bioinformatics, YYYY, pp. 1–22

doi: 10.1093/bioinformatics/xxxxx
DD Month YYYY
Manuscript

MANUSCRIPT

A-TWAS: An aggregated transcriptome-wide association study model incorporating multiple Bayesian priors

Yilan Liang,[1, †] Han Wang[2, †] and Yan Dora Zhang[1, *]

[1]Department of Statistics and Acturial Science, School of Computing and Data Science, The University of Hong Kong, Pokfulam Road, 999077, Hong Kong SAR, China and [2]College of Science, China Agricultural University, 17 Qinghua Donglu, 100083, Beijing, China
[†]Co-first authors.[*]Corresponding author. doraz@hku.hk
FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

Motivation: Transcriptome-wide association study (TWAS) is a significant methodology utilized for identifying associations between genes and diseases by integrating transcriptome and genome-wide association studies (GWAS) data. The approach has been successful in pinpointing risk genes for various diseases, including Alzheimer's disease, schizophrenia, and different types of cancers. TWAS typically involves two key steps: imputation and association analysis. In the imputation step, the original PrediXcan employs the elastic-net model, while subsequent research endeavors have delved into more intricate models such as FUSION and TIGAR. Despite these advancements, the existing individual methods may not capture the intricate genotype-expression relationships in a comprehensive way. Given the complexity of genetic contributions of genotypes to gene expression, sophisticated modeling techniques are imperative. In response to this, we have introduced Aggregated-TWAS (A-TWAS), a comprehensive tool that amalgamates multiple imputation models to accommodate diverse genetic architectures. A-TWAS utilizes several continuous shrinkage priors, including Laplace, Horseshoe, and Horseshoe+, to ensure precise imputation results while incorporating the benefits of Bayesian variable selection methods. In the association phase of TWAS, we employ the aggregated Cauchy association test (ACAT) to obtain an omnibus p-value.
Result: We demonstrate the effectiveness of A-TWAS through comprehensive simulations and real data analyses. Simulation studies highlight that A-TWAS produces substantial improvements in predictive R^2 and statistical power, while maintaining the type error I at a low level. Furthermore, TWASs are conducted on schizophrenia and obsessive-compulsive symptoms datasets. In comparison to standard methods, the Bayesian imputation model demonstrates superior accuracy, and A-TWAS, by integrating multiple Bayesian priors, identifies a notably greater number of disease-relevant genes.
Availability and Implementation: The source code of A-TWAS is available at `https://github.com/Yilan-Liang/A-TWAS`.

Key words: TWAS, GWAS, Bayessian Lasso, Horseshoe, Horseshoe+

Introduction

Genome-wide association studies (GWAS) have identified tens of thousands of genetic variants associated with complex traits and diseases, offering unprecedented insights into genetic architectures and disease mechanisms (Uffelmann et al., 2021). However, GWAS results primarily identify associations at the variant level, which complicates biological interpretation and limits their utility in understanding gene-level mechanisms. Transcriptome-wide association studies (TWAS) have emerged as a powerful tool to bridge this gap by integrating transcriptomic and GWAS data to identify gene-trait associations (Gusev et al., 2016a). TWAS leverages gene expression as an intermediate phenotype, linking genotype to disease phenotypes through imputed gene expression, and has been successful in identifying risk genes for numerous diseases, such as Alzheimer's disease (Sun et al., 2021; Harwood et al., 2021), schizophrenia (Gusev et al., 2018), and multiple types of cancers (Mancuso et al., 2018; Wu et al., 2018; Bhattacharya et al., 2020; Bossé et al., 2020).

The implementation of TWAS involves two main stages (Gusev et al., 2016a). The first stage trains predictive models to impute gene expression based on genotype data, while the second stage tests for associations between imputed gene expression and phenotypes using GWAS summary statistics. Despite its utility, a key challenge in TWAS lies in accurately modeling the complex relationship between genotype and gene expression. This relationship is influenced by various factors, including cis-regulatory elements, tissue specificity, and

© The Author YYYY. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

disease-specific mechanisms, making it highly variable across tissues and traits. Existing models often assume simplified or fixed structures for this relationship, which may fail to comprehensively capture the underlying genetic architecture of gene expression. This limitation motivates the development of more flexible and robust approaches to improve imputation accuracy and downstream gene-trait association detection.

Classic TWAS methods, such as PrediXcan (Gamazon et al., 2015), utilized elastic-net regression to model the cis-regulatory effects of single nucleotide polymorphisms (SNPs) on gene expression. Subsequent methods have employed more sophisticated statistical frameworks to improve imputation accuracy. For example, FUSION (Gusev et al., 2016b) uses Bayesian Sparse Linear Mixed Models (BSLMM) to capture both sparse and polygenic genetic effects, while TIGAR (Nagpal et al., 2019) adopts a nonparametric Bayesian model to flexibly model local expression quantitative trait expression quantitative trait loci (cis-eQTL) effect sizes. Despite these advancements, single-model approaches are limited by their reliance on specific model assumptions, which may not align with the true underlying genetic structure of gene expression. This misalignment can result in suboptimal imputation performance, particularly when the genetic architecture varies across tissues, diseases, or environmental contexts (Mancuso et al., 2019; Bryois et al., 2020).

To address these limitations, we propose Aggregated-TWAS (A-TWAS), an omnibus framework that integrates multiple Bayesian models which better capture the diverse and complex relationship between genotype and gene expression through two main enhancements. For the first enhancement, inspired by the flexibility of neuronized priors in Bayesian regression (Shin and Liu, 2021), A-TWAS aggregates information from multiple imputation models, each assuming different underlying structures for the cis-eQTL effect sizes. Specifically, we incorporate Bayesian Lasso (Park and Casella, 2008), Horseshoe (Carvalho et al., 2009), and Horseshoe+ priors (Bhadra et al., 2017) to represent varying degrees of sparsity and shrinkage in the cis-eQTL effect size distribution. These priors are particularly well-suited for high-dimensional transcriptomic data, as they assume that only a subset of genetic variants have strong regulatory effects on gene expression, aligning with the sparse nature of transcriptomic regulation.

The second enhancement is that we aggregate the association p-values obtained from multiple imputation models using the Adaptive Combination of Evidence Test (ACAT) (Liu et al., 2019) in the second stage of A-TWAS, producing a unified and comprehensive gene-trait association p-value. This approach combines the strengths of different Bayesian models while mitigating the biases introduced by any single model's assumptions. By leveraging the complementary strengths of multiple priors and integrating their results in a statistically principled manner, A-TWAS provides a more robust and flexible framework for TWAS analysis, particularly when the genetic architecture of gene expression is complex or varies across tissues and traits.

To evaluate the performance of A-TWAS, we conducted extensive simulations and real-world data analyses. The simulations demonstrate that A-TWAS improves imputation accuracy and statistical power while maintaining type-I error at a controlled level. Using GWAS summary statistics for schizophrenia and obsessive-compulsive disorder (OCD), we show that A-TWAS identifies substantially more disease-relevant genes compared to single-model TWAS approaches. Our results highlight the adaptability and robustness of A-TWAS in capturing diverse genetic architectures and underscore its utility as a powerful tool for gene-trait association studies.

Methods and Material

Overview of TWAS

Traditional TWAS methods mainly consist of the imputation step and the association step. The first imputation step utilizes individual-level genotype and transcriptome data from a reference panel to fit a regression model for each gene g, estimating the effect size of gene g's cis-SNP on the expression level. Concretely, for each gene g, consider the following model:

$$\mathbf{e}_g = \mathbf{X}_g \mathbf{w} + \boldsymbol{\epsilon}_g, \boldsymbol{\epsilon}_g \sim N\left(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}\right). \tag{1}$$

where $\mathbf{e}_g$ is an n-vector denoting the expression levels for gene g. $\mathbf{X}_g$ is an $n \times m$ genotype matrix recording the information of m cis-SNPs for gene g, only cis-SNPs within 1 MB of the flanking 5' and 3' ends are included. $\mathbf{w}$ is the effect size vector for m cis-SNPs of gene g, $\boldsymbol{\epsilon}_g$ is the random error term following a normal distribution. The intercept term is ignored in the model since $\mathbf{e}_g$ and $\mathbf{X}_g$ are assumed to be standardized.

In the second step of TWAS, if individual-level GWAS data is available, by using the effect size $\hat{\mathbf{w}}$ estimated from the first step, the genetically regulated expression (GReX) for an independent GWAS data can be imputed by $\widehat{\mathbf{GReX}}_g = \mathbf{X}_{\text{new}}\, \hat{\mathbf{w}}$, where $\mathbf{X}_{\text{new}}$ is the genotype information for gene g in the GWAS data. Then a generalized linear model (GLM) can be adopted to model the association between $\widehat{\mathbf{GReX}}_g$ and phenotype $\mathbf{Y}$:

$$E[\mathbf{Y}|\widehat{\mathbf{GReX}}_g] = g^{-1}(\widehat{\mathbf{GReX}}_g \beta)$$

where g is the link function. For a quantitative phenotype $\mathbf{Y}$, g can be the identity function; for a dichotomous phenotype Y, g can be set as the logit function. The association step evaluates the gene-level association by testing $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$. If only summary-level GWAS data is available, the S-PrediXcan (Barbeira et al., 2018) test statistics could be used:

$$Z_{g,\ \text{S-PrediXcan}} = \frac{\sum_{j=1}^{m} \hat{w}_j \hat{\sigma}_j Z_j}{\sqrt{\hat{\mathbf{w}}' \mathbf{V} \hat{\mathbf{w}}}}, \tag{2}$$

where Z_j is the Z-score test statistic for the j-th cis-SNP of gene g from summary-level GWAS data, $\hat{\mathbf{w}}$ is estimated from the TWAS first step, σ_j is the standard deviation of the genotype information for the j-th cis-SNP, $\mathbf{V}$ is the genotype covariance matrix of gene g's cis-SNPs, which could be approximated from reference panel used in the imputation step or other reference database from the same ancestry.

Baseline method

In stage I TWAS, multiple methods can be used to estimate the cis-eQTL effect size. In PrediXcan (Gamazon et al., 2015), Elastic-net (Zou and Hastie, 2005) is adopted for imputation. Elastic-net is a regularized regression method that addresses sparsity by adding L_1 and L_2 norms as penalties. The Elastic-net model is defined as follows:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \left(||\mathbf{y} - \mathbf{X}\mathbf{w}||^2 + \lambda \left((1-\alpha)||\mathbf{w}||_1 + \alpha ||\mathbf{w}||^2 \right) \right), \tag{3}$$

where α is a hyperparameter with value lies in $[0, 1]$ that adjusts the shape of the penalty, and λ is a parameter controlling the

strength of the penalty. In this work, α is manually set to 0.5, and λ is determined through cross-validation. Based on using Elastic-net for TWAS stage I imputation, the method employing GWAS summary data for stage II association testing via Equation 2 is S-PrediXcan (Barbeira et al., 2018), which is chosen as the baseline method in our study.

Bayesian Lasso

Unlike Elastic-net, which originates from the frequentist perspective, Bayesian Lasso (Park and Casella, 2008) is a penalization method derived from the Bayesian perspective. It introduces a Laplace prior distribution on the weights for inference. In the context of this research, Bayesian Lasso places a Laplace prior on the cis-eQTL weights:

$$\pi(\mathbf{w}) = \prod_{j=1}^{n} \frac{\lambda}{2} e^{-\lambda|\mathbf{w}_j|}, \tag{4}$$

where λ is the hyperparameter controlling the strength of shrinkage. This parameter can be estimated using a Gibbs sampler and does not need to be set manually.

Horseshoe Method

The horseshoe prior developed by Carvalho et al. (2009) is another powerful tool in high-dimensional regression and variable selection tasks. It exhibits strong shrinkage for small or irrelevant signals, effectively pulling them toward zero, while preserving large signals by means of the heavy tails. Moreover, unlike traditional Bayesian variable selection methods that typically impose discrete priors, which can be computationally expensive and challenging to apply in high-dimensional settings, the horseshoe prior sidesteps this issue by treating sparsity as a continuous property, enabling seamless inference without requiring hard thresholding. Specifically, the horseshoe model in the TWAS imputation step can be formulated as:

$$\begin{aligned} \mathbf{e}_g \mid \boldsymbol{X}, \mathbf{w}, \sigma^2 &\sim N_m\left(\boldsymbol{X}\mathbf{w}, \sigma^2 \boldsymbol{I}_m\right) \\ w_j \mid \lambda_j, \tau &\sim N\left(0, \lambda_j^2 \tau^2\right) \\ \lambda_j &\sim \text{Cauchy}^+(0,1) \\ \tau &\sim \text{Cauchy}^+(0,\tau_0) \end{aligned} \tag{5}$$

where $\text{Cauchy}^+(0,a)$ is the standard half-Cauchy distribution defined on R^+ with scale parameter a. λ_j is the local shrinkage parameter and τ_0 is the global shrinkage parameter.

Horseshoe+ Method

The Horseshoe+ prior(Bhadra et al., 2017) was proposed based on Horseshoe. It extends the horseshoe prior by introducing an additional layer of hierarchy to further adapt the shrinkage behavior. Assuming the model as 1 after standardization for each term, the horseshoe+ prior for the cis-eQTL weights is:

$$\begin{aligned} w_j \mid \lambda_j, \tau &\sim N\left(0, \nu_j^2 \lambda_j^2 \tau^2\right) \\ \nu_j &\sim \text{Cauchy}^+(0,1) \\ \lambda_j &\sim \text{Cauchy}^+(0,1) \\ \tau &\sim \text{Cauchy}^+(0,\tau_0) \end{aligned} \tag{6}$$

here ν_j is a new auxiliary shrinkage parameter, allowing finer control of shrinkage at the local level, offering greater adaptability for different sparsity patterns. Since the SNP signals are usually very sparse in the genes, Horseshoe and Horseshoe+ are naturally a good choice for imputing the cis-eQTL weights in the first stage of TWAS. The solution of Horseshoe and Horseshoe+ models could be attained by reformulating them into hierarchical models, and calculating the posterior estimation through Monte Carlo Markov Chain(MCMC).

A-TWAS and ACAT Test

We introduce Aggregated-TWAS (A-TWAS), a framework that integrates multiple Bayesian models to more effectively capture the complex and diverse relationships between genotype and gene expression. A-TWAS first involves training imputation models using Bayesian Lasso, Horseshoe, and Horseshoe+ for each gene. Next, the association between genotype and phenotype is tested using Equation 2 for each set of cis-eQTL weights, resulting in a set of TWAS p-values corresponding to each Bayesian method.

To integrate the p-values obtained from the three Bayesian imputation methods, we employ ACAT (Liu et al., 2019; Liu and Xie, 2020). Let p_i represent the p-value from the i-th Bayesian method, and let d denote the total number of Bayesian methods used for integration. The Cauchy combination test statistics of ACAT is defined as:

$$T = \sum_{i=1}^{d} \tan\left\{(0.5 - p_i)\,\pi\right\}/d \tag{7}$$

Since the null distribution of T can be well approximated by a Cauchy distribution under arbitrary dependency structures for p_i, the combined p-value for three Bayesian methods can be calculated as $0.5 - \arctan(T)/\pi$. This combined p-value returned by the ACAT test is just the TWAS p-value for our A-TWAS method.

ROS/MAP Dataset

For the real data analysis, we performed the first imputation step of TWAS on the Religious Orders Study and Memory and Aging Project (ROS/MAP) dataset(Bennett et al., 2012)(Bennett et al., 2018). We used the R package "bigsnpr" to preprocess the genotype data, retaining variants with a minor allele frequency (MAF) $> 5\%$ and a Hardy-Weinberg equilibrium Fisher's exact test p-value $< 1 \times 10^{-7}$. With respect to the gene expression data, for a specific gene, if more than 70% of the samples in the cohort had low FPKM values (the threshold is set to 1 in this study) for a given gene, we excluded that gene from the TWAS analysis. Furthermore, we used linear regression to adjust for confounding factors, including sex, age, study cohort (ROS or MAP), and the top three principal components of the genotype data. After data preprocessing, a total of 10,940 genes were included for the imputation step of the TWAS analysis.

Simulation Study

To evaluate the performance of Bayesian Lasso, Horseshoe and Horseshoe+, together with their aggregated performance(A-TWAS), we conducted simulation experiments to compare their test R^2, TWAS power, and type-I error. We utilized genotype information from the ROS/MAP dataset by using a randomly selected gene $CCDC104$ from chromosome 2 to construct the genotype matrix in simulation, while the gene expression and

phenotype data were artificially simulated. 50% of the samples were randomly selected for training, and the remaining 50% were used for evaluating the imputation performance by test R^2.

We varied the fraction of causal SNPs p_{cs} (0.01, 0.05, 0.1, 0.3) and gene expression heritability h_e^2 (0.01, 0.05, 0.1) to simulate gene expression data, where the gene expression heritability refers to the proportion of variance explained by causal SNPs. Under the model 1, $\boldsymbol{\epsilon}_g \sim N\left(0, \left(1-h_e^2\right)\mathbf{I}\right)$, the effect size for causal SNPs are sampled from $N(0,1)$, and then we re-scaled the effect size to ensure the expression heritability to be h_e^2.

In order to access the performance of TWAS power, we assumed the phenotype data are simulated by

$$\mathbf{Y} = h_p \mathbf{X}_g \mathbf{w} + \epsilon_p, \quad \epsilon_p \sim N(0, \mathbf{I})$$

where h_p^2 is the phenotypic heritability, namely the proportion of phenotypic variance explained by gene expression. We set $h_p^2 = 0.1$ in this study. Under this model assumption for the phenotype data, GWAS z-scores of the summary statistics used in stage II of TWAS are generated from a multivariate normal distribution $\mathbf{Z} \sim \text{MVN}\left(\boldsymbol{\Sigma}_g \mathbf{w}\sqrt{n_{\text{gwas}} h_p^2}, \boldsymbol{\Sigma}_g\right)$ (Feng et al., 2021), where $\boldsymbol{\Sigma}_g$ is the correlation matrix of the standardized genotype matrix, and n_{gwas} is the assumed GWAS sample size, which is varied with value $n_{\text{gwas}} = (20000, 50000, 80000)$ in this study. Each simulation scenario was repeated 500 times.

To evaluate the type I error of each method, we assumed the phenotype was independent of gene expression and generated the phenotype directly from $N(0,1)$. For each (p_{cs}, h_e^2) scenario, the phenotype was simulated with 500 replications.

Simulation Result

For the simulation results, A-TWAS achieves the highest statistical power under most scenarios when the significance threshold for the power calculation is $p = 0.05/500 = 1 \times 10^{-4}$ (adjusted using the Bonferroni correction). Details of the simulation results can be found in Supplementary Figs. S1, S2, S3, and S4. A-TWAS robustly outperforms the baseline S-PrediXcan method and is slightly inferior to other individual Bayesian methods only in a few cases, highlighting the effectiveness of A-TWAS.

For the individual Bayesian methods, they consistently achieve higher power than S-PrediXcan. When $h_e^2 = 0.01$, Horseshoe and Horseshoe+ outperform Bayesian Lasso, but their performance achieve a similar level when $h_e^2 = 0.05$ and 0.1. All methods exhibit comparable performance in test R^2. Only when $p_{cs} \geq 0.05$ and $h_e^2 \geq 0.05$, the Bayesian methods perform slightly better than S-PrediXcan, which employs the frequentist Elastic-net.

Regarding type-I error, the error rate is defined as the proportion of TWAS p-values falling below the threshold $p = 0.05$. Among the methods, the type-I error rates for all Bayesian methods are similar, while the rate for S-PrediXcan is stable and slightly lower than that of the Bayesian methods. Furthermore, the aggregation of Bayesian methods (A-TWAS) increases power without inflating the type-I error. Overall, all methods maintain an acceptable type-I error rate.

Real Application

To evaluate the performance of A-TWAS and the individual Bayesian methods on real data, we applied these methods, along with S-PrediXcan as a baseline, to identify significant genes. The ROS/MAP dataset was used to train GReX imputation models, and GWAS summary data for Schizophrenia (Pardiñas et al., 2018) and Obsessive-Compulsive Symptoms (Strom et al., 2024) from the Psychiatric Genomics Consortium (PGC) were utilized for $\widehat{\textbf{GReX}}$ and phenotype association testing.

In stage I of TWAS, the genotype data used to train the GReX imputation model was derived from the intersection of SNPs in both the ROS/MAP dataset (quality-controlled as described in Section 2.7) and the GWAS summary statistics for the disease. The gene expression data also came from the ROS/MAP dataset with appropriate adjustments (detailed in Section 2.7). Half of the samples were used for training, and the other half were reserved for testing.

In stage II of TWAS, the genotype information for the entire sample set was used as the reference LD panel. Imputation models with test $R^2 > 0.01$ were deemed effective and selected for the subsequent stage of TWAS. If the imputation model was null, both its train R^2 and test R^2 were set to 0.

Table 1. Effective Imputation Model Counts (test $R^2 > 0.01$).

Method	Disease	
	Schizophrenia	OCS
S-PrediXcan	5341	5437
BLasso	6330	6432
Horseshoe	6408	6546
Horseshoe+	6410	6543
A-TWAS	7411	7417

Schizophrenia

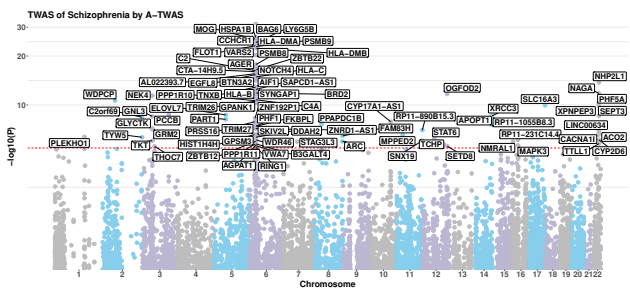


Fig. 1. Manhattan plot for Schizophrenia.

The Schizophrenia GWAS summary dataset (Pardiñas et al., 2018) contains 11,260 cases and 24,542 controls from the UK. We calculated its heritability and the proportion of causal SNPs using the two-component model of GENESIS (Zhang et al., 2018) and found that the heritability (h^2) and p_{cs} of this dataset are 0.707 and 0.0142, respectively. The number of genes containing SNP information in the Schizophrenia GWAS summary data is 10,893. Among these genes, 7,411 imputation models qualified for TWAS analysis using A-TWAS, and more than 6,000 models are qualified for Bayesian methods, compared to 5,341 models in S-PrediXcan (Table 1).

Figure S5 shows the train R^2 and test R^2 results for Schizophrenia. The train R^2 values for Bayesian methods are significantly higher than those for S-PrediXcan. For test R^2,

all methods perform similarly to each other, with Horseshoe and Horseshoe+ slightly outperforming the others, which aligns with our simulation results.

In the TWAS results, A-TWAS identified 89 significant genes with p-values reaching the Bonferroni-corrected significance level, the highest among all methods. Given the 10,893 genes containing SNP information in the Schizophrenia dataset, the corrected p-value threshold is 4.59×10^{-6}. The number of significant genes identified by other methods is as follows: S-PrediXcan 75, BLasso 43, Horseshoe 79, and Horseshoe+ 77. The Manhattan plot for A-TWAS results is shown in Fig. 1, while Manhattan plots for other methods are provided in Supplementary Fig. S6.

Among the genes identified by A-TWAS, 24 are listed in the GWAS Catalog under the trait "Schizophrenia", the most among all methods, followed by Horseshoe method (S-PrediXcan 23, BLasso 12, Horseshoe 20, Horseshoe+ 18). Details of these results are presented in Supplementary Table S1. Enrichment analysis of the significant genes identified by A-TWAS successfully identified the Schizophrenia pathway (KEGG disease) with a p-value of 0.0295, further demonstrating the efficacy of A-TWAS.

Obsessive-Compulsive Symptoms

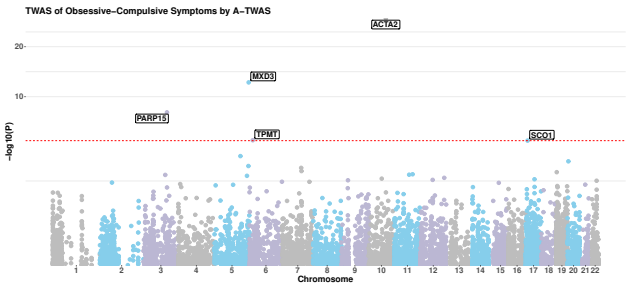


Fig. 2. Manhattan plot for Obsessive-Compulsive Symptoms.

The study of Obsessive-Compulsive Symptoms (OCS) includes 33,943 samples of European ancestry (Strom et al., 2024). The GWAS results from this study found no SNPs or genes significantly associated with OCS, as none reached the Bonferroni-corrected significance level. Using GENESIS, the p_{cs} and heritability for this GWAS summary data were calculated to be 8.592×10^{-4} and 0.0133, respectively. There are 11,014 genes containing SNP information in the OCS dataset. Among these genes, 5,437 attained valid imputation models using S-PrediXcan, which is fewer than the 7,417 identified by A-TWAS and >6,000 models identified by Bayesian methods (Table 1).

The patterns of train R^2 and test R^2 (Supplementary Fig. S8) for the OCS data are similar to those observed for Schizophrenia, as the training sources for the imputation models are largely the same. The only difference arises from variations in the SNPs recorded in the two sets of summary statistics.

In the TWAS results, A-TWAS successfully identified five new significant genes for OCS (Fig. 2), while S-PrediXcan identified none. The corrected p-value threshold is 4.54×10^{-6}. Manhattan plots for the other methods are shown in Supplementary Fig. S9, and the p-values for the significant genes are provided in Supplementary Table S2.Overall, these results from real data demonstrate the effectiveness of A-TWAS.

Conclusion and Discussion

In this work, A-TWAS was proposed to better capture the complex relationship between genotype and gene expression by combining multiple imputation models with different Bayesian priors. It enhances the detection of genetic risk factors by integrating various Bayesian models, improving power over the baseline method. We chose Bayesian Lasso, Horseshoe, and Horseshoe+ as the individual imputation models for combination after proving their efficacy. Detailed simulation experiments and real data analysis were performed for A-TWAS, together with each single Bayesian method and the baseline method S-PrediXcan, to investigate their performance.

In the simulation study, A-TWAS robustly outperformed other methods regardless of the underlying genetic architecture of the data, demonstrating the advantage of the omnibus method in capturing a more comprehensive genetic structure compared to single methods. For the single methods, Horseshoe and Horseshoe+ performed better than Bayesian Lasso when the causal SNPs in the gene were sparse. All these methods outperform the baseline S-PrediXcan. For test R^2, no significant differences were observed between the methods. Regarding type-I error, although A-TWAS and all Bayesian methods did not reach the performance of S-PrediXcan, when considering the balance between power and type-I error, all of these methods achieved relatively well-controlled type-I error rates.

For the real data analysis, GWAS summary data for the diseases Schizophrenia and Obsessive-Compulsive Symptoms (OCS) were adopted to test the performance of the methods. Among the methods, A-TWAS showed its advantage in that it could attain most of the imputation models enabling the next stage of TWAS, since it is the union of results attained by the three Bayesian methods. In Schizophrenia, A-TWAS successfully identified most of the significant genes; among the significant genes, the set of gene detected by A-TWAS includes the most that have been proven to be related to Schizophrenia and reported in the GWAS Catalog. The significant genes selected by A-TWAS also successfully identified the Schizophrenia disease pathway through enrichment analysis. For single Bayesian methods, although slightly inferior to A-TWAS, Horseshoe and Horseshoe+ have a comparable performance which is better than other single methods regarding these criteria. In the analysis of OCS, this group of GWAS summary data originally reported no significant SNPs or genes. By incorporating this data into TWAS methods, no significant genes were detected by S-PrediXcan and Bayesian Lasso. However, A-TWAS successfully identified 5 genes (Horseshoe 5 and Horseshoe+ 2). All of these evidences prove the superiority of using the continuous shrinkage prior Horseshoe and Horseshoe+ for imputation model, and the efficacy of the omnibus method A-TWAS.

From the results obtained in our work, A-TWAS clearly demonstrates its ability in identifying gene-trait relationships. However, since all Bayesian methods obtain their solutions using MCMC, which requires a large number of iterations, the computation speed is much slower than the frequentist S-PrediXcan, limiting the efficiency of A-TWAS. For the implementation of A-TWAS (GitHub website: `https://github.`

com/Yilan-Liang/A-TWAS.), our code provides functions for calculating imputation based on an individual Bayesian method or any combination of the Bayesian methods, enabling the trade-off between power and efficiency. Either the Horseshoe or Horseshoe+ model is recommended if only an individual model is required, as they perform robustly better than Bayesian Lasso.

Code and Data Availability

Code Availability

A-TWAS: https://github.com/Yilan-Liang/A-TWAS.
S-PrediXcan: https://github.com/hakyimlab/MetaXcan.
bayesreg (the R package bayesreg support the calculation of Bayesian Lasso, Horseshoe and Horseshoe+): https://cran.r-project.org/web/packages/bayesreg/index.html.

Data Availability

The data that support the findings of this study can be downloaded from the following sources:
All protected data of the GTEx project are available through the database of Genotypes and Phenotypes (dbGaP) (accession number phs000424.v8.p2) at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v8.p2.
ROSMAP: https://www.synapse.org/#!Synapse:syn3219045.
GWAS summary statistics for Schizophrenia: https://figshare.com/articles/dataset/scz2018clozuk/14681220.
GWAS summary statistics for Obsessive-Compulsive Symptoms: https://figshare.com/articles/dataset/ocs2024/26352157.

Author contributions statement

Y.D.Z. conceived and designed the model. Y.L and H.W. designed the algorithm, implemented the software, conducted analysis of simulation and real data, and wrote the original draft of the manuscript. Y.D.Z. supervised the study and contributed to the critical revision of the manuscript. Y.L. and H.W. made equal contributions.

Acknowledgments

This work was supported, in part, by Hong Kong Research Grants Council General Research Fund [17307324]. 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.

References

Alvaro N Barbeira, GTEx Consortium, Scott P Dickinson, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.*, 9(1), December 2018.

David A Bennett, Julie A Schneider, Zoe Arvanitakis, and Robert S Wilson. Overview and findings from the religious orders study. *Curr. Alzheimer Res.*, 9(6):628–645, July 2012.

David A Bennett, Aron S Buchman, Patricia A Boyle, et al. Religious orders study and rush memory and aging project. *J. Alzheimers. Dis.*, 64(s1):S161–S189, 2018.

Anindya Bhadra, Jyotishka Datta, Nicholas G Polson, and Brandon Willard. The horseshoe+ estimator of ultra-sparse signals. *Bayesian Anal.*, 12(4):1105–1131, December 2017.

Arjun Bhattacharya, Montserrat García-Closas, Andrew F Olshan, et al. A framework for transcriptome-wide association studies in breast cancer in diverse study populations. *Genome Biol.*, 21(1):42, February 2020.

Yohan Bossé, Zhonglin Li, Jun Xia, et al. Transcriptome-wide association study reveals candidate causal genes for lung cancer. *Int. J. Cancer*, 146(7):1862–1878, April 2020.

Carlos M. Carvalho, Nicholas G. Polson, and James G. Scott. Handling sparsity via the horseshoe. In David van Dyk and Max Welling, editors, *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 73–80, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 16–18 Apr 2009. PMLR. URL https://proceedings.mlr.press/v5/carvalho09a.html.

Helian Feng, Nicholas Mancuso, Alexander Gusev, et al. Leveraging expression from multiple tissues using sparse canonical correlation analysis and aggregate tests improves the power of transcriptome-wide association studies. *PLoS Genet.*, 17(4):e1008973, April 2021.

Naomi A Fineberg, Michael P Hengartner, Carmel E Bergbaum, et al. A prospective population-based cohort study of the prevalence, incidence and impact of obsessive-compulsive symptomatology. *Int. J. Psychiatry Clin. Pract.*, 17(3):170–178, August 2013.

Eric R Gamazon, Heather E Wheeler, Kaanan P Shah, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.*, 47(9):1091–1098, September 2015.

Alexander Gusev, Arthur Ko, Huwenbo Shi, Gaurav Bhatia, Wonil Chung, Brenda W J H Penninx, Rick Jansen, Eco J C de Geus, Dorret I Boomsma, Fred A Wright, Patrick F Sullivan, Elina Nikkola, Marcus Alvarez, Mete Civelek, Aldons J Lusis, Terho Lehtimäki, Emma Raitoharju, Mika Kähönen, Ilkka Seppälä, Olli T Raitakari, Johanna Kuusisto, Markku Laakso, Alkes L Price, Päivi Pajukanta, and Bogdan Pasaniuc. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.*, 48(3):245–252, March 2016a.

Alexander Gusev, Arthur Ko, Huwenbo Shi, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.*, 48(3):245–252, March 2016b.

Alexander Gusev, Nicholas Mancuso, Hyejung Won, et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat. Genet.*, 50(4):538–548, April 2018.

Janet C Harwood, Ganna Leonenko, Rebecca Sims, et al. Defining functional variants associated with alzheimer's disease in the induced immune response. *Brain Commun.*, 3(2):fcab083, April 2021.

Yaowu Liu and Jun Xie. Cauchy combination test: A powerful test with analytic p-value calculation under arbitrary dependency structures. *J. Am. Stat. Assoc.*, 115(529):393–402, January 2020.

Yaowu Liu, Sixing Chen, Zilin Li, Alanna C Morrison, Eric Boerwinkle, and Xihong Lin. ACAT: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. *Am. J. Hum. Genet.*, 104(3):410–421, March 2019.

Nicholas Mancuso, Simon Gayther, Alexander Gusev, et al. Large-scale transcriptome-wide association study identifies

new prostate cancer risk regions. *Nat. Commun.*, 9(1):4079, October 2018.

Sini Nagpal, Xiaoran Meng, Michael P Epstein, et al. TIGAR: An improved bayesian tool for transcriptomic data imputation enhances gene mapping of complex traits. *Am. J. Hum. Genet.*, 105(2):258–266, August 2019.

Antonio F Pardiñas, Peter Holmans, Andrew J Pocklington, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.*, 50(3):381–389, March 2018.

Trevor Park and George Casella. The bayesian lasso. *J. Am. Stat. Assoc.*, 103(482):681–686, June 2008.

Nora I Strom, Christie L Burton, Conrad Iyegbe, et al. Genome-Wide association study of Obsessive-Compulsive symptoms including 33,943 individuals from the general population. *Mol. Psychiatry*, 29(9):2714–2723, September 2024.

Yanfa Sun, Jingjing Zhu, Dan Zhou, et al. A transcriptome-wide association study of alzheimer's disease using prediction models of relevant tissues identifies novel candidate susceptibility genes. *Genome Med.*, 13(1):141, September 2021.

Emil Uffelmann, Qin Qin Huang, Nchangwi Syntia Munung, et al. Genome-wide association studies. *Nat. Rev. Methods Primers*, 1(1), August 2021.

Lang Wu, Wei Shi, Jirong Long, et al. A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. *Nat. Genet.*, 50(7): 968–978, July 2018.

Yan Zhang, Guanghao Qi, Ju-Hyun Park, and Nilanjan Chatterjee. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat. Genet.*, 50(9): 1318–1326, September 2018.

Hui Zou and Trevor Hastie. Addendum: Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Series B Stat. Methodol.*, 67(5):768–768, November 2005.

Supplemental Figures

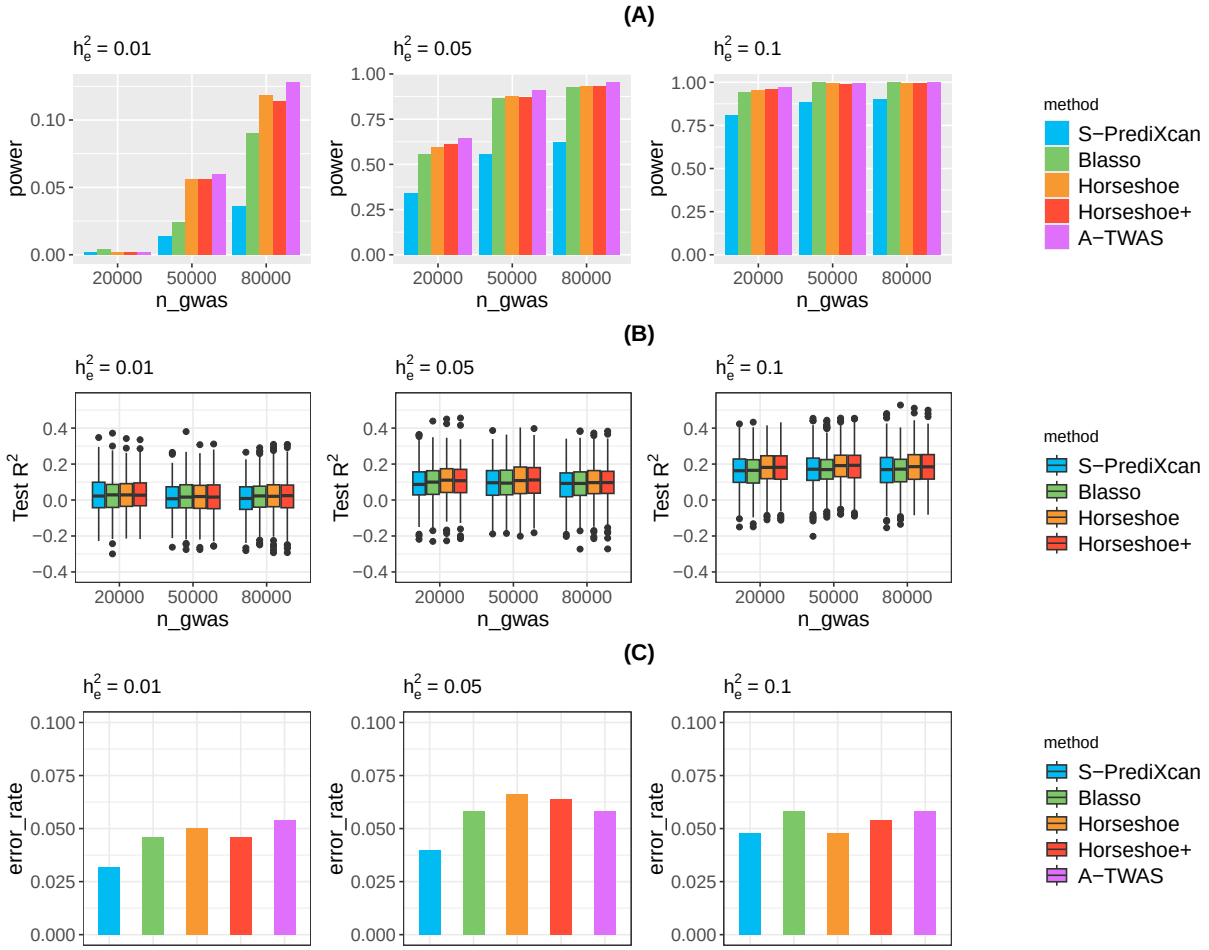


Fig. S1. simulation result under $p_{cs} = 0.01$ scenario. (A) TWAS power with threshold of p-value $< 10^{-4}$ (Bonferroni corrected significance level) vary by h_e and GWAS sample size given by S-PrediXcan and all proposed Bayesian method. **(B)** The test R^2 of S-PrediXcan, Bayesian Lasso, Horseshoe and Horseshoe+ in stage I TWAS vary by h_e and GWAS sample size. **(C)** The box plot of the p-value under the simulation of type-one error evaluation.

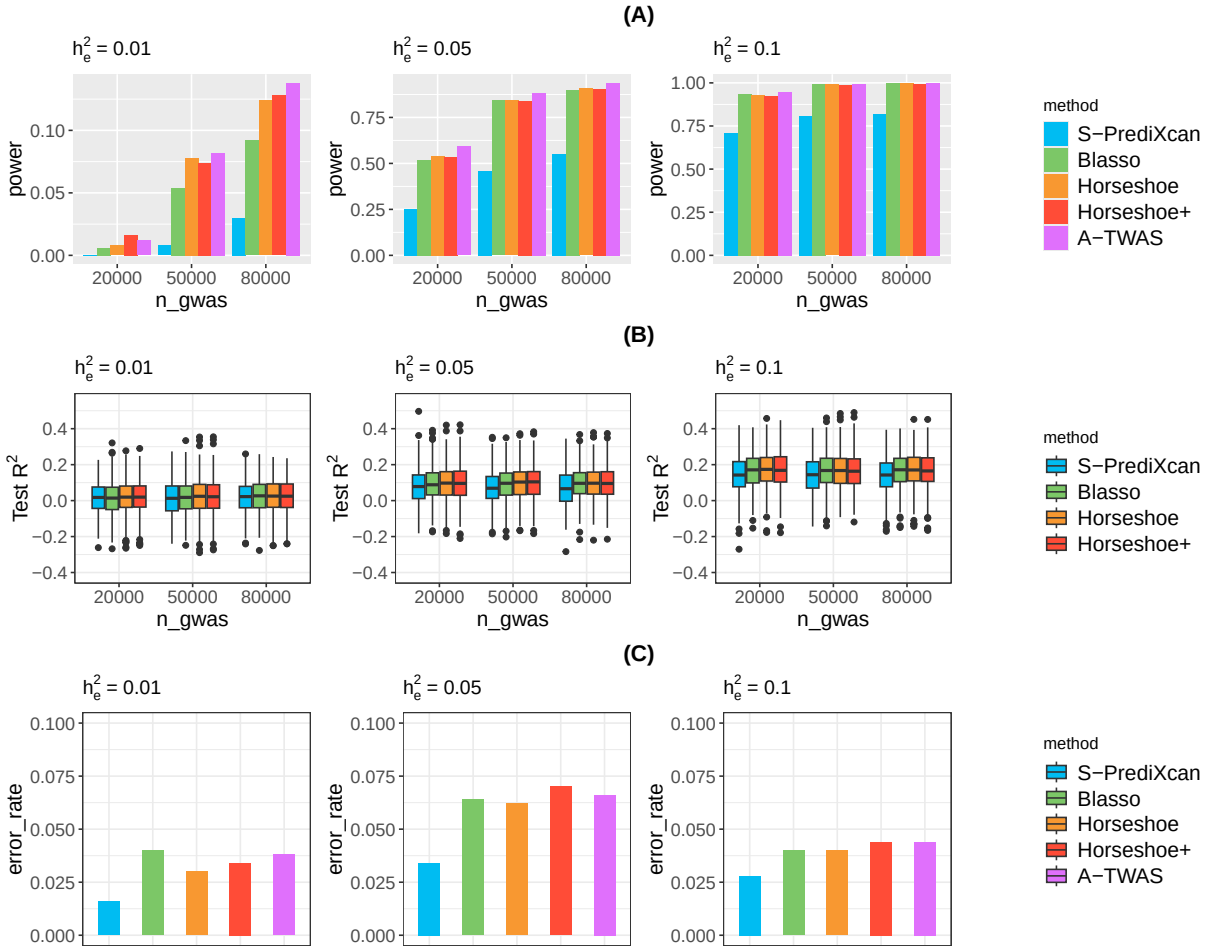


Fig. S2. simulation result under $p_{cs} = 0.05$ scenario. (A) TWAS power with threshold of p-value $< 10^{-4}$ (Bonferroni corrected significance level) vary by h_e and GWAS sample size given by S-PrediXcan and all proposed Bayesian method. **(B)** The test R^2 of S-PrediXcan, Bayesian Lasso, Horseshoe and Horseshoe+ in stage I TWAS vary by h_e and GWAS sample size. **(C)** The box plot of the p-value under the simulation of type-one error evaluation.

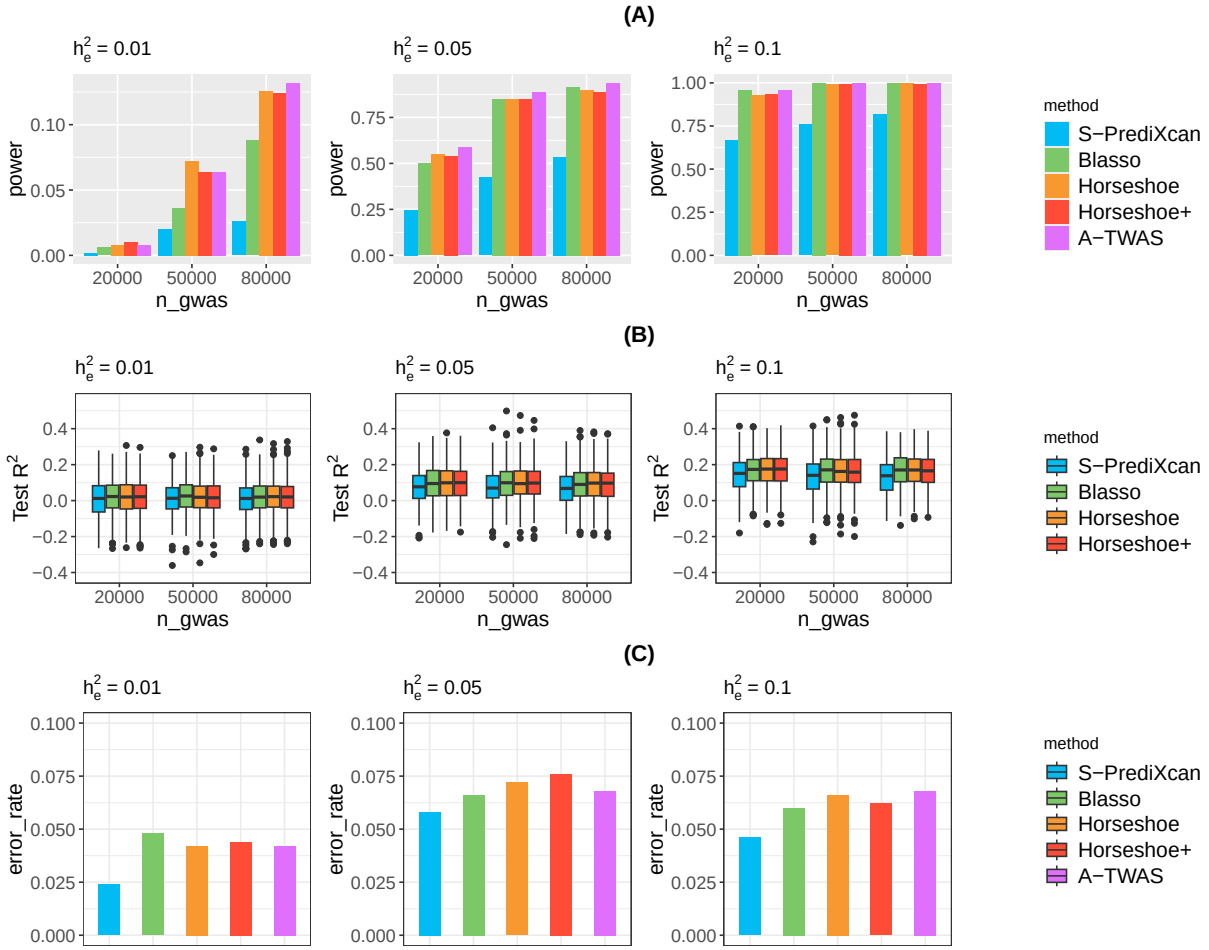


Fig. S3. simulation result under $p_{cs} = 0.1$ scenario. (A) TWAS power with threshold of p-value $< 10^{-4}$ (Bonferroni corrected significance level) vary by h_e and GWAS sample size given by S-PrediXcan and all proposed Bayesian method. **(B)** The test R^2 of S-PrediXcan, Bayesian Lasso, Horseshoe and Horseshoe+ in stage I TWAS vary by h_e and GWAS sample size. **(C)** The box plot of the p-value under the simulation of type-one error evaluation.

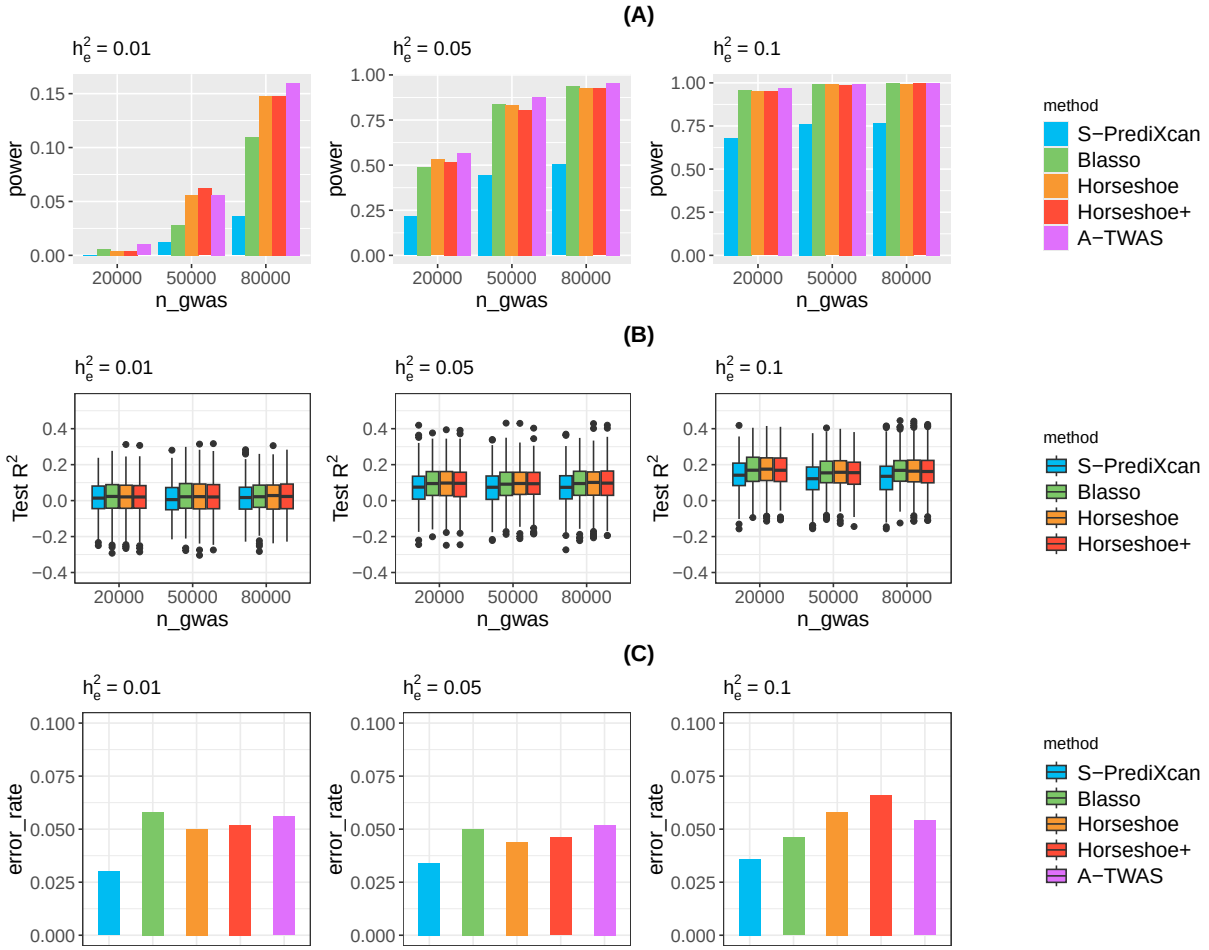
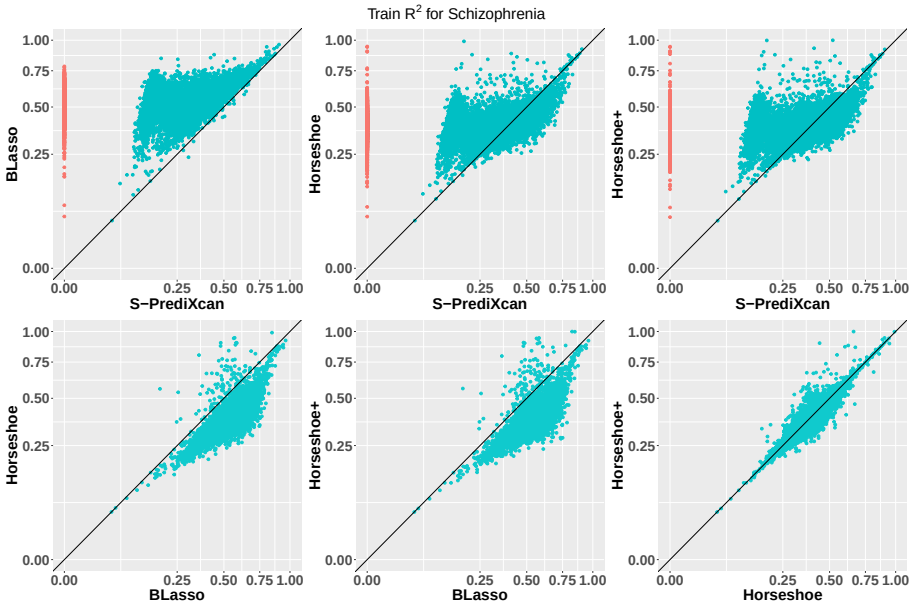
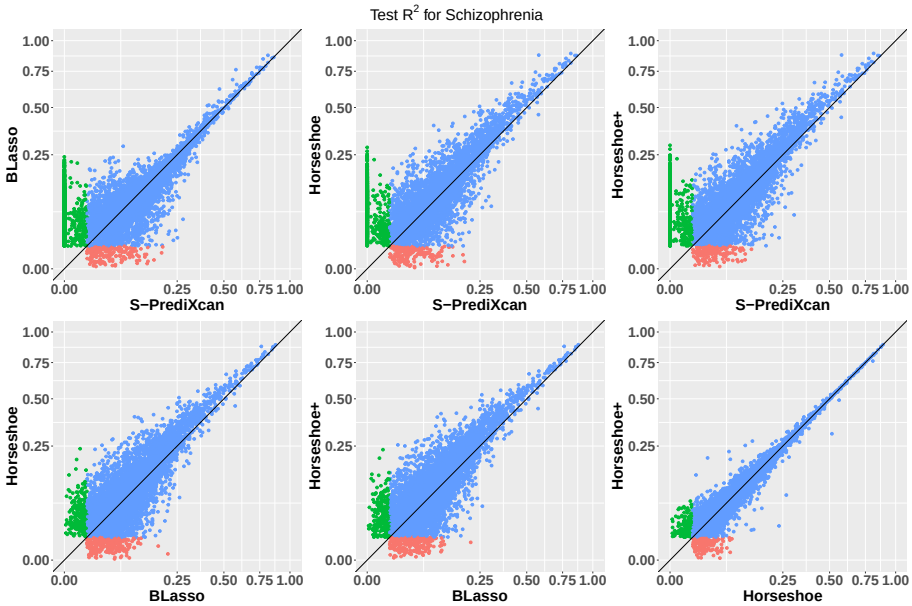


Fig. S4. simulation result under $p_{cs} = 0.3$ scenario. (A) TWAS power with threshold of p-value $< 10^{-4}$ (Bonferroni corrected significance level) vary by h_e and GWAS sample size given by S-PrediXcan and all proposed Bayesian method. **(B)** The test R^2 of S-PrediXcan, Bayesian Lasso, Horseshoe and Horseshoe+ in stage I TWAS vary by h_e and GWAS sample size. **(C)** The box plot of the p-value under the simulation of type-one error evaluation.

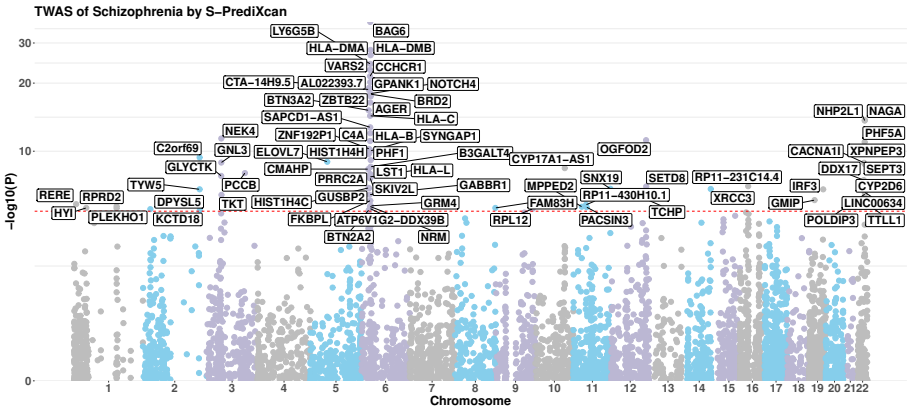


(a)

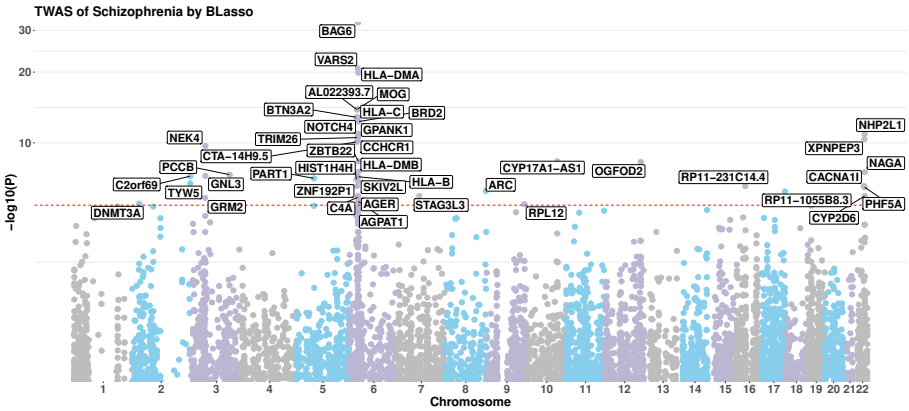


(b)

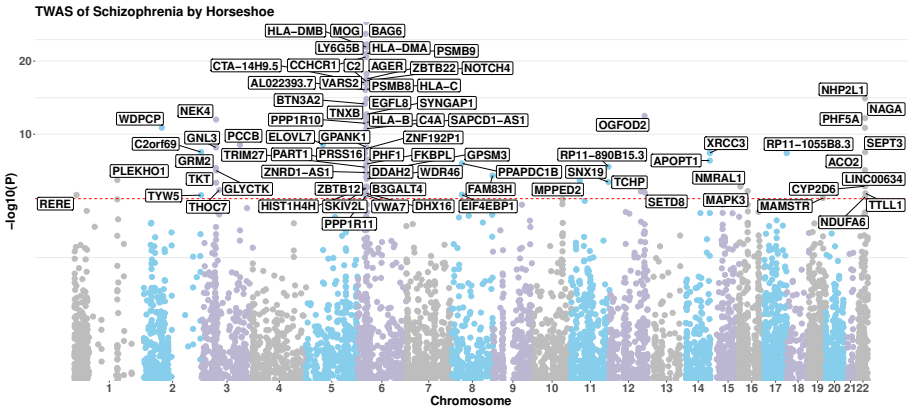
Fig. S5. testR^2 and trainR^2 for Schizophrenia. In train R^2, the red points indicating models with $R^2 > 0.01$ for method on y-axis while < 0.01 for method on x-axis; the blue points indicating models with $R^2 > 0.01$ for methods on both axis. In test R^2, the red points indicating models with $R^2 > 0.01$ for method on x-axis while < 0.01 for method on y-axis; the green points indicating models with $R^2 > 0.01$ for method on y-axis while < 0.01 for method on x-axis; and the green points indicating models with $R^2 > 0.01$ for methods on both axis.


TWAS of Schizophrenia by S−PrediXcan
−log10(P)
Chromosome
LY6G5B
BAG6
HLA−DMA
HLA−DMB
VARS2
CCHCR1
CTA−14H9.5
AL022393.7
GPANK1
NOTCH4
BTN3A2
ZBTB22
BRD2
AGER
SAPCD1−AS1
HLA−C
NEK4
ZNF192P1
C4A
HLA−B
SYNGAP1
C2orf69
GNL3
ELOVL7
HIST1H4H
PHF1
B3GALT4
CYP17A1−AS1
OGFOD2
GLYCTK
CMAHP
LST1
HLA−L
TYW5
PCCB
PRRC2A
SKIV2L
GABBR1
MPPED2
SNX19
SETD8
RP11−231C14.4
RERE
RPRD2
DPYSL5
TKT
HIST1H4C
GUSBP2
GRM4
FAM83H
RP11−430H10.1
XRCC3
HYI
PLEKHO1
KCTD18
FKBPL
ATP6V1G2−DDX39B
RPL12
PACSIN3
TCHP
BTN2A2
NRM
NHP2L1
NAGA
PHF5A
CACNA1I
XPNPEP3
DDX17
SEPT3
IRF3
CYP2D6
GMIP
LINC00634
POLDIP3
TTLL1

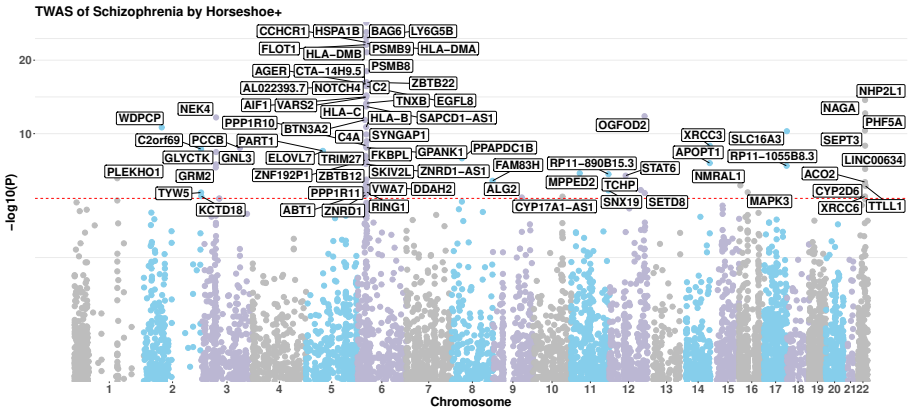
(a)


TWAS of Schizophrenia by BLasso
−log10(P)
Chromosome
BAG6
VARS2
HLA−DMA
AL022393.7
MOG
BTN3A2
HLA−C
BRD2
NOTCH4
GPANK1
NEK4
TRIM26
CCHCR1
ZBTB22
CTA−14H9.5
HLA−DMB
PCCB
HIST1H4H
PART1
CYP17A1−AS1
OGFOD2
C2orf69
GNL3
TYW5
ZNF192P1
SKIV2L
HLA−B
ARC
RP11−231C14.4
DNMT3A
GRM2
C4A
AGER
STAG3L3
RPL12
AGPAT1
NHP2L1
XPNPEP3
NAGA
CACNA1I
RP11−1055B8.3
PHF5A
CYP2D6

(b)



(c)



(d)

Fig. S6. Manhattan plots for Schizophrenia

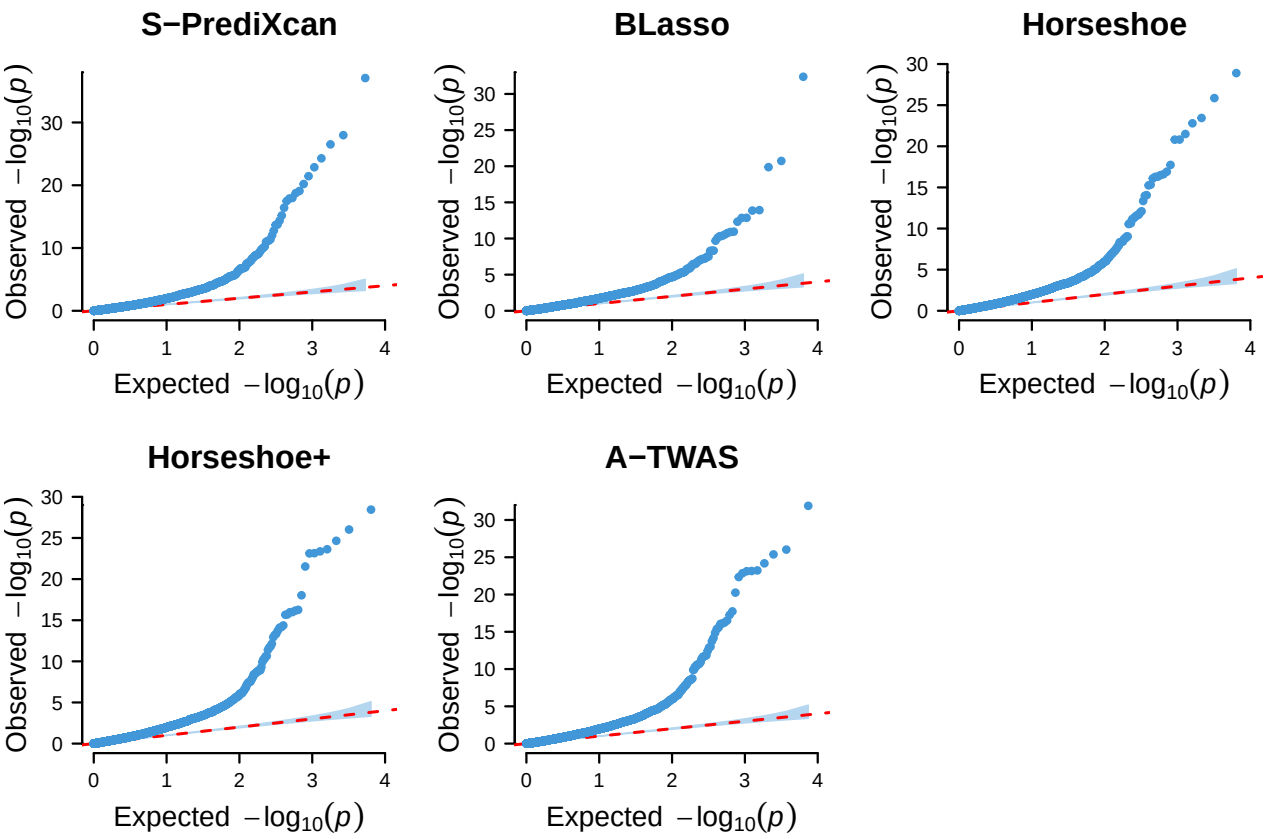
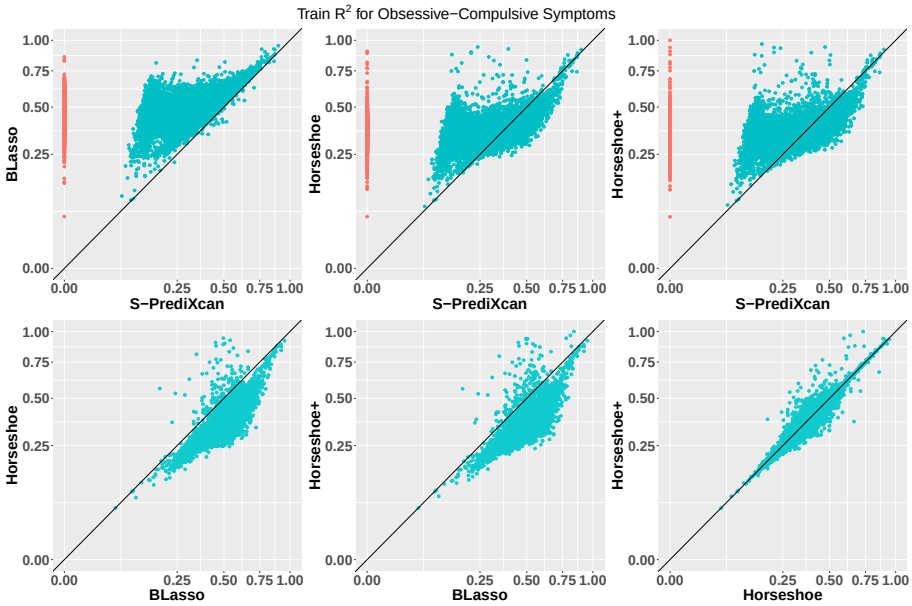
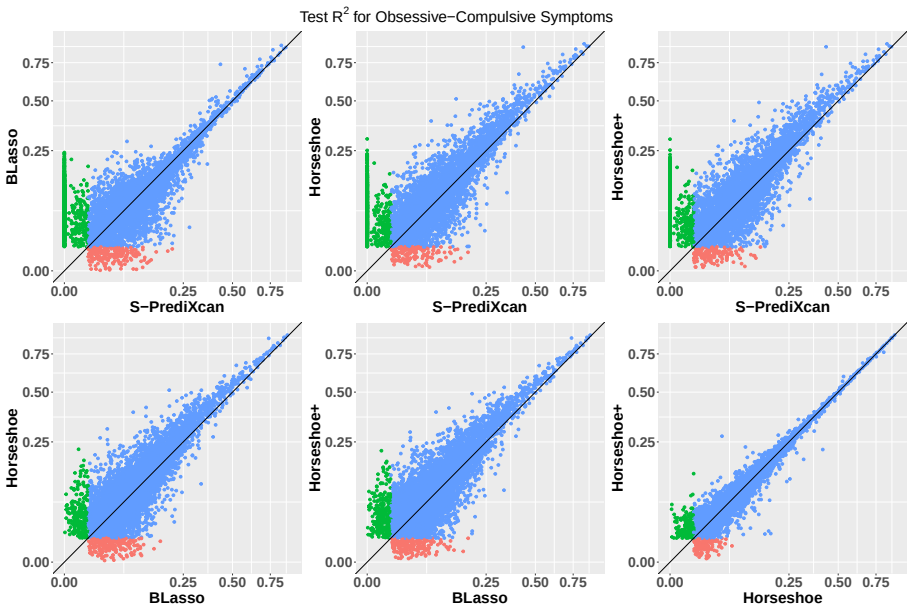


Fig. S7. Quantile−Quantile plot of TWAS p−values under uniform distribution from Schizophrenia

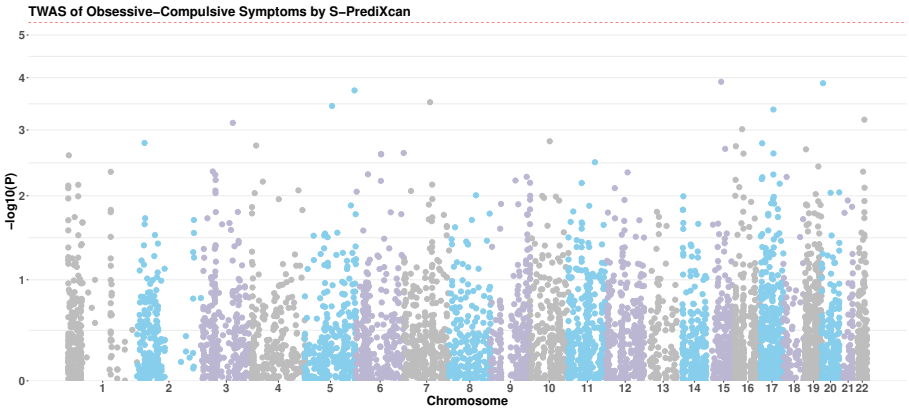


(a)

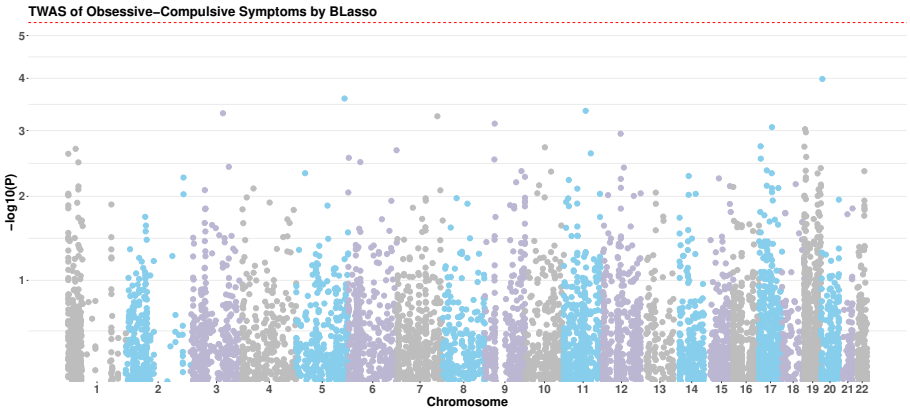


(b)

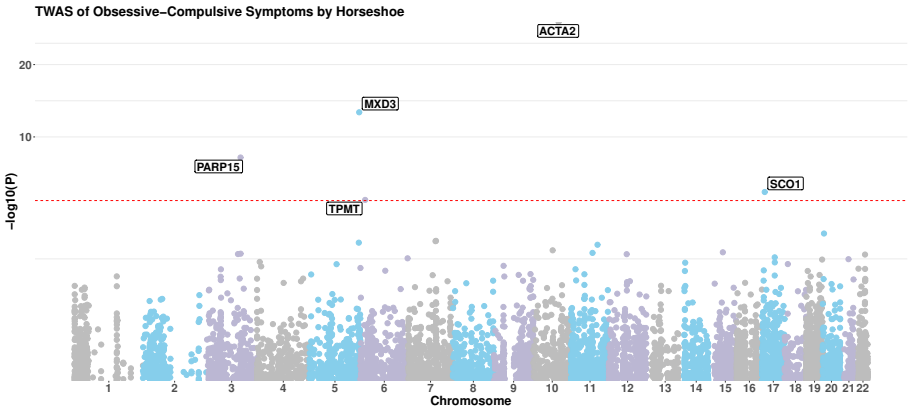
Fig. S8. testR^2 and trainR^2 for Obsessive-Compulsive Symptoms. In train R^2, the red points indicating models with $R^2 > 0.01$ for method on y-axis while < 0.01 for method on x-axis; the blue points indicating models with $R^2 > 0.01$ for methods on both axis. In test R^2, the red points indicating models with $R^2 > 0.01$ for method on x-axis while < 0.01 for method on y-axis; the green points indicating models with $R^2 > 0.01$ for method on y-axis while < 0.01 for method on x-axis; and the green points indicating models with $R^2 > 0.01$ for methods on both axis.


TWAS of Obsessive−Compulsive Symptoms by S−PrediXcan
−log10(P)
Chromosome

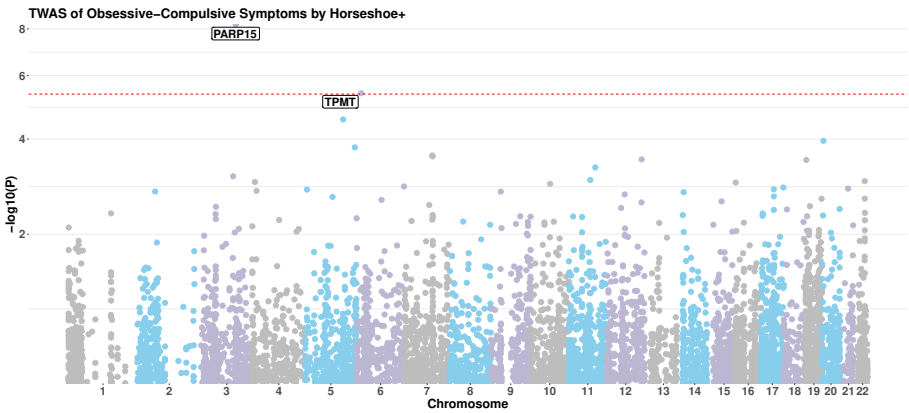
(a)


TWAS of Obsessive−Compulsive Symptoms by BLasso
−log10(P)
Chromosome

(b)



(c)



(d)

Fig. S9. Manhattan plots for Obsessive−Compulsive Symptoms

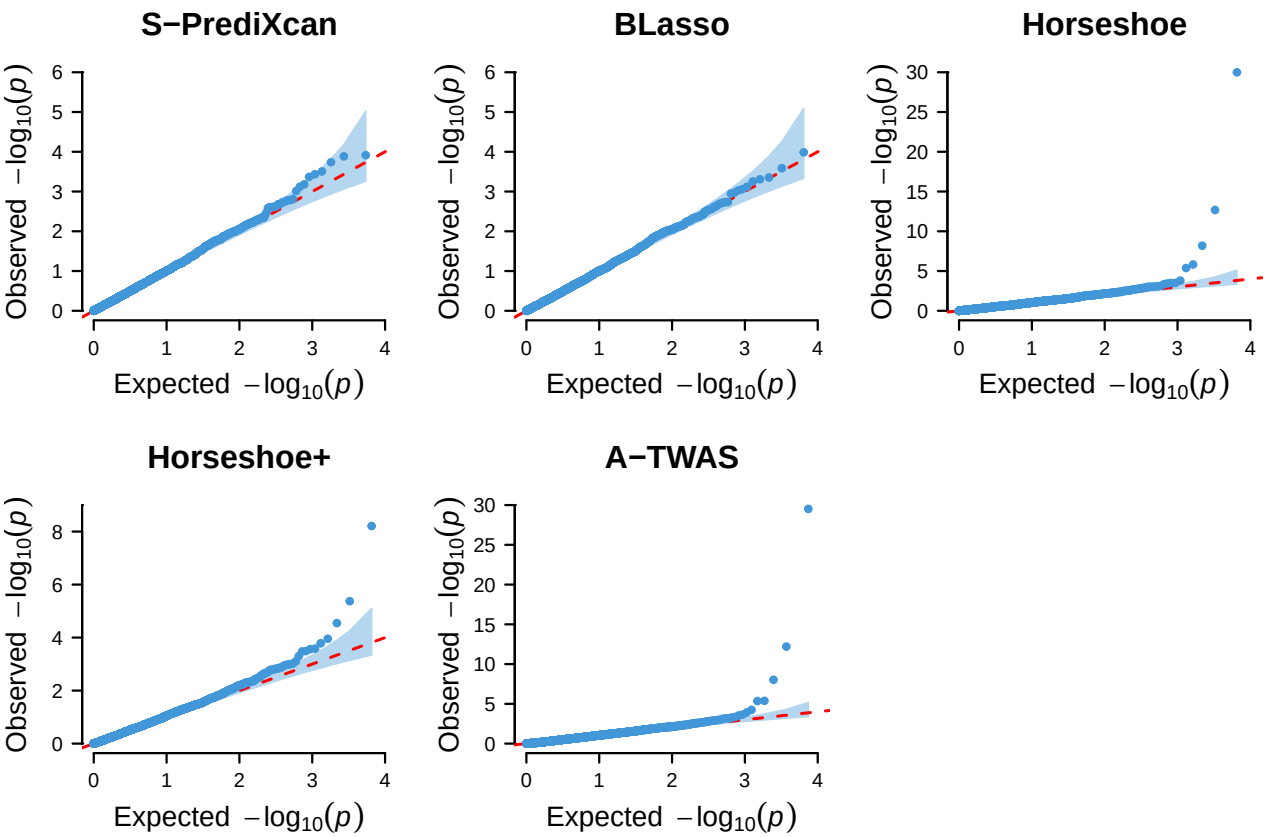


Fig. S10. Quantile−Quantile plot of TWAS p−values under uniform distribution from Obsessive−Compulsive Symptoms

Supplemental Tables

Table S1. Significant Genes with TWAS p-value less than Bonferroni corrected significance level for Schizophrenia.

CHR	GENE	S-PrediXcan	BLasso	Horseshoe	Horseshoe+	A-TWAS
chr6	BAG6	8.39E-38	4.21E-33	1.26E-29	3.58E-29	1.26E-32
chr6	HSPA1B	-	-	-	9.40E-27	9.40E-27
chr6	MOG	-	1.18E-14	1.38E-26	4.15E-01	4.15E-26
chr6	LY6G5B	5.16E-25	5.75E-06	1.52E-21	2.21E-25	6.62E-25
chr6	HLA-DMA	1.05E-28	1.34E-20	3.63E-24	4.26E-24	5.88E-24
chr6	CCHCR1	3.59E-22	4.78E-11	1.56E-21	2.36E-24	7.06E-24
chr6	FLOT1	-	-	-	7.56E-24	7.56E-24
chr6	PSMB9	-	-	3.19E-22	7.11E-24	1.39E-23
chr6	HLA-DMB	3.08E-27	3.15E-08	1.60E-23	2.94E-22	4.56E-23
chr6	VARS2	1.37E-23	1.87E-21	5.26E-17	8.10E-15	5.60E-21
chr6	PSMB8	-	-	8.04E-17	9.29E-19	1.84E-18
chr6	AGER	3.82E-17	1.31E-06	1.84E-18	1.03E-16	5.42E-18
chr6	ZBTB22	3.31E-18	4.74E-09	1.24E-17	5.50E-17	3.03E-17
chr6	C2	-	-	3.21E-17	1.96E-16	5.51E-17
chr6	NOTCH4	1.06E-18	1.12E-11	2.51E-17	4.34E-15	7.49E-17
chr6	CTA-14H9.5	1.80E-19	8.09E-11	4.98E-17	6.90E-17	8.68E-17
chr6	AL022393.7	8.14E-20	1.37E-14	5.65E-16	1.06E-16	2.67E-16
chr6	EGFL8	6.03E-04	-	1.04E-14	2.16E-16	4.24E-16
chr6	HLA-C	4.42E-15	1.42E-13	4.65E-16	6.18E-14	1.38E-15
chr6	AIF1	-	-	-	6.83E-15	6.88E-15
chr22	NHP2L1	2.11E-14	1.26E-11	8.79E-15	1.78E-14	1.76E-14
chr6	BTN3A2	6.65E-16	1.38E-13	4.48E-14	4.00E-12	1.00E-13
chr6	TNXB	-	4.05E-04	2.39E-12	4.49E-14	1.32E-13
chr6	SAPCD1-AS1	1.70E-13	3.74E-03	2.85E-11	1.17E-13	3.50E-13
chr6	BRD2	1.32E-18	4.72E-13	9.90E-01	3.28E-05	1.42E-12
chr22	-GA	1.83E-14	3.37E-08	2.18E-12	7.46E-13	1.67E-12
chr12	OGFOD2	6.33E-12	5.34E-09	1.21E-12	1.57E-12	2.05E-12
chr6	SYNGAP1	9.09E-11	3.09E-05	7.87E-13	4.95E-10	2.36E-12
chr3	NEK4	4.08E-12	2.03E-10	3.16E-12	2.18E-12	3.84E-12
chr6	HLA-B	7.97E-12	7.37E-08	4.55E-12	2.23E-11	1.13E-11
chr6	C4A	9.50E-13	1.61E-06	6.05E-12	1.19E-10	1.73E-11
chr2	WDPCP	-	-	2.28E-11	2.44E-11	2.36E-11
chr6	PPP1R10	1.07E-03	7.26E-03	7.99E-12	1.19E-09	2.38E-11
chr6	TRIM26	-	2.92E-11	-	-	2.92E-11
chr22	PHF5A	1.04E-11	3.82E-07	2.32E-11	5.00E-11	4.76E-11
chr6	GPANK1	6.35E-21	1.76E-11	1.46E-09	1.76E-09	5.18E-11
chr17	SLC16A3	-	-	1.42E-01	5.82E-11	1.16E-10
chr22	XPNPEP3	1.29E-10	4.38E-11	5.23E-05	7.07E-05	1.31E-10
chr3	PCCB	1.05E-08	5.20E-08	9.25E-10	1.91E-09	1.85E-09
chr5	ELOVL7	1.07E-09	4.96E-06	9.17E-10	3.68E-09	2.20E-09
chr22	SEPT3	1.76E-09	4.40E-05	3.80E-09	1.22E-09	2.78E-09
chr14	XRCC3	1.74E-07	8.29E-06	4.44E-09	1.24E-09	2.90E-09
chr3	GNL3	1.27E-09	6.35E-08	1.59E-09	4.09E-09	3.37E-09
chr2	C2orf69	4.17E-10	6.66E-08	3.76E-09	2.34E-09	4.23E-09
chr6	ZNF192P1	5.78E-11	3.19E-07	2.02E-09	1.73E-08	5.38E-09
chr5	PART1	6.96E-01	9.59E-08	1.13E-08	3.29E-09	7.45E-09
chr17	RP11-1055B8.3	1.41E-02	7.09E-07	4.61E-09	4.50E-08	1.25E-08
chr10	CYP17A1-AS1	3.99E-09	4.72E-09	5.16E-06	3.68E-06	1.41E-08
chr6	PRSS16	-	3.30E-03	4.75E-09	7.16E-02	1.43E-08
chr6	TRIM27	-	5.27E-03	5.30E-08	8.32E-09	2.16E-08
chr8	PPAPDC1B	1.15E-05	2.33E-05	2.82E-08	1.26E-08	2.61E-08
chr14	APOPT1	-	6.71E-04	1.73E-08	2.86E-08	3.24E-08
chr6	PHF1	7.54E-10	9.52E-05	1.43E-08	1.36E-05	4.27E-08
chr6	FKBPL	1.06E-06	1.19E-05	3.11E-08	3.00E-08	4.57E-08
chr3	GLYCTK	1.78E-08	1.68E-05	8.92E-08	3.58E-08	7.65E-08
chr6	ZNRD1-AS1	2.75E-05	3.82E-04	1.23E-07	3.39E-08	7.96E-08
chr3	GRM2	-	1.76E-06	6.09E-08	5.89E-08	8.83E-08
chr11	RP11-890B15.3	6.27E-05	2.51E-03	4.84E-08	1.69E-07	1.13E-07
chr22	LINC00634	3.00E-07	4.63E-05	1.58E-07	7.51E-08	1.53E-07
chr22	ACO2	-	2.55E-04	8.47E-08	5.34E-07	2.19E-07
chr6	SKIV2L	2.78E-08	1.50E-07	1.13E-06	3.72E-07	2.93E-07
chr6	HIST1H4H	2.01E-10	1.13E-07	1.14E-06	1.67E-05	3.07E-07
chr22	CAC-1I	1.34E-09	3.19E-07	-	-	3.19E-07
chr11	MPPED2	2.51E-06	5.55E-05	4.03E-07	1.47E-07	3.22E-07
chr6	GPSM3	2.40E-01	4.56E-01	1.34E-07	7.75E-06	3.95E-07
chr8	FAM83H	2.96E-06	9.47E-05	1.98E-07	4.47E-07	4.11E-07
chr1	PLEKHO1	3.11E-06	5.79E-06	3.80E-07	3.09E-07	4.97E-07
chr6	DDAH2	-	7.77E-06	2.49E-07	7.61E-07	5.49E-07
chr2	TYW5	1.84E-07	2.14E-07	2.96E-06	2.19E-06	5.50E-07
chr12	STAT6	1.30E-05	5.53E-03	2.14E-05	2.09E-07	6.22E-07
chr22	CYP2D6	1.79E-07	1.36E-06	9.09E-07	7.52E-07	9.48E-07
chr6	WDR46	1.50E-05	7.00E-02	3.21E-07	1.42E-04	9.60E-07
chr16	RP11-231C14.4	1.10E-07	3.31E-07	1.65E-04	1.43E-04	9.90E-07
chr11	SNX19	1.67E-07	9.09E-06	5.08E-07	1.67E-06	1.12E-06

CHR	GENE	S-PrediXcan	BLasso	Horseshoe	Horseshoe+	A-TWAS
chr3	TKT	**4.45E-07**	1.43E-03	**5.62E-07**	-	**1.12E-06**
chr6	ZBTB12	-	9.73E-05	**9.32E-07**	**7.86E-07**	**1.27E-06**
chr8	ARC	2.09E-04	**6.63E-07**	2.91E-05	-	**1.30E-06**
chr16	NMRAL1	4.42E-03	4.49E-03	**9.52E-07**	**8.54E-07**	**1.35E-06**
chr7	STAG3L3	-	**1.36E-06**	-	-	**1.36E-06**
chr22	TTLL1	**1.44E-06**	1.67E-04	**2.36E-06**	**9.17E-07**	**1.97E-06**
chr6	B3GALT4	**3.03E-09**	2.35E-04	**1.07E-06**	-	**2.13E-06**
chr6	PPP1R11	-	2.66E-02	**3.28E-06**	**9.29E-07**	**2.17E-06**
chr12	TCHP	**6.73E-07**	1.55E-03	**1.81E-06**	**1.59E-06**	**2.54E-06**
chr6	AGPAT1	9.59E-06	**2.79E-06**	-	-	**2.79E-06**
chr16	MAPK3	1.39E-05	8.07E-05	**1.74E-06**	**2.14E-06**	**2.84E-06**
chr6	VWA7	-	2.50E-05	**3.03E-06**	**1.58E-06**	**2.99E-06**
chr12	SETD8	**1.15E-07**	5.77E-03	**2.08E-06**	**2.33E-06**	**3.30E-06**
chr3	THOC7	5.90E-05	5.05E-04	**1.52E-06**	4.62E-06	**3.43E-06**
chr6	RING1	1.68E-02	9.38E-01	-	**2.04E-06**	**4.09E-06**
chr2	KCTD18	**4.12E-06**	-	8.78E-06	**3.49E-06**	4.99E-06
chr1	RPRD2	**2.35E-06**	5.49E-06	5.01E-06	5.95E-06	5.46E-06
chr6	ABT1	6.28E-06	2.05E-02	1.70E-05	**2.18E-06**	5.80E-06
chr8	EIF4EBP1	5.38E-03	7.20E-02	**2.76E-06**	6.80E-06	5.89E-06
chr1	RERE	**1.75E-06**	5.08E-05	**2.85E-06**	7.46E-06	5.95E-06
chr22	NDUFA6	3.00E-03	6.28E-02	**3.71E-06**	5.14E-06	6.46E-06
chr6	DHX16	3.45E-01	1.95E-02	**2.33E-06**	3.70E-05	6.56E-06
chr9	ALG2	5.47E-06	1.10E-05	9.50E-06	**4.09E-06**	6.81E-06
chr6	ZNRD1	4.17E-05	3.71E-01	6.00E-04	**2.98E-06**	8.90E-06
chr22	XRCC6	2.45E-05	1.42E-02	2.23E-05	**3.62E-06**	9.33E-06
chr19	MAMSTR	1.25E-04	7.25E-03	**4.24E-06**	1.37E-05	9.72E-06
chr2	DNMT3A	-	**3.78E-06**	7.20E-05	8.52E-05	1.03E-05
chr9	RPL12	**3.39E-06**	**4.02E-06**	5.38E-05	7.90E-05	1.07E-05
chr6	HLA-L	**2.15E-08**	1.68E-05	-	-	1.68E-05
chr6	ATP6V1G2-DDX39B	**3.23E-06**	9.52E-03	9.83E-06	2.19E-05	2.03E-05
chr2	DPYSL5	**3.49E-06**	1.28E-05	6.70E-05	2.37E-05	2.22E-05
chr1	HYI	**2.96E-06**	1.20E-04	1.91E-05	1.72E-05	2.53E-05
chr6	GABBR1	**4.33E-07**	1.04E-05	2.00E-02	2.91E-02	3.12E-05
chr6	HIST1H4C	**1.46E-07**	1.58E-05	7.56E-05	7.76E-05	3.35E-05
chr11	RP11-430H10.1	**1.91E-06**	1.16E-05	1.81E-03	1.79E-03	3.43E-05
chr6	GRM4	**2.21E-06**	2.92E-05	2.19E-05	4.42E-04	3.65E-05
chr6	CMAHP	**4.87E-09**	2.21E-04	3.05E-04	5.80E-04	3.15E-04
chr6	BTN2A2	**4.02E-06**	3.08E-04	5.68E-04	9.35E-04	4.94E-04
chr6	GUSBP2	**1.55E-07**	2.23E-03	1.37E-03	3.96E-04	8.09E-04
chr19	IRF3	**1.85E-07**	4.11E-02	6.75E-04	1.03E-03	1.21E-03
chr6	NRM	**2.40E-06**	1.62E-02	5.48E-03	1.87E-03	3.85E-03
chr22	DDX17	**3.84E-08**	5.22E-02	1.31E-02	5.70E-03	1.11E-02
chr22	POLDIP3	**3.95E-07**	1.95E-02	-	-	1.95E-02
chr6	PRRC2A	**2.72E-08**	-	7.65E-02	1.77E-02	2.88E-02
chr11	PACSIN3	**2.66E-06**	4.52E-01	1.49E-01	1.25E-01	1.86E-01
chr6	LST1	**1.25E-08**	-	-	-	-
chr19	GMIP	**9.81E-07**	-	-	-	-

The table display all the significant genes selected by the methods. For the column "GENE", the genes highlighted with red color indicating it have been included in the GWAS Catalog under the trait Schizophrenia. The columnns of each method record the TWAS p-value of each gene by this methood. Value marked in bold indicating it reach the significant level of corrected p-value. The value omitted as "-" indicating that the method did not attain a effective imputation model with test$R^2 > 0.01$ for that gene.

Table S2. Significant Genes with TWAS p-value less than Bonferroni corrected significance level for Obsessive-Compulsive Symptoms.

CHR	GENE	S-PrediXcan	BLasso	Horseshoe	Horseshoe+	A-TWAS
chr6	**TPMT**	-	-	**4.20E-06**	**4.24E-06**	**4.22E-06**
chr10	**ACTA2**	7.84E-01	7.18E-01	**1.02E-30**	1.69E-01	**3.06E-30**
chr3	**PARP15**	-	3.17E-01	**6.49E-09**	**6.16E-09**	**9.48E-09**
chr5	**MXD3**	1.74E-01	1.93E-01	**2.10E-13**	2.65E-01	**6.31E-13**
chr17	**SCO1**	9.67E-01	9.00E-01	**1.50E-06**	7.54E-01	**4.50E-06**

The table display all the significant genes selected by the methods. The columnns of each method record the TWAS p-value of each gene by this methood. Value marked in bold indicating it reach the significant level of corrected p-value. The value omitted as "-" indicating that the method did not attain a effective imputation model with test$R^2 > 0.01$ for that gene.