

# Statistical power of transcriptome-wide association studies

Ruoyu He<sup>1,2</sup> | Haoran Xue<sup>2</sup> | Wei Pan<sup>2</sup>  | for the Alzheimer's Disease Neuroimaging Initiative

<sup>1</sup>School of Statistics, University of Minnesota, Minneapolis, Minnesota, USA

<sup>2</sup>University of Minnesota, Division of Biostatistics, School of Public Health, Minneapolis, Minnesota, USA

## Correspondence

Wei Pan, 420 Delaware St SE, A460 Mayo Building, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA.

Email: [panxx014@umn.edu](mailto:panxx014@umn.edu)

## Funding information

National Institutes of Health, Grant/Award Numbers: R01AG065636, R01HL116720, RF1AG067924, U01AG073079

## Abstract

Transcriptome-Wide Association Studies (TWASs) have become increasingly popular in identifying genes (or other endophenotypes or exposures) associated with complex traits. In TWAS, one first builds a predictive model for gene expressions using an expression quantitative trait loci (eQTL) data set in stage 1, then tests the association between the predicted gene expression and a trait based on a large, independent genome-wide association study (GWAS) data set in stage 2. However, since the sample size of the eQTL data set is usually small and the coefficient of multiple determination (i.e.,  $R^2$ ) of the model for many genes is also small, a question of interest is to what extent these factors affect the statistical power of TWAS. In addition, in contrast to a standard (univariate) TWAS (UV-TWAS) considering only a single gene at a time, multivariate TWAS (MV-TWAS) methods have recently emerged to account for the effects of multiple genes, or a gene's nonlinear effects, simultaneously. With the absence of the power analysis for these MV-TWAS methods, it would be of interest to investigate whether one can gain or lose power by using the newly proposed MV-TWAS instead of UV-TWAS. In this paper, we first outline a general method for sample size/power calculations for two-sample TWAS, then use real data—the Alzheimer's Disease Neuroimaging Initiative (ADNI) expression quantitative trait loci (eQTL) data and the Genotype-Tissue Expression (GTEx) eQTL data for stage 1, the International Genomics of Alzheimer's Project Alzheimer's disease (AD) GWAS summary data and UK Biobank (UKB) individual-level data for stage 2—to empirically address these questions. Our most important conclusions are the following. First, a sample size of a few thousands (~8000) would suffice in stage 1, where the power of TWAS would be more determined by *cis*-heritability of gene expression. Second, as in the general

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: <http://adni.loni.usc.edu/wp-content/uploads/howtoapply/ADNIAcknowledgementListSep23.pdf>.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Genetic Epidemiology* published by Wiley Periodicals LLC.

case of simple regression versus multiple regression, the power of MV-TWAS may be higher or lower than that of UV-TWAS, depending on the specific relationships among the GWAS trait and multiple genes (or linear and nonlinear terms of the same gene's expression levels), such as their correlations and effect sizes. Interestingly, several top genes with large power gains in MV-TWAS (over that in UV-TWAS) were known to be (and in our data more significantly) associated with AD. We also reached similar conclusions in an application to the GTEx whole blood gene expression data and UKB GWAS data of high-density lipoprotein cholesterol. The proposed method and the conclusions are expected to be useful in planning and designing future TWAS and other related studies (e.g., Proteome- or Metabolome-Wide Association Studies) when determining the sample sizes for the two stages.

#### KEYWORDS

2SLS, Alzheimer's disease, causal inference, sample size, TWAS

## 1 | INTRODUCTION

Transcriptome-Wide Association Studies (TWASs) have been increasingly applied to identify genes associated with complex traits (Gusev et al., 2016). The statistical principle underlying TWAS is (two-sample) two-stage least squares (2SLS; Gamazon et al., 2015; Gusev et al., 2016). In stage 1, for a candidate gene, a regression model is often trained on a small eQTL data set to predict its expression level using its *cis*-nucleotide polymorphism (*cis*-SNPs). Then in stage 2, the pretrained model is used to predict the gene expression (of the candidate gene) using the SNPs from a much larger genome-wide association study (GWAS) data set that is independent of the eQTL data set in stage 1; then the predicted gene expression is tested for association with the trait of the GWAS data to determine whether there is an association between the gene and the trait. If an association is established, under the framework of instrumental variable (IV) regression and its related valid IV assumptions, the gene can be interpreted as (putative) causal, similar to Mendelian randomization (MR; Angrist et al., 1996; Xue & Pan, 2020). For stage 1, various methods, including stepwise variable selection coupled with ordinary least squares (OLS), lasso and elastic net penalized regression, and some Bayesian methods, have been applied to select *cis*-SNPs to be IVs while building a predictive model for gene expression (Gamazon et al., 2015; Gusev et al., 2016; Xue & Pan, 2020). In contrast, in stage 2 typically a simple regression or a univariate association test is conducted. It is also notable that GWAS summary data, not necessarily GWAS individual-level, can be applied in stage 2, largely facilitating the wide applicability of TWAS (Gusev et al., 2016).

Despite the success of TWAS in discovering important gene-trait associations (and putative causal genes for traits; Gamazon et al., 2015; Gusev et al., 2016), some important questions remain open. First, although the GWAS data set used in stage 2 usually contains from tens to hundreds of thousands of individuals, the eQTL data set is often quite small with a sample size of only a few hundreds or thousands. To what extent does the difference in sample sizes affect the power of TWAS? We therefore study the effects of stages 1 and 2 sample sizes on the power of TWAS. Second, the predictive power as measured by the coefficient of multiple determination,  $R^2$ , of a pretrained model in stage 1 is usually very low with  $R^2 \leq 0.05$ . Part of the reason is that the predictive power is upper bounded by the gene's expression heritability (Gamazon et al., 2015). Assuming that a lower  $R^2$  is partly due to the small sample size (and thus large estimation errors), if we were able to increase  $R^2$ , how much would that boost the power of TWAS? Lastly, in contrast to a standard (univariate) TWAS (UV-TWAS) considering only a single gene at a time, multivariate TWAS (MV-TWAS) has recently emerged to account for the effects of multiple genes, or of a gene's linear and nonlinear effects (Knutson et al., 2020; Lin et al., 2022). We are not aware of any studies on the power of MV-TWAS. In particular, it would be of interest to investigate whether one can gain or lose power by using MV-TWAS as compared with UV-TWAS: while accounting for multiple related genes' (predicted) expression levels may boost power, the expected presence of their correlations may reduce the precision of the estimates and thus statistical power.

We will use real data, mainly the Alzheimer's Disease Neuroimaging Initiative (ADNI) gene expression data

(Shen et al., 2013) and the International Genomics of Alzheimer's Project (IGAP) Alzheimer's disease (AD) GWAS summary data (Li et al., 2002), to empirically answer these questions. More specifically, we will use the fitted gene expression imputation models with the ADNI eQTL data and the fitted TWAS models with the IGAP GWAS data as the true models (with the true parameter values) in power/sample size calculations, which are more realistic than some arbitrarily chosen simulation models or parameter values to mimic unknown truths, leading to more meaningful conclusions to guide the study design for future TWAS. In addition, we will also apply the methods to the GTEx gene expression data and UK Biobank (UKB) GWAS data (GTEx Consortium, 2020; Sudlow et al., 2015). We will propose a general approach to sample size/power calculations for both univariate and multivariate TWAS, which can be applied to plan future TWAS and other related studies (such as Proteome- or Metabolome-Wide Association Studies). In particular, since both TWAS and MR are special cases of IV regression (Burgess et al., 2015; Xue & Pan, 2020), one may wonder whether we could simply follow an approach used for MR to calculate the power for TWAS. The short answer is no. Most power calculation methods for MR either account for only a single IV or focus on one-sample problems (Brion et al., 2012; Burgess, 2014; Freeman et al., 2013; Pierce et al., 2010). L. Deng et al. (2020) proposed a general procedure allowing for both two-sample setups and multiple (independent) IVs, but they assumed a single exposure. In contrast, we aim to inspect multivariate TWAS models with multiple (possibly correlated) exposures and multiple correlated IVs/SNPs, comparing their power to that of the standard/univariate TWAS. We note that our proposed sample size/power calculation method is general for two-sample 2SLS, applicable not only to (two-sample) TWAS (and related Proteome- or Metabolome-Wide Association Studies), but also to (two-sample) multivariable MR with possibly correlated SNPs/IVs (Burgess et al., 2015; Burgess & Thompson, 2015; Porcu et al., 2019).

The paper is organized as follows: in Section 2, we first introduce the general TWAS model along with necessary notations, then three specific TWAS models, whose power will be studied. Next, we propose a general sample size/power calculation procedure that can be applied to TWAS analyses, followed by the real data sets to be used. In Section 3, we apply the proposed method to the real data sets, addressing the three questions mentioned above. In Section 4, we summarize the main results, discuss the importance of the findings, and point out some potential limitations of the current study.

## 2 | MATERIALS AND METHODS

### 2.1 | A general TWAS model

Let  $n_1$  and  $n_2$  be the sample sizes of the eQTL and GWAS data in stages 1 and 2, respectively. We assume that there are no overlaps between the two data sets; that is, the two samples are independent as required by the two-sample 2SLS. For a given gene of interest, let  $Z_1$  be an  $n_1 \times p$  matrix coding for its *cis*-SNPs as IVs in the eQTL data set,  $Z_2$  be an  $n_2 \times p$  matrix for the IVs in the GWAS data set,  $x_0$  be an  $n_1 \times 1$  vector of the observed expression levels of the gene in the eQTL data set, and  $Y$  be an  $n_2 \times 1$  vector of the GWAS trait;  $p$  is the numbers of the IVs/SNPs used to predict the gene's expression. Following these notations and the standard valid IV assumptions, a general TWAS model is

$$X = Z_1\beta + \nu_1, \quad (1)$$

$$X_2 = Z_2\beta + \nu_2, \quad (2)$$

$$Y = X_2\theta + u = Z_2\beta\theta + \nu_2\theta + u = Z_2\beta\theta + \omega, \quad (3)$$

where  $X = (x_0, x_1, \dots, x_{(q-1)})$  with  $x_1, \dots, x_{(q-1)}$  being either higher-order terms of  $x_0$  or expression levels of other genes related to the gene of interest in the eQTL data;  $X_2$  would be the gene's expression levels in the GWAS data, though not actually observed, and  $\hat{X}_2 = Z_2\hat{\beta}$  is the imputed gene expression levels;  $\theta = (\theta_0, \dots, \theta_{(q-1)})^T$  is a  $q \times 1$  vector of the (unknown) causal effects of interest, and  $\beta = (\beta_0, \dots, \beta_{(q-1)})$  is a  $p \times q$  matrix for the (unknown) regression coefficients;  $\nu_1$  and  $\nu_2$  are independent error terms as  $n_1 \times q$  and  $n_2 \times q$  matrices, each row of which is independently and normally distributed with mean 0 and covariance matrix  $\Sigma_\nu$ . Independent of  $\nu_1$ , the error term  $u$  is an  $n_2 \times 1$  vector following a normal distribution with mean 0 and covariance matrix  $\sigma_u^2 I$ , and the error term  $\omega$  is an  $n_2 \times 1$  vector following a normal distribution with mean 0 and covariance matrix  $\sigma_\omega^2 I$ .

For simplicity of notation, we assume that  $X$  and  $Y$  are already suitably adjusted for covariates, and variables  $Z_1$ ,  $Z_2$ ,  $X$ , and  $Y$  are standardized to have their sample means all equal to 0 and sample variances to 1. Furthermore, here we assume that the GWAS individual-level data are available for simplicity and clarity of presentation, but in practice only GWAS summary data are required in stage 2 as to be shown in our numerical examples.

This general TWAS model allows us to accommodate more general cases than the standard TWAS where genes are imputed and tested only one by one

(i.e., with  $q = 1$ ). For example, in contrast to the standard TWAS of imputing and testing only a linear term  $x_0$ , one can impute  $x_0$  and  $x_0^2$  in stage 1 and test both terms in stage 2 to account for possibly nonlinear effects of gene expression; doing so can gain power and identify additional genes (Lin et al., 2022). When  $q$  is 1 and larger than 1, the general TWAS model represents the UV-TWAS and MV-TWAS, respectively.

Applying (two-sample) 2SLS (i.e., OLS in each stage), we obtain the estimators for the unknown parameters in the general TWAS model as

$$\begin{aligned}\hat{\beta} &= (Z_1^T Z_1)^{-1} Z_1^T X, & \hat{X}_2 &= Z_2 \hat{\beta}, \\ \hat{\theta} &= (\hat{X}_2^T \hat{X}_2)^{-1} \hat{X}_2^T Y, \\ \hat{\Sigma}_v &= (X - Z_1 \hat{\beta})^T (X - Z_1 \hat{\beta}) / (n_1 - 1), & \hat{\sigma}_\omega^2 &= \|Y - \hat{X}_2 \hat{\theta}\|_2^2 / n_2.\end{aligned}$$

Note that our proposed methods require the use of OLS estimator (OLSE), or more generally, maximum likelihood estimator (MLE), in each stage as used above, because the (corrected) covariance matrix formula for  $\hat{\theta}$  to be shown later is based on the OLSE (or MLE).

## 2.2 | Some specific TWAS methods

### 2.2.1 | UV-TWAS

We first consider the standard (univariate) TWAS, testing the linear effects of the genes one by one with  $X = x_0$ , denoted as *UV-L*. For this model, we focused on studying the impact of the sample size and the stage 1 pretrained model's predictive capability as measured by the coefficient of multiple determination,  $R^2$ , on the power of TWAS. We increased the sample size of stage 1 by a factor of 1, 2, 5, and 10, the sample size of stage 2 by a factor of 1, 1.2, 1.5, and 2, and the  $R^2$  of the fitted imputation model by 0.1, 0.2, and 0.3 (but capped the final increased  $R^2$  at  $\leq 0.95$ ) to examine their influences on the power. Note that we increase the  $R^2$  by a fixed amount, not relative to the original  $R^2$ , to see the direct effects of  $R^2$ . The calculation of  $R^2$  can be found in Section 2.2.4.

As noted in Lin et al. (2022), the quadratic effect of gene expression can be regarded as the influence of a gene's expression variability on the trait. We therefore also consider another UV-TWAS model, denoted as *UV-Q*, in which we test the quadratic term of gene expression on a trait with  $X = x_0^2$  (i.e., with only a quadratic term and no linear term). We will compare the

power of this model to that of MV-TWAS to be introduced later.

### 2.2.2 | TWAS-LQ

As shown in Lin et al. (2022), testing for quadratic effects of gene expression may unveil additional genes whose expression levels are nonlinearly associated with a trait. In this extended MV-TWAS model, denoted *TWAS-LQ*, we consider both linear and quadratic terms of gene expression. In other words, we have  $X = (x_0, x_0^2)$  and test the corresponding  $\theta$  in the model as described in (1)–(7). For power calculations, we set  $\tilde{\theta} = \hat{\theta}$  as estimated from the IGAP AD GWAS summary data, and used the corrected covariance matrix formula. Depending on which components of  $\theta$  to be tested, we can have three approaches, called *TWAS-L*, *TWAS-Q*, and *TWAS-LQ*: We test only the linear component, the quadratic component, and both linear and quadratic components, respectively.

### 2.2.3 | Other MV-TWAS

One possible downside of MV-TWAS with multiple genes is that their correlated gene expression levels may lead to the loss of power as compared with UV-TWAS. Here we consider a possibly more extreme situation where we expect a greater loss of power: because many physically neighboring genes have correlated expression levels (and their expression levels are often more highly correlated if they are closer to each other), we include every two neighboring genes in the MV-TWAS model. That is, if  $x_0$  and  $x_1$  are the expression levels of two neighboring genes, we have  $X = (x_0, x_1)$  in the MV-TWAS model. This setup is both more challenging and useful for TWAS because we often would like to identify which gene (among several with correlated expression) in a GWAS trait-associated locus is indeed causal. Similar to *TWAS-LQ*, in MV-TWAS we consider the power for three tests, called *MV-Target*, *MV-Alternative*, and *MV-Joint*, to test the (linear) effects of the first gene, of its neighbor, and of both the genes, respectively.

### 2.2.4 | Coefficient of multiple determination

For each UV-TWAS model, the coefficient of multiple determination  $R^2$  was calculated as



$$\begin{aligned} R^2 &= 1 - \text{RSS}/\text{TSS} = 1 - \text{RSS}/(n_1 - 1) \\ &= 1 - \hat{\Sigma}_v, \end{aligned}$$

where  $\text{RSS}$  and  $\text{TSS}$  were the residual sum of squares and the total sum of squares for the model; the second equality in the above equation follows from the fact that each  $X$  was standardized to have a sample variance of 1; that is,  $\text{Var}(X) = \text{TSS}/(n_1 - 1) = 1$ . To assess the effects of different  $R^2$  on the power of TWAS, we simply changed the value of  $\hat{\Sigma}_v$ : for example, to increase  $R^2$  by 0.1, we set the new  $\hat{\Sigma}_v$  to  $\max(0.05, \hat{\Sigma}_v - 0.1)$  and then used the new  $\hat{\Sigma}_v$  in the power calculation. Note that this capped  $R^2$  at 0.95.

The  $R^2$  can be interpreted as the (estimated) *cis*-heritability of the gene's expression levels based on the gene's *cis*-SNPs, which is expected to be lower than the gene's heritability based on the genome-wide SNPs (i.e., both *cis*- and *trans*-SNPs).

## 2.3 | Power calculations

The standard TWAS estimate of the variance of  $\hat{\theta}$  ignores the estimation errors with  $\hat{X}$ , or equivalently, with  $\hat{\beta}$ , in stage 1, thereby underestimating the true variance. We adopt the formula in Inoue and Solon (2010) for a corrected covariance matrix estimate of  $\hat{\theta}$  as

$$\hat{\text{Var}}_c(\hat{\theta}) = \hat{\text{Var}}_s(\hat{\theta}) \left( 1 + \frac{n_2}{n_1} \hat{\theta}^T \hat{\Sigma}_v \hat{\theta} / \hat{\sigma}_\omega^2 \right), \quad (4)$$

where  $\hat{\text{Var}}_s(\hat{\theta}) = (\hat{X}_2^T \hat{X}_2)^{-1} \hat{\sigma}_\omega^2$  is the standard/naive variance estimate, which ignores the estimation variability in stage 1 and thus would be obtained under the assumption of  $n_1 = \infty$ . The formula clearly illustrates that if the sample size in stage 1 is much smaller than that in stage 2, then we will have an inflated variance and thus reduced power in stage 2. With the corrected estimate of the covariance matrix, we now outline the hypothesis testing procedure as follows. Let  $q$  be the number of exposures (columns of  $X$ ),  $r$  be the number of exposures we want to test,  $A = \{A_1, \dots, A_r\}$  be a given subset of  $\{1, 2, \dots, q\}$  with each  $A_i \in \{1, \dots, q\}$  and  $1 \leq r \leq q$ . For each gene, we aim to perform the following test in stage 2 with

$$H_0 : \theta^A = 0 \text{ versus } H_1 : \theta^A = \tilde{\theta}^A \neq 0, \quad (5)$$

where  $\theta^A = (\theta_{A_1}, \dots, \theta_{A_r})^T$  is a subvector of  $\theta$  to be tested, and here we let  $\tilde{\theta}^A = \hat{\theta}^A$ , that is, we use the estimated

effects as the real effects under the alternative. Then  $\widehat{\text{Var}}_c(\hat{\theta}^A)$ , the estimated covariance matrix of  $\hat{\theta}^A$ , is just the corresponding submatrix of  $\widehat{\text{Var}}_c(\hat{\theta})$ . The TWAS test statistic is

$$T = \hat{\theta}^{A^T} (\widehat{\text{Var}}_c(\hat{\theta}^A))^{-1} \hat{\theta}^A, \quad (6)$$

which has a central  $\chi_r^2$  distribution under the null  $H_0$ ; however, under the alternative  $H_1$ ,  $T$  follows a non-central  $\chi^2$  distribution with  $r$  degrees of freedom and noncentrality parameter  $\tilde{\theta}^{A^T} (\widehat{\text{Var}}_c(\hat{\theta}^A))^{-1} \tilde{\theta}^A$ . The power is given by

$$P(T \geq \chi_{r, 1-\alpha/m}^2 | H_1), \quad (7)$$

where  $\alpha$  is the nominal significance level,  $m$  is the number of genes to be tested, and  $\chi_{r, 1-\alpha/m}^2$  is the  $(1 - \alpha/m)$ th quantile of a central  $\chi^2$  distribution with  $r$  degrees of freedom. Here we follow the popular practice of using the Bonferroni correction to adjust for multiple testing in TWAS.

## 2.4 | Data

### 2.4.1 | The ADNI eQTL data

Data used in the preparation of this article were obtained from the ADNI database (<http://adni.loni.usc.edu>). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations, as a 60-million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California-San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the United States

and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55–90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow-up duration of each group is specified in the protocols for ADNI-1, ADNI-2, and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see [www.adni-info.org](http://www.adni-info.org).

We used the ADNI whole-genome sequencing (WGS) and gene expression data (Shen et al., 2013) as the eQTL data in stage 1. After removing 37 individuals with any missing data, we had  $n_1 = 711$  individuals and 17,256 genes. Following previous ADNI studies (Lin et al., 2022), we adjusted for five covariates, namely, age, gender, year of education, handedness (left or right handed), and intracranial volume. We regressed the gene expression ( $X$ ) on these covariates and then used the standardized residuals as the new  $X$  to remove any potential effects they may have on AD, as shown by previous studies (Doody et al., 1999; van Loenhoud et al., 2022). For each gene, we extracted the SNPs from its *cis*-region by expanding 100 kb upstream and downstream of its coding region. SNPs with minor allele frequency  $\leq 0.05$  or any missing values were also excluded. We further pruned the SNPs so that the absolute value of the pairwise Pearson correlation between any two SNPs was less than 0.8. If the number of SNPs was still larger than 50 after pruning, we would keep the top 50 SNPs with the highest correlations (in their absolute values) with the gene's expression levels  $X$ . If  $q > 1$  (as in the MV-TWAS models), for each column  $i$  of  $X$ , we would take the top 50 SNPs with the highest correlations (in absolute values) with the column as  $I_i$  and use  $I = \cup_i I_i$  as the SNPs for stage 1. Next, the data were normalized to have mean 0 and variance 1, and we used backward selection with AIC as the criterion to select SNPs. The final model would be a linear regression model given by the backward selection procedure. For each gene, the  $F$  test was performed to assess the fit of the final model; if the  $p$  value was  $\geq 0.001$ , the gene was discarded (i.e., not used in the subsequent analysis). The choice of the tuning parameters (such as 0.8, 50, and 0.001 above) was somewhat arbitrary and largely followed a previous study (Xue & Pan, 2020) with some adjustments and the following rationale: highly correlated SNPs are not expected to help much to predict gene expression while causing the multicollinearity problem, thus we pruned out highly correlated SNPs; given the relatively small sample size, to avoid large estimation errors (due to large variability), we decided to use a relatively simple linear

model for effective prediction; and to satisfy the IV relevance assumption in 2SLS, we only chose *cis*-SNPs that were likely to be associated with a gene's expression level as IVs.

#### 2.4.2 | The IGAP AD GWAS summary data

For the GWAS data in stage 2, we used the summary data set released by the IGAP in 2013, which contained 54,162 individuals (Li et al., 2002). The IGAP AD GWAS summary data set was used to estimate the parameters in stage 2, namely,  $\hat{\theta}$  and  $\hat{\sigma}_\omega^2$ . To estimate  $\hat{\theta}$ , notice that

$$\hat{\theta} = (\hat{X}_2^T \hat{X}_2)^{-1} \hat{X}_2^T Y = (\hat{\beta}^T Z_2^T Z_2 \hat{\beta})^{-1} \hat{\beta}^T Z_2^T Y.$$

Since we did not have individual-level data for stage 2, we used  $Z_1$  to estimate  $Z_2^T Z_2$  and the summary statistics to estimate  $Z_2^T Y$ . For  $\hat{\sigma}_\omega^2$ , we have

$$\hat{\sigma}_\omega^2 = \frac{\|Y - \hat{X}_2 \hat{\beta}\|_2^2}{n_2} = 1 - 2 \frac{Y^T Z_2 \hat{\beta} \hat{\theta}}{n_2} + \frac{\|Z_2 \hat{\beta} \hat{\theta}\|_2^2}{n_2}.$$

We substituted  $Z_1$  for  $Z_2$  in the last term, and used the summary statistics to estimate  $Y^T Z_2$  in the second term. To calculate the power, we assume that under the alternative hypothesis, the true  $\theta$  is equal to the estimated one, that is,  $\theta = \hat{\theta}$ .

#### 2.4.3 | The GTEx and UKB data

To check our conclusions further, we used the Genotype-Tissue Expression (GTEx) data set and the UKB data (GTEx Consortium, 2020; Sudlow et al., 2015). For our purposes, we used the GTEx v8 whole blood data with 19,696 genes and  $n_1 = 670$  in stage 1. We regressed the gene expression on some covariates (including the first five genotype principal components, WGS sequencing platform, WGS library construction protocol, donor sex, and PEER factors) provided in the data set, and used the standardized residuals as the gene expression (i.e.,  $x_0$ ) as in a previous TWAS (Lin et al., 2022). The rest of the data preprocessing and quality control procedures are the same as for the ADNI data.

For stage 2, we used the individual-level UKB data with the high-density lipoprotein cholesterol as the trait of interest. Only the data from the individuals of white British ancestry were used. Any individual who might be a close relative of another (kinship  $> 0$ ) or had any missing values for the SNPs selected in stage 1 was also

removed. The sample sizes for most of the genes were about 200,000—the exact sample size for a gene depended on the number of individuals removed due to missing data—and the total number of SNPs was around 800,000.

For the GTEx data, we obtained the estimated expression heritabilities and their 95% confidence intervals (CIs) from (Wheeler et al., 2016). Therefore, we used the estimated expression heritabilities as the upper bounds of the stage 1  $R^2$  to demonstrate the effects of increasing  $R^2$  on the power of TWAS under more realistic conditions. The genes with estimated heritability smaller than the  $R^2$  of the stage 1 fitted model were thus removed. In addition, if the lower bound of the 95% CI of the estimated heritability was 0, then the  $R^2$  of the fitted model was used as the lower bound because a gene with heritability 0 cannot be genetically associated with any trait.

### 3 | RESULTS

#### 3.1 | Effect of the sample size on TWAS power

In Figure 1, we show how power varies with the increasing sample size in stage 1. Except for a few genes for which the power does not increase at all, in general increasing the sample size leads to higher power. When the stage 1 sample size  $n_1$  is about 8000, that is, a little over eleven times the original size, the power nearly reaches its maximum for most of the genes as if  $n_1 = \infty$ . Overall, for most genes, there seems to be only diminishing power gains by increasing  $n_1$ ; that is, the power only increased barely or only slightly. This result suggests that the current practice of having a sample size of eQTL data ranging from a few hundreds to tens of thousands (Gamazon et al., 2015; Gusev et al., 2016) is not expected to be a severe limiting factor of the power of TWAS.

For comparison, in Figure 2, we show how power varies with the increasing sample size in stage 2. Note that the naive estimator,  $\widehat{\text{Var}}_s(\hat{\theta}) = (\hat{X}_2^T \hat{X}_2)^{-1} \hat{\sigma}_\omega^2$ , is inversely proportional to  $n_2$  since as  $n_2 \rightarrow \infty$ ,  $(\hat{X}_2^T \hat{X}_2)^{-1} \rightarrow 0$ , while  $\hat{\sigma}_\omega^2 \rightarrow \sigma_\omega^2$ . As a result, increasing the stage 2 sample size will lower the naive variance estimate; at the same time, although the inflation factor,  $1 + \frac{n_2}{n_1} \hat{\theta}^T \hat{\Sigma}_v \hat{\theta} / \hat{\sigma}_\omega^2$ , will increase with  $n_2$ , the final variance  $\widehat{\text{Var}}_c(\hat{\theta})$  of the estimated causal effect  $\hat{\theta}$  will decrease since  $\hat{\theta}^T \hat{\Sigma}_v \hat{\theta} / \hat{\sigma}_\omega^2$  is usually very small ( $\leq 0.03$ ). It is clear that unless the original power is already close to 1, there

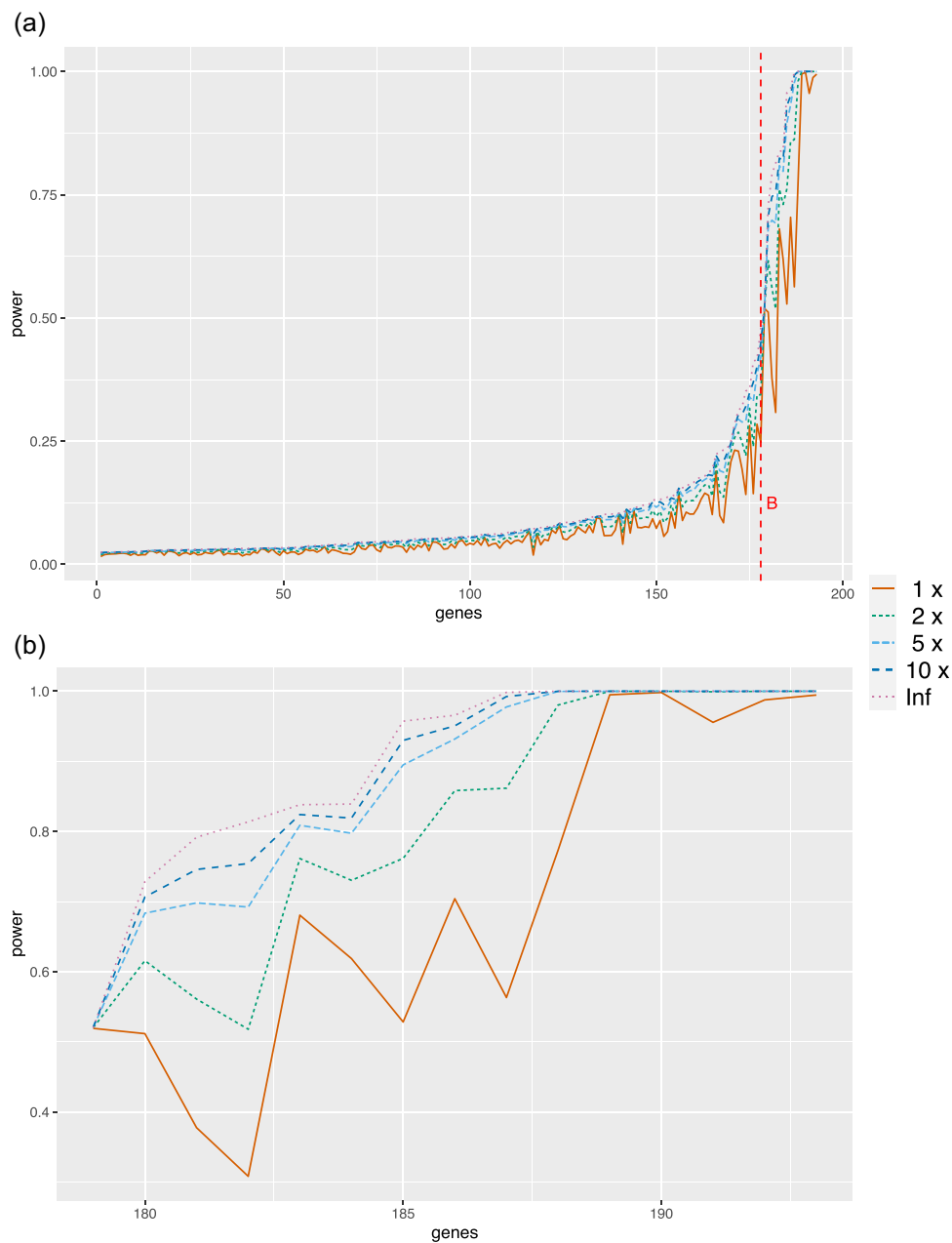
will be notable power gains for almost every gene by increasing  $n_2$ , unlike the situation for increasing  $n_1$ .

#### 3.2 | Effect of $R^2$ on the power of TWAS

A more important factor determining TWAS power is the predictive power of gene expression by SNPs in stage 1, which can be measured by multiple coefficient of determination,  $R^2$ . In Figure 3, we show that how TWAS power varies with  $R^2$ ; the genes are ordered by the power after increasing the observed  $R^2$  by 0.3. We can see that even 0.1 increase in  $R^2$  boosts the power by a large margin; a 0.3 increase of  $R^2$  leads to high power for most of the genes. Note that for some genes, there is little power gain because their original  $R^2$ 's are already large (close to 0.95), leaving barely any room to increase their  $R^2$ 's (since  $R^2$  cannot be larger than 1 and we capped any increased  $R^2$  at 0.95); in addition, the power of TWAS depends on not only the  $R^2$  in stage 1, but other parameters in stage 2, such as the sample size, prediction accuracy of the stage 2 model, and the size and variance of the estimated causal effect,  $\hat{\theta}$ . On the other hand, it is worth pointing out that, by definition, the true  $R^2$  for any gene is biologically upper bounded by its heritability. In other words, if a gene's true  $R^2$  is small, regardless of the sample size  $n_1$ , any analysis method will always be relatively low powered to detect its association with a trait in stage 2.

#### 3.3 | Power analysis for TWAS-LQ and MV-TWAS

Figure 4 compares the power for different tests in MV-TWAS and TWAS-LQ with that in UV-TWAS. It can be seen that, in general, MV-TWAS/TWAS-LQ may substantially increase or decrease the power as compared with the standard (linear) UV-TWAS denoted as UV-L. For example, when the underlying relationship between the gene expression and the trait can be approximated by a quadratic function (when TWAS-Q or UV-Q gives a small  $p$  value), including a quadratic term of the gene expression in TWAS-LQ will greatly increase the power, as indicated by panel C in Figure 4. However, in the case where the quadratic term is not relevant/significant, for example, the underlying relationship is approximately linear, UV-L outperforms TWAS-LQ. We observe a similar result for the MV-TWAS model. When the expression of the neighbor gene nearest to the target gene has a great contribution to the trait, MV-TWAS outperforms UV-TWAS; otherwise, we lose power by including it in the model.

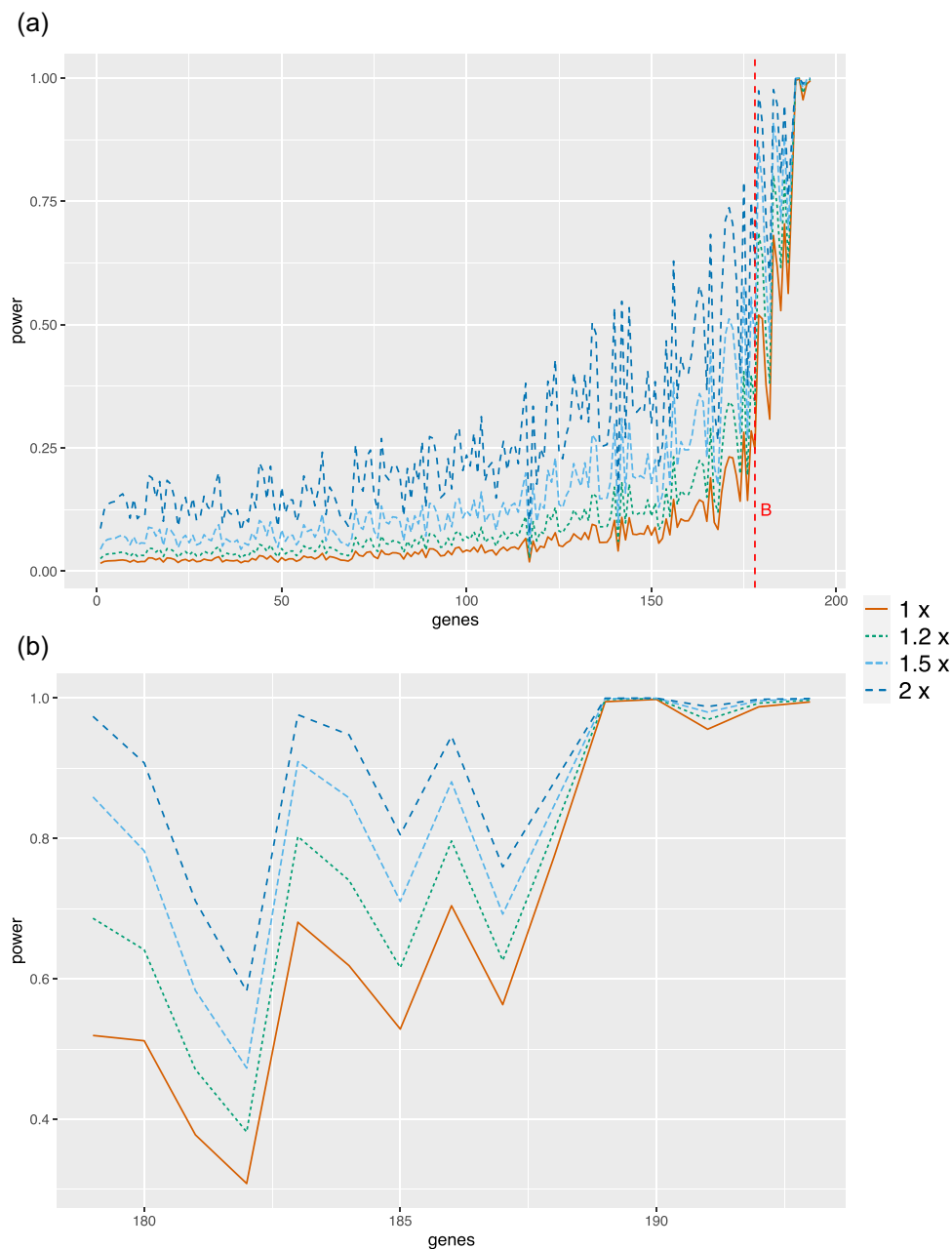


**FIGURE 1** Power of UV-L with different sample sizes  $n_1$  for the ADNI/IGAP data: (a) for all genes with stage 2  $p \leq 0.01$  (in stage 2 for their association with AD); (b) for 15 top genes (marked by the vertical line in panel (a)) with the most significant  $p \leq 5.6e - 6$ , which is the significance level after the Bonferroni correction. The genes are ordered by the power for “Inf” (i.e.,  $n_1 = \infty$ ); the lines correspond to the factors by which we increase  $n_1$ . AD, Alzheimer’s disease; ADNI, Alzheimer’s Disease Neuroimaging Initiative; IGAP, International Genomics of Alzheimer’s Project; UV-L, univariate linear.

As an example, consider gene *HLA-DRB5* and its nearest neighbor gene *HLA-DRB1* in Figure 4(c). When considered separately, that is, applying the UV-L model to each of the two genes, *HLA-DRB5* is significantly associated with AD with a  $p$  value of  $4.4e - 6$  that can be detected with power 0.52, while *HLA-DRB1* has a  $p$  value of  $3e - 5$  and power 0.28. Pearson’s correlation between the predicted expressions of the two genes is 0.21. Given the small  $p$  value of the neighbor gene and the low

correlation between their predicted expressions, incorporating the neighbor gene into the MV-TWAS model boosts the power from 0.52 to 0.83. On the other hand, gene *GLT8D2* has a  $p$  value of  $2.02e - 10$  and power 0.7 under the UV-L model. However, since its nearest gene, *TDG*, has a much larger  $p$  value of 0.64, including it in the MV-TWAS model decreases the power to 0.03. Interestingly, there is rarely any notable power decrease when the high correlation between a target gene and its



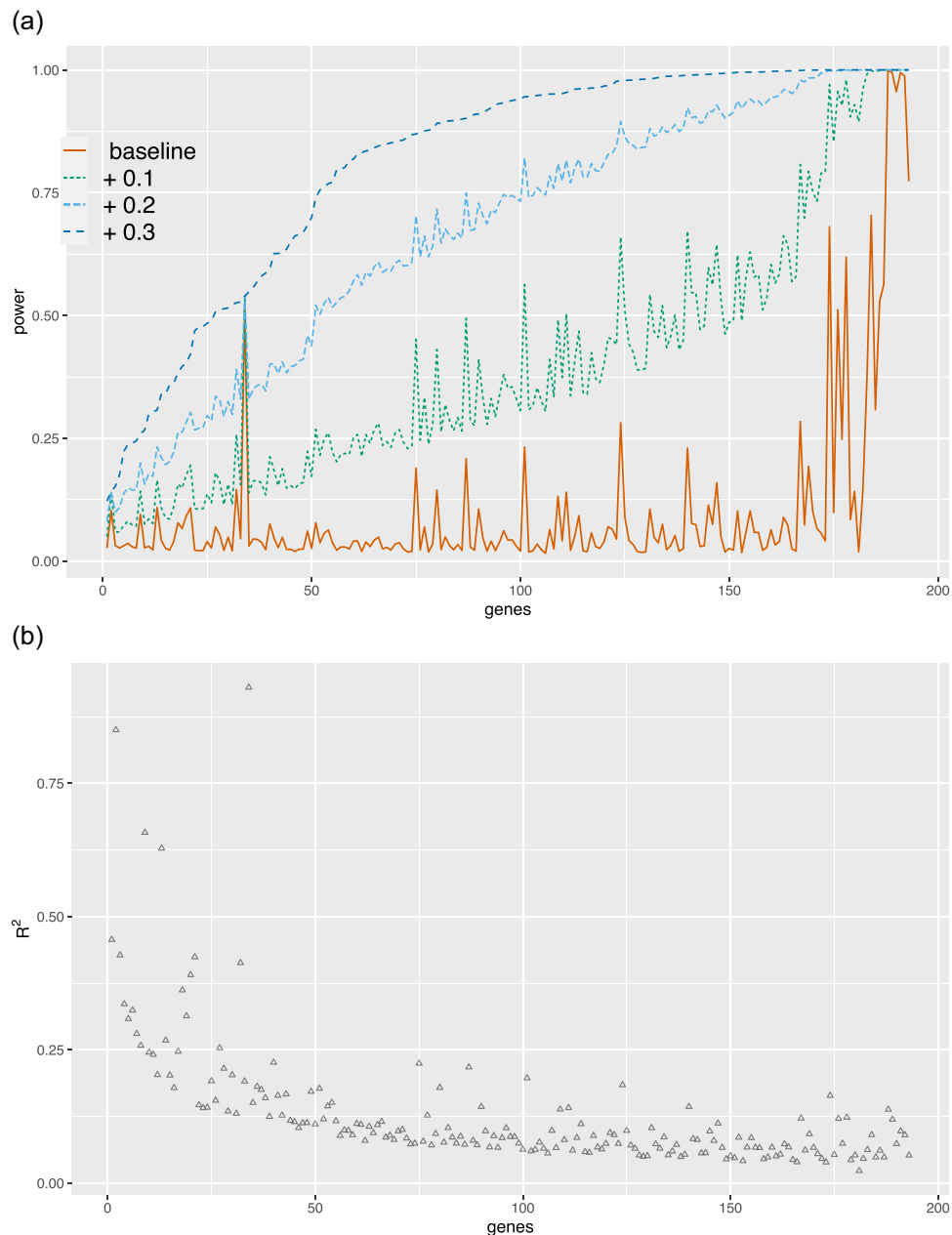


**FIGURE 2** Power of UV-L with different sample sizes  $n_2$  for the ADNI/IGAP data: (a) for all genes with stage 2  $p \leq 0.01$  (in stage 2 for their association with AD); (b) for 15 top genes with the most significant  $p \leq 5.6e-6$ .  $5.6e-6$  is given by the Bonferroni correction. For comparison, the order of the genes remains the same as in Figure 1 (ordered by the power for “Inf” in Figure 1); the different lines correspond to the factors by which we increase  $n_2$ . AD, Alzheimer’s disease; ADNI, Alzheimer’s Disease Neuroimaging Initiative; IGAP, International Genomics of Alzheimer’s Project; UV-L, univariate linear.

neighbor gene is the only reason. In fact, if both genes have  $p$  values less than or equal to 0.01 under the UV-L model, the correlation between their predicted expressions is less than 0.3 in absolute values except for gene *RNASE3*, for which the correlation is 0.87. Under the UV-L model, *RNASE3* and its neighboring gene *RNASE2* have  $p$  values of  $5.6e-3$  and  $8.5e-4$ , respectively, and corresponding power of 0.04 and 0.11; in contrast, the power for their MV-TWAS model is 0.09, an increase for

*RNASE3* but a decrease for *RNASE2*. The results suggest that for MV-TWAS, the relevance of the genes to the trait has a larger impact on the power than their correlation.

Tables 1 and 2 list the genes for which there are relatively larger improvements of power by TWAS-LQ and MV-TWAS over that of UV-TWAS. The  $p$  values, the power, and the power gained by using TWAS-LQ/MV-TWAS are also listed. As detailed in Section 4, the



**FIGURE 3** (a) Power of UV-L with different  $R^2$  for the genes with stage 2  $p \leq 0.01$  for the ADNI/IGAP data; (b) the corresponding baseline  $R^2$  for the genes. The genes are ordered by their power for “+ 0.3.” AD, Alzheimer’s disease; ADNI, Alzheimer’s Disease Neuroimaging Initiative; IGAP, International Genomics of Alzheimer’s Project; UV-L, univariate linear.

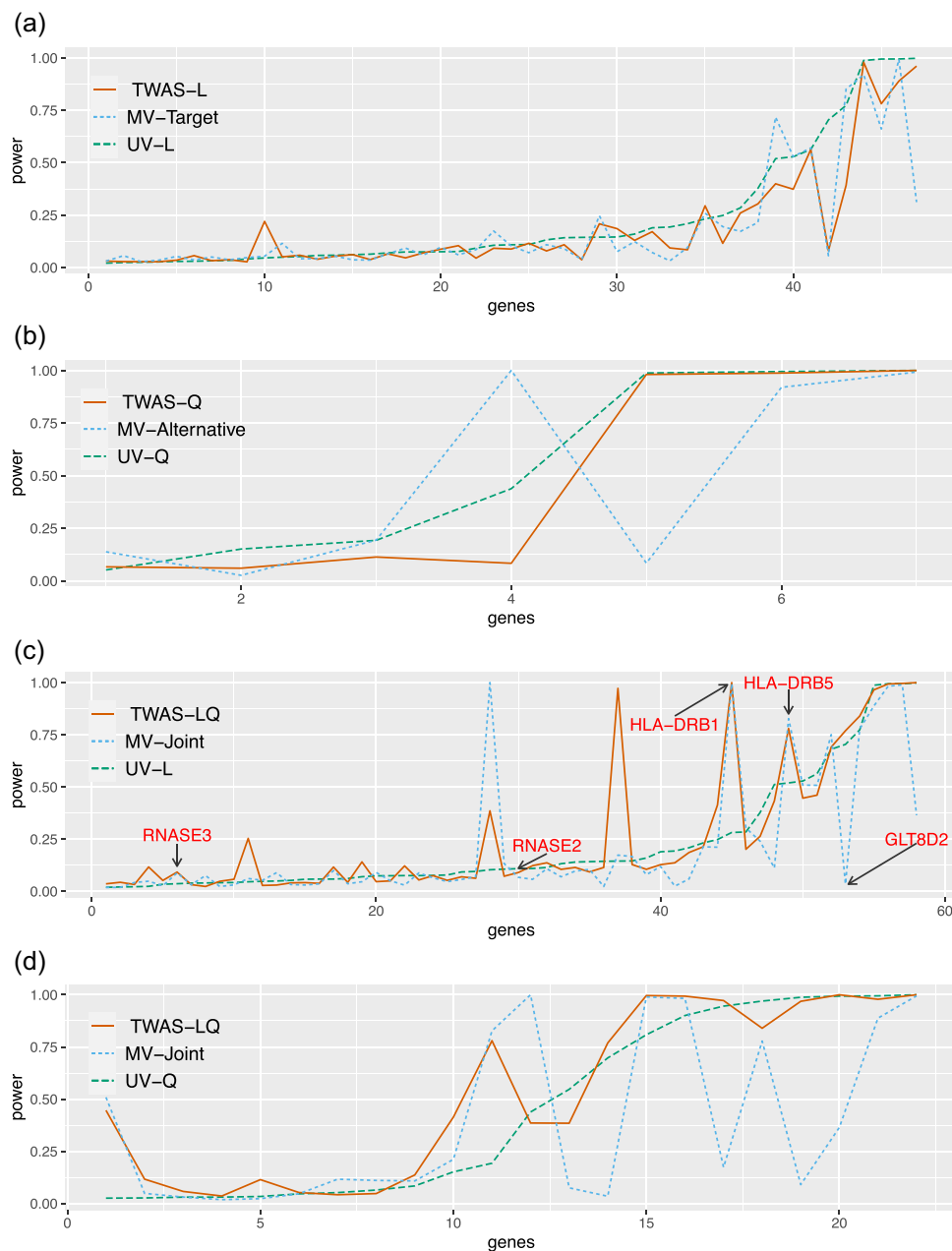
relevance of these genes to AD has been discussed in the literature.

### 3.4 | Results for the GTEx and UKB data

We focused on analyzing the effects of sample sizes and  $R^2$  with the GTEx and UKB data, and reached the same conclusions as before. We observed power increases as the stage 1 sample size was multiplied by a factor of 2, 5, 10, and 50 (Figure 5). Note that the power increased much less from

“5 $\times$ ” to “10 $\times$ ” than from “1 $\times$ ” to “2 $\times$ ” or from “2 $\times$ ” to “5 $\times$ ,” the same effect on diminishing power gains discussed in Section 3.1. Recall that the stage 1 sample size  $n_1$  was 670, and the stage 2 sample size  $n_2$  was around 200,000. Due to the large difference in the sample sizes, we needed to increase  $n_1$  by a factor of no more than 50 for the power to be almost equal to its maximum as if  $n_1 = \infty$ . The ratio  $n_1/n_2$  at which the power nearly attained the maximum was similar for both data set—roughly 1/6 or 1/7.

The case for  $n_2$  was the same as for the ADNI/IGAP data, see Figure 6. Regardless of the ratio  $n_1/n_2$ ,



**FIGURE 4** Power comparison for genes with stage 2  $p \leq 0.01$  for all models with the ADNI/IGAP data: (a) UV-L versus TWAS-L versus MV-Target; (b) UV-Q versus TWAS-Q versus MV-Alternative; (c) UV-L versus TWAS-LQ versus MV-Joint; (d) UV-Q versus TWAS-LQ versus MV-Joint. The genes are ordered by the power for the UV-TWAS models. AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; IGAP, International Genomics of Alzheimer's Project; LQ, linear and quadratic; MV, multivariate; TWAS-L, Transcriptome-Wide Association Study linear; UV-L, univariate linear; UV-Q, univariate quadratic.

increasing  $n_2$  always resulted in a notable power gain unless the power was close to 1, in contrast to diminishing power gains of increasing  $n_1$ .

The availability of the estimated expression heritability allowed for a more realistic expectation of the upper bound of the  $R^2$  in stage 1 (Wheeler et al., 2016), as compared with arbitrarily increasing  $R^2$  as done for the ADNI data. In Figure 7 A, we plotted the power calculated with the estimated heritability and the 95%

CI bounds against the power calculated with the actual  $R^2$  of the fitted model. Panel B compares the estimated heritability to the  $R^2$  of the fitted model. For most genes, their fitted  $R^2$  values were quite close to their heritability estimates, leading to their similar power curves. Similar to that observed for the ADNI data, a larger  $R^2$  would always result in a gain in power; how much a larger  $R^2$  would improve the power was also dependent on other factors as discussed in Section 3.2.

**TABLE 1** Some AD-associated genes with relatively large power gains in TWAS-LQ over UV-TWAS (UV-L or UV-Q) for the ADNI/IGAP data

Gene	Chromosome	<i>p</i> value	Power	Power gain
<i>FAM117B</i>	2	5.2e – 5	0.21	0.06
<i>HLA-DQA2</i>	6	3.3e – 26	0.99	0.19
<i>HLA-DQB1</i>	6	2.8e – 5	0.39	0.28
<i>HLA-DRB1</i>	6	0	1	0.72
<i>HLA-DRB5</i>	6	8.1e – 8	0.78	0.26
<i>BCL3</i>	19	2.7e – 8	0.41	0.26
<i>CNN2</i>	19	9.2e – 4	0.19	0.04

Note: The *p* value/power refers to the stage 2 *p* value/power of TWAS-L, TWAS-Q, or TWAS-LQ, whichever resulted in a large power gain, which was calculated by subtracting the power of UV-TWAS from that of TWAS-LQ.

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; IGAP, International Genomics of Alzheimer's Project; LQ, linear and quadratic; TWAS-L, Transcriptome-Wide Association Study linear; TWAS-Q, Transcriptome-Wide Association Study quadratic; UV-L, univariate linear; UV-Q, univariate quadratic.

**TABLE 2** Some AD-associated genes with relatively large power gains in MV-TWAS over UV-TWAS (UV-L or UV-Q) for the ADNI/IGAP data

Gene	Chromosome	<i>p</i> value	Power	Power gain
<i>BIN1</i>	2	3.8e – 15	0.75	0.07
<i>HLA-DQA2</i>	6	7.1e – 30	0.99	0.18
<i>HLA-DQB1</i>	6	5.3e – 22	0.99	0.89
<i>HLA-DRB1</i>	6	1.6e – 16	0.99	0.71
<i>HLA-DRB5</i>	6	1.1e – 8	0.83	0.31
<i>APOC1P1</i>	19	1.01e – 61	1	0.08

Note: The *p* value/power refers to the stage 2 *p* value/power of MV-Target, MV-Alternative, or MV-Joint, whichever resulted in a large power gain, which was calculated by subtracting the power of UV-TWAS from that of MV-TWAS.

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; IGAP, International Genomics of Alzheimer's Project; LQ, linear and quadratic; MV, multivariate; TWAS-L, Transcriptome-Wide Association Study linear; TWAS-Q, Transcriptome-Wide Association Study quadratic; UV-L, univariate linear; UV-Q, univariate quadratic.

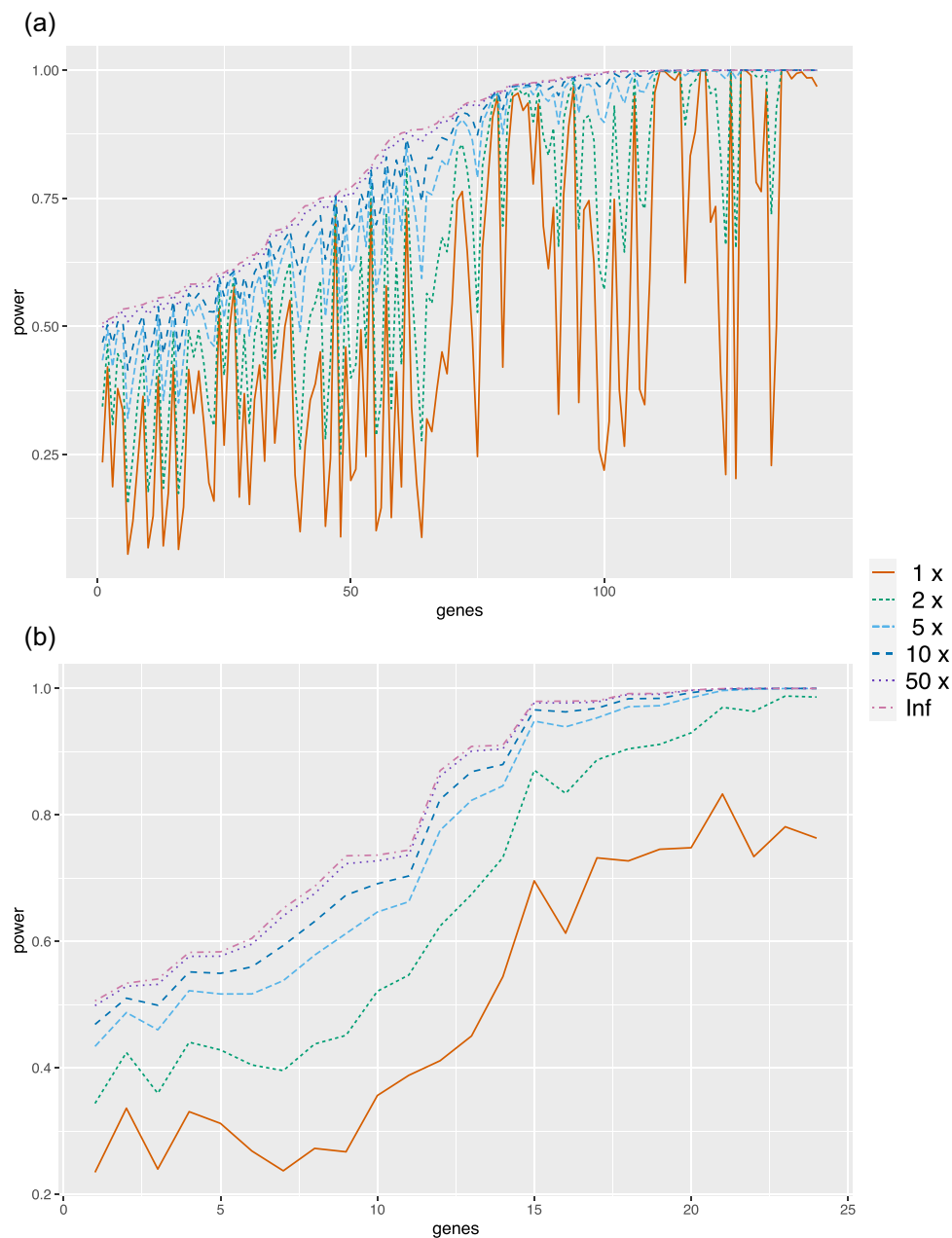
## 4 | DISCUSSION

In summary, we have studied the effects of the sample size and  $R^2$  (or more generally, gene expression heritability) in stage 1 on the power of the standard/univariate TWAS, and the potential power gains/losses of multivariate TWAS. The small sample size of the eQTL

data in TWAS may raise concerns on its possibly diminishing the power of TWAS by inflating the variance of the causal estimate  $\hat{\theta}$  in stage 2. We have shown that, as expected, increasing the stage 1 sample size would increase the power, but, perhaps surprisingly, only to a limited extent: a sample size  $n_1$  of eight thousands, only about 1/7 of the stage 2 sample size, seemed to nearly reach the power upper bound (when  $n_1 = \infty$ ) for almost all the genes in our ADNI/IGAP real data example. This result is surprising because this sample size  $n_1$  in stage 1 is still much smaller than  $n_2$  in stage 2. On the other hand, the effect of the sample size  $n_2$  in stage 2 appears to be of much higher impact. The same conclusions on the differing effects of the sample sizes in the two stages in TWAS were reached in the application to the GTEx and UKB data. These results are in agreement with a recent empirical study using other omic and GWAS data (Baranger et al., 2022): using smaller GWAS data in stage 2 may dramatically hinder any new discovery at the end for TWAS.

The case for  $R^2$ , however, is more complicated than that for the sample size. Whether and how much the power would increase as a result of a larger  $R^2$  depended on (1) the original (or baseline)  $R^2$ , (2) the estimated variance of the causal effect, and (3) the original power. Generally, if the original power and original  $R^2$  were small, and the estimated variance was large, then one would benefit more from a larger  $R^2$  than otherwise. However, it is noted that the  $R^2$  for any gene is upper bounded by the heritability of its expression, so it cannot really be manipulated for real data to increase power.

The multivariate TWAS analysis suggested that compared with the standard/univariate TWAS model, including higher-order terms or information from other genes could greatly increase the power (or decrease the power), at least under some realistic configurations (since the unknown true parameters are estimated from the real data) as we demonstrated with the real data. Interestingly, as shown in Tables 1 and 2, many of the genes with substantial power gains by TWAS-LQ and MV-TWAS over that of the standard UV-TWAS are related to AD as discussed in the literature. For example, the genes in the human leukocyte antigen (HLA) complex have been considered risk factors of late-onset AD (Mansouri et al., 2015; Steele et al., 2017; Wang et al., 2020). *APOC1* is another widely known gene, related to *APOE*, that affects the risk of AD (Q. Zhou & Zhao, 2014; X. Zhou et al., 2019). On the other hand, it has been proposed that *BIN1* mediates the risk of AD (Chapuis et al., 2013). The results justified the use of MV-TWAS models to identify additional genes that could be missed by UV-TWAS, so researchers may consider incorporating them into their studies.



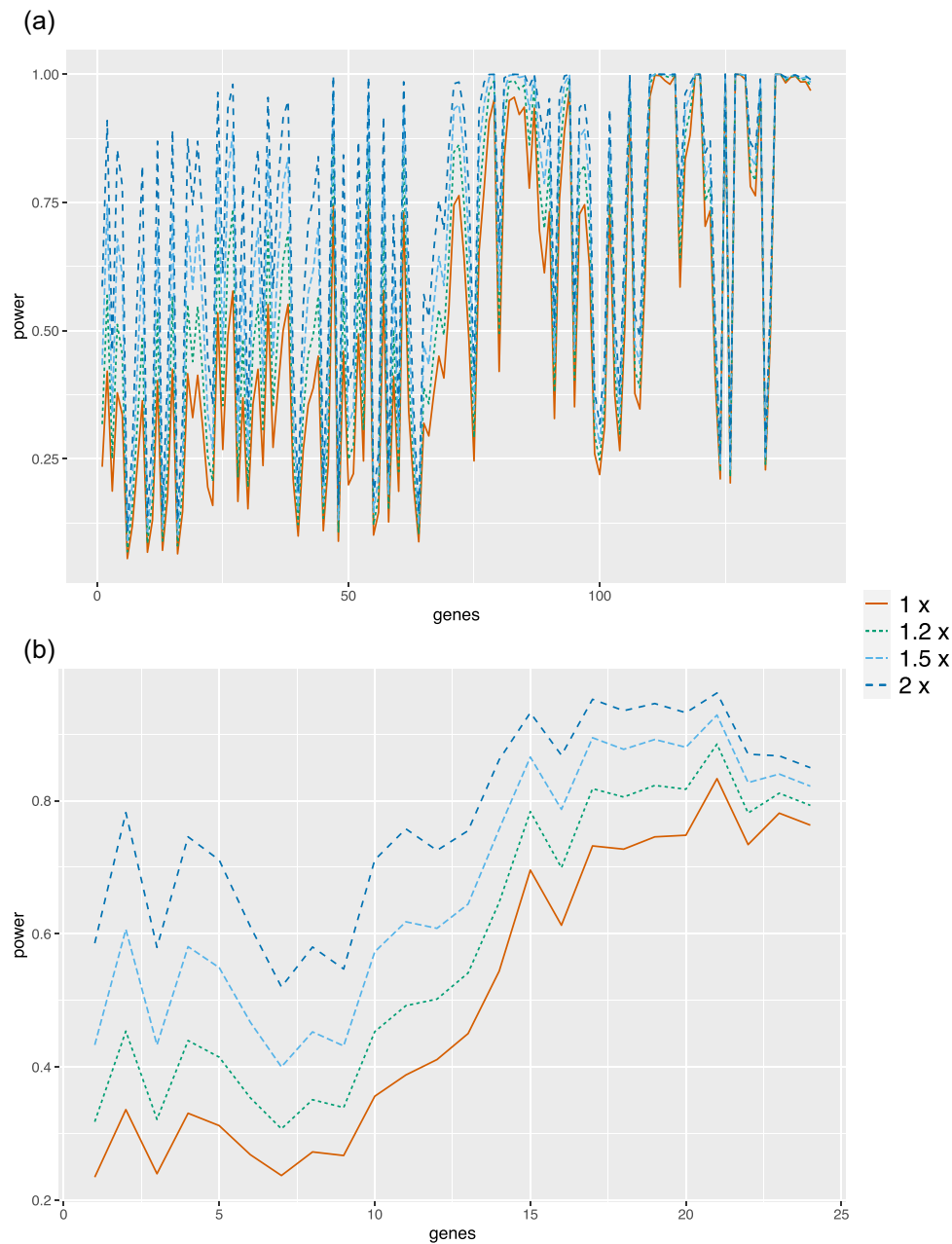
**FIGURE 5** Power of UV-L with different sample sizes  $n_1$  for the GTEx/UKB data. (a) Top 142 genes with the significant  $p$  values ( $\leq 4.36e - 6$ ). (b) A selected set of 24 genes with the significant  $p$  values, leading to a relatively smooth baseline power curve for better visualization.  $4.3e - 6$  is the significance level after the Bonferroni correction. The genes are ordered by the power for “Inf” (i.e.,  $n_1 = \infty$ ); the lines correspond to the factors by which we increase  $n_1$ . GTEx, Genotype-Tissue Expression; UKB, UK Biobank; UV-L, univariate linear.

Although power analysis for TWAS methods has gained some attention in the literature, to the best of our knowledge, none of the studies addressed the issues discussed in this paper. For example, a recent study inspected the effects of prediction accuracy of the stage 1 model on the TWAS power (Cao et al., 2021). However, they did not directly examine the effect of  $R^2$  on the power, but instead focused on the influence of replacing predicted gene expression with observed gene expression. Other studies either compared TWAS methods with

GWAS, or focused on closely related MR analyses (L. Deng et al., 2020; Veturi & Ritchie, 2018). In particular, these studies assumed a univariate, linear stage 2 model and conducted simulations, instead of using real data to generate more realistic parameters for analysis. Importantly, we expect that the current study will be useful in offering a general approach to sample size/power calculations for both univariate and multivariate TWAS.

There are a few limitations of this study. First, we did not account for potential horizontal pleiotropy

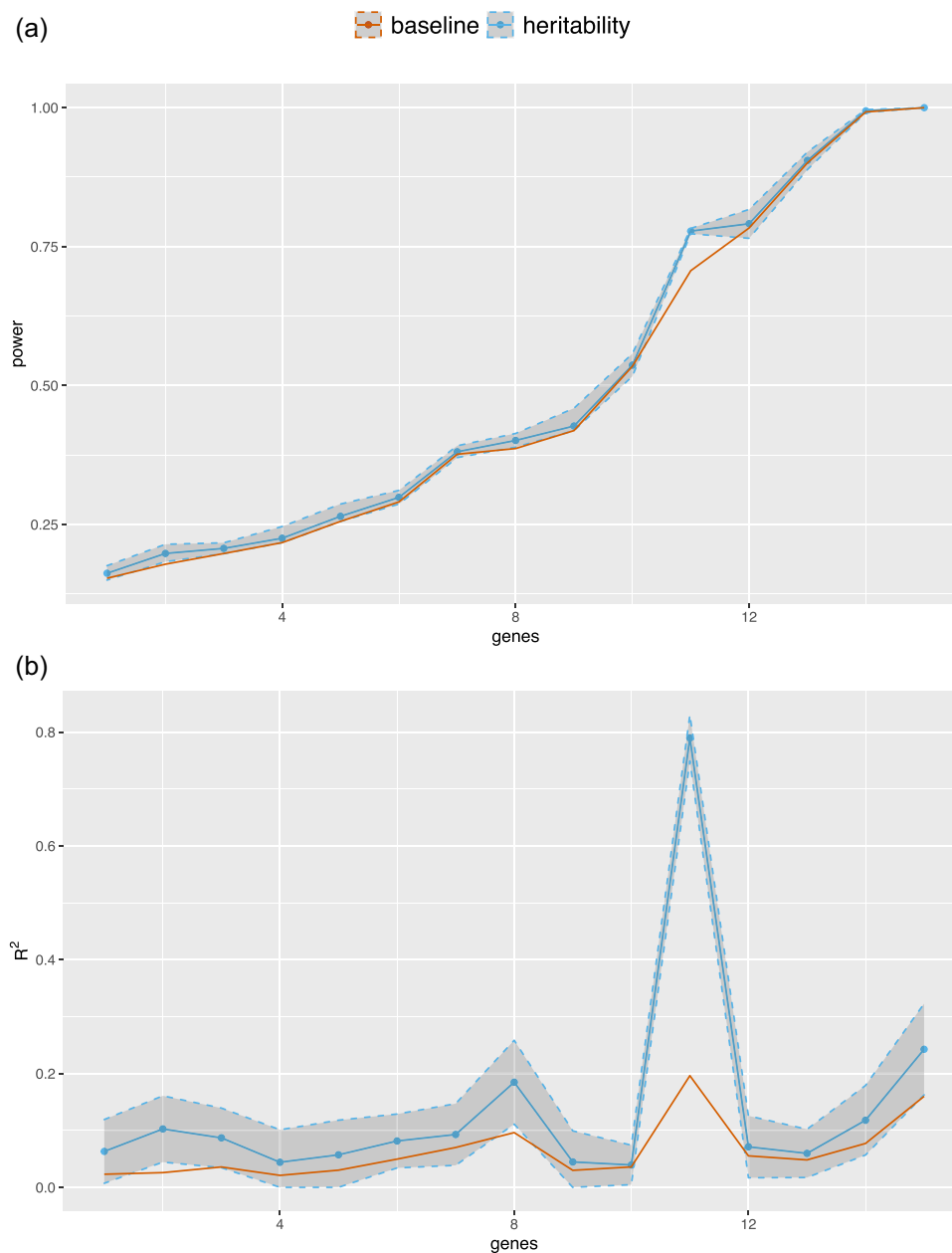




**FIGURE 6** Power of UV-L with different sample sizes  $n_2$  for the GTEx/UKB data. (a) Top 142 genes with the significant  $p \leq 4.36e - 6$ . (b) A selected set of 24 genes with the significant  $p$  values, leading to a relatively smooth baseline power curve for better visualization.  $4.3e - 6$  is the significance level after the Bonferroni correction; For comparison with Figure 5, the order of the genes remains the same (ordered by the power for “Inf” in Figure 5); the different lines correspond to the factors by which we increase  $n_2$ . GTEx, Genotype-Tissue Expression; UKB, UK Biobank; UV-L, univariate linear.

where genetic variants contribute to both gene expression levels and traits directly, leading to a biased causal estimate if not suitably accounted (Y. Deng & Pan, 2021; Lin et al., 2022). Second, one could consider nonlinear or nonparametric models like random forest (Breiman, 2001), gradient boosting machines (Friedman, 2001), or other machine learning methods (Okoro et al., 2021) to boost the prediction accuracy in stage 1, to which the

analysis presented here no longer applies because of our adopted linear models in stage 1. Third, perhaps most importantly, some specific results (e.g., required  $n_1$  or the ratio of  $n_1/n_2$ ) may change if different eQTL and GWAS data are considered, though we do not expect that our general conclusions (as discussed in the first three paragraphs in this section) will change dramatically.



**FIGURE 7** (a) Power of UV-L with different  $R^2$  for the GTEx/UKB data. For “baseline,” the power was calculated with the  $R^2$  of the fitted stage 1 model; for “heritability,” the power was calculated using the estimated heritability and the upper/lower bounds of its 95% CI as the  $R^2$ . (b)  $R^2$  of the fitted model (“baseline”), the estimated heritability and its 95% CI (“heritability”). Only the significant genes ( $p \leq 5.63e - 5$ ) with their estimated heritability larger than the fitted  $R^2$  are shown. The genes are ordered by the power for “baseline.” CI, confidence interval; GTEx, Genotype-Tissue Expression; UKB, UK Biobank; UV-L, univariate linear.

## ACKNOWLEDGMENTS

Data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's

Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd., and its affiliated company Genentech Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co. Inc.;

Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from dbGaP Project #26511. The access to the UKB data was approved through UKB Application #35107. This study was supported by NIH grants R01AG065636, RF1AG067924, U01AG073079, and R01HL116720, and by the MSI. We thank the reviewers for many helpful and constructive comments.

#### DATA AVAILABILITY STATEMENT

The ADNI data are available to the approved user at the ADNI website <http://adni.loni.usc.edu>; see the ADNI website for information about the application procedure. The IGAP AD GWAS summary data can be downloaded at <https://www.ebi.ac.uk/gwas/studies/GCST002245>. The GTEx data are available to the approved user at <https://www.ncbi.nlm.nih.gov/gap/>, and the UKB data are available to the approved user at <https://www.ukbiobank.ac.uk/>. The R code can be found at [https://github.com/RuoyuHe/TWAS\\_Power\\_Analysis](https://github.com/RuoyuHe/TWAS_Power_Analysis). All the analyses were performed on the MSI (<https://www.msi.umn.edu/>) server with a single AMD ROME compute node. The run time was roughly 4 h for every 2000 genes.

#### ORCID

Wei Pan  <http://orcid.org/0000-0002-1159-0582>

#### REFERENCES

- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455. <https://doi.org/10.1080/01621459.1996.10476902>
- Baranger, D. A., Hatoum, A. S., Polimanti, R., Gelernter, J., Edenberg, H. J., Bogdan, R., & Agrawal, A. (2022). *Multi-omics analyses cannot identify true-positive novel associations from underpowered genome-wide association studies of four brain-related traits*. bioRxiv. <https://doi.org/10.1101/2022.04.13.487655>
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. <https://doi.org/10.1023/A:1010950718922>
- Brion, M.-J. A., Shakhbazov, K., & Visscher, P. M. (2012). Calculating statistical power in Mendelian randomization studies. *International Journal of Epidemiology*, *42*(5), 1497–1501. <https://doi.org/10.1093/ije/dyt179>
- Burgess, S. (2014). Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *International Journal of Epidemiology*, *43*(3), 922–929. <https://doi.org/10.1093/ije/dyu005>
- Burgess, S., Small, D. S., & Thompson, S. G. (2015). A review of instrumental variable estimators for Mendelian randomization. *Statistical Methods in Medical Research*, *26*(5), 2333–2355. <https://doi.org/10.1177/0962280215597579>
- Burgess, S., & Thompson, S. G. (2015). Multivariable Mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects. *American Journal of Epidemiology*, *181*(4), 251–260. <https://doi.org/10.1093/aje/kwu283>
- Cao, C., Ding, B., Li, Q., Kwok, D., Wu, J., & Long, Q. (2021). Power analysis of transcriptome-wide association study: Implications for practical protocol choice. *PLOS Genetics*, *17*(2), e1009405. <https://doi.org/10.1371/journal.pgen.1009405>
- Chapuis, J., Hansmannel, F., Gistelink, M., Mounier, A., Cauwenbergh, C. V., Kolen, K. V., Geller, F., Sottejeau, Y., Harold, D., Dourlen, P., Grenier-Boley, B., Kamatani, Y., Delepine, B., Demiautte, F., Zelenika, D., Zommer, N., Hamdane, M., Bellenguez, C., Dartigues, J.-F., ... GERAD Consortium (2013). Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Molecular Psychiatry*, *18*(11), 1225–1234. <https://doi.org/10.1038/mp.2013.1>
- Deng, L., Zhang, H., & Yu, K. (2020). Power calculation for the general two-sample Mendelian randomization analysis. *Genetic Epidemiology*, *44*(3), 290–299. <https://doi.org/10.1002/gepi.22284>
- Deng, Y., & Pan, W. (2021). Model checking via testing for direct effects in Mendelian randomization and transcriptome-wide association studies. *PLOS Computational Biology*, *17*(8), e1009266. <https://doi.org/10.1371/journal.pcbi.1009266>
- Doody, R. S., Vacca, J. L., Massman, P. J., & Liao, T.-y. (1999). The influence of handedness on the clinical presentation and neuropsychology of Alzheimer disease. *Archives of Neurology*, *56*(9), 1133. <https://doi.org/10.1001/archneur.56.9.1133>
- Freeman, G., Cowling, B. J., & Schooling, C. M. (2013). Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *International Journal of Epidemiology*, *42*(4), 1157–1163. <https://doi.org/10.1093/ije/dyt110>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- 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. (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics*, *47*(9), 1091–1098. <https://doi.org/10.1038/ng.3367>
- GTEx Consortium. (2020). The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science*, *369*(6509), 1318–1330. <https://doi.org/10.1126/science.aaz1776>

- Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W. J. H., Jansen, R., de Geus, E. J. C., Boomsma, D. I., Wright, F. A., Sullivan, P. F., Nikkola, E., Alvarez, M., Civelek, M., Lusi, A. J., Lehtimäki, T., Raitoharju, E., Kähönen, M., Seppälä, I., ... Pasaniuc, B. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics*, *48*(3), 245–252. <https://doi.org/10.1038/ng.3506>
- Inoue, A., & Solon, G. (2010). Two-sample instrumental variables estimators. *Review of Economics and Statistics*, *92*(3), 557–561. [https://doi.org/10.1162/rest\\_a\\_00011](https://doi.org/10.1162/rest_a_00011)
- Knutson, K. A., Deng, Y., & Pan, W. (2020). Implicating causal brain imaging endophenotypes in Alzheimer's disease using multi-variable IWAS and GWAS summary data. *NeuroImage*, *223*, 117347. <https://doi.org/10.1016/j.neuroimage.2020.117347>
- Li, K. K. W., Ng, I. O. L., Fan, S. T., Albrecht, J. H., Yamashita, K., & Poon, R. Y. C. (2002). Activation of cyclin-dependent kinases CDC2 and CDK2 in hepatocellular carcinoma. *Liver*, *22*(3), 259–268. <https://doi.org/10.1046/j.0106-9543.2002.01629.x>
- Lin, Z., Xue, H., Malakhov, M. M., Knutson, K. A., & Pan, W. (2022). Accounting for nonlinear effects of gene expression identifies additional associated genes in transcriptome-wide association studies. *Human Molecular Genetics*. Advance online publication. <https://doi.org/10.1093/hmg/ddac015>
- Mansouri, L., Klai, S., Gritli, N., Fekih-Mrissa, N., Messalmani, M., Bedoui, I., Derbali, H., & Mrissa, R. (2015). Association of HLA-DR/DQ polymorphism with Alzheimer's disease. *The American Journal of the Medical Sciences*, *349*(4), 334–337. <https://doi.org/10.1097/maj.0000000000000416>
- Okoro, P. C., Schubert, R., Guo, X., Johnson, W. C., Rotter, J. I., Hoeschele, I., Liu, Y., Im, H. K., Luke, A., Dugas, L. R., & Wheeler, H. E. (2021). Transcriptome prediction performance across machine learning models and diverse ancestries. *Human Genetics and Genomics Advances*, *2*(2), 100019. <https://doi.org/10.1016/j.xhgg.2020.100019>
- Pierce, B. L., Ahsan, H., & VanderWeele, T. J. (2010). Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *International Journal of Epidemiology*, *40*(3), 740–752. <https://doi.org/10.1093/ije/dyq151>
- Porcu, E., Rüeger, S., Lepik, K., Santoni, F. A., Reymond, A., & Kutalik, Z. (2019). Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nature Communications*, *10*(1), 3300. <https://doi.org/10.1038/s41467-019-10936-0>
- Shen, L., Thompson, P. M., Potkin, S. G., Bertram, L., Farrer, L. A., Foroud, T. M., Green, R. C., Hu, X., Huentelman, M. J., Kim, S., Kauwe, J. S. K., Li, Q., Liu, E., Macciardi, F., Moore, J. H., Munsie, L., Nho, K., Ramanan, V. K., Risacher, S. L., ... Saykin, A. (2013). Genetic analysis of quantitative phenotypes in AD and MCI: Imaging, cognition and biomarkers. *Brain Imaging and Behavior*, *8*(2), 183–207. <https://doi.org/10.1007/s11682-013-9262-z>
- Steele, N. Z. R., Carr, J. S., Bonham, L. W., Geier, E. G., Damotte, V., Miller, Z. A., Desikan, R. S., Boehme, K. L., Mukherjee, S., Crane, P. K., Kauwe, J. S. K., Kramer, J. H., Miller, B. L., Coppola, G., Hollenbach, J. A., Huang, Y., & Yokoyama, J. S. (2017). Fine-mapping of the human leukocyte antigen locus as a risk factor for Alzheimer disease: A case-control study. *PLOS Medicine*, *14*(3), e1002272. <https://doi.org/10.1371/journal.pmed.1002272>
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*, *12*(3), e1001779. <https://doi.org/10.1371/journal.pmed.1001779>
- van Loenhoud, A. C., Groot, C., Bocancea, D. I., Barkhof, F., Teunissen, C., Scheltens, P., van de Flier, W. M., & Ossenkoppele, R. (2022). Association of education and intracranial volume with cognitive trajectories and mortality rates across the Alzheimer disease continuum. *Neurology*, *98*(16), e1679–e1691. <https://doi.org/10.1212/wnl.000000000000200116>
- Veturi, Y., & Ritchie, M. D. (2018). How powerful are summary-based methods for identifying expression-trait associations under different genetic architectures? *Pacific Symposium on Biocomputing*, *23*, 228–239. World Scientific. [https://doi.org/10.1142/9789813235533\\_0021](https://doi.org/10.1142/9789813235533_0021)
- Wang, Z.-X., Wan, Q., & Xing, A. (2020). HLA in Alzheimer's disease: Genetic association and possible pathogenic roles. *NeuroMolecular Medicine*, *22*(4), 464–473. <https://doi.org/10.1007/s12017-020-08612-4>
- Wheeler, H. E., Shah, K. P., Brenner, J., Garcia, T., Aquino-Michaels, K., GTEx Consortium, Cox, N. J., Nicolae, D. L., & Im, H. K. (2016). Survey of the heritability and sparse architecture of gene expression traits across human tissues. *PLOS Genetics*, *12*(11), e1006423. <https://doi.org/10.1371/journal.pgen.1006423>
- Xue, H., Pan, W., & Alzheimer's Disease Neuroimaging Initiative. (2020). Some statistical consideration in transcriptome-wide association studies. *Genetic Epidemiology*, *44*(3), 221–232. <https://doi.org/10.1002/gepi.22274>
- Zhou, Q., Zhao, F., Lv, Z.-p., Zheng, C.-g., Zheng, W.-d., Sun, L., Wang, N.-n., Pang, S., de Andrade, F. M., Fu, M., He, X.-h., Hui, J., Jiang, W.-y., Yang, C.-y., Shi, X.-h., Zhu, X.-q., Pang, G.-f., Yang, Y.-g., Xie, H.-q., ... Yang, Z. (2014). Association between APOC1 polymorphism and Alzheimer's disease: A case-control study and meta-analysis. *PLOS One*, *9*(1), e87017. <https://doi.org/10.1371/journal.pone.0087017>
- Zhou, X., Chen, Y., Mok, K. Y., Kwok, T. C. Y., Mok, V. C. T., Guo, Q., Ip, F. C., Chen, Y., Mullapudi, N., Alzheimer's Disease Neuroimaging Initiative, Giusti-Rodríguez, P., Sullivan, P. F., Hardy, J., Fu, A. K. Y., Li, Y., & Ip, N. Y. (2019). noncoding variability at the APOE locus contributes to the Alzheimer's risk. *Nature Communications*, *10*(1), 3310. <https://doi.org/10.1038/s41467-019-10945-z>

**How to cite this article:** He, R., Xue, H., & Pan, W. (2022). Statistical power of transcriptome-wide association studies. *Genetic Epidemiology*, *46*, 572–588. <https://doi.org/10.1002/gepi.22491>