/lymphatic disease.

Data Availability

eQTLGen consortium data are available from their portal website (https://www.eqtlgen.org). UK Biobank summary-level GWAS data are available through the Alkes Group (https://alkesgroup.broadinstitute.org/UKBB). Individual-level GTEx reference data are available through dbGap (Accession phs000424.v8.p2). ROS/MAP/MSBB WGS data used in our simulation studies are available through Synapse with data access application (https://www.synapse.org/#!Synapse:syn10901595). All source code and scripts used in this study are available through OTTERS Github page (https://github.com/daiqile96/OTTERS).

Ethics Approval

All data used in this study are de-identified genotype data and summary level eQTL and GWAS data. ROS/MAP genotype data were collected with ethics approval from the IRB at Rush University and all participants consented to participate.

Acknowledgements

The authors thank Dr. Greg Gibson from Georgia Tech for his insightful comments and discussion that help the development and improve the quality of this manuscript. This work was supported by National Institutes of Health grant awards R35GM138313 (QD, JY), RF1AG071170 (QD, MPE), and Estonian Research Council Grant PUT (PRG1291) for TE. NIH/NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, R01AG56352, U01AG32984, U01AG46152, U01AG61356, the Illinois Department of Public Health, the Translational Genomics Research Institute support the generation of the ROS/MAP data led by ASB, PLDJ and DAB. The Young Finns Study has been financially supported by the Academy of Finland: grants 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 255381, 256474, 283115, 319060, 320297, 314389, 338395, 330809, and 104821, 129378 (Salve), 117797 (Gendi), and 141071 (Skidi), the Social Insurance Institution of Finland, Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001), Juho Vainio Foundation, Paavo Nurmi Foundation, Finnish Foundation for Cardiovascular Research, Finnish Cultural Foundation, The Sigrid Juselius Foundation, Tampere Tuberculosis Foundation, Emil Aaltonen Foundation, Yrjö Jahnsson Foundation, Signe and Ane Gyllenberg Foundation, Diabetes Research Foundation of Finnish Diabetes Association, EU Horizon 2020 (grant 755320 for TAXINOMISIS and grant 848146 for To Aition), European Research Council (grant 742927 for MULTIEPIGEN project), Tampere University Hospital Supporting Foundation, and Finnish Society of Clinical Chemistry and the Cancer Foundation Finland. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

QD conducted data analysis and drafted the manuscript; JY and MPE conceptualized and led the project, and edited the manuscript; GZ and HZ consulted data analysis and edited the manuscript; UV, LF, AB, AT, TL, OR, TE, contributed reference eQTL summary data and edited the manuscript; eQTLGen Consortium contributed eQTLGen summary data.

Competing interests

The authors declare no competing interests and consent for publication.

Materials & Correspondence

Correspondence and material requests should be sent to M.P.E. (mpepste@emory.edu) and J.Y. (jingjing.yang@emory.edu).

Description of Supplemental Data

Supplemental data include 12 figures, 4 tables, and supplemental methods.

Web Resources

ACAT: https://github.com/yaowuliu/ACAT

eQTLGen: https://www.eqtlgen.org

FUSION: http://gusevlab.org/projects/fusion

GoDMC: http://www.godmc.org.uk

lassosum: https://github.com/tshmak/lassosum

OTTERS: https://github.com/daiqile96/OTTERS

PLINK: https://www.cog-genomics.org/plink

PRS-CS: https://github.com/getian107/PRScs

SDPR: https://github.com/eldronzhou/SDPR

TWAS hub: http://twas-hub.org

Tables

Table 1. Test R^2 in 315 whole blood tissue samples from GTEx V8.

	P+T(0.001)	P+T(0.05)	lassosum	SDPR	PRS-CS	FUSION[b]
# Genes with R^2 > 0.01	9,816	9,662	8,718	9,670	10,337	4,704
Median R^2[a]	0.044	0.0430	0.0416	0.0418	0.0517	0.0367

a: Median R^2 among genes with test $R^2 > 0.01$ per method.

b: FUSION was trained on GTEx V6 blood samples while all other training methods were trained using eQTLGen summary statistics (n=31,684) and reference LD from GTEx V8 samples.

Table 2. Independent TWAS risk genes of cardiovascular disease identified by OTTERS. Reference eQTL summary data from eQTLGen consortium and GWAS summary data from UKBB were used. The corresponding TWAS p-values by 5 individual PRS methods and OTTERS are shown in the table with significant p-values in bold, and those for genes with test GReX $R^2 \leqslant 0.01$ were shown as a dash.

CHROM	ID	OTTERS	P+T(0.001)	P+T(0.05)	lassosum	SDPR	PRS-CS
1	CLCN6[a]	**5.75E-15**	**4.94E-09**	**5.40E-08**	**8.77E-09**	**1.19E-15**	**1.43E-09**
1	NPPA[b]	**4.32E-08**	**1.55E-08**	**2.14E-07**	**-**	-	6.71E-06
1	PSRC1[a]	**8.37E-20**	**5.68E-08**	**8.46E-07**	**6.26E-11**	**1.67E-20**	**1.41E-12**
2	RP11-378A13.1[a]	**9.78E-09**	3.97E-02	4.98E-02	1.62E-05	**1.96E-09**	1.15E-04
4	LINC01093[c]	**2.57E-09**	9.85E-02	5.31E-02	**5.13E-10**	1.08E-02	2.41E-02
5	CPEB4[b]	**3.05E-14**	1.26E-02	2.05E-02	2.70E-05	**6.05E-15**	**1.60E-07**
6	SERPINB6[c]	**1.47E-07**	2.12E-01	2.24E-01	7.56E-03	**2.95E-08**	7.53E-04
6	CARMIL1[c]	**9.23E-09**	5.34E-03	3.41E-03	4.15E-03	**1.85E-09**	1.72E-03
6	ZSCAN12P1[c]	**1.84E-08**	6.00E-01	5.75E-01	4.62E-01	**3.67E-09**	3.10E-01
6	HCG4P7[c]	**8.93E-50**	3.70E-01	3.69E-01	2.30E-01	**1.79E-50**	7.26E-01
6	HCG4P3[c]	**5.33E-20**	4.20E-01	4.05E-01	5.03E-04	**1.07E-20**	2.42E-03
6	HLA-S[c]	**4.57E-07**	7.13E-01	7.31E-01	3.02E-01	**9.14E-08**	2.33E-01
7	PSPHP1[c]	**1.21E-09**	2.17E-01	2.26E-01	9.65E-03	**2.43E-10**	1.10E-01
8	LPL[c]	**5.73E-07**	1.78E-03	3.26E-03	4.44E-02	**1.15E-07**	1.05E-04
8	PTP4A3[c]	**1.28E-06**	8.13E-02	8.33E-02	6.23E-05	**2.58E-07**	1.67E-03
10	CAMK1D[a]	**2.51E-09**	3.83E-02	4.97E-02	1.23E-03	**5.03E-10**	4.97E-05
10	NT5C2[b]	**1.21E-07**	**1.69E-06**	**2.92E-06**	1.64E-05	**3.15E-07**	**2.69E-08**
11	TNNT3[b]	**1.67E-10**	**1.09E-06**	3.33E-06	**2.03E-09**	**3.40E-11**	**4.01E-07**
11	C11orf49[b]	**2.28E-06**	**8.55E-07**	**1.78E-06**	5.44E-05	-	2.93E-04
11	SIDT2[a]	**7.26E-09**	6.14E-05	1.33E-04	3.66E-05	**1.46E-09**	**3.81E-07**
15	CSK[b]	**2.30E-09**	**1.70E-07**	**2.15E-06**	**7.41E-10**	**2.80E-09**	**2.17E-09**
15	FES[b]	**2.87E-32**	**4.78E-08**	**1.23E-06**	**9.13E-24**	**5.75E-33**	**1.94E-15**
15	SLCO3A1[c]	**3.78E-08**	1.85E-02	3.15E-02	4.65E-05	**7.57E-09**	1.14E-03
16	MBTPS1[b]	**5.80E-08**	2.62E-01	3.05E-01	9.15E-04	**1.16E-08**	2.34E-03
16	MTHFSD[a]	**4.65E-07**	5.16E-02	5.94E-02	1.65E-02	**9.30E-08**	3.20E-03
17	ACE[b]	**9.42E-07**	4.93E-06	1.03E-05	4.23E-06	**9.66E-07**	**2.68E-07**
18	RALBP1[c]	**1.40E-06**	1.48E-01	1.54E-01	2.12E-04	**2.81E-07**	5.55E-03
19	MRI1[b]	**8.38E-09**	8.34E-03	1.60E-02	7.79E-03	**1.68E-09**	2.65E-03
19	HAUS8[b]	**1.60E-07**	**4.41E-08**	**1.38E-07**	**1.67E-06**	**1.42E-06**	3.29E-05
19	SULT2B1[c]	**2.32E-06**	**7.73E-07**	**-**	-	2.97E-02	1.10E-02
19	NTN5[a]	**9.03E-10**	**2.75E-08**	**1.16E-07**	6.23E-06	**1.85E-10**	**9.73E-09**
19	RPL28[b]	**3.76E-07**	7.33E-02	1.16E-01	6.64E-03	**7.52E-08**	4.23E-03
20	CTSZ[b]	**3.32E-09**	2.57E-02	1.99E-02	**3.40E-09**	**8.25E-10**	1.04E-01
20	EDN3[c]	**1.29E-07**	**3.61E-08**	**9.15E-08**	8.60E-06	5.90E-03	1.58E-02
20	ZBTB46[c]	**1.07E-06**	**2.83E-07**	8.35E-06	-	1.81E-03	1.27E-05
20	OPRL1[a]	**5.84E-07**	**3.44E-07**	**2.69E-06**	1.85E-03	5.51E-05	**1.90E-07**
21	FAM3B[c]	**1.08E-10**	2.28E-02	2.58E-02	8.07E-06	**2.17E-11**	1.04E-05
21	MX1[c]	**6.04E-22**	4.36E-01	3.83E-01	**3.16E-07**	**1.21E-22**	1.24E-03

a: Risk gene of UKBB cardiovascular disease in TWAS-hub identified using GTEx whole blood tissue.

b: Risk genes of UKBB cardiovascular disease in TWAS-hub identified using other GTEx tissue types.

c: Novel risk gene

Figure Titles and Legends

Figure 1. OTTERS framework.

OTTERS estimates cis-eQTL weights from eQTL summary data and reference LD panel using four imputation models (Stage I), and conducts ACAT-O test to combine gene-based association test p-values from individual methods with individual/summary level test GWAS data (Stage II).

Figure 2. Test R^2 (A) and TWAS power (B) comparison in simulation studies

Various proportions of true causal cis-eQTL $p_{causal} = (0.001, 0.01)$ and gene expression heritability $h_e^2 = (0.01, 0.05, 0.1)$ were considered in the simulation studies. The GWAS sample size was chosen with respect to h_e^2 values. The proportion of phenotype variance explained by gene expression (h_p^2) was set to be 0.025. TWAS was conducted using simulated GWAS Z-scores.

Figure 3. Test R^2 by *PRS-CS* versus *P+T(0.001), P+T(0.05), lassosum, SDPR, FUSION.*

Test R^2 by *PRS-CS* versus *P+T(0.001) (A)*, *P+T(0.05) (B), lassosum (C)*, *SDPR (D), and FUSION (E)* with 315 GTEx V8 test samples, with different colors denoting whether test R^2 > 0.01 only by PRS-CS (red), only by the y axis method (green), or both methods (blue). Genes with test $R^2 >$ 0.01 by at least one method were included in the plot.

Figure 4. Manhattan plot of TWAS results by OTTERS.

Manhattan plot of TWAS results by OTTERS using GWAS summary-level statistics of cardiovascular disease and imputation models fitted based on eQTLGen summary statistics. Independently significant TWAS risk genes are labeled.

eQTLGen Consortium Author List

Mawussé Agbessi[1], Habibul Ahsan[2], Isabel Alves[1], Anand Kumar Andiappan[3], Wibowo Arindrarto[4], Philip Awadalla[1], Alexis Battle[5,6], Frank Beutner[7], Marc Jan Bonder[8,9], Dorret I. Boomsma[10], Mark W. Christiansen[11], Annique Claringbould[8,12], Patrick Deelen[8,13,12,14], Tõnu Esko[15], Marie-Julie Favé[1], Lude Franke[8,12], Timothy Frayling[16], Sina A. Gharib[11,17], Greg Gibson[18], Bastiaan T. Heijmans[4], Gibran Hemani[19], Rick Jansen[20], Mika Kähönen[21], Anette Kalnapenkis[15], Silva Kasela[15], Johannes Kettunen[22], Yungil Kim[23,5], Holger Kirsten[24], Peter Kovacs[25], Knut Krohn[26], Jaanika Kronberg[15], Viktorija Kukushkina[15], Zoltan Kutalik[27], Bernett Lee[3], Terho Lehtimäki[28], Markus Loeffler[24], Urko M. Marigorta[18,29,30], Hailang Mei[31], Lili Milani[15], Grant W. Montgomery[32], Martina Müller-Nurasyid[33,34,35], Matthias Nauck[36,37], Michel G. Nivard[38], Brenda Penninx[20], Markus Perola[39], Natalia Pervjakova[15], Brandon L. Pierce[2], Joseph Powell[40], Holger Prokisch[41,42], Bruce M. Psaty[11,43], Olli T. Raitakari[44], Samuli Ripatti[45], Olaf Rotzschke[3], Sina Rüeger[27], Ashis Saha[5], Markus Scholz[24], Katharina Schramm[46,34], Ilkka Seppälä[28], Eline P. Slagboom[4], Coen D.A. Stehouwer[47], Michael Stumvoll[48], Patrick Sullivan[49], Peter A.C. 't Hoen[50], Alexander Teumer[51], Joachim Thiery[52], Lin Tong[2], Anke Tönjes[48], Jenny van Dongen[10], Maarten van Iterson[4], Joyce van Meurs[53], Jan H. Veldink[54], Joost Verlouw[53], Peter M. Visscher[32], Uwe Völker[55], Urmo Võsa[8,15], Harm-Jan Westra[8,12], Cisca Wijmenga[8], Hanieh Yaghootkar[16,56,57], Jian Yang[32,58], Biao Zeng[18], Futao Zhang[32]

Author list is ordered alphabetically

1. Computational Biology, Ontario Institute for Cancer Research, Toronto, Canada
2. Department of Public Health Sciences, University of Chicago, Chicago, United States of America
3. Singapore Immunology Network, Agency for Science, Technology and Research, Singapore, Singapore
4. Leiden University Medical Center, Leiden, The Netherlands
5. Department of Computer Science, Johns Hopkins University, Baltimore, United States of America
6. Departments of Biomedical Engineering, Johns Hopkins University, Baltimore, United States of America
7. Heart Center Leipzig, Universität Leipzig, Leipzig, Germany
8. Department of Genetics, University Medical Centre Groningen, Groningen, The Netherlands
9. European Molecular Biology Laboratory, Genome Biology Unit, 69117 Heidelberg, Germany
10. Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam Public Health research institute and Amsterdam Neuroscience, the Netherlands
11. Cardiovascular Health Research Unit, University of Washington, Seattle, United States of America
12. Oncode Institute
13. Genomics Coordination Center, University Medical Centre Groningen, Groningen, The Netherlands
14. Department of Genetics, University Medical Centre Utrecht, P.O. Box 85500, 3508 GA, Utrecht, The Netherlands
15. Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu 51010, Estonia
16. Genetics of Complex Traits, University of Exeter Medical School, Royal Devon & Exeter Hospital, Exeter, United Kingdom
17. Department of Medicine, University of Washington, Seattle, United States of America
18. School of Biological Sciences, Georgia Tech, Atlanta, United States of America

19. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom
20. Amsterdam UMC, Vrije Universiteit, Department of Psychiatry, Amsterdam Public Health research institute and Amsterdam Neuroscience, The Netherlands
21. Department of Clinical Physiology, Tampere University Hospital and Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
22. University of Helsinki, Helsinki, Finland
23. Genetics and Genomic Science Department, Icahn School of Medicine at Mount Sinai, New York, United States of America
24. Institut für Medizinische InformatiK, Statistik und Epidemiologie, LIFE – Leipzig Research Center for Civilization Diseases, Universität Leipzig, Leipzig, Germany
25. IFB Adiposity Diseases, Universität Leipzig, Leipzig, Germany
26. Interdisciplinary Center for Clinical Research, Faculty of Medicine, Universität Leipzig, Leipzig, Germany
27. Lausanne University Hospital, Lausanne, Switzerland
28. Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
29. Integrative Genomics Lab, CIC bioGUNE, Bizkaia Science and Technology Park, Derio, Bizkaia, Basque Country, Spain
30. IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
31. Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands
32. Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia
33. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
34. Department of Medicine I, University Hospital Munich, Ludwig Maximilian's University, München, Germany
35. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany
36. Institute of Clinical Chemistry and Laboratory Medicine, Greifswald University Hospital, Greifswald, Germany
37. German Center for Cardiovascular Research (partner site Greifswald), Greifswald, Germany
38. Department of Biological Psychology, Faculty of Behaviour and Movement Sciences, VU, Amsterdam, The Netherlands
39. National Institute for Health and Welfare, University of Helsinki, Helsinki, Finland
40. Garvan Institute of Medical Research, Garvan-Weizmann Centre for Cellular Genomics, Sydney, Australia
41. Institute of Neurogenomics, Helmholtz Zentrum München, Neuherberg, Germany
42. Institute of Human Genetics, Technical University Munich, Munich, Germany
43. Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States of America
44. Centre for Population Health Research, Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital and University of Turku, Turku, Finland
45. Statistical and Translational Genetics, University of Helsinki, Helsinki, Finland
46. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
47. Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, The Netherlands
48. Department of Medicine, Universität Leipzig, Leipzig, Germany
49. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

50. Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center Nijmegen, Nijmegen, The Netherlands
51. Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
52. Institute for Laboratory Medicine, LIFE – Leipzig Research Center for Civilization Diseases, Universität Leipzig, Leipzig, Germany
53. Department of Internal Medicine, Erasmus Medical Centre, Rotterdam, The Netherlands
54. UMC Utrecht Brain Center, University Medical Center Utrecht, Department of Neurology, Utrecht University, Utrecht, The Netherlands
55. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
56. School of Life Sciences, College of Liberal Arts and Science, University of Westminster, 115 New Cavendish Street, London, United Kingdom
57. Division of Medical Sciences, Department of Health Sciences, Luleå University of Technology, Luleå, Sweden
58. Institute for Advanced Research, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China

References

1. Mancuso, N. *et al.* Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *Am J Hum Genet* **100**, 473–487 (2017).
2. Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat Genet* **50**, 538–548 (2018).
3. Mancuso, N. *et al.* Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nat Commun* **9**, 4079 (2018).
4. Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* **51**, 592–599 (2019).
5. Strunz, T., Lauwen, S., Kiel, C., Hollander, A. den & Weber, B. H. F. A transcriptome-wide association study based on 27 tissues identifies 106 genes potentially relevant for disease pathology in age-related macular degeneration. *Sci Rep* **10**, 1584 (2020).
6. Raj, T. *et al.* Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat Genet* **50**, 1584–1592 (2018).
7. Hao, S., Wang, R., Zhang, Y. & Zhan, H. Prediction of Alzheimer's Disease-Associated Genes by Integration of GWAS Summary Data and Expression Data. *Frontiers in Genetics* **9**, (2019).
8. Luningham, J. M. *et al.* Bayesian Genome-wide TWAS Method to Leverage both cis- and trans-eQTL Information through Summary Statistics. *The American Journal of Human Genetics* **107**, 714–726 (2020).
9. Hoffman, J. D. *et al.* Cis-eQTL-based trans-ethnic meta-analysis reveals novel genes associated with breast cancer risk. *PLoS Genet* **13**, e1006690 (2017).
10. Wu, L. *et al.* A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. *Nat Genet* **50**, 968–978 (2018).

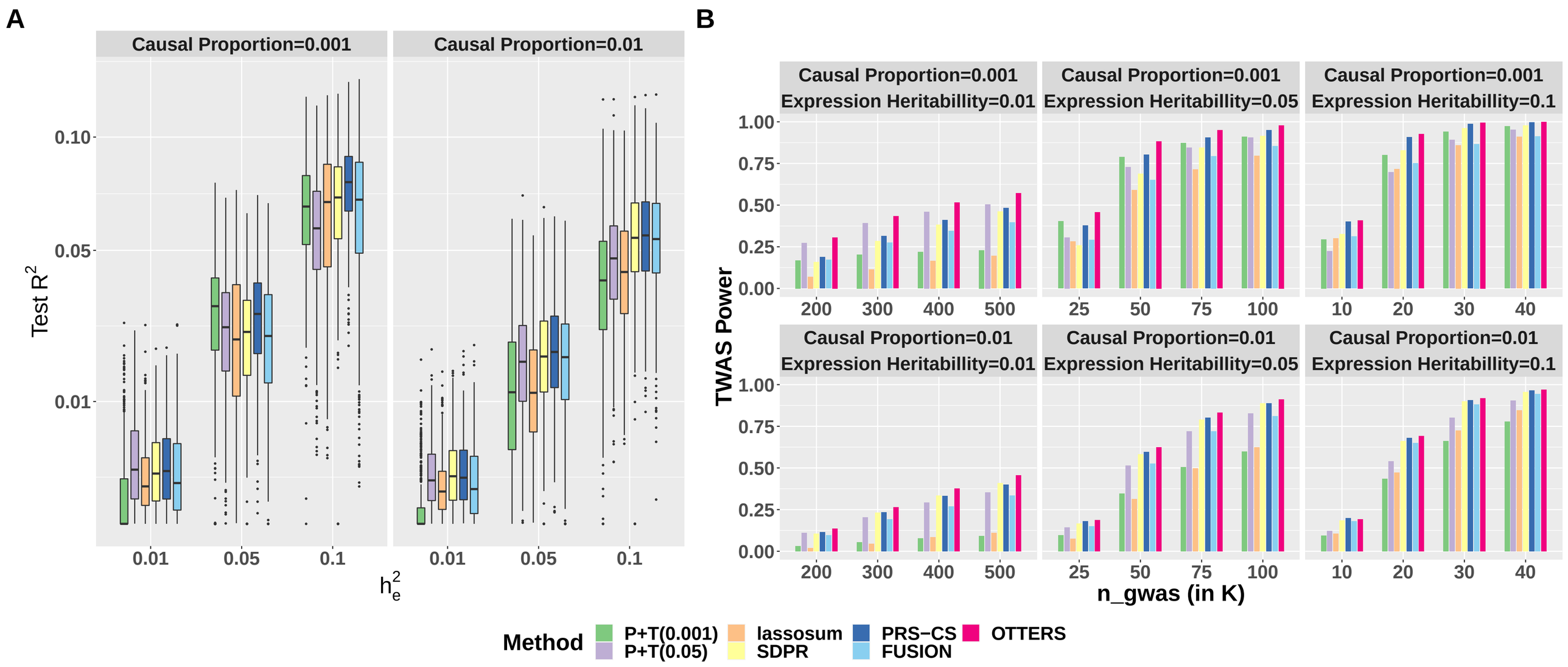
856 11. Bhattacharya, A. *et al.* A framework for transcriptome-wide association studies in breast cancer in
857 diverse study populations. *Genome Biol* **21**, 42 (2020).
858 12. Gusev, A. *et al.* A transcriptome-wide association study of high-grade serous epithelial ovarian
859 cancer identifies new susceptibility genes and splice variants. *Nat Genet* **51**, 815–823 (2019).
860 13. Parrish, R. L., Gibson, G. C., Epstein, M. P. & Yang, J. TIGAR-V2: Efficient TWAS tool with
861 nonparametric Bayesian eQTL weights of 49 tissue types from GTEx V8. *Human Genetics and*
862 *Genomics Advances* **3**, 100068 (2022).
863 14. Thériault, S. *et al.* Genetic Association Analyses Highlight IL6, ALPL, and NAV1 As 3 New
864 Susceptibility Genes Underlying Calcific Aortic Valve Stenosis. *Circulation: Genomic and Precision*
865 *Medicine* **12**, e002617 (2019).
866 15. Zhu, Z. *et al.* Genetic overlap of chronic obstructive pulmonary disease and cardiovascular disease-
867 related traits: a large-scale genome-wide cross-trait analysis. *Respiratory Research* **20**, 64 (2019).
868 16. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580–585 (2013).
869 17. THE GTEX CONSORTIUM. The GTEx Consortium atlas of genetic regulatory effects across human
870 tissues. *Science* **369**, 1318–1330 (2020).
871 18. Gibbs, J. R. *et al.* Abundant Quantitative Trait Loci Exist for DNA Methylation and Gene Expression in
872 Human Brain. *PLOS Genetics* **6**, e1000952 (2010).
873 19. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference
874 transcriptome data. *Nat Genet* **47**, 1091–1098 (2015).
875 20. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat*
876 *Genet* **48**, 245–252 (2016).
877 21. Tang, S. *et al.* Novel Variance-Component TWAS method for studying complex human diseases with
878 applications to Alzheimer's dementia. *PLOS Genetics* **17**, e1009482 (2021).

879 22. Nagpal, S. *et al.* TIGAR: An Improved Bayesian Tool for Transcriptomic Data Imputation Enhances
880 Gene Mapping of Complex Traits. *The American Journal of Human Genetics* **105**, 258–266 (2019).
881 23. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and
882 polygenic scores that regulate blood gene expression. *Nat Genet* **53**, 1300–1310 (2021).
883 24. The CommonMind Consortium (CMC) *et al.* Large eQTL meta-analysis reveals differing patterns
884 between cerebral cortical and cerebellar brain regions. *Sci Data* **7**, 340 (2020).
885 25. Cao, C. *et al.* Power analysis of transcriptome-wide association study: Implications for practical
886 protocol choice. *PLOS Genetics* **17**, e1009405 (2021).
887 26. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar
888 disorder. *Nature* **460**, 748–752 (2009).
889 27. Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X. & Sham, P. C. Polygenic scores via penalized
890 regression on summary statistics. *Genetic Epidemiology* **41**, 469–480 (2017).
891 28. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression
892 and continuous shrinkage priors. *Nat Commun* **10**, 1776 (2019).
893 29. Zhou, G. & Zhao, H. A fast and robust Bayesian nonparametric method for prediction of complex
894 traits using summary statistics. *PLoS Genet* **17**, e1009697 (2021).
895 30. Zhang, Q., Privé, F., Vilhjálmsson, B. & Speed, D. Improved genetic prediction of complex traits from
896 individual-level data or summary statistics. *Nat Commun* **12**, 4192 (2021).
897 31. Zhao, Z. *et al.* PUMAS: fine-tuning polygenic risk scores with GWAS summary statistics. *Genome*
898 *Biology* **22**, 257 (2021).
899 32. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical*
900 *Society: Series B (Methodological)* **58**, 267–288 (1996).
901 33. Zeng, P. & Zhou, X. Non-parametric genetic prediction of complex traits with latent Dirichlet process
902 regression models. *Nat Commun* **8**, 456 (2017).

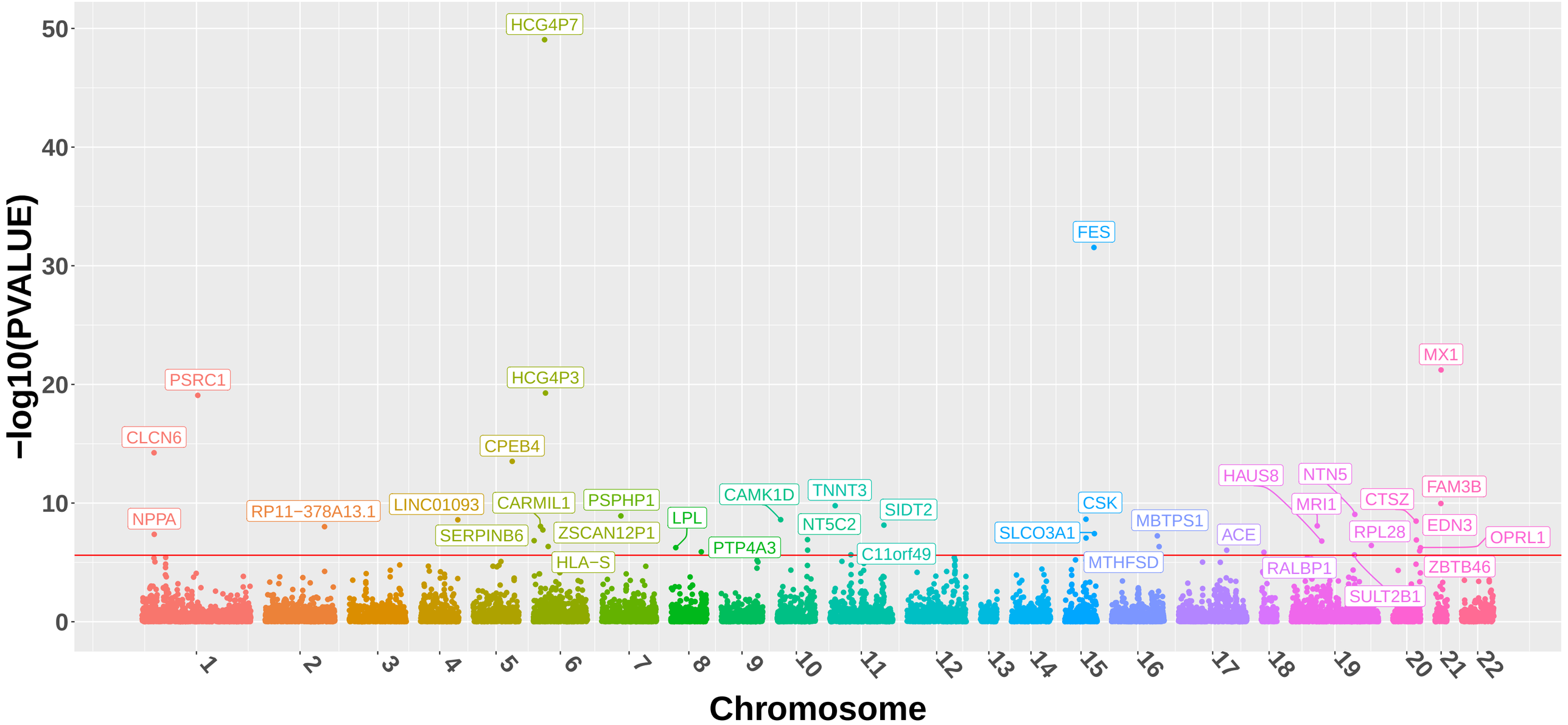
903 34. Liu, Y. *et al.* ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in
904 Sequencing Studies. *The American Journal of Human Genetics* **104**, 410–421 (2019).
905 35. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-
906 scale datasets. *Nat Genet* **50**, 906–908 (2018).
907 36. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference
908 transcriptome data. *Nat Genet* **47**, 1091–1098 (2015).
909 37. Li, X. *et al.* Dynamic incorporation of multiple in silico functional annotations empowers rare variant
910 association analysis of large whole-genome sequencing studies at scale. *Nat Genet* **52**, 969–983
911 (2020).
912 38. Feng, H. *et al.* Leveraging expression from multiple tissues using sparse canonical correlation
913 analysis and aggregate tests improves the power of transcriptome-wide association studies. *PLOS*
914 *Genetics* **17**, e1008973 (2021).
915 39. Wang, T., Ionita-Laza, I. & Wei, Y. Integrated Quantile RAnk Test (iQRAT) for gene-level associations.
916 *arXiv:1910.10102 [stat]* (2020).
917 40. Bennett, D. A., Schneider, J. A., Arvanitakis, Z. & Wilson, R. S. OVERVIEW AND FINDINGS FROM THE
918 RELIGIOUS ORDERS STUDY. *Curr Alzheimer Res* **9**, 628–645 (2012).
919 41. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J Alzheimers Dis* **64**,
920 S161–S189 (2018).
921 42. Wang, M. *et al.* The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data
922 in Alzheimer's disease. *Sci Data* **5**, 180185 (2018).
923 43. Bhattacharya, A., Li, Y. & Love, M. I. MOSTWAS: Multi-Omic Strategies for Transcriptome-Wide
924 Association Studies. *PLOS Genetics* **17**, e1009398 (2021).
925 44. Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on gene expression
926 across human tissues. *Nature* **550**, 204–213 (2017).

927 45. Devlin, B., Roeder, K. & Wasserman, L. Genomic Control, a New Approach to Genetic-Based
928 Association Studies. *Theoretical Population Biology* **60**, 155–166 (2001).
929 46. Fuchs, F. D. & Whelton, P. K. High Blood Pressure and Cardiovascular Disease. *Hypertension* **75**, 285–
930 292 (2020).
931 47. Masaki, T. The endothelin family: an overview. *J Cardiovasc Pharmacol* **35**, S3-5 (2000).
932 48. Xue, H., Pan, W. & Initiative, for the A. D. N. Some statistical consideration in transcriptome-wide
933 association studies. *Genetic Epidemiology* **44**, 221–232 (2020).
934 49. Liu, A. E. & Kang, H. M. Meta-imputation of transcriptome from genotypes across multiple datasets
935 by leveraging publicly available summary-level data. *PLOS Genetics* **18**, e1009571 (2022).
936 50. Yang, Y. *et al.* CoMM-S2: a collaborative mixed model using summary statistics in transcriptome-
937 wide association studies. *Bioinformatics* **36**, 2009–2016 (2020).
938 51. Yuan, Z. *et al.* Testing and controlling for horizontal pleiotropy with probabilistic Mendelian
939 randomization in transcriptome-wide association studies. *Nat Commun* **11**, 3861 (2020).
940 52. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in
941 genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
942 53. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene
943 targets. *Nat Genet* **48**, 481–487 (2016).
944 54. Li, H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* **27**,
945 718–719 (2011).
946 55. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets.
947 *Gigascience* **4**, 7 (2015).
948 56. Keys, K. L. *et al.* On the cross-population generalizability of gene expression prediction models. *PLOS*
949 *Genetics* **16**, e1008927 (2020).

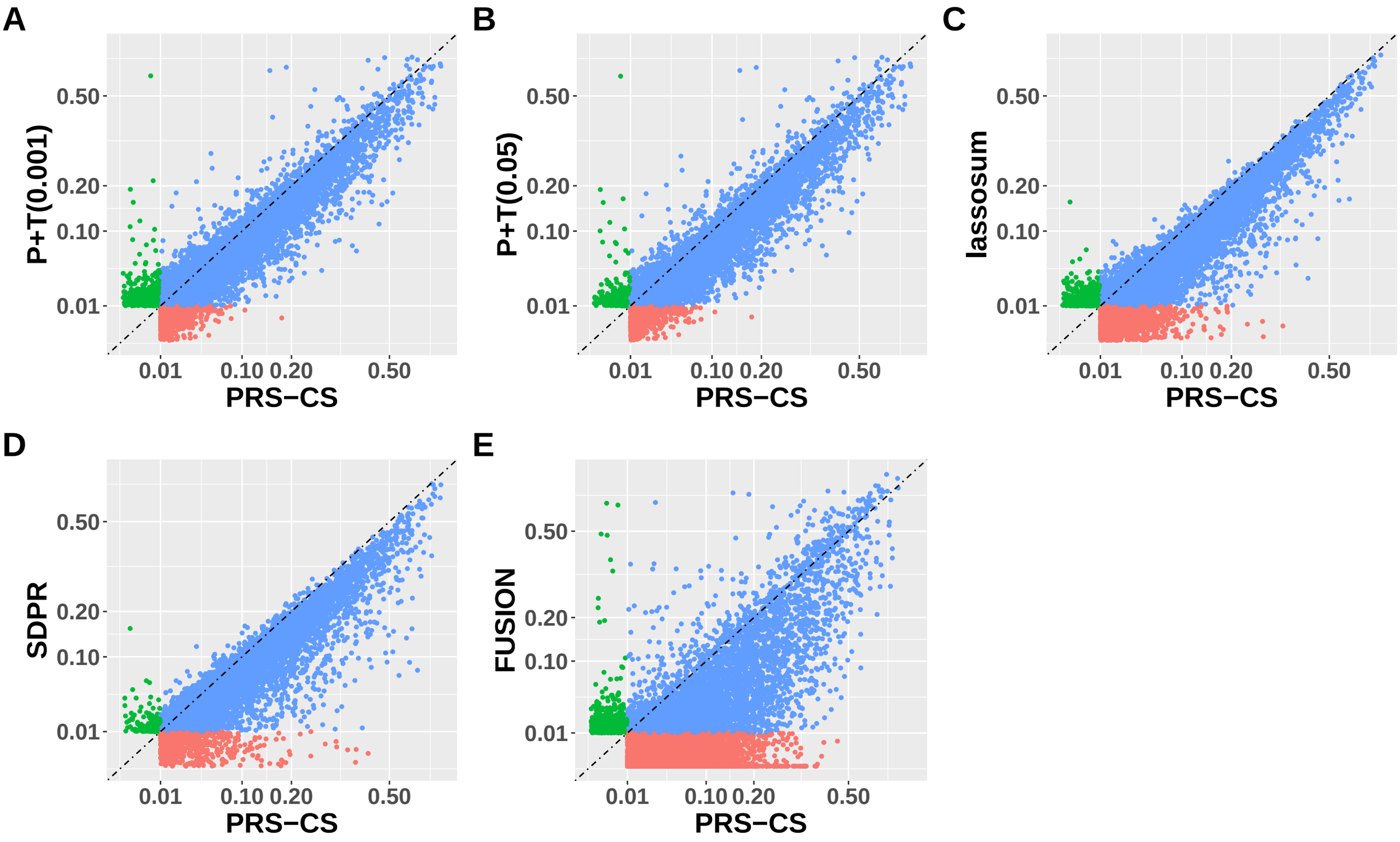
950 57. Folkersen, L. *et al.* Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931
951 individuals. *Nat Metab* **2**, 1135–1148 (2020).
952 58. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
953 59. Lijoi, A., Prünster, I. & Walker, S. G. On Consistency of Nonparametric Normal Mixtures for Bayesian
954 Density Estimation. *Journal of the American Statistical Association* **100**, 1292–1296 (2005).
955 60. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human
956 populations. *Bioinformatics* **32**, 283–285 (2016).
957 61. Robinson, G. K. That BLUP is a Good Thing: The Estimation of Random Effects. *Statistical Science* **6**,
958 15–32 (1991).
959 62. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal*
960 *Statistical Society: Series B (Statistical Methodology)* **67**, 301–320 (2005).
961 63. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large
962 cohorts. *Nat Genet* **47**, 284–290 (2015).
963

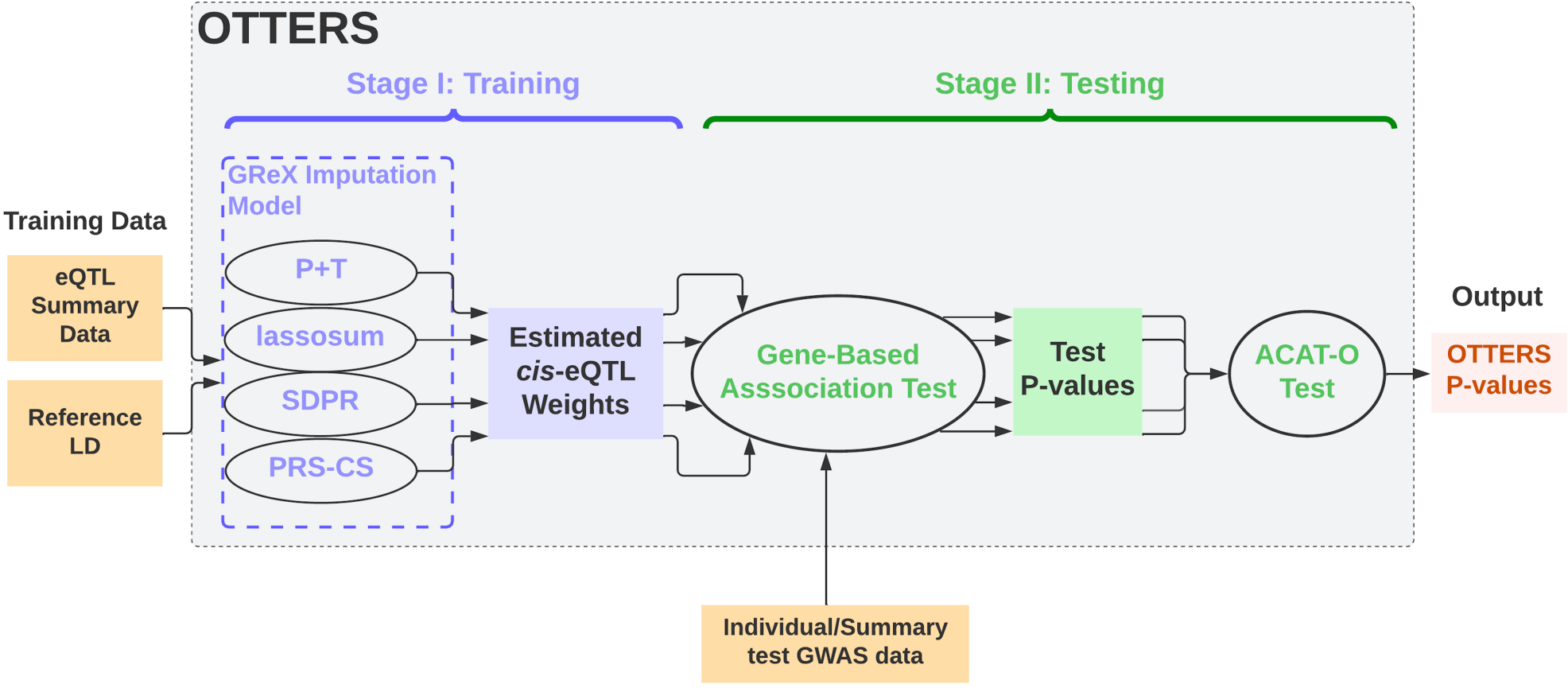

A
Causal Proportion=0.001
Causal Proportion=0.01
Test R^2
0.10
0.05
0.01
0.01
0.05
0.1
h_e^2
B
Causal Proportion=0.001
Expression Heritabillity=0.01
Causal Proportion=0.001
Expression Heritabillity=0.05
Causal Proportion=0.001
Expression Heritabillity=0.1
Causal Proportion=0.01
Expression Heritabillity=0.01
Causal Proportion=0.01
Expression Heritabillity=0.05
Causal Proportion=0.01
Expression Heritabillity=0.1
TWAS Power
1.00
0.75
0.50
0.25
0.00
200
300
400
500
25
50
75
100
10
20
30
40
n_gwas (in K)
Method
P+T(0.001)
P+T(0.05)
lassosum
SDPR
PRS−CS
FUSION
OTTERS

TWAS of Cardiovascular Disease by OTTERS



Test R^2


A
P+T(0.001)
0.50
0.20
0.10
0.01
0.01
0.10 0.20
0.50
PRS−CS
B
P+T(0.05)
0.50
0.20
0.10
0.01
0.01
0.10 0.20
0.50
PRS−CS
C
lassosum
0.50
0.20
0.10
0.01
0.01
0.10 0.20
0.50
PRS−CS
D
SDPR
0.50
0.20
0.10
0.01
0.01
0.10 0.20
0.50
PRS−CS
E
FUSION
0.50
0.20
0.10
0.01
0.01
0.10 0.20
0.50
PRS−CS


OTTERS
Stage I: Training
Stage II: Testing
GReX Imputation Model
Training Data
eQTL Summary Data
Reference LD
P+T
lassosum
SDPR
PRS-CS
Estimated *cis*-eQTL Weights
Gene-Based Asssociation Test
Test P-values
ACAT-O Test
Output
OTTERS P-values
Individual/Summary test GWAS data