

# Multi-trait transcriptome-wide association studies with probabilistic Mendelian randomization

Lu Liu,<sup>1</sup> Ping Zeng,<sup>2</sup> Fuzhong Xue,<sup>1</sup> Zhongshang Yuan,<sup>1,\*</sup> and Xiang Zhou<sup>3,4,\*</sup>

## Summary

A transcriptome-wide association study (TWAS) integrates data from genome-wide association studies and gene expression mapping studies for investigating the gene regulatory mechanisms underlying diseases. Existing TWAS methods are primarily univariate in nature, focusing on analyzing one outcome trait at a time. However, many complex traits are correlated with each other and share a common genetic basis. Consequently, analyzing multiple traits jointly through multivariate analysis can potentially improve the power of TWASs. Here, we develop a method, moPMR-Egger (multiple outcome probabilistic Mendelian randomization with Egger assumption), for analyzing multiple outcome traits in TWAS applications. moPMR-Egger examines one gene at a time, relies on its *cis*-SNPs that are in potential linkage disequilibrium with each other to serve as instrumental variables, and tests its causal effects on multiple traits jointly. A key feature of moPMR-Egger is its ability to test and control for potential horizontal pleiotropic effects from instruments, thus maximizing power while minimizing false associations for TWASs. In simulations, moPMR-Egger provides calibrated type I error control for both causal effects testing and horizontal pleiotropic effects testing and is more powerful than existing univariate TWAS approaches in detecting causal associations. We apply moPMR-Egger to analyze 11 traits from 5 trait categories in the UK Biobank. In the analysis, moPMR-Egger identified 13.15% more gene associations than univariate approaches across trait categories and revealed distinct regulatory mechanisms underlying systolic and diastolic blood pressures.

## Introduction

Transcriptome-wide association studies (TWASs) are widely applied to integrate genome-wide association studies (GWASs) with gene expression mapping studies for investigating the causal molecular mechanisms underlying diseases and disease-related complex traits.<sup>1,2</sup> While TWASs were originally proposed either as a weighted SNP set test<sup>2</sup> or a test for various relationships among SNPs, gene expression, and an outcome trait,<sup>1</sup> TWASs are closely related to Mendelian randomization (MR) analysis per its detailed algorithmic formulation, with one of the outcomes effectively testing the causal effect of a gene on the GWAS trait by treating the *cis*-SNPs of the gene as its instrumental variables.<sup>3,4</sup> Many statistical methods have been recently developed for TWASs and exemplary methods include PrediXcan,<sup>2</sup> TWAS,<sup>1</sup> DPR,<sup>5</sup> TIGAR,<sup>6</sup> SMR,<sup>7</sup> PMR-Egger,<sup>4</sup> FOCUS,<sup>8</sup> and UTMOST,<sup>9</sup> to name a few. Different TWAS methods differ in their ways of using *cis*-SNPs (i.e., some use one *cis*-SNP<sup>7</sup> while others use all *cis*-SNPs), modeling SNP effects on gene expression (i.e., some make a sparse effect assumption<sup>7</sup> while others make different polygenic modeling assumptions), fitting models (i.e., some use a likelihood-based algorithm<sup>4</sup> while others use two-stage regression based algorithms), and accounting for horizontal pleiotropy (i.e., some account for it<sup>4</sup> while others do not)—horizontal pleiotropy occurs when genetic variants affect the GWAS trait through pathways other than or in addition to the gene of focus and is partic-

ularly important to control for as it is widespread in TWAS applications.<sup>4</sup>

Despite the above technical differences, almost all existing TWAS methods are univariate in nature and analyze one GWAS trait at a time. However, GWASs often collect multiple correlated phenotypes<sup>10</sup> that share a common genetic basis.<sup>11</sup> Indeed, many loci have been recently identified to have pleiotropic associations with multiple phenotypes. Exemplary pleiotropic gene associations include *CACNA1C* (MIM: 114205) for both bipolar disorder (MIM: 125480) and schizophrenia (MIM: 181500),<sup>12</sup> *SLC39A8* (MIM: 608732) for schizophrenia and Parkinson disease (MIM: 168600),<sup>13</sup> and *RGS12* (MIM: 602512) for serum lipids and inflammatory bowel disease (MIM: 266600).<sup>14</sup> Consequently, performing multivariate analysis to test gene associations with multiple traits jointly may lead to an appreciable power gain. The benefits of multivariate analysis over univariate analysis have been well documented in other analytic settings such as association tests in GWASs.<sup>15–18</sup> There, by modeling multiple traits together, multivariate analysis can increase power over univariate analysis in identifying pleiotropic loci that affect multiple traits simultaneously. In addition, by explicitly modeling phenotypic correlation, multivariate analysis can also increase power over univariate analysis to identify loci that affect only one trait, because of its ability to control for the other correlated traits. Therefore, it is appealing to develop statistical methods for analyzing multiple traits jointly for TWAS applications.

<sup>1</sup>Department of Biostatistics, School of Public Health, CheeLo College of Medicine, Shandong University, Jinan, Shandong 250012, China; <sup>2</sup>Department of Epidemiology and Biostatistics, Xuzhou Medical University, Xuzhou, Jiangsu 221004, China; <sup>3</sup>Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA; <sup>4</sup>Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA

\*Correspondence: [yuanzhongshang@sdu.edu.cn](mailto:yuanzhongshang@sdu.edu.cn) (Z.Y.), [xzhousph@umich.edu](mailto:xzhousph@umich.edu) (X.Z.)

<https://doi.org/10.1016/j.ajhg.2020.12.006>

© 2020 American Society of Human Genetics.



Here, we develop such a method to identify genes associated with multiple correlated traits in TWASs. Our method builds upon a previously developed likelihood inference framework for TWAS analysis<sup>4</sup> and extends it to analyze multiple outcome traits. Our method explicitly accounts for the correlation structure among multiple traits, accommodates *cis*-SNPs that are in linkage disequilibrium (LD) with each other, relies on the widely used Egger assumption to model horizontal pleiotropic effects, and performs inference based on a maximum likelihood framework. We refer to our method as the multiple outcome probabilistic Mendelian randomization with Egger assumption (moPMR-Egger). With simulations and real data applications, we show that moPMR-Egger provides calibrated type I error control for both causal effects testing and horizontal pleiotropic effects testing, yields substantial power gain over univariate approaches, and is computationally efficient for biobank scale datasets.

## Material and Methods

### moPMR-Egger Overview

We provide an overview of moPMR-Egger here, with its details supplied in the [Supplemental Material and Methods](#). moPMR-Egger is developed for identifying genes causally associated with multiple outcome traits in TWAS applications. moPMR-Egger builds upon the two-sample MR framework, which aims to estimate and test for the causal effect of an exposure on an outcome in the setting where the exposure and outcome are measured in two separate studies with no sample overlap. In the TWAS setting we consider here, the exposure is gene expression level that is measured in a gene expression study, while the outcomes are multiple correlated quantitative traits that are measured in a GWAS. In moPMR-Egger, we examine one gene at a time and treat its *cis*-SNPs as instrumental variables. With the instrumental variables, we estimate and test the causal effects of the gene on multiple outcome traits together. We do so by performing joint analysis of gene expression and multiple outcome traits in a likelihood framework while properly accounting for potential horizontal pleiotropic effects. An illustrative diagram of the moPMR-Egger model is displayed in [Figure 1A](#).

Technically, we follow existing TWAS approaches and use all *cis*-SNPs that are in LD with each other as instruments. We denote  $\mathbf{x}$  as an  $n_1$ -vector of gene expression levels that are measured on  $n_1$  individuals in the gene expression study and denote  $\mathbf{Z}_x$  as an  $n_1$  by  $p$  matrix of genotypes for  $p$  instruments (i.e., *cis*-SNPs) in the same study. We denote  $\mathbf{Y}$  as a  $k$  by  $n_2$  matrix of  $k$  outcome traits measured on  $n_2$  individuals in the GWAS and denote  $\mathbf{Z}_Y$  as an  $n_2$  by  $p$  matrix of genotypes for the same  $p$  instruments there. We assume  $\mathbf{x}$  and each column of  $\mathbf{Y}^T$ ,  $\mathbf{Z}_x$ , and  $\mathbf{Z}_Y$  have all been standardized to have a mean of zero and a standard deviation of one. We consider three linear regressions to model the two studies separately,

$$\mathbf{x} = \mathbf{1}_{n_1}\mu_x + \mathbf{Z}_x\boldsymbol{\beta} + \boldsymbol{\varepsilon}_x, \quad (\text{Equation 1})$$

$$\tilde{\mathbf{x}} = \mathbf{1}_{n_2}\mu_x + \mathbf{Z}_Y\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{\tilde{x}}, \quad (\text{Equation 2})$$

$$\mathbf{Y} = \tilde{\boldsymbol{\mu}}_Y\mathbf{1}_{n_2}^T + \boldsymbol{\alpha}\tilde{\mathbf{x}}^T + \boldsymbol{\Gamma}\mathbf{Z}_Y^T + \tilde{\mathbf{E}}, \quad (\text{Equation 3})$$

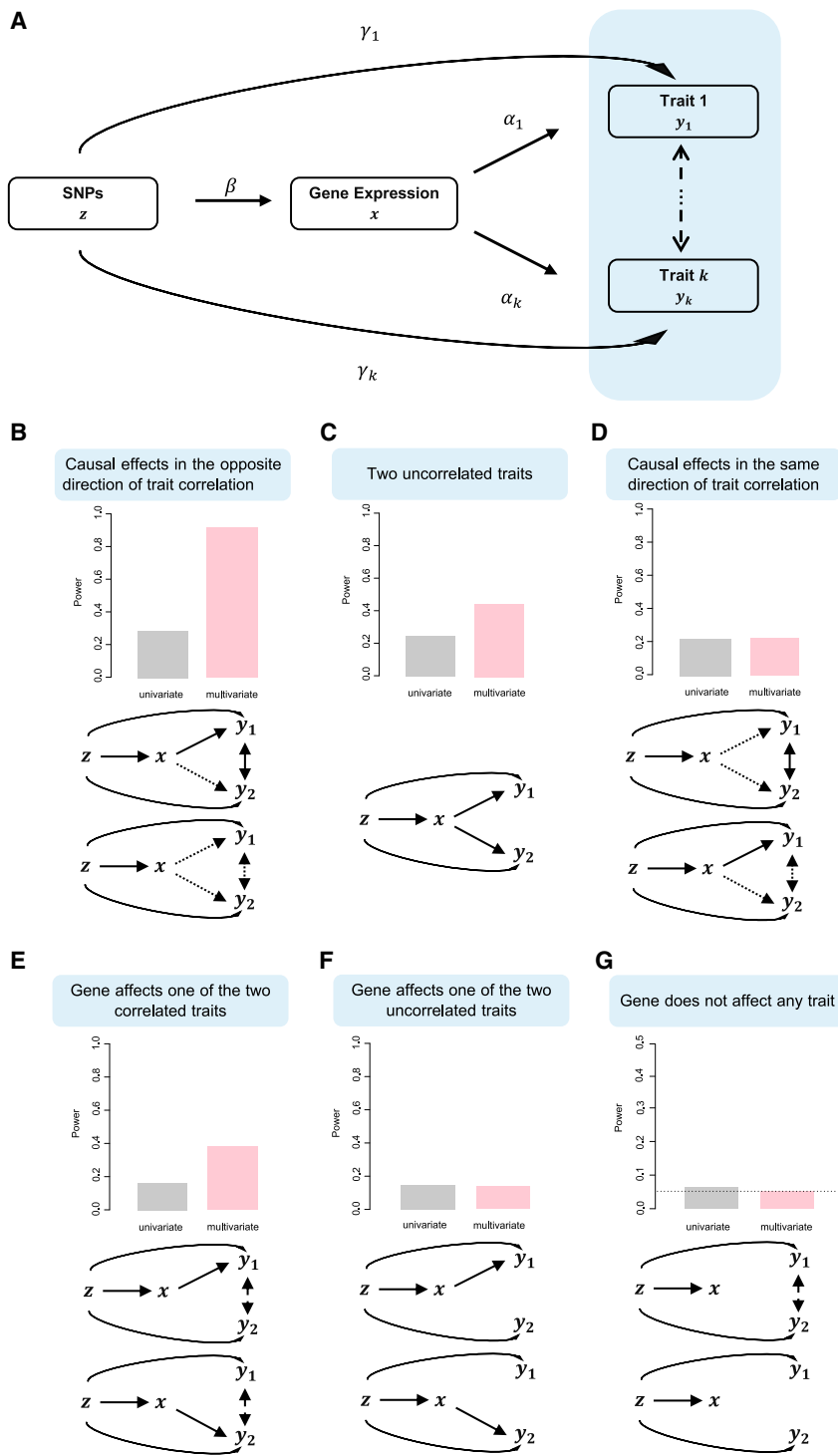
where [Equation 1](#) is for the gene expression study and [Equations 2](#) and [3](#) are for the GWAS. Above, both  $\mu_x$  (a scalar) and  $\tilde{\boldsymbol{\mu}}_Y$  (a  $k$ -vector) are intercepts;  $\tilde{\mathbf{x}}$  is an unobserved  $n_2$ -vector of gene expression for GWAS individuals;  $\boldsymbol{\beta}$  is a  $p$ -vector of SNP effect sizes on gene expression;  $\boldsymbol{\alpha}$  is a  $k$ -vector and represents the  $k$  causal effects of gene expression on the  $k$  outcomes;  $\boldsymbol{\Gamma}$  is a  $k$  by  $p$  matrix of horizontal pleiotropic effects, representing the pleiotropic effects of  $p$  instruments on the  $k$  outcomes;  $\boldsymbol{\varepsilon}_x$  is an  $n_1$ -vector of residual error with each element independently and identically distributed from a normal distribution  $N(0, \sigma_x^2)$ ;  $\boldsymbol{\varepsilon}_{\tilde{x}}$  is an  $n_2$ -vector of residual error with each element independently and identically distributed from the same normal distribution  $N(0, \sigma_x^2)$ ; and  $\tilde{\mathbf{E}}$  is a  $k$  by  $n_2$  matrix of residual error, with each column following a multivariate normal distribution  $MVN_k(0, \Omega)$ , where  $\Omega$  is a  $k$  by  $k$  covariance matrix that accounts for the correlation structure among  $k$  outcomes. Note that, while the above three equations are specified based on two separate studies, they are joined together with the common parameter  $\boldsymbol{\beta}$  and the unobserved gene expression measurements  $\tilde{\mathbf{x}}$ . [Equations 2](#) and [3](#) can also be combined into

$$\mathbf{Y} = \tilde{\boldsymbol{\mu}}_Y\mathbf{1}_{n_2}^T + \boldsymbol{\alpha}(\mathbf{Z}_Y\boldsymbol{\beta})^T + \boldsymbol{\Gamma}\mathbf{Z}_Y^T + \mathbf{E}, \quad (\text{Equation 4})$$

where  $\mathbf{E} = \boldsymbol{\alpha}\boldsymbol{\varepsilon}_{\tilde{x}}^T + \tilde{\mathbf{E}}$  and  $\tilde{\boldsymbol{\mu}}_Y = \boldsymbol{\alpha}\mu_x + \tilde{\boldsymbol{\mu}}_Y$ . Importantly, we emphasize that the above [Equations 1](#), [2](#), [3](#), and [4](#) define a data generative model, which determines how gene expression and outcome traits are generated based on *cis*-SNPs. In addition, the inclusion of the horizontal pleiotropic effects term  $\boldsymbol{\Gamma}\mathbf{Z}_Y^T$  in [Equations 3](#) and [4](#), when further paired with the equal horizontal pleiotropic effects assumption made on  $\boldsymbol{\Gamma}$  detailed below, effectively extends the commonly used MR-Egger modeling framework<sup>19</sup> toward accommodating multiple correlated instruments and multiple outcome traits.

Because  $p$  is often larger than  $n_1$ , we will need to make additional modeling assumptions on  $\boldsymbol{\beta}$  to make the model identifiable. In addition, we will need to make additional modeling assumptions on  $\boldsymbol{\Gamma}$ ; otherwise the two instrumental effect terms defined in [Equation 4](#)—the vertical pleiotropic effects  $\boldsymbol{\alpha}(\mathbf{Z}_Y\boldsymbol{\beta})^T$  and the horizontal pleiotropic effects  $\boldsymbol{\Gamma}\mathbf{Z}_Y^T$ —are also not identifiable from each other. Here, we follow the standard omnigenic modeling assumption and assume that all elements in  $\boldsymbol{\beta}$  are non-zero and each follows a normal distribution  $N(0, \sigma_\beta^2)$ . In addition, we follow the PMR-Egger assumption<sup>4</sup> and assume equal horizontal pleiotropic effects across SNPs for each trait  $i$ :  $\Gamma_{ij} = \gamma_i$  for  $j = 1, \dots, p$ , with  $\gamma_i$  being a scalar of horizontal pleiotropic effect. The assumption on equal horizontal pleiotropic effects across SNPs on a trait is widely applied in other TWAS<sup>8</sup> and robust MR studies<sup>19,20</sup> and is equivalent to assuming  $\boldsymbol{\Gamma} = \boldsymbol{\gamma}\mathbf{1}_p^T$ , with  $\mathbf{1}_p$  being a  $p$ -vector of ones and  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_k)^T$ .

Our key parameters of interest in the above joint model are the causal effects  $\boldsymbol{\alpha}$  of the gene on multiple outcome traits. The causal interpretation and identification of  $\boldsymbol{\alpha}$  can be derived based on the decision-theoretic framework of causal inference<sup>21–24</sup> (details in [Supplemental Material and Methods](#)). Such causal interpretation of  $\boldsymbol{\alpha}$  requires at least two assumptions of MR to hold: (1) instruments are associated with gene expression; (2) instruments are not associated with any other confounders that may be associated with both gene expression and each outcome. moPMR-Egger no longer requires the general exclusion restriction condition of traditional MR (i.e., instruments only influence each outcome through the path of gene expression), as we make an explicit modeling assumption on the horizontal pleiotropy effects  $\boldsymbol{\Gamma}$ . However, our explicit modeling assumption on  $\boldsymbol{\Gamma}$  follows (3) the InSIDE



**Figure 1. A method schematic for moPMR-Egger and illustration of its power as compared to the univariate counterpart in various scenarios**

(A) moPMR-Egger applies to the TWAS setting and attempts to estimate the causal effects ( $\alpha_1, \dots, \alpha_k$ ) of gene expression ( $x$ ) on multiple traits of interest ( $y_1, \dots, y_k$ ) in the presence of confounding factors ( $U$ , not shown) by using *cis*-SNPs ( $z$ ) as instrumental variables. moPMR-Egger relies on a joint likelihood framework and effectively accounts for the horizontal pleiotropy ( $\gamma_1, \dots, \gamma_k$ ) and the correlation among multiple traits.

(B–G) Multi-trait modeling with moPMR-Egger (pink) is beneficial under various scenarios as compared to univariate method PMR-Egger that analyzes one trait at a time (gray). Six scenarios are examined using simple illustrative simulations on two traits ( $y_1, y_2$ ): gene affects two correlated traits, with its causal effects on the two traits in the opposite direction of trait correlation (B); gene affects two uncorrelated traits (C); gene affects two correlated traits, with its causal effects on the two traits in the same direction of trait correlation (D); gene affects one of the two correlated traits (E); gene affects one of the two uncorrelated traits (F); or gene does not affect any trait (G). In these scenarios, the gene has either non-zero causal effects on two traits (two arrows from gene to traits) or one trait (one arrow) or no trait (no arrow), with causal effects in the same direction (two dotted arrows) or in the opposite direction (one solid arrow and one dotted arrow) or in either directions (two solid arrows). The traits are positively correlated (solid double-head arrows) or negative correlated (dotted double-head arrows) or either (dashed double-head arrows) or uncorrelated (no double-head arrows).  $y$  axis shows power in (B)–(F) and type I error in (G). Type I error is calculated as the percentage of discoveries among 1 million null simulation replicates. Power is calculated as the percentage of discoveries among the 1,000 alternative simulation replicates. For power calculation, discoveries are declared based on a Bonferroni corrected p value threshold: 0.05/1,000 for the multivariate approach and 0.05/2,000 for the univariate approach that tested each of the two traits separately.

assumption that the instrument-gene expression effects and instrument-outcome effects are independent of each other, which is sometimes referred to as the weak exclusion restriction condition.<sup>19</sup> Consequently, the causal effect interpretation of  $\alpha$  depends on MR assumptions as well as the other explicit modeling assumptions. Many of these assumptions are not testable in practice as an exhaust list of confounding factors is often unknown. Therefore, while we follow standard MR analysis and use the term “causal effect” throughout the text, we only intend to use this term to

emphasize that the  $\alpha$  estimates here are more likely to be closer to the causal estimates than the effect size estimates from a standard multivariate regression of  $\mathbf{Y}$  on  $\bar{\mathbf{x}}$ .

In the above model, we are interested in estimating the causal effects  $\alpha$  and testing the null hypothesis  $H_0 : \alpha = \mathbf{0}$  in the presence of the horizontal pleiotropic effects  $\gamma$ . Rejecting the null hypothesis of  $\alpha = \mathbf{0}$  would suggest that the gene of focus has non-zero causal effects on at least one trait. In addition, we are interested in estimating the horizontal pleiotropic effect size  $\gamma$  and testing

the null hypothesis  $H_0 : \gamma = \mathbf{0}$ . Rejecting the null hypothesis of  $\gamma = \mathbf{0}$  would suggest that the *cis*-SNPs have non-zero horizontal pleiotropic effects on at least one trait. We accomplish both tasks under the maximum likelihood inference framework, in direct contrast to the standard MR-Egger inference framework<sup>19</sup> and most previous TWAS approaches<sup>1,2,5</sup> that use two-stage regression-based algorithms. Compared to the two-stage regression-based algorithms, maximum likelihood-based inference can properly account for the uncertainty in parameter estimates in the first regression stage, thus potentially improving statistical power.<sup>4</sup> To enable maximum likelihood-based inference, we develop a parameter expansion version of the expectation maximization (EM) algorithm by maximizing the joint likelihood defined based on Equations 1 and 4 (details in the Supplemental Material and Methods). The EM algorithm allows us to obtain the maximum likelihood of the joint model, together with maximum likelihood estimates for both  $\alpha$  and  $\gamma$ . In addition to the joint model, we apply the EM algorithm to two reduced models, one without  $\alpha$  and the other without  $\gamma$ , to obtain the corresponding maximum likelihoods there. Afterward, we perform likelihood ratio tests for either  $H_0 : \alpha = \mathbf{0}$  or  $H_0 : \gamma = \mathbf{0}$ , by contrasting the maximum likelihood obtained from the joint model to that obtained from each of the two reduced models, respectively.

We refer to our model and algorithm together as the two-sample probabilistic Mendelian randomization with Egger regression for multiple outcomes (moPMR-Egger). The term “mo” is referred to the modeling of multiple outcomes of interest. The term “probabilistic” is referred to the data generative nature of our model and the maximum likelihood inference procedure as explained above. The term “Egger” is referred to the horizontal pleiotropic assumption made on  $\Gamma$  that effectively generalizes the MR-Egger regression assumption to both correlated instruments and multiple outcomes.

### Extensions to summary statistics

While we have presented moPMR-Egger based on individual-level data, moPMR-Egger can be easily extended to perform inference using summary statistics only. To do so, we denote  $\Sigma_1$  as the SNP-SNP correlation matrix (i.e., LDmatrix) among *cis*-SNPs of the gene of focus in the gene expression study. We denote  $\Sigma_2$  as the corresponding SNP correlation matrix in the GWAS data. Both matrices are of dimensionality  $p$  by  $p$  and are positive semi-definite. Note that both  $\Sigma_1$  and  $\Sigma_2$  can be estimated using individuals of corresponding ethnicity from an LD reference panel (e.g., the 1000 Genomes project). With the definition of  $\Sigma_1$  and  $\Sigma_2$ , we can re-express the moPMR-Egger model in terms of summary statistics as

$$\hat{\beta}_x = \Sigma_1 \beta + \mathbf{e}_x, \quad (\text{Equation 5})$$

$$\hat{\mathbf{B}}_y = \alpha(\Sigma_2 \beta)^T + \Gamma \Sigma_2 + \mathbf{E}_y, \quad (\text{Equation 6})$$

where  $\hat{\beta}_x$  is a  $p$ -vector of estimates for the standardized marginal SNP effect sizes on the gene expression;  $\hat{\mathbf{B}}_y$  is a  $k$  by  $p$  matrix of estimates for the standardized marginal SNP effect sizes on the  $k$  outcomes;  $\mathbf{e}_x$  is a  $p$ -vector of estimation errors that follow a multivariate normal distribution  $N(0, \Sigma_1 \sigma_x^2 / (n_1 - 1))$ ; and  $\mathbf{E}_y$  is a  $k$  by  $p$  matrix of estimation errors that follow a matrix normal distribution  $MN(0, \Omega, \Sigma_2 / (n_2 - 1))$  with  $\Omega$  being a  $k$  by  $k$  row covariance matrix and  $\Sigma_2 / (n_2 - 1)$  being a  $p$  by  $p$  column covariance matrix. Again, the matrix  $\Omega$  is used to model the covariance structure among multiple correlated outcome traits. We adapt our EM algo-

rithm for individual-level data to perform estimation and inference using summary statistics (details in the Supplemental Material and Methods). The estimation and inference procedures are all based on the maximum likelihood framework and are largely similar to what has been described in the previous section.

### Simulations

We performed cross-gene based simulations to assess the performance of moPMR-Egger and compare it with existing approaches. To do so, we randomly selected 10,000 genes from GEUVADIS.<sup>25</sup> We extracted *cis*-SNPs for these 10,000 genes, obtaining a median of 576 *cis*-SNPs per gene (min = 11; max = 7,409). For each gene in turn, we simulated the gene expression level and the outcome traits using the genotype data obtained from the gene expression study and the GWAS, respectively. Specifically, we first obtained genotypes for  $p$  *cis*-SNPs of the gene of focus from the GEUVADIS data. We standardized the genotype vector of each SNP to have a zero mean and a unit standard deviation. With the standardized genotype matrix  $\mathbf{Z}_x$ , we simulated the SNP effect sizes  $\beta$  from a normal distribution  $N(0, PVE_{zx} / p)$ , where  $PVE_{zx}$  represents the proportion of gene expression variance explained by genetic effects. We summed the genetic effects across all *cis*-SNPs as  $\mathbf{Z}_x \beta$ . In addition, we simulated the residual errors  $\mathbf{e}_x$  from a normal distribution  $N(0, 1 - PVE_{zx})$ . We finally summed the genetic effects and residual errors to yield the simulated gene expression level.

Next, we obtained genotypes  $\mathbf{Z}_y$  for the same  $p$  SNPs from 2,000 randomly selected control individuals in the Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging (GERA).<sup>26,27</sup> We standardized the genotype vector of each SNP to have a zero mean and a unit standard deviation. With the standardized genotype matrix  $\mathbf{Z}_y$ , we simulated the causal effects of the gene of focus on four outcome traits. Specifically, we used the same SNP effect sizes  $\beta$  in the gene expression data and set the four causal effects  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)^T = \sqrt{PVE_{zy} / PVE_{zx}}$ . Here,  $PVE_{zy}$  is a vector of size four and each of its elements represents the proportion of phenotypic variance explained by the causal effect of SNPs for the corresponding phenotype. Afterward, we simulated the residual errors for the four phenotypes for each individual from a multivariate normal distribution  $MVN(0, \Omega)$ , where we used the correlation matrix of four lipid traits (total cholesterol [TC], low-density lipoprotein [LDL], high-density lipoprotein [HDL], and triglyceride [TG]) calculated in the NFB1966 database<sup>28</sup> to serve as  $\Omega$ . Specifically, the correlations are 0.13 between TC and HDL, 0.88 between TC and LDL, 0.41 between TC and TG, 0.09 between HDL and LDL, -0.44 between HDL and TG, and 0.29 between LDL and TG (Table S1). We also simulated the horizontal pleiotropic effects  $\gamma$  of these SNPs on the four traits, with details described in the next paragraph. We finally summed the horizontal pleiotropic effects, vertical pleiotropic effects, and residual errors to yield the four simulated traits.

In the simulations, we first examined a baseline simulation setting where we set  $n_1 = 465$ ,  $n_2 = 2,000$ ,  $PVE_{zx} = 10\%$ , and  $PVE_{zy} = (0,0,0,0)^T$ , with all  $\Gamma_{kj} = 0$  ( $k = 1, \dots, 4; j = 1, \dots, p$ ). On top of the baseline setting, we varied one parameter at a time to examine the influence of various parameters. For  $PVE_{zx}$ , we set it to be either 1%, 5%, or 10%; the last value is close to the median gene expression heritability estimates across genes.<sup>29,30</sup> For  $\beta$ , we examined alternative SNP effect size distributions that deviate from the omnigenic assumption. Specifically, we considered sparse settings where we randomly selected either one SNP or three SNPs to have non-zero effects on gene expression, as well

as polygenic settings where 1% or 10% of the SNPs were randomly selected to have non-zero effects. Again, these non-zero effects were simulated from a normal distribution to explain a fixed  $PVE_{zx}$ . For the horizontal pleiotropic effects  $\Gamma$ , for each trait in turn, the trait-specific horizontal pleiotropic effect size is randomly selected for each gene among five choices: 0, small ( $1 \times 10^{-4}$ ), moderate ( $5 \times 10^{-4}$ ,  $1 \times 10^{-3}$ ), or large ( $2 \times 10^{-3}$ ), which corresponds approximately to the 0%, 50%, 70%, 90%, or 95% quantiles of horizontal pleiotropic effect estimates from real data,<sup>4</sup> respectively. Note that the large horizontal pleiotropic effect of  $2 \times 10^{-3}$  is also large in our real data applications: across 11 analyzed traits, on average, only 5.27 genes among 13,513 genes have an absolute horizontal pleiotropic effect size greater than  $2 \times 10^{-3}$  (Figure S1). The same trait-specific horizontal pleiotropic effect is assigned to all SNPs with non-zero horizontal pleiotropic effects. The proportion of non-zero effect SNPs for each trait is further determined randomly with replacement from five choices (0%, 10%, 30%, 50%, or 100%) and thus may vary across genes and traits. In addition, we examined directional pleiotropy setting (the ratio of SNPs with negative versus positive horizontal pleiotropic effects is 0:10), approximately directional pleiotropy settings (1:9 or 3:7), and balanced pleiotropy setting (5:5). For  $\Omega$ , in addition to using the estimated  $\Omega$  from the real data, we also examined settings where  $\Omega$  is an identity matrix and where  $\Omega$  has an exponential covariance structure with each element  $\Omega_{ij} = \rho^{|i-j|}$  with  $\rho$  being either 0.5, 0.7, or 0.9. The trait correlation in the exponential covariance structure ranges from 0.125 to 0.9, closely resembling the estimates obtained from real data (Table S1). We also examined the setting where  $\beta$  are correlated. In this setting, we simulated the SNP effects on gene expression from a multivariate normal distribution with covariance matrix  $w\Sigma$ . Here,  $\Sigma$  is the LD matrix among SNPs and  $w$  is a scalar that is chosen to ensure that  $PVE_{zx}$  equals to 10%. For  $\mathbf{PVE}_{zy}$ , in addition to the baseline setting with zero causal effects, we also examined homogeneous causal effect settings and heterogeneous causal effect settings. In the homogeneous causal effect settings, we set all elements of  $\mathbf{PVE}_{zy}$  to the same value  $v$ , with  $v$  being either 0.5%/4, 1%/4, 1.5%/4, or 2%/4 in different settings. In the heterogeneous causal effect settings, we set the first element of  $\mathbf{PVE}_{zy}$  as  $v$  and the other elements of  $\mathbf{PVE}_{zy}$  as  $0.15v$ ,  $0.85v$ , and  $0.50v$ , respectively, with  $v$  being either 0.5%, 1%, 1.5%, or 2% in different settings. In terms of causal effect size direction, we randomly selected one, two, three, or four traits to be affected by gene expression both in the homogeneous and heterogeneous causal effect settings. When the gene expression affected two phenotypes, its effect on one of the two traits was in the opposite direction of its effect on the other. When the gene expression affected three phenotypes, its effects on two randomly selected traits were in the opposite direction of its effect on the third trait. When the gene expression affected four phenotypes, its effects on two or three randomly selected traits were in the opposite direction of its effects on the other two or one trait.

For type I error control examination, we performed 10,000 simulation replicates for each simulation scenario described above. For power calculation, we examined two different approaches depending on the methods we compare to (more details in the next section). In the first approach, we performed 1,000 alternative simulations and compared power based on a Bonferroni corrected p value threshold (0.05/1,000 for the multivariate approach and 0.05/4,000 for the univariate approaches that tested each of the four traits separately). In the second approach, we performed 100 alternative simulations together with 900 null simula-

tions for each simulation scenario and calculated power based on a false discovery rate (FDR) of 0.05. While we mainly focus on using individual-level data for simulations, we also validate the implementation of the summary statistics based moPMR-Egger in a subset of simulations (details in Discussion).

Besides the above comprehensive simulations, we also conducted a set of simple simulations on two traits to illustrate the benefits of multivariate modeling over univariate modeling. Here, we set  $n_1 = 465$ ,  $n_2 = 2,000$ ,  $PVE_{zx} = 10\%$ , and  $\gamma = (0.0001, 0.0005)^T$ , and set the correlation between the two traits to be either 0.66 (correlated) or 0 (independent). The value 0.66 corresponds to the correlation estimate between systolic blood pressure (SBP) and diastolic blood pressure (DBP) in UK Biobank (Table S1). In the simulations, we examined three causal effect settings: a setting where the gene causally affects both traits with  $\mathbf{PVE}_{zy} = (0.01, 0.01)^T$ , with effect sizes either in the same direction or in the opposite direction; a setting where the gene causally affects one trait with  $\mathbf{PVE}_{zy} = (0.01, 0)^T$ ; and a null setting where the gene affects neither traits causally with  $\mathbf{PVE}_{zy} = (0, 0)^T$ . In total, we examined six simulation scenarios that combine three causal effects settings and two trait correlation settings. We performed 1,000 simulation replicates for each alternative scenario for comparing power at the Bonferroni thresholds and performed 1 million simulation replicates for each null scenario for examining type I error control. We only compared moPMR-Egger with the univariate method PMR-Egger in these simple simulations for illustrative purposes.

## Compared Methods

No other multivariate methods have been developed so far for analyzing multiple traits in TWAS applications. Therefore, we examined three univariate TWAS methods that include the following. (1) PMR-Egger, which tests and controls for horizontal pleiotropy using multiple correlated instruments, for which we used all *cis*-SNPs for the model and used PMR\_individual function implemented in the R package PMR to test the causal effect and pleiotropic effect. (2) PrediXcan, which uses multiple correlated instruments but does not control for horizontal pleiotropy. For PrediXcan, we used all *cis*-SNPs for the model and used ElasticNet implemented in the R package glmnet to obtain the coefficient estimates for the *cis*-SNP effects on gene expression. (3) TWAS, which uses multiple correlated instruments but does not control for horizontal pleiotropy. For TWAS, we used all *cis*-SNPs for the model and used BSLMM<sup>31</sup> implemented in the GEMMA<sup>32</sup> software to obtain coefficient estimates for the *cis*-SNP effects on gene expression. All these methods are suitable for two-sample design and yield p values for testing the causal effects. These three methods differ in their prior assumptions on  $\beta$ : PrediXcan relies on the ElasticNet assumption; TWAS relies on the BSLMM<sup>31</sup> assumption; and PMR-Egger relies on the normal assumption. In addition, PrediXcan and TWAS rely on a two-stage regression procedure while PMR-Egger is based on maximum likelihood. In the analysis, we applied each univariate TWAS method to analyze one trait at a time. We then declared gene association significance based on a Bonferroni corrected p value threshold that adjusted for both the number of genes tested and the number of traits tested.

We also modified the above three univariate TWAS methods into *ad hoc* multivariate TWAS procedures for analyzing multiple outcome traits. Specifically, for each univariate method, we examined one gene at a time and obtained the minimal p value across multiple traits to serve as the association evidence for the gene of focus. Because the null distribution of minimum p values is

not trivial to obtain, we compared the power of these minimal p value approaches based on FDR. Specifically, for simulations, we computed the power of different methods based on an FDR cutoff of 0.05. For real data applications, we first performed 10 random permutations on the individuals while maintaining the correlation among traits. We then applied each method on the permuted data to obtain the null distribution of minimum p values, with which we calculated empirical FDR and declared gene association significance based on an FDR cutoff of 0.05.

Besides the above methods, we also compared moPMR-Egger with TWMR<sup>33</sup> in simulations. TWMR is a two-stage based MR method that incorporates multiple genes as exposure variables. TWMR analyzes one outcome trait at a time and uses only independent SNPs to serve as instruments. Here, we considered both the standard univariate procedure of TWMR for analyzing one trait a time and the minimum p value procedure of TWMR similar to what is described above for multiple trait analysis. We examined the type I error control of TWMR in the baseline simulation settings, with parameters  $n_1 = 465$ ,  $n_2 = 2,000$ ,  $PVE_{zx} = 10\%$ ,  $\mathbf{PVE}_{zy} = (0,0,0,0)^T$ , and  $\gamma = (0, 0, 0, 0)^T$  in the absence of horizontal pleiotropy or  $\gamma = (1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 2 \times 10^{-3})^T$  in the presence of horizontal pleiotropy. We also assessed power of TWMR for testing causal effects in alternative simulation settings using the minimum p value approach of TWMR in the homogeneous causal effects settings where each element of  $\mathbf{PVE}_{zy}$  is set to be 2%/4. In the TWMR analysis, we examined one gene at a time while controlling for its neighboring genes as covariates. To do so, following Porcu et al.,<sup>33</sup> for each gene in turn, we used PLINK software<sup>34</sup> (v.1.90b6.13) to perform linear regression analysis on its *cis*-SNPs and obtain *cis*-SNPs that are significantly associated with gene expression level at an FDR threshold of 0.05. We termed these significant *cis*-SNPs as *cis*-eQTLs. Next, we retained genes that have at least one *cis*-eQTL and retained SNPs that are *cis*-eQTLs for at least one of these genes for analysis. We further pruned these SNPs using PLINK (with  $r^2 < 0.1$ , which is the default setting recommended in TWMR) to retain an independent set of SNPs to serve as instruments. Due to these stringent filtering steps required by TWMR,<sup>33</sup> we were only able to analyze an average of 81.61% of genes in the simulations. TWMR results are described in the [Discussion](#) section.

## Real data applications

We applied moPMR-Egger to perform multi-trait TWAS by integrating gene expression data from GEUVADIS<sup>25</sup> with GWASs from UK Biobank.<sup>35</sup> Specifically, we obtained the GEUVADIS data as the gene expression data and examined 11 traits from the UK Biobank. The traits were selected based on previous studies<sup>36,37</sup> and all traits have a SNP heritability greater than 0.2. These outcome traits can be roughly divided into the following five trait categories: (1) blood pressure (SBP and DBP), (2) physical measures including height (MIM: 606255), body mass index (BMI [MIM: 606641]), forced vital capacity (FVC), and FEV1-FVC ratio, (3) blood count (platelet count, red blood cell count, eosinophils count, and white blood cell count), (4) white blood cell indices (eosinophils count and white blood cell count), and (5) red blood cell indices (red blood cell count and red blood cell distribution width [RDW]). Note that the white blood cell indices category represents a subset of the blood count category. The correlations among all analyzed trait pairs within each category are listed in [Table S1](#). For each of these five trait categories in turn, we applied moPMR-Egger and the univariate approach PMR-Egger to analyze

the traits in the trait category. The detailed data processing steps for the GEUVADIS data and UK Biobank data are described below.

The GEUVADIS data<sup>25</sup> contains gene expression measurements for 465 individuals collected from five different populations that include CEPH (CEU), Finns (FIN), British (GBR), Toscani (TSI), and Yoruba (YRI). In the expression data, we only focused on protein coding genes and lincRNAs that are annotated in GENCODE (release 12).<sup>38,39</sup> Among these genes, we removed lowly expressed genes that have zero counts in at least half of the individuals to obtain a final set of 15,810 genes. We performed PEER normalization to remove confounding effects and unwanted variations following previous studies.<sup>5,40</sup> Afterward, following Zeng and Zhou,<sup>5</sup> to remove the remaining population stratification, we quantile normalized the gene expression measurements across individuals in each population to a standard normal distribution, and then further quantile normalized the gene expression measurements to a standard normal distribution across individuals from all five populations. Besides the expression data, all individuals in GEUVADIS also have their genotypes sequenced in the 1000 Genomes Project. We obtained genotype data from the 1000 Genomes Project phase 3. We filtered out SNPs that have a Hardy-Weinberg equilibrium (HWE) p value  $< 10^{-4}$ , a genotype call rate  $< 95\%$ , or a minor allele frequency (MAF)  $< 0.01$ . We retained a total of 7,072,917 SNPs for analysis.

The UK Biobank data consists of 487,298 individuals and 92,693,895 imputed SNPs.<sup>35</sup> We followed the same sample QC procedure in Neale lab ([Web Resources](#)) to retain a total of 337,129 individuals of European ancestry. We filtered out SNPs with an HWE p value  $< 10^{-7}$ , a genotype call rate  $< 95\%$ , or an MAF  $< 0.001$  to obtain a total of 13,876,958 SNPs. For each trait in turn, we regressed the resulting standardized phenotypes on sex and top ten genotype principal components (PCs) to obtain the residuals, standardized the residuals to have a mean of zero and a standard deviation of one, and finally used these scaled residuals to conduct TWAS analysis.

We integrated the GEUVADIS data with GWASs from UK Biobank for TWAS analysis. For each gene in turn in the GEUVADIS data, we extracted *cis*-SNPs that are within either 100 kb upstream of the transcription start site (TSS) or 100 kb downstream of the transcription end site (TES). We overlapped these SNPs in GEUVADIS with the SNPs obtained from UK Biobank to obtain common sets of SNPs. The mean number of the overlapped *cis*-SNPs between GEUVADIS and UK Biobank is 497 (median = 443, min = 1, max = 8,085). Afterward, for each pair of gene (from GEUVADIS) and trait category (from GWAS) in turn, we run the multivariate approach moPMR-Egger and the univariate approach PMR-Egger to examine the causal relationship between gene expression and multiple traits in the category. For comparison between moPMR-Egger and PMR-Egger, we declared significance based on the corresponding Bonferroni corrected thresholds: 0.05/15,810 for moPMR-Egger and 0.05/(15,810\*k) for PMR-Egger, where  $k$  is the number of the outcome traits in the specific category.

Among the five trait categories, we focused on the blood pressure category further for an in-depth analysis. The blood pressure category contains two traits, SBP and DBP, with an estimated correlation of 0.66 between them. Blood pressure is a complex trait with heritability estimated in the range of 0.3–0.5.<sup>36</sup> Many large-scale GWASs have been conducted to investigate the genetic architecture underlying blood pressure.<sup>41–45</sup> Elevated blood pressure is a strong and modifiable<sup>46–49</sup> driver for risk of stroke (MIM: 601367) and coronary artery disease (MIM: 608320), which are leading causes of mortality and morbidity globally.<sup>50,51</sup> For the blood

pressure category, in addition to moPMR-Egger and PMR-Egger, we also applied the minimum p value approaches of PMR-Egger, TWAS and PrediXcan. For these minimum p value approaches, we applied each method to examine one gene at a time. For each gene in turn, we analyzed each trait in the trait category separately and obtained the minimum p value across these traits as the association evidence for the gene of focus. We performed 10 random permutations of individuals to obtain a null distribution of minimum p values, with which we calculated FDR. Note that the correlation between the two phenotypes remains after such permutation. In the moPMR-Egger analysis, we also divided the identified genes based on their effect sizes on the two traits into two gene groups: genes that have the same effect signs on both traits and genes that have opposite effect signs on both traits. We examined these two gene sets carefully by performing gene set enrichment analysis using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways with the clusterProfiler R package.<sup>52</sup> We declared gene set significance based on a q-value threshold of 0.05.

For the in-depth analysis of the blood pressure category traits, we also considered two other TWAS methods: FOCUS<sup>8</sup> and TWMR.<sup>33</sup> FOCUS is a TWAS fine mapping method and relies on a Bayesian framework to analyze genomic regions that contain at least one significant gene detected by any standard TWAS method. Therefore, we paired FOCUS with moPMR-Egger or each of the three univariate TWAS methods to analyze regions that contain at least one significant gene detected by the corresponding TWAS method. Following Mancuso et al.,<sup>8</sup> we obtained a set of independent non-overlapping genomic regions termed as LD blocks from LDetect.<sup>53</sup> We removed genomic regions that overlap with the MHC region due to the extensive LD structure there. Also following Mancuso et al.,<sup>8</sup> we conducted our analysis on a subset of regions that harbor at least one genome-wide-significant SNP ( $p < 5 \times 10^{-8}$ ) and at least one significant TWAS gene identified by the corresponding TWAS method based on an FDR threshold of 0.05. Due to these stringent filtering steps required by FOCUS, we were able to analyze only 4,692 genes in 246 regions. We declared genes as significant if they are in the 90% credible set output from FOCUS. We then compared the performance of moPMR-Egger and each of the three univariate TWAS methods by examining how consistent the significant genes are before and after FOCUS analysis. For TWMR, we followed the same procedure described in the compared methods subsection to extract multiple independent *cis*-eQTLs to serve as instrument variables. We retained 398,996 *cis*-eQTLs and 5,573 genes for TWMR analysis. We analyzed the two blood pressure traits with the minimum p value approach of TWMR and declared gene significance based on a Bonferroni corrected threshold.

## Results

### Method overview

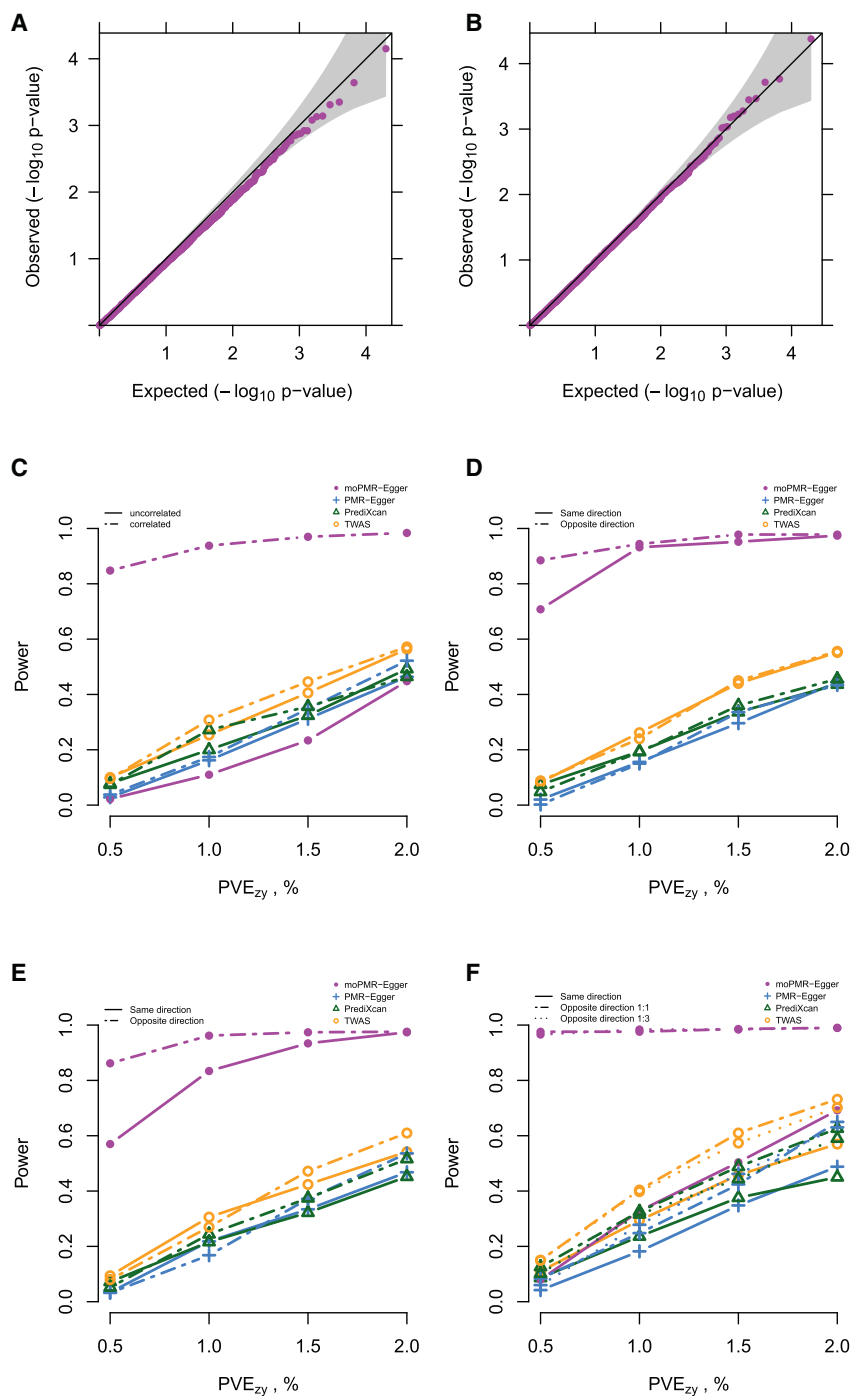
moPMR-Egger is described in the [Material and Methods](#), with technical details provided in the [Supplemental Material and Methods](#). For TWAS applications, moPMR-Egger examines one gene at a time and estimates and tests its causal effects on multiple outcome traits together. Different from many existing TWAS approaches, moPMR-Egger models multiple SNP instruments that are in LD with each other, performs TWAS in a maximum like-

lihood inference framework, is capable of testing and controlling for horizontal pleiotropic effects commonly encountered in TWAS, while jointly analyzing multiple outcome traits ([Figure 1A](#)). By modeling multiple correlated traits together, moPMR-Egger can improve the power of TWAS.

We first performed simple simulations to develop intuition and illustrate the benefits of modeling multiple traits together. Briefly, we relied on real genotype data and simulated gene expression along with two outcome traits ([Material and Methods](#)). We compared moPMR-Egger with its univariate counterpart PMR-Egger. In the simulations, we found that multi-trait modeling is particularly beneficial in the pleiotropic causal effects setting where the gene causally affects both traits, regardless whether the two traits are correlated ([Figure 1B](#)) or not ([Figure 1C](#)). The power gain brought by multi-trait modeling in the absence of trait correlation is presumably due to its ability to properly account for the correlation between test statistics on the two traits there—the univariate TWAS test statistics on the two traits remain correlated (correlation coefficient = 0.58) due to the shared underlying gene expression predictor, even though the two traits are not correlated with each other. The power gain brought by multi-trait modeling is especially apparent when the gene effects on the two traits are in the opposite direction as the trait correlation ([Figure 1B](#)), but becomes much less obvious when the gene effects on the two traits are in the same direction as the trait correlation ([Figure 1D](#)) where it becomes harder to disentangle the causal effects from the trait correlations.<sup>16</sup> Multi-trait modeling is also beneficial in the case where the gene only affects one trait, as long as the two traits are correlated with each other ([Figure 1E](#)). After all, multivariate analysis can also improve power to detect genes associated with only one trait by effectively controlling for the other correlated traits and thus reducing error variance as previously shown in other settings.<sup>15,16</sup> Certainly, the power of multi-trait modeling is similar to or slightly lower than univariate trait modeling when the gene only affects one trait and when the two traits are also independent of each other ([Figure 1F](#)). The power gain brought by multivariate modeling across majority of simulation scenarios is accompanied with effective type I error control ([Figure 1G](#)).

### Simulations: testing and estimating the causal effects

We performed comprehensive simulations to carefully examine the effectiveness of moPMR-Egger and compared it with existing TWAS approaches in realistic scenarios. Briefly, we relied on real genotype data and simulated gene expression along with four outcome traits ([Material and Methods](#)). We compared moPMR-Egger with three existing univariate TWAS methods that include PMR-Egger, PrediXcan, and TWAS. We examined type I error control and power of moPMR-Egger for both causal effects testing and horizontal pleiotropic effects testing across a total of 287 simulation scenarios (25 null and 152 alternative



**Figure 2. Type I error control and power for testing the causal effects under various simulation scenarios**

(A and B) Quantile-quantile plot of  $-\log_{10}$  p values for testing the causal effects either in the absence or in the presence of horizontal pleiotropic effects under null simulations. Null simulations are performed under different horizontal pleiotropic effect sizes: (A)  $\gamma = (0, 0, 0, 0)^T$ ; (B)  $\gamma_j$  randomly selected from  $(0, 1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 2 \times 10^{-3})$ ,  $j = 1, 2, 3, 4$ . p values from moPMR-Egger are on the expected diagonal line across a range of horizontal pleiotropic effect sizes.

(C–F) Power (y axis) at a Bonferroni adjusted threshold to detect the causal effects is plotted against different causal effect sizes characterized by  $\mathbf{PVE}_{zy}$  for the first trait (x axis) in the heterogeneous causal effect settings, where the  $\mathbf{PVE}_{zy}$  for the remaining traits are 15%, 85%, 50% of the  $\mathbf{PVE}_{zy}$  for the first trait. Compared methods include moPMR-Egger (magenta), PMR-Egger (blue), PrediXcan (green), and TWAS (orange). Different line symbols represent whether the four traits are correlated or not in (C) and the direction of causal effects in (D)–(F). Simulations are performed under different number of affected traits being from one to four (C–F) in the absence of horizontal pleiotropic effect.

horizontal pleiotropic effects, regardless of the horizontal pleiotropic effect sizes. The null p value distribution from moPMR-Egger remains largely similar regardless whether the genetic architecture underlying gene expression is sparse or polygenic (Figure S2), regardless of the gene expression heritability (Figure S3), regardless whether the SNP effects on gene expression are simulated to be correlated with respect to LD or not (Figure S4), and regardless whether the multiple traits are correlated or not (Figure S5). Note that the p values of moPMR-Egger are slightly conservative when the gene expression heritability is low, likely

scenarios for causal effects testing; 22 null and 88 alternative scenarios for horizontal pleiotropic effects testing). These simulation scenarios are summarized in Table S2 and include both those simulated under the model and those simulated with various model misspecifications.

Our first set of simulations is focused on the causal effects test. We first examined type I error control of moPMR-Egger under the null, where the gene has no causal effects on any of the four traits. We found that moPMR-Egger provides well-calibrated type I error control both in the absence (Figure 2A) and presence (Figure 2B) of

horizontal pleiotropic effects, regardless of the horizontal pleiotropic effect sizes. The null p value distribution from moPMR-Egger remains largely similar regardless whether the genetic architecture underlying gene expression is sparse or polygenic (Figure S2), regardless of the gene expression heritability (Figure S3), regardless whether the SNP effects on gene expression are simulated to be correlated with respect to LD or not (Figure S4), and regardless whether the multiple traits are correlated or not (Figure S5). Note that the p values of moPMR-Egger are slightly conservative when the gene expression heritability is low, likely

due to the fact that the joint likelihood is no longer informative on the causal effects and thus cannot be approximated well by an asymptotic normal distribution (Figure S3). moPMR-Egger makes a relatively strong modeling assumption on the horizontal pleiotropy and assumes that for a given trait all SNPs have the same horizontal pleiotropic effects. Such horizontal pleiotropic modeling assumption follows that of the Egger regression. To examine the robustness of such assumption, we randomly selected a fixed proportion of SNPs, instead of all of them,



to exhibit horizontal pleiotropic effects. We found that *p* values of moPMR-Egger remain calibrated regardless of the sparsity of the horizontal pleiotropic SNPs (Figure S6). Besides the directional pleiotropy settings where the ratio of SNPs with negative versus positive effects is set to be 0:10, we also examined two approximately directional pleiotropy settings (1:9 or 3:7) and one balanced setting (5:5) by randomly assigning the corresponding proportion of SNPs to have negative versus positive effects. We found that *p* values of moPMR-Egger remain calibrated in either the approximately directional pleiotropy settings or in the balanced setting when the horizontal pleiotropic effects for each trait is small or moderate (Figure S7A). However, when horizontal pleiotropic effect for one of the traits is large, as one might expect,<sup>4</sup> moPMR-Egger *p* values become inflated (Figure S7B). Overall, the *p* values of moPMR-Egger for testing causal effects generally adhere to the diagonal line under the null with various moderate model misspecifications.

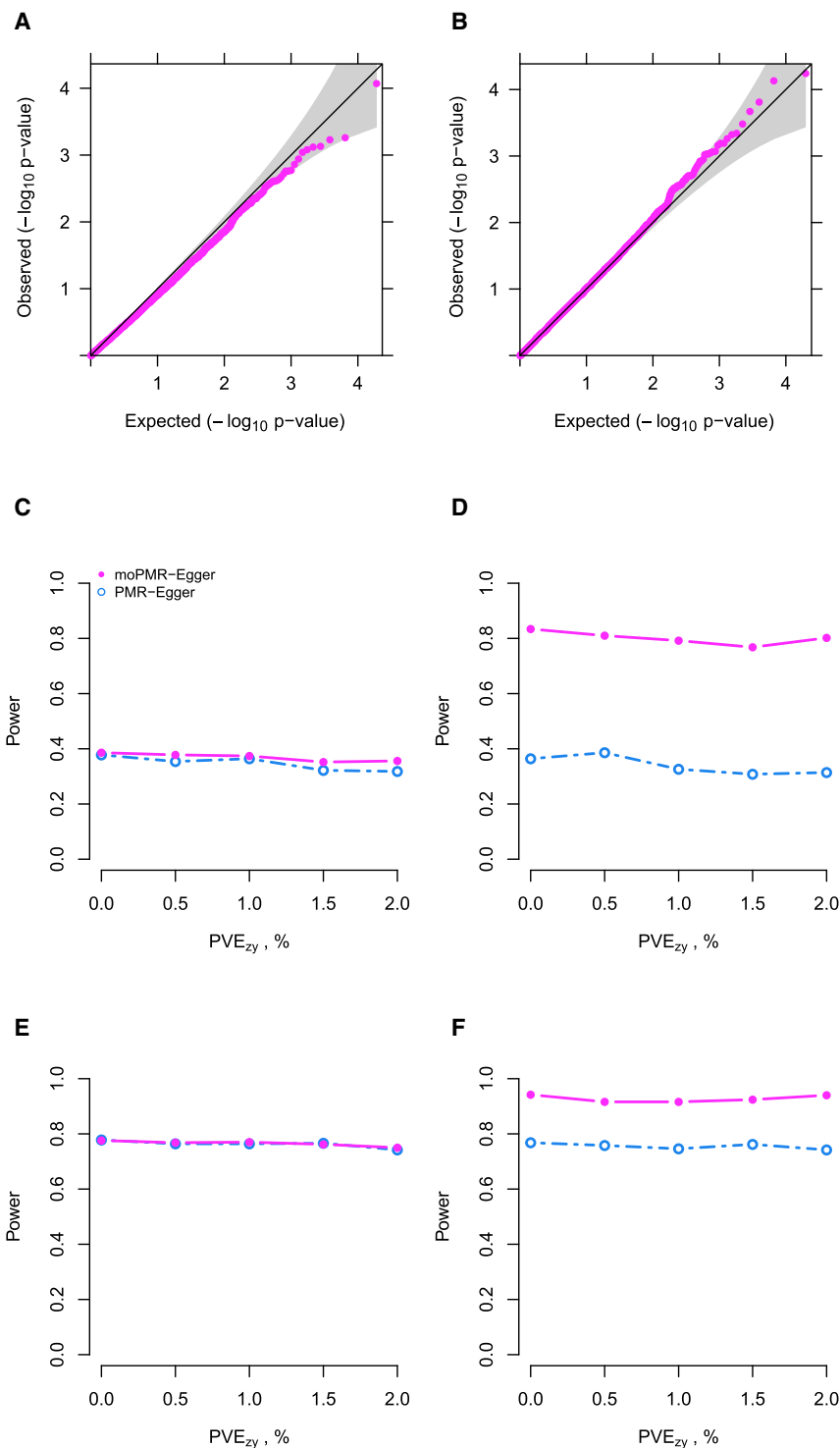
Next, we examined the power of moPMR-Egger and compared it with three univariate methods to identify non-zero causal effects. As expected, moPMR-Egger is more powerful than the univariate TWAS methods in most simulation scenarios. Specifically, across a total of 152 simulation scenarios, moPMR-Egger achieves an average of 53.12%, 42.40%, and 36.79% power gain as compared to PMR-Egger, PrediXcan, and TWAS, respectively. The power gain by moPMR-Egger remains substantial regardless whether the causal effects on different traits are homogeneous (Figures S8 and S9) or heterogeneous (Figures 2C–2F and S10), whether the gene affects one trait or multiple traits (Figure S11), how these traits are correlated with each other (Figure S12), and in the absence (Figures 2C–2F and S8) or presence of horizontal pleiotropic effects (Figures S9 and S10). The only exception is the simulation scenario where four traits are all independent of each other and where the gene is only associated with one of the four traits (Figures 2C and S8A–S10A). The lower power of moPMR-Egger in this scenario is presumably because moPMR-Egger uses extra parameters to model the non-zero causal effects on multiple correlated traits, thus suffering from a loss of degrees of freedom and subsequent loss of power there. However, even when the gene is associated with only one of the four traits, moPMR-Egger still has substantially more power than the other methods when traits are correlated with each other (Figures 2C, S8A, S9A, and S10A). Note that both moPMR-Egger and PMR-Egger control for horizontal pleiotropy while PrediXcan and TWAS do not. Consequently, in the presence of horizontal pleiotropy, the *p* values from PrediXcan and TWAS are known to be inflated and fail to control for type I error.<sup>4</sup> As a result, PrediXcan and TWAS appear to have higher power than PMR-Egger if we treat all *p* values as calibrated and rely on a nominal *p* value threshold instead of the corrected type I error threshold. However, even in the presence of

horizontal pleiotropy and inflated *p* values, the apparent power based on a nominal *p* value threshold for PrediXcan and TWAS is lower than that for moPMR-Egger, highlighting the importance of multi-trait modeling.

Power of different methods and the power gain brought by moPMR-Egger depend on several important parameters. First, the power of all methods reduces as the causal effects decrease, regardless whether the horizontal pleiotropy is absent or present. However, the power reduction of moPMR-Egger is the lowest among all methods, supporting its robust performance (Figure S11). Second, moPMR-Egger models multiple traits jointly and explicitly accounts for the trait correlations. Consequently, the power of moPMR-Egger increases as the correlation among the traits increases, while the power of the other methods does not change much (Figure S12). Third, as in illustrative simulations, the relative power gain brought up by modeling multiple traits depends on whether the causal effects on traits are in the same direction as the trait correlation or not. Specifically, when two traits are positively correlated with each other, then the power gain by moPMR-Egger is larger when the causal effect on one trait is in the opposite direction of that on the other as compared to being in the same direction, and vice versa (Figures 2D, S8B, S9B, and S10B). Finally, moPMR-Egger explicitly models the horizontal pleiotropic effects of SNPs on multiple traits separately. Consequently, moPMR-Egger remains powerful with an increased number of traits influenced by horizontal pleiotropy or with an increased horizontal pleiotropic effect size, whereas the other methods suffer (Figures S11C and S11D).

In the above simulations, we have primarily compared moPMR-Egger with the univariate approach of three existing TWAS methods. Here, we also compared moPMR-Egger with a minimum *p* value modification of these univariate methods for adapting them to analyze multiple traits together. Specially, we obtained the minimum *p* values across traits as the association evidence for the given gene and computed power of different methods based on an FDR of 0.05. Consistent with previous simulation results, the power improvement by moPMR-Egger over PMR-Egger, PrediXcan, and TWAS remains substantial regardless whether the causal effects on different traits are homogeneous (Figures S13A and S13C) or heterogeneous (Figures S13B and S13D) and whether there is an absence (Figures S13A and S13B) or presence (Figures S13C and S13D) of horizontal pleiotropy effects. Note that different from the above results under the Bonferroni adjusted threshold, both PrediXcan and TWAS display lower power in the presence of horizontal pleiotropy effect conditional on fixed FDR as one would expect (Figures S13C and S13D).

Finally, besides testing, moPMR-Egger produces accurate estimates of the causal effects, both under the null and under various alternatives, in the absence or presence of horizontal pleiotropic effects (Figure S14A).



**Figure 3. Type I error control and power for testing the horizontal pleiotropic effects under various simulation scenarios**

(A and B) Quantile-quantile plot of  $-\log_{10}$  p values from moPMR-Egger for testing the horizontal pleiotropic effects either in the absence or presence of causal effect under null simulations. Null simulations are performed under different causal effect sizes characterized by different  $\mathbf{PVE}_{zy}$ : (A)  $\mathbf{PVE}_{zy} = (0, 0, 0, 0)^T$ ; (B)  $\mathbf{PVE}_{zy, j}$  randomly selected from  $(0, 0.005, 0.01, 0.015, 0.02)^T$ ,  $j = 1, 2, 3, 4$ . p values from moPMR-Egger are on the expected diagonal line across a range of causal effect sizes.

(C–F) Power (y axis) at a Bonferroni adjusted threshold to detect the pleiotropic effects is plotted against different causal effect sizes characterized by  $\mathbf{PVE}_{zy}$  (x axis). Simulations are performed under either correlated traits (D and F) or independent traits (C and E), in the presence of horizontal pleiotropic effect with  $\gamma = (0, 1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3})^T$  in (C) and (D) and  $\gamma = (1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 2 \times 10^{-3})^T$  in (E) and (F).

where no horizontal pleiotropic effect exists for any of the four traits. We found that the p values from moPMR-Egger on testing horizontal pleiotropy are well calibrated, either in the absence or presence of causal effects (Figures 3A and 3B), regardless of the correlation among multiple traits (Figure S15) and regardless whether the genetic architecture underlying gene expression is sparse or polygenic as long as causal effects are absent (Figure S16). The only setting where moPMR-Egger fails is when its modeling assumptions are violated in multiple ways. For example, when the genetic architecture underlying gene expression is sparse and when the gene affects more than two traits, then the p values of moPMR-Egger become inflated (Figures S16C and S16D). Overall, the p value of moPMR-Egger for testing horizontal

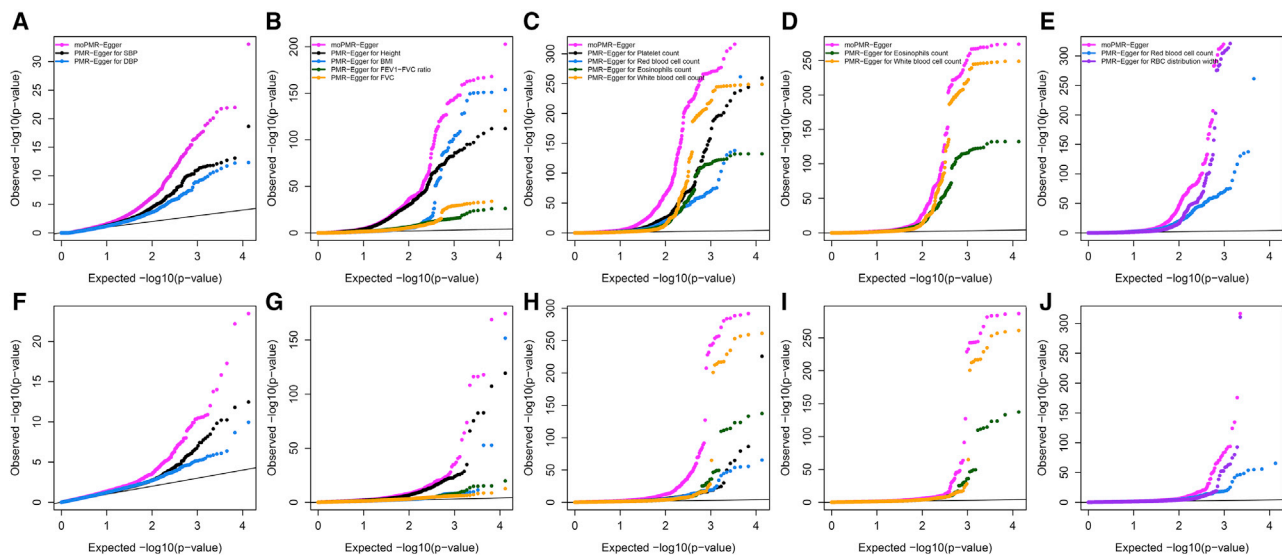
### Simulations: testing and estimating horizontal pleiotropic effects

Our second set of simulations is focused on horizontal pleiotropic effects testing. A benefit of moPMR-Egger, as compared to the usual TWAS/MR methods, is its ability to test whether SNPs exhibit non-zero horizontal pleiotropic effects on any outcome traits.

We first examined type I error control of moPMR-Egger on horizontal pleiotropic effects testing under the null,

pleiotropic effects under the null is well calibrated across the majority of scenarios.

Next, we examined the power of moPMR-Egger in detecting non-zero horizontal pleiotropic effects. Here, we compared the performance of moPMR-Egger with PMR-Egger, which is the only existing method that can provide a calibrated test for horizontal pleiotropic effect.<sup>4</sup> We first compared with the univariate approach of PMR-Egger. In the simulations, we found that the power of PMR-Egger



**Figure 4. TWAS analysis results for five trait categories in the UK Biobank using moPMR-Egger and PMR-Egger**

Quantile-quantile plots of  $-\log_{10}$  p values for testing the causal effects are shown for the blood pressure trait category (A), the physical measures trait category (B), the blood count trait category (C), the white blood cell indices trait category (D), and the red blood cell indices trait category (E). Quantile-quantile plot of  $-\log_{10}$  p values for testing the horizontal pleiotropic effects are shown for the blood pressure trait category (F), the physical measures trait category (G), the blood count trait category (H), the white blood cell indices trait category (I), and the red blood cell indices trait category (J). Compared methods include the multi-trait method moPMR-Egger (magenta) and univariate method PMR-Egger (different colors for different traits). SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; FVC, forced vital capacity; RBC, red blood cell count.

and moPMR-Egger increases with increasing horizontal pleiotropy, with moPMR-Egger outperforming PMR-Egger across a range of settings (Figure 3C versus Figure 3E and Figure 3D versus Figure 3F). While moPMR-Egger has comparable power with PMR-Egger when all four traits are uncorrelated with each other (Figures 3C and 3E), moPMR-Egger outperforms PMR-Egger in the presence of trait correlation (Figures 3D and 3F). In addition, the power of both methods reduces with increasing sparsity of the horizontal pleiotropic effects, though moPMR-Egger remains more powerful than PMR-Egger across a range of sparsity levels either in the absence or presence of causal effects (Figure S17). Consistent with the simulations on causal effects testing, the power of both methods on pleiotropic effects testing suffers in the absence of directional pleiotropic effects (Figure S18). Besides the univariate approach of PMR-Egger, we also compared moPMR-Egger with the minimum p value approach of PMR-Egger and obtained consistent results (Figure S19).

Finally, moPMR-Egger can estimate the horizontal pleiotropic effects accurately in the presence of directional pleiotropic effects. However, in the absence of directional pleiotropic effects, as expected, the pleiotropic effect estimates become downward biased (Figure S14B).

### Real data applications

We applied moPMR-Egger for TWAS analysis to integrate gene expression data from GEUVADIS with GWAS data on 11 traits from 5 trait categories in the UK Biobank (details in Material and Methods). The gene expression data from GEUVADIS study includes 15,810 genes. The five trait

categories in UK Biobank include the blood pressure category, physical measures category, blood count category, white blood cell indices category, and red blood cell indices category. We examined one trait category at a time to identify genes associated with traits in the category.

The p values for testing the causal effects of each gene on the traits from moPMR-Egger are displayed for each trait category, along with the p values from PMR-Egger for each trait in the category (Figures 4A–4E) as well as the corresponding minimum p values across traits in the category (Figure S20). We did not apply the univariate approaches of PrediXcan and TWAS as both do not control for horizontal pleiotropic effects that are prevalent in the data as demonstrated before.<sup>4</sup> Indeed, both approaches produce inflated p values under the null simulations when the absolute horizontal pleiotropic effect exceeds 0.0001 (Figure S21), which happens on an average of 1,636.46 genes across 11 traits examined here (ranges from 1,320 for BMI to 2,632 for height; Figure S1). Therefore, instead, we compared the minimum p value approaches of these two methods in an in-depth analysis described in the next paragraph. Consistent with simulations, moPMR-Egger identified more genes than the univariate PMR-Egger at the corresponding Bonferroni corrected transcriptome-wide thresholds (Table S3). moPMR-Egger identified a total of 13.15% more genes as compared to PMR-Egger, highlighting the power of analyzing multiple traits jointly in TWAS. Majority of genes (89.29%) identified by PMR-Egger are also identified by moPMR-Egger but with increased association significance (Figure S22A). For example, *TFRC* (MIM: 190010) is identified to be associated with RDW

(PMR-Egger  $p = 3.22 \times 10^{-17}$ ), but with increased association evidence when all traits in the red blood cell indices category are modeled together (moPMR-Egger  $p = 5.72 \times 10^{-22}$ ). *TFRC* encodes the classical transferrin receptor that is involved in cellular iron uptake.<sup>54</sup> Multiple SNPs in *TFRC* have been established to be associated with various erythrocyte phenotypes in GWASs.<sup>55</sup> These associated erythrocyte phenotypes include the mean corpuscular hemoglobin (MCH) and mean corpuscular volume (MCV, the average volume of red blood cells) which is closely related to RDW.<sup>54,55</sup> The variants in *TFRC* likely lead to decreased iron availability for red cell precursors, as has been observed in mice deficient in *Tfrc*, thus resulting in a compensatory increase of red blood cell size as measured by RDW.<sup>56</sup> In addition, and perhaps more importantly, moPMR-Egger detected many likely causal genes that are missed by PMR-Egger. For example, *EPHB4* (MIM: 600011) is identified by moPMR-Egger only in the blood pressure category (moPMR-Egger  $p = 1.41 \times 10^{-8}$ , PMR-Egger  $p = 0.98$  for SBP and  $p = 5.33 \times 10^{-4}$  for DBP). *EPHB4* encodes the Ephrin type-B receptor 4 that binds to ephrin-B2 and plays an essential role in vascular development and angiogenesis.<sup>57,58</sup> Previous studies have demonstrated that deletion of *Ephb4* in mice leads to hypotension,<sup>59</sup> supporting the causal role of *EPHB4* in regulating blood pressure. We list the regional association plots for *TFRC* and *EPHB4* in [Figures S23–S25](#).

We performed an in-depth analysis on the blood pressure category, which contains only two traits (SBP and DBP) and is thus easy to explore in detail. Here, in addition to comparing with the univariate approach of PMR-Egger as described above, we compared moPMR-Egger with the minimum p value approaches of PMR-Egger, PrediXcan, and TWAS. We calculated the power of different methods based on an empirical FDR of either 0.01, 0.025, 0.05, 0.075, or 0.1 using permutations. Consistent with simulations, moPMR-Egger identified more genes than the other methods, highlighting the importance of multi-trait modeling ([Figure S26](#)). Specifically, the number of significant genes detected by moPMR-Egger, PMR-Egger, PrediXcan, and TWAS based on an empirical FDR of 0.05 are 765, 691, 90, and 83, respectively. The performance of moPMR-Egger is followed closely by PMR-Egger, supporting the previous observation that likelihood-based inference as used in these two methods is more powerful than the two-stage based inference used as in PrediXcan and TWAS.<sup>4,60</sup> Among the genes identified by moPMR-Egger, approximately 40% (301/765) of significant genes identified by moPMR-Egger have opposite causal effect directions on the two traits. GO and KEGG pathway enrichment analyses ([Figure 5](#)) show that genes with the same causal effect direction on the two traits are significantly enriched in the pathways of antigen processing and presentation ( $q = 1.96 \times 10^{-4}$ ; [Figure 5A](#)), human T cell leukemia virus 1 infection ( $q = 0.03$ ), viral myocarditis ( $q = 0.03$ ), and intestinal immune network for IgA production ( $q = 0.04$ , [Figure 5B](#)). Such enrichment in inflammation and induced

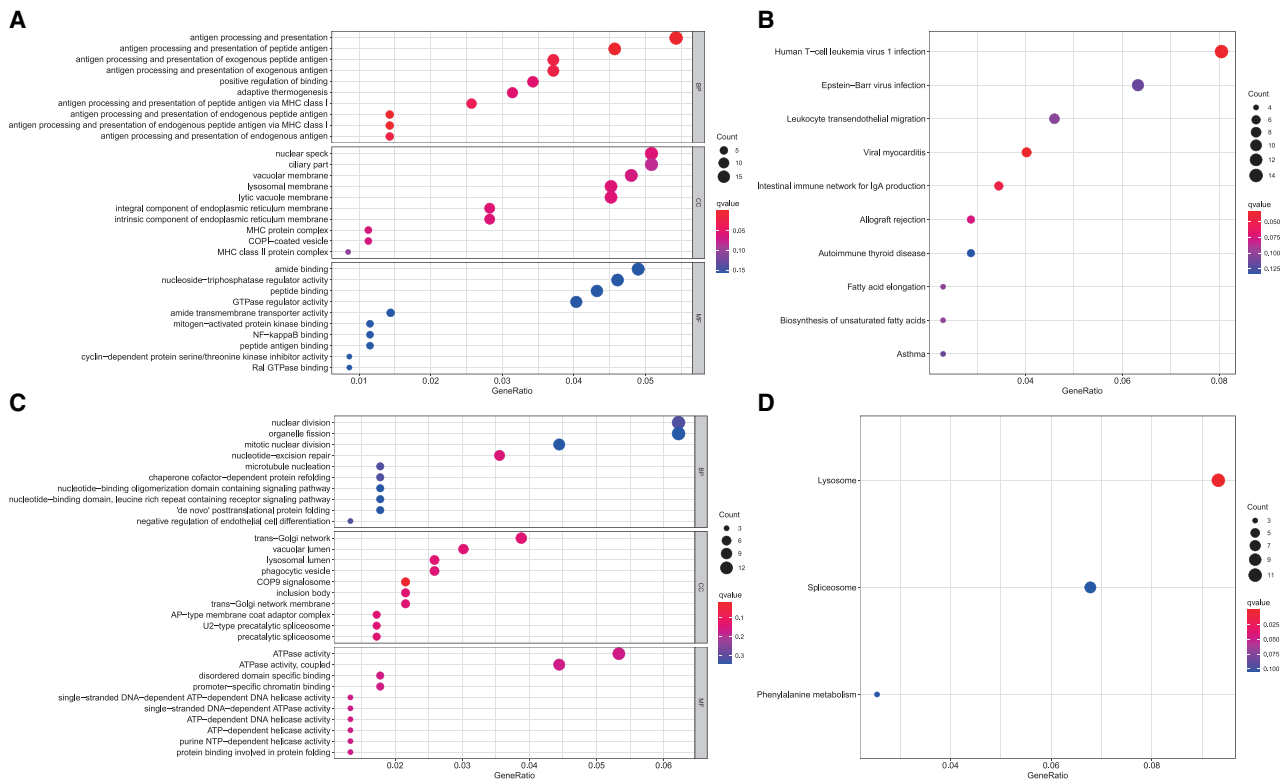
immune response pathways supports their recently revealed roles in regulating blood pressure.<sup>61–63</sup> Indeed, inflammation activates innate and adaptive immune responses, resulting in alterations in the vasculature, kidneys, and sympathetic nervous system (SNS) that can eventually lead to chronically elevated blood pressures.<sup>64</sup> In contrast, genes with opposite causal effect directions on the two traits are significantly enriched in the pathways of COP9 signalosome ( $q = 0.02$ ; [Figure 5C](#)) and lysosome ( $q = 5.14 \times 10^{-4}$ , [Figure 5D](#)). The enrichment in COP9 signalosome and lysosome pathways supports their critical functions in maintaining homeostasis and plasticity of vasculature and subsequent regulation of blood pressure. Indeed, both COP9 signalosome and lysosome control ubiquitination in the vasculature and the subsequent modulation of protein turnover.<sup>65</sup> Protein ubiquitination and turnover in the vasculature determine the vascular tone and stiffness, which can affect SBP and DBP differently during cardiac cycles of cell contraction and relaxation.<sup>66</sup>

Next, we shifted our focus to testing horizontal pleiotropic effects. The p values for testing the horizontal pleiotropic effects of each gene from moPMR-Egger and PMR-Egger are shown for the five trait categories ([Figures 4F–4J](#)). Consistent with simulations, moPMR-Egger detected a total of 17.10% more genes with horizontal pleiotropy than PMR-Egger ([Table S3](#)). Specifically, moPMR-Egger detected 56 genes in the blood pressure category, 275 in the physical measures category, 383 in the blood cell count category, 146 in the white blood cell indices category, and 153 in the red blood cell indices category, with gene overlaps showing in [Figure S27](#). In contrast, PMR-Egger detected 26, 196, 266, 106, and 136, respectively. The majority of genes (71.53%) identified by PMR-Egger are also identified by moPMR-Egger ([Figure S22B](#)). Importantly, 36.16% of genes with significant horizontal pleiotropic effects identified by moPMR-Egger have significant causal effects, while 11.64% of genes with significant causal effects have horizontal pleiotropic effects. The noticeable overlap between genes with horizontal pleiotropic effects and genes with causal effects highlights the importance of modeling both effects in TWAS. In-depth analysis in the blood pressure category also revealed that genes with significant causal effects on DBP and SBP are enriched in pathways of antigen processing and presentation as well as protein refolding ([Figures S28A and S28B](#)), while genes with the significant horizontal pleiotropic effects are enriched in various metabolic processes and protein export ([Figures S28C and S28D](#)).

Finally, we note that moPMR-Egger is also computationally efficient, with similar computing time and physical memory requirement as existing TWAS methods ([Table S4](#)).

## Discussion

We have presented moPMR-Egger, a method that extends the univariate PMR-Egger for analyzing multiple outcome



**Figure 5. GO function and KEGG pathway enrichment analysis on genes identified by moPMR-Egger and PMR-Egger in the blood pressure trait category**

Dot plots show the top ten enriched GO BP, CC, and MF terms for identified genes that have the same causal effect direction on SBP and DBP (A) and that have the opposite causal effect directions on SBP and DBP (C). Dot plots show the top ten enriched KEGG pathway terms for genes that have the same causal effect direction on SBP and DBP (B) and that have the opposite causal effect directions on SBP and DBP (D). Dot color represents statistical significance of enrichment analysis based on q-value while dot size represents the fraction of genes annotated to each term. SBP, systolic blood pressure; DBP, diastolic blood pressure; GO, Gene Ontology; BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes.

traits in TWAS applications. moPMR-Egger accounts for the correlation structure among multiple traits, takes advantage of all *cis*-SNPs that are in LD with each other, tests and controls for horizontal pleiotropic effects, and performs inference under a likelihood framework. In both simulations and real data applications, moPMR-Egger yields calibrated p values across a wide range of scenarios and substantially improves power over existing univariate approaches.

One important modeling assumption we made in moPMR-Egger is that SNPs exhibit the same horizontal pleiotropic effect on the same trait. Such equal effect size assumption for each trait directly follows that of the Egger assumption<sup>4,19,20</sup> and is analogous to the burden effect size assumption commonly used for rare variant tests. Consistent with previous studies,<sup>4,19,20</sup> we found that the equal effect size assumption employed in moPMR-Egger appears to work reasonably robust for causal effect estimations and testing with respect to a range of model misspecifications and appears to be effective in the real data applications examined here. However, we do acknowledge that our equal effect size assumption in moPMR-Egger can be overly restrictive in many settings. Therefore, while we view moPMR-Egger as an important first step toward multi-trait

TWAS applications, we emphasize that imposing more realistic modeling assumptions on the horizontal pleiotropic effects in the future will likely be beneficial.

We have primarily illustrated the benefits of moPMR-Egger for analyzing individual-level GWAS data. moPMR-Egger can be easily extended to take input in the form of summary statistics alone (details in [Material and Methods](#)). Specifically, the summary statistics version of moPMR-Egger requires standardized marginal SNP effect size estimates or marginal z-scores, both on the gene expression and on the multiple outcome traits of interest. Moreover, it requires an LD correlation matrix among *cis*-SNPs that can be constructed based on a reference panel. Note that when a reference panel is used to construct the LD correlation matrix, one needs to ensure that the ethnicity of the reference panel matches that of the study data to avoid potentially biased inference results.<sup>4</sup> In addition, it requires an estimated correlation matrix among multiple traits. The trait correlation matrix can be easily estimated by using the marginal z-scores of genome-wide SNPs for each trait: because the correlation between two traits equals approximately to the correlation between the marginal z-scores for the two traits under the null where SNPs are not associated with any trait, we can select SNPs with a

p value greater than a predefined significance threshold (e.g.,  $10^{-5}$ ) for any trait and use the correlation among these marginal z-scores as an estimate for the trait correlation.<sup>67</sup> We have validated the implementation of the summary statistics version of moPMR-Egger in the simulations and found that the results are largely consistent with that of moPMR-Egger fitted using individual level data (Figures S29, S30). Being capable of making use of summary statistics extends the applicability of moPMR-Egger to datasets where individual-level genotype or phenotype is not available.

We have mainly compared moPMR-Egger with three existing univariate TWAS methods that include PMR-Egger, PrediXcan, and TWAS. These three univariate TWAS methods and moPMR-Egger all examine one gene at a time. We note, however, that other univariate TWAS methods have been recently developed for modeling multiple genes jointly. For example, TWMR analyzes one gene at a time while controlling for its neighboring genes as covariates. In the null simulations, we found that the p values from the univariate approach of TWMR are overly conservative while the p values from the minimum p value approach of TWMR are inflated (Figure S31A), even in the absence of horizontal pleiotropy. The p values from both the univariate approach and the minimum p value approach of TWMR are overly inflated in the presence of horizontal pleiotropy (Figure S31B). In the alternative simulations, we found that TWMR is much less powerful than moPMR-Egger, resulting in an average of around 20-fold power loss across simulation scenarios: the power of TWMR is only around 0.08 in the absence of horizontal pleiotropy and only around 0.03 in the presence of horizontal pleiotropy. The low power of TWMR is likely due to the relatively small sample size in GEUVADIS as brought up by Porcu et al.<sup>33</sup> We also applied TWMR to analyze the two blood pressure traits. Due to the stringent filtering criteria recommend by Porcu et al.,<sup>33</sup> we analyzed a total of 5,573 genes. In the real data analysis, TWMR identified only 8 significant genes associated with at least one trait while moPMR-Egger identified 67 in the same analyzed gene set, supporting the low power of TWMR as observed in simulations. As another example, FOCUS<sup>8</sup> is a TWAS fine mapping method that examines one genomic region at a time to identify the causal gene among a list of candidate ones in the region. FOCUS is commonly applied to only examine genomic regions that contain at least one significant gene detected by a standard TWAS method. Consequently, FOCUS can be effectively paired with any univariate TWAS method including moPMR-Egger as a pre-selection step. We followed Mancuso et al.<sup>8</sup> to pair FOCUS with moPMR-Egger and each of the three univariate TWAS methods to analyze the two blood pressure traits. Due to the stringent filtering criteria recommend by Mancuso et al.,<sup>8</sup> we analyzed a total of 246 genomic regions and 4,692 genes. In the analysis, we found that the significant genes identified by moPMR-Egger are largely consistent with those identified after FOCUS analysis, more so than

the other methods (Figure S32), supporting the effectiveness of moPMR-Egger. Integrating multi-trait modeling with multi-gene modeling is an important future research direction.

## Data and code availability

No data were generated in the present study. The GEUVADIS gene expression data are publicly available online. The GERA data is publicly available on dbGaP with accession number phs000788. The UK Biobank data is from UK Biobank resource under application number 30686.

The moPMR-Egger is implemented in the R package PMR, freely available on GitHub.

## Supplemental Information

Supplemental Information can be found online at <https://doi.org/10.1016/j.ajhg.2020.12.006>.

## Acknowledgments

This study was supported by the National Natural Science Foundation of China (81673272 and 81872712), the Natural Science Foundation of Shandong Province (ZR2019ZD02), and the Young Scholars Program of Shandong University (2016WLJH23), all awarded to Z.Y. X.Z. is supported by the University of Michigan, Ann Arbor, US. The GERA data (dbGaP: phs000788) came from a grant, the Resource for Genetic Epidemiology Research in Adult Health and Aging (RC2 AG033067; Schaefer and Risch, PIs) awarded to the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and the UCSF Institute for Human Genetics. The RPGEH was supported by grants from the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, Kaiser Permanente Northern California, and the Kaiser Permanente National and Northern California Community Benefit Programs. The RPGEH and the Resource for Genetic Epidemiology Research in Adult Health and Aging are described on the corresponding dbGaP website. This study has been conducted using UK Biobank resource under Application Number 30686. UK Biobank was established by the Wellcome Trust medical charity, Medical Research Council, Department of Health, Scottish Government, and the Northwest Regional Development Agency. It has also had funding from the Welsh Assembly Government, British Heart Foundation, and Diabetes UK.

## Declaration of Interests

The authors declare no competing interests.

Received: September 9, 2020

Accepted: December 7, 2020

Published: January 11, 2021

## Web Resources

dbGaP, <https://www.ncbi.nlm.nih.gov/gap>

GEMMA, <http://www.xzlab.org/software.html>

GEUVADIS, <http://www.geuvadis.org>

moPMR-Egger, <https://github.com/yuanzhongshang/PMR>  
OMIM, <https://www.omim.org/>  
PLINK, <https://www.cog-genomics.org/plink/>  
Sample QC procedure in Neale lab, [https://github.com/Nealelab/UK\\_Biobank\\_GWAS/tree/master/imputed-v2-gwas](https://github.com/Nealelab/UK_Biobank_GWAS/tree/master/imputed-v2-gwas)  
UK Biobank, <https://www.ukbiobank.ac.uk/>

## References

- Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., de Geus, E.J., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* *48*, 245–252.
- 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., Im, H.K.; and GTEx Consortium (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* *47*, 1091–1098.
- Zhu, H.H., and Zhou, X. (2020). Transcriptome-wide association studies: a view from Mendelian randomization. *Quant. Biol.* Published online June 17, 2020. <https://doi.org/10.1007/s40484-020-0207-4>.
- Yuan, Z., Zhu, H., Zeng, P., Yang, S., Sun, S., Yang, C., Liu, J., and Zhou, X. (2020). Testing and controlling for horizontal pleiotropy with probabilistic Mendelian randomization in transcriptome-wide association studies. *Nat. Commun.* *11*, 3861.
- Zeng, P., and Zhou, X. (2017). Non-parametric genetic prediction of complex traits with latent Dirichlet process regression models. *Nat. Commun.* *8*, 456.
- Nagpal, S., Meng, X., Epstein, M.P., Tsoi, L.C., Patrick, M., Gibson, G., De Jager, P.L., Bennett, D.A., Wingo, A.P., Wingo, T.S., and Yang, J. (2019). TIGAR: An Improved Bayesian Tool for Transcriptomic Data Imputation Enhances Gene Mapping of Complex Traits. *Am. J. Hum. Genet.* *105*, 258–266.
- Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* *48*, 481–487.
- Mancuso, N., Freund, M.K., Johnson, R., Shi, H., Kichaev, G., Gusev, A., and Pasaniuc, B. (2019). Probabilistic fine-mapping of transcriptome-wide association studies. *Nat. Genet.* *51*, 675–682.
- Hu, Y., Li, M., Lu, Q., Weng, H., Wang, J., Zekavat, S.M., Yu, Z., Li, B., Gu, J., Muchnik, S., et al.; Alzheimer's Disease Genetics Consortium (2019). A statistical framework for cross-tissue transcriptome-wide association analysis. *Nat. Genet.* *51*, 568–576.
- Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M., and Smoller, J.W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.* *14*, 483–495.
- Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., et al.; ReproGen Consortium; Psychiatric Genomics Consortium; and Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3 (2015). An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* *47*, 1236–1241.
- Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* *381*, 1371–1379.
- Pickrell, J.K., Berisa, T., Liu, J.Z., Séguirel, L., Tung, J.Y., and Hinds, D.A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* *48*, 709–717.
- Kanai, M., Akiyama, M., Takahashi, A., Matoba, N., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., et al. (2018). Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* *50*, 390–400.
- Stephens, M. (2013). A unified framework for association analysis with multiple related phenotypes. *PLoS ONE* *8*, e65245.
- Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* *11*, 407–409.
- Baselmans, B.M.L., Jansen, R., Ip, H.F., van Dongen, J., Abdellaoui, A., van de Weijer, M.P., Bao, Y., Smart, M., Kumari, M., Willemsen, G., et al.; BIOS consortium; and Social Science Genetic Association Consortium (2019). Multivariate genome-wide analyses of the well-being spectrum. *Nat. Genet.* *51*, 445–451.
- Turley, P., Walters, R.K., Maghziyan, O., Okbay, A., Lee, J.J., Fontana, M.A., Nguyen-Viet, T.A., Wedow, R., Zacher, M., Furlotte, N.A., et al.; 23andMe Research Team; and Social Science Genetic Association Consortium (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* *50*, 229–237.
- Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* *44*, 512–525.
- Barfield, R., Feng, H., Gusev, A., Wu, L., Zheng, W., Pasaniuc, B., and Kraft, P. (2018). Transcriptome-wide association studies accounting for colocalization using Egger regression. *Genet. Epidemiol.* *42*, 418–433.
- Berzuini, C., Guo, H., Burgess, S., and Bernardinelli, L. (2020). A Bayesian approach to Mendelian randomization with multiple pleiotropic variants. *Biostatistics* *21*, 86–101.
- Dawid, A.P. (2000). Causal inference without counterfactuals. *J. Am. Stat. Assoc.* *95*, 407–424.
- Dawid, A.P. (2015). Statistical causality from a decision-theoretic perspective. *Annu. Rev. Stat. Appl.* *2*, 273–303.
- Berzuini, C., Dawid, P., and Bernardinelli, L. (2012). Causality: Statistical perspectives and applications (John Wiley & Sons).
- Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., González-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al.; Geuvadis Consortium (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* *501*, 506–511.
- Banda, Y., Kvale, M.N., Hoffmann, T.J., Hesselson, S.E., Rana-tunga, D., Tang, H., Sabatti, C., Croen, L.A., Dispensa, B.P., Henderson, M., et al. (2015). Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Genetics* *200*, 1285–1295.
- Kvale, M.N., Hesselson, S., Hoffmann, T.J., Cao, Y., Chan, D., Connell, S., Croen, L.A., Dispensa, B.P., Eshragh, J., Finn, A., et al. (2015). Genotyping informatics and quality control for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Genetics* *200*, 1051–1060.

28. Bennett, B.J., Farber, C.R., Orozco, L., Kang, H.M., Ghazalpour, A., Siemers, N., Neubauer, M., Neuhaus, I., Yordanova, R., Guan, B., et al. (2010). A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res.* 20, 281–290.
29. Price, A.L., Patterson, N., Hancks, D.C., Myers, S., Reich, D., Cheung, V.G., and Spielman, R.S. (2008). Effects of cis and trans genetic ancestry on gene expression in African Americans. *PLoS Genet.* 4, e1000294.
30. Price, A.L., Helgason, A., Thorleifsson, G., McCarroll, S.A., Kong, A., and Stefansson, K. (2011). Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genet.* 7, e1001317.
31. Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.* 9, e1003264.
32. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824.
33. Porcu, E., Rieger, S., Lepik, K., Santoni, F.A., Reymond, A., Kutalik, Z.; eQTLGen Consortium; and BIOS Consortium (2019). Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat. Commun.* 10, 3300.
34. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
35. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209.
36. Ehret, G.B., and Caulfield, M.J. (2013). Genes for blood pressure: an opportunity to understand hypertension. *Eur. Heart J.* 34, 951–961.
37. Loh, P.R., Kichaev, G., Gazal, S., Schoech, A.P., and Price, A.L. (2018). Mixed-model association for biobank-scale datasets. *Nat. Genet.* 50, 906–908.
38. Wen, X., Luca, F., and Pique-Regi, R. (2015). Cross-population joint analysis of eQTLs: fine mapping and functional annotation. *PLoS Genet.* 11, e1005176.
39. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Ziadna, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22, 1760–1774.
40. Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* 7, 500–507.
41. Evangelou, E., Warren, H.R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H., Ntritsos, G., Dimou, N., Cabrera, C.P., Karaman, I., et al.; Million Veteran Program (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* 50, 1412–1425.
42. Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C., Hwang, S.J., et al.; International Consortium for Blood Pressure Genome-Wide Association Studies; CARDIoGRAM consortium; CKDGen Consortium; KidneyGen Consortium; EchoGen consortium; and CHARGE-HF consortium (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 478, 103–109.
43. Wain, L.V., Vaez, A., Jansen, R., Joehanes, R., van der Most, P.J., Erzurumluoglu, A.M., O'Reilly, P.F., Cabrera, C.P., Warren, H.R., Rose, L.M., et al. (2017). Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension* 70, E4.
44. Warren, H.R., Evangelou, E., Cabrera, C.P., Gao, H., Ren, M., Mifsud, B., Ntalla, I., Surendran, P., Liu, C., Cook, J.P., et al.; International Consortium of Blood Pressure (ICBP) 1000G Analyses; BIOS Consortium; Lifelines Cohort Study; Understanding Society Scientific group; CHD Exome+ Consortium; ExomeBP Consortium; T2D-GENES Consortium; GoT2DGenes Consortium; Cohorts for Heart and Ageing Research in Genome Epidemiology (CHARGE) BP Exome Consortium; International Genomics of Blood Pressure (iGEN-BP) Consortium; and UK Biobank CardioMetabolic Consortium BP working group (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* 49, 403–415.
45. Ehret, G.B., Ferreira, T., Chasman, D.I., Jackson, A.U., Schmidt, E.M., Johnson, T., Thorleifsson, G., Luan, J., Donnelly, L.A., Kanoni, S., et al.; CHARGE-EchoGen consortium; CHARGE-HF consortium; and Wellcome Trust Case Control Consortium (2016). The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat. Genet.* 48, 1171–1184.
46. Muñoz, M., Pong-Wong, R., Canela-Xandri, O., Rawlik, K., Haley, C.S., and Tenesa, A. (2016). Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat. Genet.* 48, 980–983.
47. Feinleib, M., Garrison, R.J., Fabsitz, R., Christian, J.C., Hrubec, Z., Borhani, N.O., Kannel, W.B., Rosenman, R., Schwartz, J.T., and Wagner, J.O. (1977). The NHLBI twin study of cardiovascular disease risk factors: methodology and summary of results. *Am. J. Epidemiol.* 106, 284–285.
48. Poulter, N.R., Prabhakaran, D., and Caulfield, M. (2015). Hypertension. *Lancet* 386, 801–812.
49. Mongeau, J.G., Biron, P., and Sing, C.F. (1986). The influence of genetics and household environment upon the variability of normal blood pressure: the Montreal Adoption Survey. *Clin. Exp. Hypertens. A* 8, 653–660.
50. Forouzanfar, M.H., Alexander, L., Anderson, H.R., Bachman, V.F., Biryukov, S., Brauer, M., Burnett, R., Casey, D., Coates, M.M., Cohen, A., et al.; GBD 2013 Risk Factors Collaborators (2015). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 386, 2287–2323.
51. Sundström, J.; and Blood Pressure Lowering Treatment Trialists' Collaboration (2014). Blood pressure-lowering treatment based on cardiovascular risk: a meta-analysis of individual patient data. *Lancet* 384, 591–598.
52. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287.
53. Berisa, T., and Pickrell, J.K. (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* 32, 283–285.



54. Andrews, N.C. (2009). Genes determining blood cell traits. *Nat. Genet.* *41*, 1161–1162.
55. Ganesh, S.K., Zakai, N.A., van Rooij, F.J., Soranzo, N., Smith, A.V., Nalls, M.A., Chen, M.-H., Kottgen, A., Glazer, N.L., Dehghan, A., et al. (2009). Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat. Genet.* *41*, 1191–1198.
56. Levy, J.E., Jin, O., Fujiwara, Y., Kuo, F., and Andrews, N.C. (1999). Transferrin receptor is necessary for development of erythrocytes and the nervous system. *Nat. Genet.* *21*, 396–399.
57. Gerety, S.S., Wang, H.U., Chen, Z.F., and Anderson, D.J. (1999). Symmetrical mutant phenotypes of the receptor EphB4 and its specific transmembrane ligand ephrin-B2 in cardiovascular development. *Mol. Cell* *4*, 403–414.
58. Salvucci, O., and Tosato, G. (2012). Essential roles of EphB receptors and EphrinB ligands in endothelial cell function and angiogenesis. *Adv. Cancer Res.* *114*, 21–57.
59. Wang, Y., Thorin, E., Luo, H., Tremblay, J., Lavoie, J.L., Wu, Z., Peng, J., Qi, S., and Wu, J. (2015). EPHB4 Protein Expression in Vascular Smooth Muscle Cells Regulates Their Contractility, and EPHB4 Deletion Leads to Hypotension in Mice. *J. Biol. Chem.* *290*, 14235–14244.
60. Burgess, S., Small, D.S., and Thompson, S.G. (2017). A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* *26*, 2333–2355.
61. Rodríguez-Iturbe, B., Pons, H., and Johnson, R.J. (2017). Role of the Immune System in Hypertension. *Physiol. Rev.* *97*, 1127–1164.
62. Rodríguez-Iturbe, B., Pons, H., Quiroz, Y., and Johnson, R.J. (2014). The immunological basis of hypertension. *Am. J. Hypertens.* *27*, 1327–1337.
63. Wenzel, U., Turner, J.E., Krebs, C., Kurts, C., Harrison, D.G., and Ehmke, H. (2016). Immune Mechanisms in Arterial Hypertension. *J. Am. Soc. Nephrol.* *27*, 677–686.
64. Drummond, G.R., Vinh, A., Guzik, T.J., and Sobey, C.G. (2019). Immune mechanisms of hypertension. *Nat. Rev. Immunol.* *19*, 517–532.
65. Martin, D.S., and Wang, X. (2015). The COP9 signalosome and vascular function: intriguing possibilities? *Am. J. Cardiovasc. Dis.* *5*, 33–52.
66. Milic, J., Tian, Y., and Bernhagen, J. (2019). Role of the COP9 Signalosome (CSN) in Cardiovascular Diseases. *Biomolecules* *9*, 9.
67. Ray, D., and Boehnke, M. (2018). Methods for meta-analysis of multiple traits using GWAS summary statistics. *Genet. Epidemiol.* *42*, 134–145.