**Testing and controlling for horizontal pleiotropy with the probabilistic Mendelian randomization in transcriptome-wide association studies**

Zhongshang Yuan1,2, Huanhuan Zhu2, Ping Zeng3, Sheng Yang2, Shiquan Sun2, Can Yang4, Jin Liu5, Xiang Zhou2,6,#

1. Department of Biostatistics, School of Public Health, Shandong University, Jinan, China.

2. Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

3. Department of Epidemiology and Biostatistics, Xuzhou Medical University, Xuzhou, Jiangsu, China

4. Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, China

5. Centre for Quantitative Medicine, Program in Health Services and Systems Research, Duke-NUS Medical School, Singapore 169857

6. Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA

#: To whom correspondence should be addressed: xzhousph@umich.edu

**Abstract**

Integrating association results from both genome-wide association studies (GWASs) and expression quantitative trait locus (eQTL) mapping studies has the potential to shed light on the molecular mechanism underlying disease etiology. Several statistical methods have been recently developed to integrate GWASs with eQTL studies in the form of transcriptome-wide association studies (TWASs). These existing methods can all be viewed as two sample Mendelian randomization (MR) methods, which are also widely used in various GWASs for inferring the causal relationship among complex traits. Unfortunately, most existing TWAS and MR methods make an unrealistic modeling assumption that the instrumental variables do not exhibit horizontal pleiotropic effects. However, horizontal pleiotropic effects have been recently discovered to be wide spread across complex traits, and as we will show here, are also wide spread across gene expression traits. Therefore, allowing for no horizontal pleiotropic effects can be overall restrictive, and, as we will be show here, can lead to a substantial inflation of test statistics and subsequently false discoveries in TWAS applications. Here, we present a probabilistic MR method, which we refer to as PMR-Egger, for testing and controlling of horizontal pleiotropic effects in TWAS applications. PMR-Egger relies on a new MR likelihood framework that unifies many existing TWAS and MR methods, accommodates multiple correlated instruments, tests the causal effect of gene on trait in the presence of horizontal pleiotropy, and, with a newly developed parameter expansion version of the expectation maximization algorithm, is scalable to hundreds of thousands of individuals. With extensive simulations, we show that PMR-Egger provides calibrated type I error control for causal effect testing in the presence of horizontal pleiotropic effects, is reasonably robust for various types of horizontal pleiotropic effect mis-specifications, is more powerful than existing MR approaches, and, as a by-product, can directly test for horizontal pleiotropy. We illustrate the benefits of our method in applications to 39 diseases and complex traits obtained from three GWASs including the UK Biobank and show how PMR-Egger can lead to new biological discoveries through integrative analysis.

**Introduction**

Genome-wide association studies (GWASs) have identified many SNPs associated with common diseases or disease related traits. Parallel expression quantitative trait loci (eQTL) mapping studies have also identified many cis-acting SNPs associated with gene expression level of nearby genes. Integrating association results from both GWASs and eQTL mapping studies has the potential to shield light on the molecular mechanism underlying disease etiology. Several methods have been recently proposed to integrate GWASs with eQTL mapping studies. For example prediXcan[1](#_ENREF_1" \o "Gamazon, 2015 #57) proposes to perform a weighted SNP set test in GWAS by inferring SNP weights from eQTL studies. TWAS[2](#_ENREF_2) proposes to infer the association between gene expression and disease trait by leveraging the shared common set of cis-SNPs. SMR[3](#_ENREF_3) or GSMR[4](#_ENREF_4) directly tests the causal association between gene expression and disease trait under a Mendelian randomization (MR) framework through selecting a single instrument or multiple independent or near independent instruments. While each of these integrative methods was originally proposed to solve a different problem, all the aforementioned methods can be viewed as a two-sample MR method but with different modeling assumptions, aiming to identify genes causally associated with complex traits or diseases in the context of transcriptome-wide association studies (TWAS).

MR analysis is a form of instrumental variable analysis that was originally developed in the field of causal inference[5](#_ENREF_5). MR aims to determine the causal relationship between an exposure variable (e.g. gene expression) and an outcome variable (e.g. complex trait) in observational studies. MR treats SNPs as instrumental variables for the exposure variable of interest and uses these SNP instruments to estimate and test the causal effect of the exposure variable on the outcome variable. MR methods have been widely applied to investigate the causal relationship among various complex traits[6-9](#_ENREF_6), and, through a two-sample design, can be easily adapted to settings where the exposure and outcome are measured on two different sets of individuals[10](#_ENREF_10); [11](#_ENREF_11). However, MR analysis for TWAS is not straightforward and requires the development of new methods that can accommodate two important TWAS data features.

First, both GWASs and eQTL mapping studies collect multiple SNPs that are in high linkage disequilibrium (LD) with each other. Traditional MR methods, such as the random effects version or the fixed effect version of the inverse variance weighted regression[12](#_ENREF_12), MR-Egger[13](#_ENREF_13), median-based regression[14](#_ENREF_14), SMR[3](#_ENREF_3), or GSMR[4](#_ENREF_4), are only applicable to a single SNP instrument or multiple independent/near-independent SNP instruments. Handling only independent SNPs is restrictive, as most exposure variables/molecular traits are polygenic/omni-genic and are influenced by multiple SNPs that are in potential LD with each other. Indeed, incorporating multiple correlated SNPs often can help explain a greater proportion of variance in the exposure variable than using independent SNPs, and thus can help increase power and improve estimation accuracy of MR analysis[5](#_ENREF_5); [15-17](#_ENREF_15). Due to the benefits of multiple correlated instruments, both prediXcan and TWAS, rely on either ElasticNet[18](#_ENREF_18" \o "Zou, 2005 #150) or BSLMM[19](#_ENREF_19), incorporate all cis-SNPs that are in high LD for TWAS applications, which, as we will show below, leads to a substantial power improvement over standard MR approaches that use independent SNPs. Unfortunately, both prediXcan and TWAS rely on a two-stage MR inference procedure, estimating SNP effect sizes in the exposure study and plugging in these estimates to the outcome study for causal effect inference. Two-stage inference procedure in MR fails to account for the uncertainty in parameter estimates in the exposure study and can often lead to biased causal effect estimates and power loss especially in the presence of weak instruments[5](#_ENREF_5); [16](#_ENREF_16). Indeed, similar to what have been observed in the MR filed, our previous study also suggest that likelihood based inference can substantially improve power for TWAS[20](#_ENREF_20). Therefore, it is important to incorporate multiple correlated instruments in a likelihood inference framework for MR analysis in TWAS.

Second, perhaps more importantly, SNP instruments often exhibit pervasive horizontal pleiotropic effects[21](#_ENREF_21). Horizontal pleiotropy occurs when a genetic variant affects the outcome variable through pathways other than or in addition to the exposure variable[22](#_ENREF_22). Horizontal pleiotropy is in contrast with vertical pleiotropy, which characterizes instrument effects on the outcome variable through the path of the exposure. Horizontal pleiotropy is widely distributed across the genome, affects a wide spectrum of complex traits, and can be driven by LD and extreme polygenicity of traits[21](#_ENREF_21) [23](#_ENREF_23). Despite its wide prevalence, however, only a limited number of MR methods have been developed that can test and control for horizontal pleiotropy; even fewer are applicable for TWAS applications. For example, some existing methods (e.g. Cohran’s Q statistic[24](#_ENREF_24); [25](#_ENREF_25) and MR-PRESSO[21](#_ENREF_21)) test for horizontal pleiotropic effects without directly controlling for them. Some methods (e.g. CaMMEL[26](#_ENREF_26" \o "Park, 2017 #88)) control for horizontal pleiotropic effects without directly testing them[27](#_ENREF_27); [28](#_ENREF_28). Some methods (e.g. Egger regression[13](#_ENREF_13); [29](#_ENREF_29), GLIDE[30](#_ENREF_30), GSMR[4](#_ENREF_4), MR-median method[14](#_ENREF_14), profile score approach[31](#_ENREF_31), MRMix[32](#_ENREF_32" \o "Qi, 2018 #82) and Bayesian MR[33](#_ENREF_33); [34](#_ENREF_34)) test and control for horizontal pleiotropic effects, but can only accommodate independent or near-independent instruments. As far as we are aware, there is only one two-sample MR method currently developed for testing and controlling for pleiotropic effects in the presence of correlated instruments: LDA MR-Egger[35](#_ENREF_35). Unfortunately, as we will show below, LDA MR-Egger cannot handle realistic LD pattern among cis-SNPs for TWAS applications.

Here, we develop a generative two-sample MR method in a likelihood framework, which we refer to as the probabilistic two-sample Mendelian randomization (PMR), to perform MR analysis using multiple correlated instruments for TWAS applications. We illustrate how the PMR framework can facilitate the understanding of many existing MR approaches as well as many existing integrative analysis approaches. Within the PMR framework, we focus on a particular horizontal pleiotropy effect modeling assumption based on the burden test assumption commonly used for rare variant test, which, as we will show later, effectively generalizes the Egger regression assumption commonly used for MR analysis to correlated instruments. Our method allows us to test the causal effect in the presence of horizontal pleiotropy, and, with a parameter expansion version of the expectation maximization algorithm (PX-EM), is scalable to hundreds of thousands of individuals. We refer to our method as PMR-Egger. With simulations, we show that PMR-Egger provides calibrated type I error for causal effect testing in the presence of horizontal pleiotropic effects, is more powerful than existing MR approaches, and, as a by-product, can directly test for horizontal pleiotropy. We apply our method to perform TWAS for 39 diseases and complex traits obtained from three GWASs with sample size ranging from 4,686 to 337,199.

**Materials and Methods**

**PMR-Egger Overview**

We consider a probabilistic Mendelian randomization framework for performing two-sample Mendelian randomization analysis with correlated SNP instruments. Two-sample Mendelian randomization analysis aims to estimate and test for the causal effect of an explanatory variable on an outcome variable in the setting where the exposure and outcome variables are measured in two separate studies with no sample overlap. In the TWAS applications we consider here, the explanatory variable is gene expression level that is measured in a gene expression study, while the outcome variable is a quantitative trait or a dichotomous disease status that is measured in a GWAS. Often times, in such TWAS setting, the gene expression study and GWAS are performed on two separate samples. While we mostly focus on TWAS applications in the present study, we note that two-sample Mendelian randomization is also commonly performed in settings where both the explanatory and outcome variables are complex traits that are measured in two separate GWASs. An illustrative diagram of MR analysis is displayed in Figure S1.

We denote ***x*** as an -vector of exposure variable (i.e. gene expression measurements) that is measured on individuals in the gene expression study and denote as an by *p* matrix of genotypes for *p* selected instruments (i.e. cis-SNPs) in the same study. We denote **y** as an -vector of outcome variable (i.e. trait) that is measured on individuals in the GWAS and denote as an by *p* matrix of genotypes for the same *p* selected instruments there. We consider three linear regressions to model the two studies separately

where the equation (1) is for the gene expression data and the equations (2)-(3) are for the GWAS data. Here, and are the intercepts; is an unobserved -vector of exposure variable on the individuals in the GWAS; is a *p*-vector of instrumental effect sizes on the explanatory variable; is a scalar that represents the causal effect of the exposure variable on the outcome variable; is a *p*-vector of horizontal pleiotropic effect sizes of *p* instruments on the outcome variable; is an -vector of residual error with each element independently and identically distributed from a normal distribution ; is an -vector of residual error with each element independently and identically distributed from the same normal distribution ; and is an -vector of residual error with each element independently and identically distributed from a normal distribution . We note that while the above three equations are specified based on two separate studies, they are joined together with the common parameter and the unobserved gene expression measurements. Note that equations (2)-(3) can be combined into

where .

Our key parameter of interest in the above joint model is the causal effect . The causal interpretation of requires two assumptions of MR analysis to hold: (i) instruments are associated with the exposure; (ii) instruments are not associated with any other confounders that may be associated with both exposure and outcome. Note that our model no longer requires the exclusion restriction condition of traditional MR (i.e. instruments only influence the outcome by the path of exposure), as we directly model horizontal pleiotropy effects . With the above two assumptions of MR analysis, we derive the causal interpretation and identification of under the framework of decision-theoretic causal inference[33](#_ENREF_33); [36-38](#_ENREF_36) (details in Supplementary Text). Because the causal effect interpretation of depends on the two MR assumptions, at least one of which is often not testable in practice, MR analysis in observational studies likely provides weaker causality evidence than randomized clinical trials. Therefore, while we follow standard MR analysis and use the term “causal effect” through the text, we only intend to use this term to emphasize the fact that estimate from an MR analysis is more trustworthy than the effect size estimate in a standard linear regression of ***y*** on .

Because *p* is often larger than , we will need to make additional modeling assumptions on to make the model identifiable. In addition, the two instrumental effect terms defined in equation (4), the vertical pleiotropic effect and the horizontal pleiotropic effect , are also not identifiable from each other, unless we make additional modeling assumptions on . Here, we follow standard polygenic model and assume that all elements in are non-zero and that each follows a normal distribution . In addition, we follow the burden test assumption commonly used for rare variant test and assume that for . With the burden test assumption on , in the special case where instruments are independent and treated as fixed effects and where a two-stage estimation procedure is used for inference, our model reduces to the commonly used MR-Egger regression model. Therefore, our model effectively generalizes the Egger regression model to a likelihood inference framework with correlated instruments.

We are interested in estimating the causal effect and testing the null hypothesis in the presence of horizontal pleiotropy . In addition, we are interested in estimating the horizontal pleiotropic effect size and testing the null hypothesis . We accomplish both tasks through the maximum likelihood inference framework. In particular, we develop an expectation maximization (EM) algorithm for parameter inference by maximizing the joint likelihood defined based on equations (1) and (4) (details in the Supplementary Text). The EM algorithm allows us to obtain the maximum likelihood of the joint model, together with maximum likelihood estimates for both and . In addition, we apply the EM algorithm to two reduced models, one without and the other without , to obtain the corresponding maximum likelihoods. Afterwards, we perform likelihood ratio tests for either or , by contrasting the maximum likelihood obtained from the joint model to that obtained from each of the two reduced models, respectively. We refer to the above inference procedure as probabilistic, as we place estimation and testing into a maximum likelihood framework. Our inference procedure is in contrast to the commonly used two-stage estimation procedure (as used in, for example, Egger regression[13](#_ENREF_13); [29](#_ENREF_29), PrediXcan[1](#_ENREF_1" \o "Gamazon, 2015 #57) and TWAS[2](#_ENREF_2)), which estimates from equation (1) first and then directly plug in the estimates into equation (4) for inference. The previous two-stage estimation procedure fails to properly account for the estimation uncertainty in and is known to lose power compared to a formal likelihood inference procedure[5](#_ENREF_5); [16](#_ENREF_16); [20](#_ENREF_20).

We refer to our model and algorithm together as the two-sample probabilistic Mendelian randomization with Egger regression (PMR-Egger). As explained above, we use “probabilistic” to refer to both the data generative model and the maximum likelihood inference procedure. We use “Egger” to refer to the horizontal pleiotropic assumption on that effectively generalizes the Egger-regression assumption to correlated instruments. We also note that the joint generative Mendelian randomization model defined in equations (1) and (4) is a useful conceptual framework that unifies many existing MR methods. In particular, almost all existing MR methods are built upon on the joint model, but with different modeling assumptions on and ,and with different inference procedures (Table 1). Compared with these existing MR approaches, PMR-Egger is capable of modeling multiple correlated instruments, effectively controls for horizontal pleiotropy, and places inference into a likelihood framework.

**Simulations**

We performed simulations to assess the performance of PMR-Egger and compare it with existing approaches. To do so, we first obtained 556 cis-SNPs for the gene *BACE1* on chromosome 11 from the GEUVADIS data[39](#_ENREF_39) (data processing details in the next section) and simulated gene expression values. We used the gene *BACE1* because the number of cis-SNPs in this gene represents the median of that across all genes. With the scaled genotype data , we simulated SNP effect sizes from a normal distribution and summed the genetic effects across all cis-SNPs as . We also simulated residual errors from a normal distribution . The variances of the two normal distributions are chosen so that in expectation, the proportion of phenotypic variance explained by genetic effects is .

Next, we obtained genotypes for the same 556 SNPs from 2,000 randomly selected control individuals in the Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging (GERA)[40](#_ENREF_40); [41](#_ENREF_41) and simulated a quantitative trait. Here, we directly used from the gene expression data, which, when paired with causal effect , yielded the vertical pleiotropic effects . We set , and we simulated residual errors from a normal distribution . These parameters were selected so that in expectation, in the absence of horizontal pleiotropic effects, the proportion of phenotypic variance explained by vertical pleiotropic effects is . Afterwards, we simulated horizontal pleiotropic effects for these SNPs (more details below). We summed the horizontal pleiotropic effects, vertical pleiotropic effects and residual errors to yield the simulated trait.

In the simulations, we first examined a baseline simulation setting where we set , , with all . On top of the baseline setting, we varied one parameter at a time to examine the influence of various parameters. For , we set it to be either 1%, 5% or , close to the median gene expression heritability estimates across genes[42](#_ENREF_42); [43](#_ENREF_43). For , we examined alternative SNP effect size distributions that deviate from the polygenic assumption in the baseline setting. Specifically, we randomly selected either 1 SNP, 1%, 10% or 100% of the SNPs to have non-zero effect, while simulated their effects from a normal distribution to explain a fixed . For , we varied its value to be either 0% (for null simulations), 0.2%, 0.4% or 0.6% (for power simulations). For , we randomly assigned a fixed proportion of to be non-zero (proportion equals 10%, 30%, 50%, or 100%). Afterwards, we set the absolute value of non-zero to be the same value of , but randomly assign some of their signs to be positive and some of their signs to be negative, with the ratio of positive effects to negative effects being either 1:9, 3:7, or 5:5. Here, we set to be , , or , which corresponds to the 50%, 70%, 90%, 95% quantiles of horizontal pleiotropic effect estimates across all genes and all traits in the WTCCC data (more details below), respectively.

For null simulations and type I error control examination, we performed 10,000 simulation replicates for each simulation scenario described above. For power calculation, for each scenario, we performed 1,000 alternative simulations together with 9,000 null simulations and calculated power based on false discovery rate (FDR).

**Real Data Applications**

We applied our method to perform TWAS by integrating gene expression data with several GWASs. Specifically, we obtained GEUVADIS data[39](#_ENREF_39) as the gene expression data and examined 39 phenotypes from three GWASs. The three GWASs include the Wellcome trust case control study (WTCCC)[44](#_ENREF_44), the Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging (GERA)[40](#_ENREF_40); [41](#_ENREF_41), and the UK Biobank[45](#_ENREF_45).

The GEUVADIS data[39](#_ENREF_39) 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 [46](#_ENREF_46); [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)46; [47](#_ENREF_47)(release 12). 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[48](#_ENREF_48); [49](#_ENREF_49). Afterwards, following[49](#_ENREF_49), to remove 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 expression data, all individuals in GEUVADIS also have their genotypes sequenced in the 1,000 Genomes Project. We obtained genotype data from the 1,000 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 WTCCC data consists of about 14,000 cases from seven common diseases and 2,938 shared controls[44](#_ENREF_44). The diseases include type 1 diabetes (T1D; *n*=1,963), Crohn’s disease (CD; *n*=1,748), rheumatoid arthritis (RA; *n*=1,861), bipolar disorder (BD; *n*=1,868), type 2 diabetes (T2D; *n*=1,924), coronary artery disease (CAD; *n*=1,926), and hypertension (HT; *n*=1,952). We obtained quality controlled genotypes from WTCCC and initially imputed missing genotypes using BIMBAM[50](#_ENREF_50) to arrive at a total of 458,868 SNPs shared across all individuals. Afterwards, we further imputed SNPs using the 1,000 Genomes as the reference panel with SHAPEIT and IMPUTE2[51](#_ENREF_51). We filtered out SNPs that have an HWE p-value < 10-4, a genotype call rate <95%, or an MAF<0.01 to obtain a total of 2,793,818 imputed SNPs.

The GERA study consists of 61,953 individuals and 675,367genotyped SNPs. We filtered out SNPs that had a genotype calling rate below 0.95, MAF<0.01, or HWE p value<10-4 to yield a total of 487,609 SNPs. We then phased genotypes using SHAPEIT[52](#_ENREF_52) and imputed SNPs based on the Haplotype Reference Consortium (HRC version r1.1) reference panel[53](#_ENREF_53) on the Michigan Imputation Server using Minimac3[54](#_ENREF_54). Afterwards, we further filtered out SNPs that have a HWE p-value < 10-4, a genotype call rate <95%, an MAF<0.01, or an imputation score<0.30 to arrive at a total of 8,385,867 SNPs shared across 61,953 individuals. We examined 22 diseases in GERA that include Asthma (number of cases n=10,101), Allergic Rhinitis (n=15,193), Cardiovascular Disease (CARD, n=16,431), Cancers (n=18,714), Depressive Disorder (n=7,900), Dermatophytosis (n=8,443), Type 2 Diabetes (T2D, n=7,638), Dyslipidemia (n=33,071), Hypertension (HT, n=31,044), Hemorrhoids (n=9,922), Abdominal Hernia (n=6,876), Insomnia (n=4,357), Iron Deficiency (n=2,706), Irritable Bowel Syndrome (n=3,367), Macular Degeneration (n=4,031), Osteoarthritis (n=22,062), Osteoporosis (n=5,909), Peripheral Vascular Disease (PVD, n=4,718), Peptic Ulcer (n=1,007), Psychiatric disorders (n=9408), Stress Disorders (n=4,706), and Varicose Veins (n=2,714).

The UK Biobank data consists of 487,409 individuals and 92,693,895 imputed SNPs[45](#_ENREF_45). We followed the same sample QC procedure in Neale lab (<https://github.com/Nealelab/UK_Biobank_GWAS/tree/master/imputed-v2-gwas>) to retain a total of 337,199 individuals of European ancestry. In addition, 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. We selected 10 UK Biobank quantitative traits that have a phenotyping rate > 80%, a SNP-heritability > 0.2 and a low correlation among them following a previous study[55](#_ENREF_55). The 10 traits include Height (;), Platelet count (), Bone mineral density (), Red blood cell count (), FEV1-FVC ratio (), Body mass index (BMI, ), RBC distribution width (), Eosinophils count (), Forced vital capacity (), White blood cell count (). For each trait in turn, we standardized the phenotype to have a mean of zero and a standard deviation of one, regressed the resulting standardized phenotypes on sex and top 10 genotype principal components (PCs) to obtain the residuals, and finally used the scaled residuals to conduct TWAS analysis.

In the GEUVADIS data, for each gene in turn, 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 each of the three GWASs to obtain common sets of SNPs. The median number of the overlapped cis-SNPs between GEUVADIS and WTCCC, GERA or UK Biobank are 200, 556 or 500, respectively. Afterwards, for each pair of gene (from GEUVADIS) and trait (from GWAS) in turn, we examined the causal relationship between gene expression and trait of interest while testing and controlling for potential horizontal pleiotropic effects.

**Compared Methods**

For testing the causal effect, we compared the performance of PMR-Egger with five existing methods that include: (1) SMR that uses a single instrument but does not control for horizontal pleiotropy; here, we first performed a linear regression to choose the top associated cis-SNP to be the instrumental variable; (2) PrediXcan that uses multiple correlated instruments but does not control for horizontal pleiotropy; we rely on the R package glmnet to obtain the coefficient estimates of the cis-SNPs; (3) TWAS that uses multiple correlated instruments but does not control for horizontal pleiotropy; we rely on the BSLMM[19](#_ENREF_19) implemented in the GEMMA software[56](#_ENREF_56) to obtain coefficient estimates of the cis-SNPs; (4) CoMM that uses multiple correlated instruments but does not control for horizontal pleiotropy; CoMM is implemented in the R package CoMM; (5) LDA MR-Egger that uses multiple correlated instruments and controls for horizontal pleiotropy; we contacted the authors to obtain the method source code. All these methods are suitable for two-sample design and yield *p* values for testing the causal effect . Note that PrediXcan, TWAS and CoMM are not originally described as an MR method but conceptually rely on the same joint MR model based on equations (1) and (4). These three methods differ in their prior assumptions on : PrediXcan relies on Elastic Net assumption; TWAS relies on BSLMM[19](#_ENREF_19) assumption; while CoMM relies on the normal prior assumption. In addition, PrediXcan and TWAS rely on a two-stage regression procedure while CoMM is based on maximum likelihood. In addition, SMR is only able to use a single instrument while all other methods are capable of using multiple correlated instruments. We did not compare with GSMR, which requires multiple near independent instruments that are generally not feasible to obtain in TWAS applications. Finally, we note that we are unable to compare our method directly with the standard Egger regression, because the later requires multiple independent SNP instruments that are not readily available for TWAS.

For testing horizontal pleiotropic effect, we compared the performance of PMR-Egger with two existing methods that include (1) LDA MR-Egger; and (2) the global test in MR-PRESSO, which is implemented as an R package. Both these methods are capable of producing *p* values to test horizontal pleiotropic effect for one gene at a time.

**Software Availability**

PMR-Egger is implemented in the R package PMR, freely available on GitHub (<https://github.com/yuanzhongshang/PMR>). The code to reproduce all the analyses presented in the paper are available on GitHub (https://github.com/yuanzhongshang/PMRreproduce).

**Results**

Our method is described in the Materials and Methods, with technical details provided in the Supplementary Text. For TWAS applications, our method examines one gene at a time and estimates and tests its causal effect on a trait of interest. Our method models multiple correlated instruments, performs MR inference in a maximum likelihood inference framework, and is capable of testing and controlling for horizontal pleiotropic effects commonly encountered in TWAS. We refer to our method as the probabilistic Mendelian randomization with Egger regression (PMR-Egger), which is implemented in the R package PMR. Our method is computationally efficient and can analyze one gene with a few hundred thousand individuals in minutes (Table 2).

**Simulations: Testing and estimating causal effect**

We performed simulations to examine the effectiveness of our method and compared it with existing MR approaches. Simulation details are provided in the Materials and Methods. Briefly, we simulated gene expression values based on genotypes from 456 individuals in GEUVADIS and simulated phenotypes based on genotypes from 2,000 randomly selected individuals in GERA. In the simulations, we varied the genetic architecture underlying gene expression from sparse (one SNP or 1% of SNPs are causal) to polygenic (10% or 100% of SNPs are causal). We varied the proportion of SNPs exhibiting horizontal pleiotropic effects in a wide range (from 0%, 10%, 30%, 50% to 100%). We examined directional pleiotropy setting (the ratio of SNPs with negative vs positive horizontal pleiotropic effects is 0:10), approximately directional pleiotropy setting (1:9 or 3:7) and balanced pleiotropy settings (5:5). We varied the magnitude of horizontal pleiotropic effects to be either 1x10-4, 5x10-4, 1x10-3, or 2x10-3, which corresponds to the 50%, 70%, 90%, 95% percentiles of the horizontal pleiotropic effect estimate in real data. We also varied the magnitude of causal effect to be either 0, 0.14, 0.2 or 0.245, which corresponds to a proportion of phenotypic variance explained by vertical pleiotropic effects ) as 0, 0.2%, 0.4% and 0.6% respectively.

Our first set of simulations is focused on causal effect testing. Here, we compared PMR-Egger with five different methods that include SMR, PrediXcan, TWAS, CoMM, and LDA MR-Egger. We first examined type I error control of different methods under the null (). In the absence of horizontal pleiotropic effects, PMR-Egger, together with PrediXcan, TWAS, and CoMM, all provides calibrated type I error (Figure 1A). SMR produces overly-conservative/deflated p-values, presumably because SMR requires the selected instrument being a true causal SNP with large effect size, which is not always guaranteed in practice. LDA MR-Egger produces inflated p-values, presumably because LDA MR-Egger makes a fixed effect assumption on , whichoften does not work well in the TWAS setting where the number of SNPs are in the same order as the number of samples in the gene expression study. Subsequently, the p-values from LDA MR-Egger are sensitive to the linkage disequilibrium (LD) structure among SNPs and are not well behaved in realistic LD settings (Figure S2).

In the presence of horizontal pleiotropic effects, PMR-Egger becomes the only method that produces calibrated (or slightly conservative) p-values (Figures 1B-1D). In contrast, the p-values from all other methods become inflated, and more so with increasingly large horizontal pleiotropic effect. For example, when is 5x10-4, the genomic control factors from PMR-Egger, SMR, PrediXcan, TWAS, CoMM, and LDA MR-Egger are 0.93, 1.30, 1.33, 1.33, 1.49 and 2.61 respectively. When is increased to 1x10-3, the genomic control factors from PMR-Egger, SMR, PrediXcan, TWAS, CoMM, and LDA MR-Egger become 0.93, 2.39, 2.27, 2.46, 4.03 and 2.57 respectively. The null p-value distributions from different methods remain largely similar regardless of the genetic architecture underlying gene expression being sparse or polygenic (Figure S3). Note that, here, the p-values from SMR become less deflated when there is a sparse set of SNPs affecting gene expression; however, such deflation is not completely abolished even when one SNP has non-zero effect on gene expression, presumably because we cannot always identify the true non-zero effect SNP through eQTL mapping and may supply a tagged SNP for SMR analysis. The p-value distribution pattern for different methods under the null does not change much with reduced the gene expression heritability value . When is either 5% or 1%, PMR-Egger still produces well-calibrated *p* values (Figure S4). Finally, like standard MR-Egger regression, our PMR-Egger makes a relatively strong assumption on the horizontal pleiotropic effect and assumes that all SNPs have the same horizontal pleiotropic effect. To examine the robustness of such assumption, besides the above settings where either 0% or 100% SNPs have horizontal pleiotropic effects, we varied the proportion of horizontal pleiotropic SNPs to be either 10%, 30%, 50%. We found that PMR-Egger remain calibrated regardless of the sparsity of horizontal pleiotropic SNPs (Figure S5). In addition, besides the above directional pleiotropy setting where the ratio of SNPs with negative vs 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). We found that p-values of PMR-Egger remains calibrated in either the approximately directional pleiotropy settings or the balanced setting when horizontal pleiotropic effect is small or moderate (1x10-4, 5x10-4, or 1x10-3; Figures S6A-S6C). However, when horizontal pleiotropic effect is large (2x10-3), as one would expect, the p-values from PMR-Egger becomes inflated, with genomic control factor being 1.08, 1.31 and 1.37, for settings where the ratio is 1:9, 3:7 and 5:5, respectively (Figure S6D).

Next, we examined the power of different methods to identify the causal effect for a range of possible causal effect sizes . Because the same p-value from different methods may correspond to different type I errors, we computed power based on FDR of 0.1 instead of a nominal p-value threshold to allow for fair comparison across methods. In the absence of horizontal pleiotropic effects or in the presence of small horizontal pleiotropic effects, PMR-Egger, TWAS and CoMM have similarly power, all outperforming the other three methods (Figures 2A-2B). The power of PMR-Egger is slightly lower than the other two, presumably because PMR-Egger uses extra parameters to model horizontal pleiotropy and thus lose degree of freedom and subsequently power in the absence of horizontal pleiotropy. The power of all methods increases with , though their relative performance rank does not change. In the presence of horizontal pleiotropy, the power of all methods reduces (Figures 2C-2D). However, the power reduction from PMR-Egger is substantially smaller than all other methods. For example, when and , PMR-Egger reaches a power of 41%; the power of SMR, PrediXcan, TWAS, CoMM, and LDA MR-Egger are 7%, 24%, 31%, 33% and 1%, respectively. When but , the power of PMR-Egger remains similar and is 40%; the power of SMR, PrediXcan, TWAS, CoMM, and LDAMR-Egger reduces to 3%, 13%, 16%, 16% and 0.9%, respectively. Besides horizontal pleiotropic effects , we examined how power is influenced by the genetic architecture underlying gene expression, (Figure S7). We found that the power of different methods in the setting where 10% of SNPs have non-zero effects on gene expression are similar to the baseline setting where all SNPs have non-zero effects, both in the absence (Figure S7E vs Figure 2A) or in the presence of horizontal pleiotropic effects (Figure S7F vs Figure 2D). However, the relative performance of different methods changes when there is only one SNP or 1% SNPs having non-zero effect on gene expression. Specifically, in the absence of horizontal pleiotropic effects, the power of both PrediXcan and SMR become slightly higher than PMR-Egger, TWAS and CoMM, all of which have substantially higher power than LDA MR-Egger (Figures S7A, S7C). The higher power of PrediXcan and SMR in the sparse setting presumably is because the ElasticNet assumption employed in PrediXcan favors sparse eQTLs while SMR explicitly makes a single eQTL assumption. In the presence of horizontal pleiotropic effects, however, PMR-Egger remains the most powerful, even in the setting where only one SNP has non-zero effect on gene expression (Figures S7B, S7D). Finally, we found that PMR-Egger produces accurate estimate of the causal effect , both under the null and under various alternatives, in the presence or absence of horizontal pleiotropic effects (Figure S8).

**Simulations: Testing and estimating horizontal pleiotropic effect**

Our second set of simulations is focused on horizontal pleiotropic effect testing. Here, we compared PMR-Egger with two different methods: LDA MR-Egger and MR-PRESSO. All three methods examine one gene at a time and test whether cis-SNPs within the gene exhibit non-zero horizontal pleiotropic effects. Note that, unlike PMR-Egger and LDA MR-Egger, MR-PRESSO requires independent instruments and uses permutation to obtain the empirical p-values. Due to the heavy computational burden resulting from permutations, we are unable to apply MR-PRESSO to all simulation scenarios and also have to restrict the number of permutations to 10,000 for each explored simulation scenario. Subsequently, the lowest possible p value from MR-PRESSO is . We first examined type I error control of different methods under the null, where there is no horizontal pleiotropic effect. We found that the p-values from PMR-Egger provide calibrated type I error control under a range of causal effect sizes (Figure 3). However, p-values from both LDA MR-Egger and MR-PRESSO are inflated, and more so with increasingly large causal effect . For example, when , the genomic control factor from PMR-Egger and LDA MR-Egger are 0.96 and 2.31, respectively. When is increased to 0.6%, the genomic control factor from PMR-Egger remains 0.96, while the genomic control factors from LDA MR-Egger become 3.04. The overly inflated p-values from LDA MR-Egger is presumably due to its fixed effect modeling assumption on and the subsequent failure to control for realistic LD pattern. The inflation of MR-PRESSO p values is presumably because MR-PRESSO can only handle independent instruments and thus does not fare well in TWAS settings. Inflation of p-value on testing horizontal pleiotropy would incorrectly identify genes with no pleiotropic effects, thus likely reducing the power to detect true causal effect . Importantly, the p-values from PMR-Egger remain calibrated regardless of the genetic architecture underlying gene expression (Figure S9).

Next, we examined the power of different methods in detecting non-zero horizontal pleiotropic effect. Again, we computed power based on a FDR of 0.1 instead of the nominal p-value to allow for fair comparison across methods. We dropped MR-PRESSO for comparison here due to its heavy computational burden. We found that PMR-Egger outperforms LDA MR-Egger in a range of possible horizontal pleiotropic effect sizes, and that the power of both methods increases with increasing horizontal pleiotropy (Figure 2E-2F). For example, when and , PMR-Egger achieves a power of 1.6% while LDA MR-Egger achieves a power of 1% (note that both power are relatively small due to the small sample size used in the simulations). When but , the power of PMR-Egger increases to 58.9% while the power of LDA MR-Egger increases to 32%. In addition, the power to detect horizontal pleiotropic effects is not influenced by the sparsity level of the genetic architecture underlying gene expression (Figure S10). The power to detect horizontal pleiotropic effects does, however, depend on the sparsity level of (Figure S11A). Specifically, power of both PMR-Egger and LDA MR-Egger reduces with increasing sparsity of , though the power of PMR-Egger remains higher than LDA MR-Egger across a range of sparsity values. Similar trend can be found under various directional horizontal pleiotropic effect assumptions (Figure S11B). Finally, we note that, in the presence of directional pleiotropic effect, PMR-Egger can also estimate the horizontal pleiotropic effect size accurately (Figure S12).

**Real data applications**

We performed TWAS to detect genes causally associated with any of the 39 phenotypes collected from three GWASs (details in Materials and Methods). The examined gene expression data is obtained from the GEUVADIS study and contains 15,810 genes. The examined phenotypes include 7 common diseases from WTCCC, 22 diseases from GERA, and 10 quantitative traits from UK Biobank. The GWAS sample size ranges from 4,686 (for Crohn’s disease in WTCCC) to 337,199 (for UK Biobank). We applied PMR-Egger together with five other approaches (SMR, PrediXcan, TWAS, CoMM, and LDA MR-Egger) to examine pairs of gene and phenotype one at a time. The p-values for testing the causal effect of each gene on the phenotype are shown for WTCCC traits (Figures 4A, 4B and S13), GERA traits (Figures 5A, 5B and S14), and UK Biobank traits (Figures 6A, 6B and S15); with genomic control factors listed in Table S1 and visualized in Figures 4C, 5C and 6C. We also display qq-plots for two selected traits in each data, one with a relatively low number of gene associations and the other with a relatively high number gene associations, in Figures 4A, 4B, 5A, 5B, 6A and 6B, respectively. Among the selected six traits, the one with zero number of associated genes (BD in WTCCC; Figure 4A) and the one with one associated gene (Irritable Bowel Syndrome in GERA; Figure 5A), represent approximately null traits with no apparently associated genes. For the six selected traits, consistent with the simulations, we found that the p-values from PMR-Egger are well calibrated, at least more so than the other methods. In contrast, the p-values from CoMM, TWAS, PrediXcan and LDA MR-Egger are inflated and deviated upward from the diagonal line, while the p-values from SMR are overly conservative and lie below the diagonal line. The results observed in these exemplary traits generalize to all other examined traits. For example, the genomic control factor from PMR-Egger is the lowest among all methods in 25 out of the 39 traits, and ranges from 0.92 to 1.04 in WTCCC (Figure 4C), from 0.92 to 1.18 in GERA (Figure 5C), and from 1.12 to 1.34 in UK Biobank (Figure 6C). Note that the high genomic control factor in the large UK Biobank is expected under polygenic architecture[57](#_ENREF_57). In contrast, the genomic control factors from CoMM, TWAS, PrediXcan are often higher than that from PMR-Egger for most traits examined. For example, the genomic control factor from CoMM is often the highest among all other methods (except for LDA MR-Egger) in 22 out of the 39 traits, and ranges from 1.12 to 1.26 in WTCCC, 0.96 to 1.92 in GERA, and 1.45 to 1.90 in UK Biobank. The genomic control factor from TWAS is the highest among all other methods (except for LDA MR-Egger) in 14 out of the 39 traits, ranges from 1.17 to 1.33 in WTCCC, 1.01 to 1.45 in GERA, and 1.30 to 2.17 in UK Biobank. The genomic control factor from PrediXcan is the highest among all other methods (except for LDA MR-Egger) in 5 out of the 39 traits, and ranges from 1.21 to 1.32 in WTCCC, 1.02 to 1.38 in GERA, and 1.09 to 1.46 in UK Biobank. In addition, consistent with simulations, we observed a substantial inflation of LDA MR-Egger p-values: its genomic control factor ranges from 7.10 to 10.48 in WTCCC, 32.15 to 35.24 in GERA, and 10.48 to 16.65 in UK Biobank. Also consistent with simulations, the p-values from SMR often lies underneath the expected null, even though its genomic control factors are often well behaved (Figures 4A,4B,5A,5B,6A,6B and S13-S15).

We examined the number of associated genes detected by different methods based on a Bonferroni corrected genome-wide threshold (Figures 4D, 5D and 6D; Table S2). For this analysis, we excluded LDA MR-Egger for comparison, as its p-values are overly inflated. Comparing across the remaining methods, we found that SMR can barely detect any genes significantly associated with traits across all three data sets, much less than that detected by the other four methods. The much lower number of genes detected by SMR than the other four methods are consistent with the relatively low power of SMR observed in simulations. For the other four methods, we found that the number of gene-trait pairs detected by CoMM and PMR-Egger is higher than that detected by TWAS and PrediXcan in all three GWASs (Figures 4D, 5D and 6D; Table S2). The higher number of discoveries by both CoMM and PMR-Egger in the three GWASs is consistent with our simulations as well as previous observations that likelihood-based inference often achieves higher power than two-stage inference for MR analysis. However, we do notice that PMR-Egger detects slightly lower number of gene-trait pairs than CoMM based on the same genome-wide p-value threshold, consistent with the inflated genomic inflation factors observed for CoMM. Indeed, we found that the estimated for the common set of genes detected by both CoMM and PMR-Egger is higher than the set of genes only detected by CoMM across traits (Figure S16A and S16B). Therefore, the genes detected by CoMM but not PMR-Egger tend to have large and small , likely reflecting false associations due to horizontal pleiotropic confounding.

Overall, by controlling for horizontal pleiotropic effects, PMR-Egger detected many likely causal genes that the other methods failed to detect. As the first example, the *C2* gene (31,765,562-32,013,449 on chr 6) is only identified by PMR-Egger to be associated with T1D (), but not by the other methods (CoMM *p*=0.97; TWAS ; PrediXcan ; SMR *p*=0.082). *C2* gene belongs to the classical pathway of the complement immune system, is located within the class III MHC region, and has been confirmed by several studies to be associated with autoimmune diseases[58](#_ENREF_58); [59](#_ENREF_59). As the second example, the *HLA-F* gene (29,590,552-29,806,305 on chr 6), which is a HLA class I heavy chain paralogue, is only identified by PMR-Egger to be associated with RA (), but not by the other methods (CoMM ; TWAS *p*=0.03; PrediXcan *p*=0.02; SMR *p*=0.08). As the third example, the *HERPUD1* gene (56,865,960-57,077,798 on chr 16), which is a well-known lipid-related gene[60](#_ENREF_60); [61](#_ENREF_61), is only detected by PMR-Egger to be associated with dyslipidemia (), but not by the other methods (CoMM ; TWAS ; PrediXcan ; SMR *p*=0.10). As the fourth example, the *ADAM15* gene (154,923,042-155,135,252 on chr 1), which has been confirmed to be associated with height in the UK Biobank Pheweb, is only detected by PMR-Egger to be associated with height (), but not by the other methods (CoMM ; TWAS *p*=0.17; PrediXcan *p*=0.44; SMR *p*=0.14). Finally, the *DNAJC27-AS1* gene (25,094,259-25,362,563 on chr 2) is only identified by PMR-Egger to be associated with height (), but not by the other methods (CoMM *p*=0.99; TWAS ; PrediXcan *p*=0.02; SMR *p*=0.05). *DNAJC27-AS1* is a *DNAJC27* antisense RNA gene and can inhibit the expression of *DNAJC27*, which has been confirmed to be associated with height in the UK Biobank Pheweb and other large association studies[62](#_ENREF_62).

Next, we shift our focus to testing horizontal pleiotropic effects. The p-values for testing the causal effect of gene on phenotype are shown for WTCCC traits (Figures 4E, 4F and S13), GERA traits (Figures 5E, 5F and S17), and UK Biobank traits (Figures 6E, 6F and S18); with genomic control factors visualized in Figures 4G, 5G and 6G. We also display qq-plots for the previously selected exemplary traits in Figures 4E, 4F, 5E, 5F, 6E and 6F. Overall, consistent with simulations, the p-values from PMR-Egger are well behaved while the p-value from LDA MR-Egger display substantial inflation. For example, the genomic control factor from PMR-Egger ranges from 0.94 to 1.03 in WTCCC (Figure 4G), from 0.92 to 1.26 in GERA (Figure 5G), and from 1.13 to 1.71 in UK Biobank (Figure 6G). In contrast, the genomic control factor from LDA MR-Egger ranges from 10.16 to 15.64 in WTCCC, from 70.43 to 72.19 in GERA and from 17.75 to 29.85 in UK Biobank (Table S1). With the same Bonferroni adjusted genome-wide p-value threshold, PMR-Egger detected 33 gene-trait pairs in WTCCC in which the cis-SNPs exhibit significant horizontal pleiotropy, 37 gene-trait pairs in GERA, and 626 gene-trait pairs in the UK Biobank.

Horizontal pleiotropic effect tests can help us explain some of the discrepancy in terms of the causal associations detected by PMR-Egger and the other methods. For example, for T1D in WTCCC, the *ZKSCAN4* gene on chromosome 6 shows a significant pleiotropy effect () but displays no significant causal effect (*p*=0.99) by PMR-Egger. In contrast, *ZKSCAN4* is detected to be significantly associated with T1D by PrediXcan (), and, to a lesser extent, by TWAS (). However, *ZKSCAN4* has not been reportedto be associated with T1D by previous GWASs. A careful examination of the *ZKSCAN4* locus (28,112,401-28,327,011) revealed a neighboring MHC region (28,477,797-33,448,354), a region previously well recognized to affect T1D and other autoimmune diseases. Indeed, by controlling for the predicted expression level of the *APOM* gene (31,520,193-31,725,987) in the MHC region in the PrediXcan framework, the association between the predicted *ZKSCAN4* expression level and T1D is no longer significant (*p* = 0.38). Therefore, the causal association between *ZKSCAN4* and T1D detected by PrediXcan and TWAS likely reflect the horizontal pleiotropic effect of *ZKSCAN4* cis-SNPs on T1D through genes in the MHC region. As another example, for the trait of red blood cell count in UK Biobank, the *MAPT* gene on chromosome 17 shows a significant pleiotropy effect () but displays no significant causal effect (*p*=0.98) by PMR-Egger. In contrast, *MAPT* is detected to be significantly associated with red blood cell count by PrediXcan (), and, to a lesser extent, by TWAS (). However, no previous evidence suggests that *MAPT* is associated with red blood cell count. Indeed, we found that the genomic location of *MAPT* (43,871,748-44,205,700) is close to and partially overlapped with *KANSL1* (44,007,282-44,402,733), which has been previously identified to be associated with red blood cell traits[61](#_ENREF_61); [63](#_ENREF_63). Such association between *KANSL1* and red blood cell count is also detected by PMR-Egger (), by CoMM (), and, to a lesser extent, by TWAS () in the present study. By controlling for the predicted expression level of the *KANSL1* gene in the PrediXcan framework, the association between the predicted *MAPT* expression level and red blood cell count is no longer significant (*p* = 0.10). Therefore, the causal association between *MAPT* and red blood cell count detected by PrediXcan likely reflects the horizontal pleiotropic effect of *MAPT* cis-SNPs on red blood cell count through *KANSL1*. As a final example, for height in the UK Biobank, the pseudogene *RP11-9E13.2* (70,137,755-70,340,521) on chromosome 10 has a significant pleiotropy effect () but displays no significant causal effect (*p*=0.93) by PMR-Egger. In contrast, *RP11-9E13.2* is detected to be significantly associated with height by PrediXcan (), and, to a lesser extent, by TWAS (). The pseudogene *RP11-9E13.2* is in the neighborhood of *MYPN* (69,765,912-70,071,774), which has been previously identified to be associated with height[64](#_ENREF_64). The association between *MYPN* and height is also detected by PMR-Egger (), CoMM (), and to a lesser extent, PrediXcan () and TWAS (), in the present study. By controlling for the predicted expression level of *MYPN* gene in the PrediXcan framework, the association between the predicted *RP11-9E13.2* expression level and height is no longer significant at the genome-wide threshold (). Therefore, the causal association between the pseudogene *RP11-9E13.2* and height as detected by PrediXcan and TWAS likely reflects the horizontal pleiotropic effect of *RP11-9E13.2* cis-SNPs on height at least partially through *MYPN*. The results suggest the practical importance of testing and controlling for pleiotropic effects in TWAS applications.

Finally, we note that an important feature of PMR-Egger is its ability to test both causal effect and horizontal pleiotropy effect simultaneously. We contrast the p-values obtained from these two different tests across genes for those traits in which at least one gene is detected as significant from either of the two tests (Figures S19-S21). We found that different traits exhibit different gene association patterns. For example, some traits may only contain genes with a significant causal effect but without a significant horizontal pleiotropic effect (e.g. CD and CAD in WTCCC; Allergic Rhinitis, Irritable Bowel Syndrome and Psychiatric disorders in GERA). Some traits may only contain genes with a significant horizontal pleiotropic effect but without a significant causal effect (e.g. Dermatophytosis in GERA). Some traits may contain genes with a significant causal effect as well as genes with a significant horizontal pleiotropic effect, but with the two sets of genes being non-overlapped (e.g. Asthma, Dyslipidemia, HT, Abdominal Hernia and Macular Degeneration in GERA; Fored Vitral Capacity in UK Biobank). While the majority of traits contain genes with both a significant causal effect and a significant horizontal pleiotropic effect. The top gene which is most significant for both causal effect test and pleiotropy test is highlighted in the plots. Being capable of testing both causal effect and horizontal pleiotropy effect facilitates our understanding of the gene association pattern with various different complex traits.

**Discussion**

We have presented a data generative model and a likelihood framework for MR analysis that unifies many existing transcriptome wide association analysis methods and many existing MR methods. Under the framework, we have presented PMR-Egger, a new method that conducts MR analysis using multiple correlated instruments while properly controlling for horizontal pleiotropic effects. By properly controlling for horizontal pleiotropic effects and making inference under a likelihood framework, PMR-Egger yields calibrated p-values across a wide range of scenarios and improves power of MR analysis over existing approaches. We have illustrated the benefits of PMR-Egger through extensive simulations and multiple real data applications of TWAS.

One important modeling assumption we made in PMR-Egger is that the horizontal pleiotropic effects of all SNPs equal to each other. The equal effect size assumption directly follows the commonly used Egger regression modeling assumption for MR analysis and is analogous to the burden effect size assumption commonly used for rare variant tests. Consistent with existing literature on applications of the Egger regression and burden test, we also found that equal effect size assumption employed in PMR-Egger works reasonably robust with respect to a range of model mis-specifications and appears to be effective in several real data applications examined here. However, we do acknowledge that our equal effect size assumption in PMR-Egger can be overly restrictive in many settings. We have attempted to alleviate this restrictive modeling assumption by imposing an alternative modeling assumption on the horizontal effect sizes based on variance component assumption. In particular, we have attempted to assume that the horizontal pleiotropic effect of each SNP follows a normal distribution with mean zero and a certain variance component parameter, i.e. analogous to the SKAT test assumption[65](#_ENREF_65). Such variance component assumption is a more flexible modeling assumption than the equal effect size assumption, potentially alleviating much of the concern with respect to the sensitivity and robustness of equal effect size assumption. Unfortunately, under the variance component assumption, inference for the resulting PMR model becomes overly complicated. In particular, due to the estimation uncertainty in the hyper-parameter estimates, the p-values from the PMR variance component model becomes severely deflated even under simple null simulations (Figures S22). Such deflation of p-values has been previously observed in variance component tests for microbiome applications[66](#_ENREF_66). Only few methods exist to address such p-value in-calibration issue resulting from hyper-parameter estimation uncertainty[67](#_ENREF_67), and it is not straightforward to adapt any of these methods to our PMR variance component model. Besides the equal effect size modeling restriction, we also note that neither PMR-Egger nor the PMR variance component model is capable of accounting for correlation between horizontal pleiotropic effects and the SNP effects on gene expression . Therefore, while we view PMR-Egger as in important first step towards effective control of horizontal pleiotropic effects in TWAS applications, we emphasize that imposing more realistic modeling assumptions on the horizontal pleiotropic effects in the PMR framework will likely yield more fruitful results in the future.

We have primarily focused on modeling continue traits with PMR-Egger. For case control studies, we have followed previous approaches and directly treated binary phenotypes as continuous outcomes[19](#_ENREF_19); [49](#_ENREF_49); [68](#_ENREF_68); [69](#_ENREF_69), which appears to work well in both WTCCC and GERA data applications we examined. Treating binary phenotypes as continuous outcomes can be justified by recognizing the linear model as a first order Taylor approximation to a generalized linear model[19](#_ENREF_19). However, it would be desirable to extend PMR-Egger to accommodate case control data or other discrete data types in a principled way, by, for example, extending PMR-Egger into the generalized linear model framework. In particular, we could use a probit or a logistic link to extend PMR-Egger to directly model case control data. Extending PMR-Egger to model discrete data types using the generalized linear model framework would likely lead to wider applications of PMR-Egger and is thus an important avenue for future research.

We have primarily focused on modeling individual-level data with PMR-Egger. However, like many other linear model-based methods in statistical genetics, PMR-Egger can also be easily extended to make sure of summary statistics. The summary statistics version of PMR-Egger is described in detail in the Supplementary Text. Briefly, the summary statistics version of PMR-Egger requires marginal SNP effect size estimates and their standard errors, both on the gene expression and on the trait of interest. In addition, it requires a SNP by SNP correlation matrix that can be constructed based on a reference panel. Being able to make use of summary statistics extends the applicability of PMR-Egger to data sets where individual-level genotype or phenotype are not available.

Finally, in addition to what we have already mentioned in the Materials and Methods, we emphasize here again, that, while we have followed the previous MR literature and use “causal effect” through the text, the effect is causal only when certain MR modeling assumptions hold. These MR assumptions are often not straightforward to prove. For example, without measuring all potential confounders, it is not straightforward to argue that the SNP instruments are not associated with any other confounders that may be associated with both exposure and outcome. Therefore, we caution against the over-interpretation of causal inference in observation studies such as TWAS applications. However, we do believe MR is an important step that allows us to move beyond standard linear regressions and is an important analysis that can provide potentially more trustworthy evidence with regard to causality compared to simpler approaches.

**Acknowledgements**

This study was supported by the National Institutes of Health (NIH) grants R01HG009124 and R01HL142023, National Science Foundation (NSF) grant DMS1712933, and the National Natural Science Foundation of China (81872712 and 81673272). We thank the Wellcome Trust Centre for Human Genetics for making the heterogenous stock mouse data available online. This study makes use of data generated by the Wellcome Trust Case Control Consortium (WTCCC). A full list of the investigators who contributed to the generation of the data is available from http://www. wtccc.org.uk/. Funding for the WTCCC project was provided by the Wellcome Trust under award 076113 and 085475. The GERA Data (dbGaP accession number 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 in the following publication, Schaefer C, et al., The Kaiser Permanente Research Program on Genes, Environment and Health: Development of a Research Resource in a Multi-Ethnic Health Plan with Electronic Medical Records, in preparation, 2013. 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.

Reference

1. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquinomichaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., and Nicolae, D.L. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nature genetics 47, 1091-1098.

2. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., de Geus, E.J., Boomsma, D.I., and Wright, F.A. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nature genetics 48, 245-252.

3. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., and Visscher, P.M. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nature genetics 48, 481-487.

4. Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., Robinson, M.R., McGrath, J.J., Visscher, P.M., and Wray, N.R. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. Nature communications 9, 224.

5. Burgess, S., Small, D.S., and Thompson, S.G. (2017). A review of instrumental variable estimators for Mendelian randomization. Statistical methods in medical research 26, 2333-2355.

6. Ference, B.A., Robinson, J.G., Brook, R.D., Catapano, A.L., Chapman, M.J., Neff, D.R., Voros, S., Giugliano, R.P., Davey Smith, G., and Fazio, S. (2016). Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes. New England Journal of Medicine 375, 2144-2153.

7. Helgadottir, A., Gretarsdottir, S., Thorleifsson, G., Hjartarson, E., Sigurdsson, A., Magnusdottir, A., Jonasdottir, A., Kristjansson, H., Sulem, P., and Oddsson, A. (2016). Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. Nature genetics 48, 634-639.

8. Pingault, J.-B., O’Reilly, P.F., Schoeler, T., Ploubidis, G.B., Rijsdijk, F., and Dudbridge, F. (2018). Using genetic data to strengthen causal inference in observational research. Nature Reviews Genetics 19, 566-580.

9. Zheng, J., Baird, D., Borges, M.-C., Bowden, J., Hemani, G., Haycock, P., Evans, D.M., and Smith, G.D. (2017). Recent developments in Mendelian randomization studies. Current epidemiology reports 4, 330-345.

10. Haycock, P.C., Burgess, S., Wade, K.H., Bowden, J., Relton, C., and Davey Smith, G. (2016). Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. The American journal of clinical nutrition 103, 965-978.

11. Lawlor, D.A. (2016). Commentary: Two-sample Mendelian randomization: opportunities and challenges. International journal of epidemiology 45, 908-915.

12. Bowden, J., Del, F.G.M., Minelli, C., Davey, G.S., Sheehan, N., and Thompson, J. (2017). A framework for the investigation of pleiotropy in two‐sample summary data Mendelian randomization. Statistics in medicine 36, 1783-1802.

13. Bowden, J., Smith, G.D., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International journal of epidemiology 44, 512-525.

14. Bowden, J., Davey Smith, G., Haycock, P.C., and Burgess, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genetic epidemiology 40, 304-314.

15. Burgess, S., Butterworth, A., and Thompson, S.G. (2013). Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. Genetic epidemiology 37, 658-665.

16. Burgess, S., Dudbridge, F., and Thompson, S.G. (2016). Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. Statistics in medicine 35, 1880-1906.

17. Burgess, S., and Thompson, S.G. (2011). Bias in causal estimates from Mendelian randomization studies with weak instruments. Statistics in medicine 30, 1312-1323.

18. Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 301-320.

19. Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic Modeling with Bayesian Sparse Linear Mixed Models. Plos Genetics 9, : e1003264.

20. Yang, C., Wan, X., Lin, X., Chen, M., Zhou, X., and Liu, J. (2018). CoMM: a collaborative mixed model to dissecting genetic contributions to complex traits by leveraging regulatory information. Bioinformatics.

21. Verbanck, M., Chen, C.-Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nature genetics 50, 693-698.

22. Hemani, G., Bowden, J., and Davey Smith, G. (2018). Evaluating the potential role of pleiotropy in Mendelian randomization studies. Human molecular genetics 27, R195-R208.

23. Jordan, D.M., Verbanck, M., and Do, R. (2018). The landscape of pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. bioRxiv, 311332.

24. Bowden, J., Fabiola Del Greco, M., Minelli, C., Lawlor, D., Zhao, Q., Sheehan, N., Thompson, J., and Smith, G.D. (2018). Improving the accuracy of two-sample summary data Mendelian randomization: moving beyond the NOME assumption. International journal of epidemiology, 1-15.

25. Bowden, J., Hemani, G., and Davey Smith, G. (2018). Detecting individual and global horizontal pleiotropy in Mendelian randomization: a job for the humble heterogeneity statistic? American journal of epidemiology 187, 2681-2685.

26. Park, Y., Sarkar, A., He, L., Davila-Velderrain, J., De, P.J., and Kellis, M. (2017). A Bayesian approach to mediation analysis predicts 206 causal target genes in Alzheimer's disease. bioRxiv, 219428.

27. Kang, H., Zhang, A., Cai, T.T., and Small, D.S. (2016). Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. Journal of the American statistical Association 111, 132-144.

28. Guo, Z., Kang, H., Tony Cai, T., and Small, D.S. (2018). Confidence intervals for causal effects with invalid instruments by using two‐stage hard thresholding with voting. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80, 793-815.

29. Burgess, S., and Thompson, S.G. (2017). Interpreting findings from Mendelian randomization using the MR-Egger method. European journal of epidemiology 32, 391-392.

30. Dai, J.Y., Peters, U., Wang, X., Kocarnik, J., Chang-Claude, J., Slattery, M.L., Chan, A., Lemire, M., Berndt, S.I., and Casey, G. (2018). Diagnostics of Pleiotropy in Mendelian Randomization Studies: Global and Individual Tests for Direct Effects. American journal of epidemiology 187, 2672-2680.

31. Zhao, Q., Wang, J., Bowden, J., and Small, D.S. (2018). Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. arXiv:180109652.

32. Qi, G., and Chatterjee, N. (2018). Mendelian Randomization Analysis Using Mixture Models (MRMix) for Genetic Effect-Size-Distribution Leads to Robust Estimation of Causal Effects. bioRxiv, 367821.

33. Berzuini, C., Guo, H., Burgess, S., and Bernardinelli, L. (2018). A Bayesian approach to Mendelian randomization with multiple pleiotropic variants. Biostatistics, 1-16.

34. Li, S. (2017). Mendelian randomization when many instruments are invalid: hierarchical empirical Bayes estimation. arXiv:170601389.

35. 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. Genetic epidemiology 42, 418-433.

36. Dawid, A.P. (2000). Causal inference without counterfactuals. Journal of the American statistical Association 95, 407-424.

37. Dawid, A.P. (2015). Statistical causality from a decision-theoretic perspective. Annual Review of Statistics and Its Application 2, 273-303.

38. Berzuini, C., Dawid, P., and Bernardinell, L. (2012). Causality: Statistical perspectives and applications.(John Wiley & Sons).

39. Lappalainen, T., Sammeth, M., Friedländer, M.R., AC‘t Hoen, P., Monlong, J., Rivas, M.A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., and Ferreira, P.G. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501, 506-511.

40. Banda, Y., Kvale, M.N., Hoffmann, T.J., Hesselson, S.E., Ranatunga, D., Tang, H., Sabatti, C., Croen, L.A., Dispensa, B.P., and Henderson, M. (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.

41. Kvale, M.N., Hesselson, S., Hoffmann, T.J., Cao, Y., Chan, D., Connell, S., Croen, L.A., Dispensa, B.P., Eshragh, J., and Finn, A. (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.

42. 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 Genetics 4, e1000294.

43. 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 Genetics 7, e1001317.

44. Consortium, W.T.C.C. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661-678.

45. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., and O’Connell, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203-209.

46. Wen, X., Luca, F., and Pique-Regi, R. (2015). Cross-population joint analysis of eQTLs: fine mapping and functional annotation. Plos Genetics 11, e1005176.

47. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., and Searle, S. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome research 22, 1760-1774.

48. 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. Nature protocols 7, 500-507.

49. Zeng, P., and Zhou, X. (2017). Non-parametric genetic prediction of complex traits with latent Dirichlet process regression models. Nature communications 8, 456.

50. Guan, Y., and Stephens, M. (2008). Practical issues in imputation-based association mapping. Plos Genetics 4, e1000279.

51. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. Plos Genetics 5, e1000529.

52. Delaneau, O., Zagury, J.-F., and Marchini, J. (2012). Improved whole-chromosome phasing for disease and population genetic studies. Nature methods 10, 5-6.

53. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., and Sharp, K. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nature genetics 48, 1279-1283.

54. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., and McGue, M. (2016). Next-generation genotype imputation service and methods. Nature genetics 48, 1284-1287.

55. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A.P., and Price, A.L. (2018). Mixed-model association for biobank-scale datasets. Nature genetics 50, 906-908.

56. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nature genetics 44, 821-824.

57. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., and Visscher, P.M. (2018). Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. Hum Mol Genet 27, 3641-3649.

58. Yu, C.Y. (1998). Molecular genetics of the human MHC complement gene cluster. Experimental and clinical immunogenetics 15, 213-230.

59. Zhu, Z.B., Hsieh, S.-L., Bentley, D., Campbell, R., and Volanakis, J. (1992). A variable number of tandem repeats locus within the human complement C2 gene is associated with a retroposon derived from a human endogenous retrovirus. Journal of Experimental Medicine 175, 1783-1787.

60. Smith, E.N., Chen, W., Kähönen, M., Kettunen, J., Lehtimäki, T., Peltonen, L., Raitakari, O.T., Salem, R.M., Schork, N.J., and Shaw, M. (2010). Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. Plos Genetics 6, e1001094.

61. Kanai, M., Akiyama, M., Takahashi, A., Matoba, N., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., and Matsuda, K. (2018). Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nature genetics 50, 390-400.

62. Allen, H.L., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., and Raychaudhuri, S. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467, 832-838.

63. Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., and Kostadima, M.A. (2016). The allelic landscape of human blood cell trait variation and links to common complex disease. Cell 167, 1415-1429.

64. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J.a., and Kutalik, Z. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nature genetics 46, 1173-1186.

65. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. The American Journal of Human Genetics 89, 82-93.

66. Zhao, N., Chen, J., Carroll, I.M., Ringel-Kulka, T., Epstein, M.P., Zhou, H., Zhou, J.J., Ringel, Y., Li, H., and Wu, M.C. (2015). Testing in microbiome-profiling studies with MiRKAT, the microbiome regression-based kernel association test. The American Journal of Human Genetics 96, 797-807.

67. Chen, J., Chen, W., Zhao, N., Wu, M.C., and Schaid, D.J. (2016). Small sample kernel association tests for human genetic and microbiome association studies. Genetic epidemiology 40, 5-19.

68. Yang, J., Fritsche, L.G., Zhou, X., Abecasis, G., and Consortium, I.A.-R.M.D.G. (2017). A scalable Bayesian method for integrating functional information in genome-wide association studies. The American Journal of Human Genetics 101, 404-416.

69. Crawford, L., Zeng, P., Mukherjee, S., and Zhou, X. (2017). Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits. Plos Genetics 13, e1006869.