

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Shuyi Guo

Date

Genome-wide TWAS of brain and blood tissues identifies novel risk genes for
Alzheimer's disease dementia

By

Shuyi Guo

Master of Science in Public Health

Department of Biostatistics and Bioinformatics

Jingjing Yang, PhD

(Thesis Advisor)

Michael P. Epstein, PhD

(Reader)

Genome-wide TWAS of brain and blood tissues identifies novel risk genes for
Alzheimer's disease dementia

By

Shuyi Guo

B.S., Soochow University, 2017

Thesis Committee Chair: Jingjing Yang, PhD

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Biostatistics

2023

Abstract

Genome-wide TWAS of brain and blood tissues identifies novel risk genes for Alzheimer's disease dementia

By Shuyi Guo

Background: Transcriptome-wide association studies (TWAS) are a powerful tool for identifying novel genes associated with complex diseases, including Alzheimer's disease (AD) dementia. TWAS integrate reference genetic and transcriptomic data to identify expression quantitative trait loci (eQTLs) of target genes and estimate the genetically regulated gene expression (GR_{EX}) levels for each gene. GR_{EX} data are then integrated with the genome-wide association study (GWAS) summary statistics to assess the association between gene expression and the phenotype of interest. However, existing TWAS methods only consider *cis*-eQTL (eQTLs located within the 1 MB region of the gene) effects and miss effects of *trans*-eQTLs (outside of the 1 MB region). To overcome this limitation, we applied a Bayesian Genome-wide TWAS (BGW-TWAS) method to leverage both *cis*- and *trans*-eQTL information to improve the mapping of risk genes for AD dementia.

Methods and Materials: We applied BGW-TWAS to the Genotype-Tissue Expression (GTEx) dataset on three tissues – the prefrontal cortex, cortex, and whole blood. Then we integrated estimated eQTL effect sizes by BGW-TWAS with a summary-level GWAS dataset of AD dementia by S-PrediXcan to identify genes associated with AD dementia. We also use aggregated Cauchy association test-omnibus (ACAT-O) method to combine the TWAS *p*-values across the three tissues for each gene to obtain the combined *p*-values.

Results: Our analysis identified 37 genes significantly associated with AD dementia in the prefrontal cortex, 55 in the cortex, and 51 in the whole blood. After combining TWAS *p*-values across the three tissues by ACAT-O, we obtained 93 genes with significant combined *p*-values, including 50 novel genes not reported in previous studies, and 29 genes significant primarily due to *trans*-eQTLs. We detected 5 functional clusters comprised of both known AD risk genes and novel genes in the protein-protein association networks, and 7 enriched phenotypes in the phenotype enrichment analysis.

Conclusion: In this study, we conducted BGW-TWAS on three tissues and identified known and novel genes associated with AD dementia. Our study is the first genome-wide TWAS utilizing both *cis*- and *trans*-eQTLs for AD risk gene identification and provides new insights into the genetic basis of AD dementia.

Genome-wide TWAS of brain and blood tissues identifies novel risk genes for
Alzheimer's disease dementia

By

Shuyi Guo

B.S., Soochow University, 2017

Thesis Committee Chair: Jingjing Yang, PhD

An abstract of

A thesis submitted to the faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Biostatistics

2023

Acknowledgement

I would like to express my sincere gratitude to the Rollins School of Public Health at Emory University for their unwavering support throughout my study and research. The knowledge and experience I have gained in the field of Biostatistics here have significantly contributed to my progress and growth as a researcher.

I am deeply grateful to my thesis advisor, Jingjing Yang, for her invaluable guidance and training in the field of statistical genetics. Her patient teaching and mentorship have been instrumental in shaping my academic journey. I would also like to extend my appreciation to the Yang lab members for their collaboration, assistance, and camaraderie throughout this process.

Lastly, I would like to express my heartfelt thanks to my parents for their steadfast support as I pursued my master's degree abroad. Their unwavering belief in my abilities and encouragement have been the foundation of my success in this program.

Table of Contents

1. Introduction	1
2. Material and Methods	3
2.1 Bayesian Genome-wide TWAS (BGW-TWAS)	3
2.1.1 Bayesian Variable Selection Regression Model (BVSr)	3
2.1.2 EM-MCMC Algorithm	5
2.1.3 TWAS with summary-level GWAS data	6
2.2 Aggregated Cauchy association test-omnibus (ACAT-O)	7
2.3 Genotype-Tissue Expression (GTEx) dataset and GWAS summary statistics	8
2.4 Search Tool for the Retrieval of Interacting Genes/Proteins (STRING)	8
3. Results	9
3.1 Overview of the workflow	9
3.2 BGW-TWAS results of AD dementia in three tissues	10
3.3 ACAT-O results of AD dementia across three tissues	11
3.4 Known risk genes for AD dementia	14
3.5 Novel findings of risk genes for AD dementia	15
3.6 Protein-protein association networks and phenotype enrichment analysis by STRING	16
3.6.1 Protein-protein association networks	17
3.6.2 Phenotype enrichment analysis	19
3.7 eQTLs of the significant genes	20
4. Discussion	22
References	25

1. Introduction

Alzheimer's disease (AD) dementia is a complex neurodegenerative disorder that is characterized by progressive cognitive decline and memory loss, currently affecting 6.5 million Americans aged 65 and older, and listed as the seventh-leading cause of death [1]. Despite extensive research, the underlying molecular mechanisms of AD dementia are still not fully understood, and effective treatments are lacking [2]. Recent studies have highlighted the role of genetic factors in AD dementia [3, 4], and several genes have been identified as risk factors for the disease. However, these genes explain only a small portion of the heritability of AD dementia, suggesting that many other genetic variants may also contribute to disease risk.

Transcriptome-wide association studies (TWAS) have emerged as a powerful tool for identifying novel genes and biological pathways associated with complex diseases, including AD dementia [5, 6]. There are commonly two stages for a TWAS. In the first stage, by using profiled transcriptomic and genetic data from a reference panel, TWAS fit imputation regression models for the expression quantitative traits of target genes with nearby genotypes, and these imputation models are applied to the reference genotype data to estimate genetically regulated gene expression (GReX) levels for each gene. In the fitted imputation models, single nucleotide polymorphisms (SNPs) with non-zero effect sizes on reference transcriptome are identified as expression quantitative trait loci (eQTLs), representing genetic variants that affect gene expression. In the second stage, GReX are subsequently tested for associations with the phenotype of interest, using either individual-level or summary-level genome-wide association studies (GWAS) data of test samples.

However, one limitation of existing TWAS methods is that they only consider *cis*-eQTL effects [5, 7, 8], which refers to the effects of eQTLs located within the 1 MB region of the gene of interest. This can result in the missed effects of *trans*-eQTLs, which are eQTLs outside of the 1 MB region of the target

gene. *Trans*-eQTLs have been found to play an important role in regulating gene expression levels by affecting the activity of distal genes, which can in turn impact biological processes and disease susceptibility [9, 10]. Incorporating *trans*-eQTLs into TWAS is essential as they can reveal more complex regulatory mechanisms, contribute to the identification of novel gene-trait associations, and enhance our understanding of the genetic architecture underlying complex traits and diseases. To overcome this limitation, we employ a Bayesian Genome-wide TWAS (BGW-TWAS) method that leverage both *cis*- and *trans*-eQTL information [11]. By integrating large-scale reference transcriptome data with GWAS summary statistics, BGW-TWAS enables the identification of disease-relevant genes with improved accuracy and power.

In this study, we applied the BGW-TWAS method to the reference Genotype-Tissue Expression (GTEx) V8 dataset [12] on three tissues – the prefrontal cortex, cortex, and whole blood – to fit gene expression imputation models. We chose to conduct TWAS on these two brain tissues due to the substantial evidence highlighting the involvement of the cortex and prefrontal cortex in the progression of AD dementia [13]. We also chose to conduct TWAS on the whole blood tissue due to three reasons. First, whole blood tissue usually has greater accessibility and larger sample size compared to brain tissues. Second, recent studies have demonstrated that specific gene expression products in whole blood can serve as biomarkers for AD dementia [14, 15]. Third, there is a probable correlation between the gene expression in the whole blood and that of the brain's cortex [16]. By using S-PrediXcan [17], we integrated our estimated eQTL effect sizes by BGW-TWAS with a summary-level AD GWAS dataset with sample size of 74,004 individuals [18] to identify genes associated with AD dementia.

As a result, we identified 37 significant genes in the prefrontal cortex, 55 significant genes in the cortex, and 51 significant genes in the whole blood. We also utilized an aggregated Cauchy association test-omnibus (ACAT-O) method to combine the TWAS p -values across the three tissues for each gene. ACAT-O is an omnibus test that can combine p -values from multiple sources by transforming these p -values to

conform to a standard Cauchy distribution [19]. By ACAT-O, we obtained a total of 93 genes with significant combined p -values. We identified several well-known AD risk genes among these 93 significant genes, but also found 50 novel genes that have never been detected by other GWAS and TWAS studies of AD. Through protein-protein association networks of the 93 genes, we detected a main functional cluster comprised of *APOE* and other known AD risk genes, and found another 4 clusters comprised of both known AD risk genes and 19 novel genes. The phenotype enrichment analysis showed that several phenotypes, such as Apolipoprotein B, low-density lipoprotein cholesterol, inflammatory biomarkers, and C-reactive protein, were enriched with the 93 genes.

As a genome-wide TWAS, our study is the first to utilize both *cis*- and *trans*-eQTLs for AD risk gene identification. Our results provide new insights into the genetic basis of AD dementia and pave the way for the development of novel treatments for this disease.

2. Material and Methods

2.1 Bayesian Genome-wide TWAS (BGW-TWAS)

In this thesis project, we focus on the application of the BGW-TWAS method to perform TWAS. BGW-TWAS is a TWAS method that accounts for both *cis*- and *trans*-eQTLs based on a Bayesian variable selection regression (BVSR) model for imputing GReX.

2.1.1 Bayesian Variable Selection Regression Model (BVSR)

The following BVSR model [20] is assumed for quantitative gene expression traits:

$$\mathbf{E}_{n \times 1} = \mathbf{X}_{n \times p} \mathbf{w}_{p \times 1} + \boldsymbol{\epsilon}_{n \times 1}, w_i \sim \pi N(0, \sigma_w^2 \sigma_\epsilon^2) + (1 - \pi) \delta_0(\cdot), \epsilon_i \sim N(0, \sigma_\epsilon^2).$$

Here, $\mathbf{E}_{n \times 1}$ represents the centered quantitative expression levels vector for n samples; $\mathbf{X}_{n \times p}$ is the centered genotype matrix containing p genetic variants; ϵ_i is the residual error which follows a normal distribution $N(0, \sigma_\epsilon^2)$; and $\mathbf{w}_{p \times 1}$ follows a spike-and-slab prior distribution [20-22], which means that w_i follows the normal distribution $N(0, \sigma_w^2 \sigma_\epsilon^2)$ with probability π and the point-mass density function $\delta_0(\cdot)$ at 0 with probability $(1 - \pi)$.

The genotype matrix $\mathbf{X}_{n \times p}$ typically contains either dosage data within the range of $[0, 2]$ or genotype data with values $\{0,1,2\}$, representing the expected or genotyped number of minor alleles. By utilizing a spike-and-slab prior for $\mathbf{w}_{p \times 1}$, variable selection is enforced in the regression model. Since both $\mathbf{E}_{n \times 1}$ and columns of $\mathbf{X}_{n \times p}$ are centered, the intercept term is omitted from the regression model.

The BVSR model can be extended to account for both cis- and trans- eQTL genotype data for modeling quantitative gene expression traits. The extended model can be represented as:

$$\begin{aligned} \mathcal{E}_g &= \mathbf{X}_{cis} \mathbf{w}_{cis} + \mathbf{X}_{trans} \mathbf{w}_{trans} + \epsilon \\ w_{cis,i} &\sim \pi_{cis} N(0, \sigma_{cis}^2 \sigma_\epsilon^2) + (1 - \pi_{cis}) \delta_0(w_{cis,i}) \\ w_{trans,i} &\sim \pi_{trans} N(0, \sigma_{trans}^2 \sigma_\epsilon^2) + (1 - \pi_{trans}) \delta_0(w_{trans,i}) \\ \epsilon_i &\sim N(0, \sigma_\epsilon^2). \end{aligned}$$

Compared with the general BVSR model, we further separate the genotype matrix and corresponding parameters into two distinct parts based on the *cis*- and *trans*- SNPs. *Cis*- and *trans*- can be viewed as two non-overlapping annotations for SNPs in the BVSR model, making this model a special case of the previously developed Bayesian Functional GWAS (BFGWAS) method [23].

The following independent and conjugate hyper priors are assumed for hyper parameters in the model:

$$\pi_{cis} \sim \text{Beta}(a_{cis}, b_{cis}), \sigma_{cis}^2 \sim \text{IG}(k_1, k_2),$$

$$\pi_{trans} \sim \text{Beta}(a_{trans}, b_{trans}), \sigma_{trans}^2 \sim \text{IG}(k_3, k_4),$$

$$\sigma_{\epsilon}^2 \sim \text{IG}(k_5, k_6),$$

where the probability π for *cis*- and *trans*- SNPs is assumed to follow Beta distributions with different hyper priors, while $(\sigma_{cis}^2, \sigma_{trans}^2, \sigma_{\epsilon}^2)$ follows different Inverse Gamma distributions. In this study, we set $k_1 = k_2 = k_3 = k_4 = k_5 = k_6 = 0.1$, and $\frac{a_q}{a_q + b_q} = 10^{-6}$ with $(a_q + b_q)$ equal to the total number of variants of respective annotation $q = \{cis, trans\}$.

We also set a latent indicator vector $\boldsymbol{\gamma}_{p \times 1}$ in the extended model. Each element γ_i indicates whether the corresponding i th effect $w_{q,i}$ equals to 0 with $\gamma_i = 0$ or follows the $N(0, \sigma_q^2 \sigma_{\epsilon}^2)$ distribution with $\gamma_i = 1$. The latent indicator variable can be represented as:

$$\gamma_i \sim \text{Bernoulli}(\pi_i),$$

$$w_{-\gamma} \sim \delta_0(\cdot), w_{\gamma} \sim \text{MVN}_{|\gamma|}(0, \sigma_{\epsilon}^2 V_{\gamma}).$$

The expectation of the latent indicator variable ($E[\gamma_i]$) is the posterior probability (PP_i) for the i th SNP to be an eQTL with effect size w_i .

Given the BVSR model and hyperpriors, along with genotype and gene expression data, we can infer the posterior distribution of w and $\boldsymbol{\gamma}$ by using a scalable Expectation-Maximization Markov Chain Monte Carlo (EM-MCMC) algorithm [23, 24]. The detailed derivations of the Bayesian inference process and EM-MCMC algorithm are referred to the supplementary parts of BGW-TWAS and BFGWAS paper [11, 23].

2.1.2 EM-MCMC Algorithm

The EM-MCMC algorithm is a key component of the BGW-TWAS method, helping to reduce the computational burden and memory usage of the method. It can be summarized in the following steps:

- i. Generate summary statistics: obtain GWAS summary statistics from individual-level genotype data and gene expression data, by doing single variant analyses for genome-wide SNPs.
- ii. Prune genome blocks:
 - a. Consider blocks that consist of at least one *cis*-SNP or one potential trans-eQTL with a single-variant test p-value less than $1e-5$.
 - b. Select up to 100 blocks with minimal *p*-values, ordered from smallest to largest.
 - c. Select any remaining blocks containing *cis*-SNPs not chosen in step b.
- iii. Apply the EM-MCMC algorithm to pruned blocks:
 - a. Fix the variance of the expression residuals.
 - b. Set initial values for (π, σ^2) . In this study, we set $\pi_{cis} = \pi_{trans} = 1 \times 10^{-6}$ and $\sigma_{cis}^2 = \sigma_{trans}^2 = 0.1$.
 - c. E-step: Conditioned on the current estimates of (π, σ^2) , estimate $(\mathbf{w}, \mathbf{PP})$ by using MCMC algorithm per block with summary statistics.
 - d. M-step: Update the estimates of (π, σ^2) by their maximum a posteriori estimates (MAPs) [25], conditioned on estimates from the previous E-step.
 - e. Repeat steps c and d until the MAPs of (π, σ^2) converge (2 EM steps in this study).
- iv. Keep the estimates of $(\mathbf{w}, \mathbf{PP})$ from the last E-step for further imputation of GREX.

2.1.3 TWAS with summary-level GWAS data

By multiplying the estimated effect sizes \mathbf{w} with the posterior probabilities \mathbf{PP} of eQTLs, we can establish their weights. Utilizing the genotype and weight information and GWAS summary data produced by single variant tests, we employed the S-PrediXcan [17] test statistic to acquire burden TWAS Z-score test statistics. The burden TWAS Z-score test statistic for gene *g* is expressed as follows:

$$Z_g = \sum_{l \in \text{Model}_g} \widehat{w}_{lg} \frac{\widehat{\sigma}_l}{\widehat{\sigma}_g} \frac{\widehat{\beta}_l}{SE(\widehat{\beta}_l)} = \sum_{l \in \text{Model}_g} \widehat{w}_{lg} \frac{\widehat{\sigma}_l}{\widehat{\sigma}_g} Z_l = \sum_{l \in \text{Model}_g} \frac{(\widehat{w}_{lg} \widehat{\sigma}_l) Z_l}{\sqrt{\widehat{w} \mathbf{V} \widehat{w}}},$$

$$\widehat{\sigma}_l^2 = \text{Var}(x_l), \widehat{\sigma}_g^2 = \widehat{w} \mathbf{V} \widehat{w}, \mathbf{V} = \text{Cov}(\mathbf{X}),$$

where $\widehat{\beta}_l$ denotes the effect size of SNP l from GWAS, Z_l denote the Z-score statistic by single variant test, and $\widehat{w}_{lg} = \widehat{P} \widehat{P}_l \widehat{w}_i$ is the weight of SNP l which have non-zero eQTL effect size from the BVSR model.

Here, \mathbf{X} is the genotype matrix from reference panels, and \mathbf{V} is the genotype covariance matrix.

We subsequently calculated two-tailed p -values from the TWAS Z-score test statistics. If the TWAS p -value for a specific gene within a particular tissue is lower than 0.05 divided by the total number of genes in that tissue, we conclude that the GReX of this gene is significantly associated with AD dementia in the respective tissue.

2.2 Aggregated Cauchy association test-omnibus (ACAT-O)

ACAT-O is an omnibus test that combines the p -values of different set-based tests, and this method is based on aggregated Cauchy association test (ACAT) [19]. ACAT employs a linear combination of transformed p -values as the test statistic, with these p -values modified to follow a standard Cauchy distribution under the null hypothesis. This combination permits the use of flexible weights. The ACAT-O test statistics is as follows:

$$T_{ACAT-O} = \frac{1}{K} \sum_{i=1}^K \tan \{(0.5 - p_i)\pi\},$$

where p_i s are the p -values of the K tests, and the tests are treated equally in the combination.

We used the ACAT-O method to combine TWAS p -values from three tissue types per gene. In this study, ACAT-O was conducted by the “ACATO” function from R package “sumFREGAT”.

2.3 Genotype-Tissue Expression (GTEx) dataset and GWAS summary statistics

The genotype and gene expression data we used is from the GTEx Version 8 (V8) dataset, which consists of 838 donors and 17,382 samples from 52 tissues and two cell lines [12]. Specifically, we used the gene expression data of the prefrontal cortex (158 samples), cortex (184 samples) and whole blood (574 samples). We only studied 3 tissue types of the GTEx V8 dataset because of the large computational burden of BGW-TWAS.

The summary-level GWAS data used in this study were generated by the most recent large-scale GWAS for AD dementia with a sample size of 74,004 [18].

2.4 Search Tool for the Retrieval of Interacting Genes/Proteins (STRING)

STRING (version 11.5) [26] is a bioinformatics online tool that provides information on protein-protein interactions and networks, as well as functional characterization of genes and proteins. The tool integrates different types of evidence, such as genomic context, high-throughput experiments, and previous knowledge from other databases, to generate reliable predictions of protein interactions. We used the tool to study the protein-protein association networks and determine which phenotypes may be enriched with genes identified from our study.

3. Results

3.1 Overview of the workflow

To identify genes associated with AD dementia, we performed a BGW-TWAS on GTEx V8 dataset across 3 tissues, which are the prefrontal cortex, cortex, and whole blood. Briefly, this approach applies the BVSR model to estimate weights for both *cis*- and *trans*-eQTLs across the genome for the target genes.

Based on the BGW-TWAS weights, reference genotype data and AD GWAS summary data, we utilized the S-PrediXcan approach to obtain burden TWAS Z-score test statistics and derived 2-tailed p -values.

ACAT-O method was used to combine TWAS p -values across three tissues, resulting in ACAT-O combined p -values for each gene.

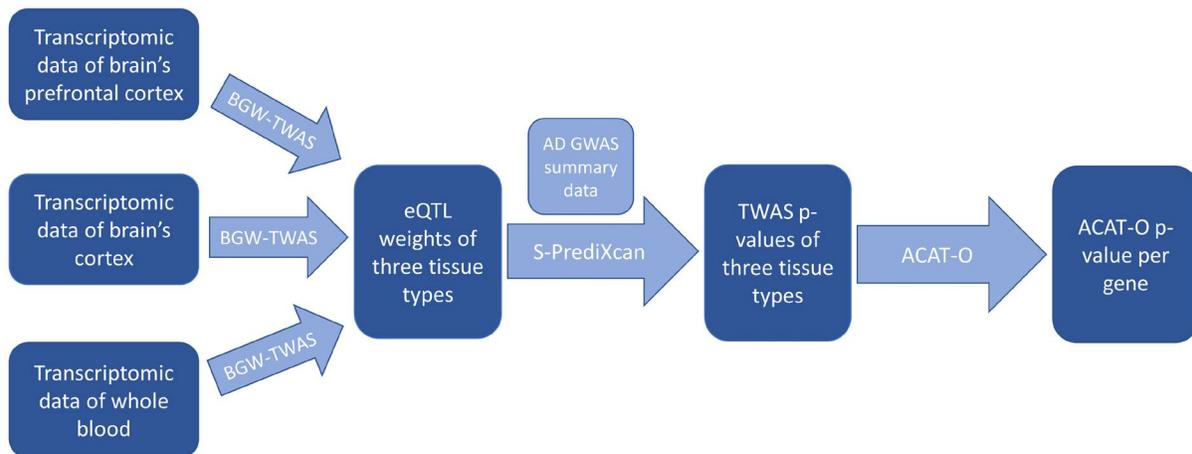


Figure 1. Workflow for the BGW-TWAS process.

3.2 BGW-TWAS results of AD dementia in three tissues

We performed BGW-TWAS on a total of 23,721 genes expressed in the prefrontal cortex, 23,864 genes expressed in the cortex, and 19,514 genes expressed in the whole blood. Our analysis identified 37, 55, and 51 genes with significant p -values in the prefrontal cortex, cortex, and whole blood, respectively. Of these, 15 genes were significant in both the prefrontal cortex and cortex, six were significant in both the prefrontal cortex and whole blood, eight were significant in both the cortex and whole blood, and three were significant across all three tissue types. These results were summarized in Figure 2 as a Venn plot.

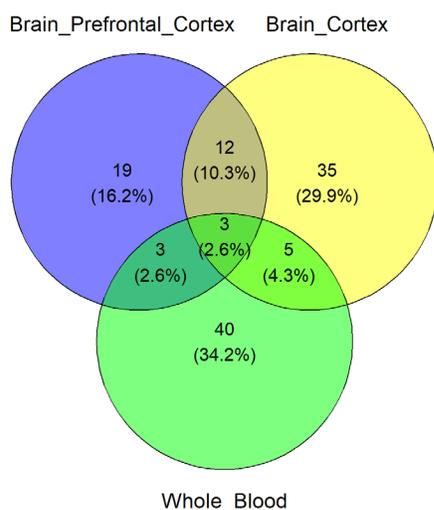


Figure 2. Venn plot of numbers of significant genes in three tissues.

We also summarized the proportion of *trans*-eQTLs of each significant gene in three tissues, as shown in Figure 3. As a result, there were 12 (32.4%), 24 (43.6%), and 16 (31.4%) significant genes in the prefrontal cortex, cortex, and whole blood, respectively, that have a proportion of *trans*-eQTLs above 0.5. These results demonstrated that *trans*-eQTLs contributed to a large proportion of the significance of many of the genes in three tissues.

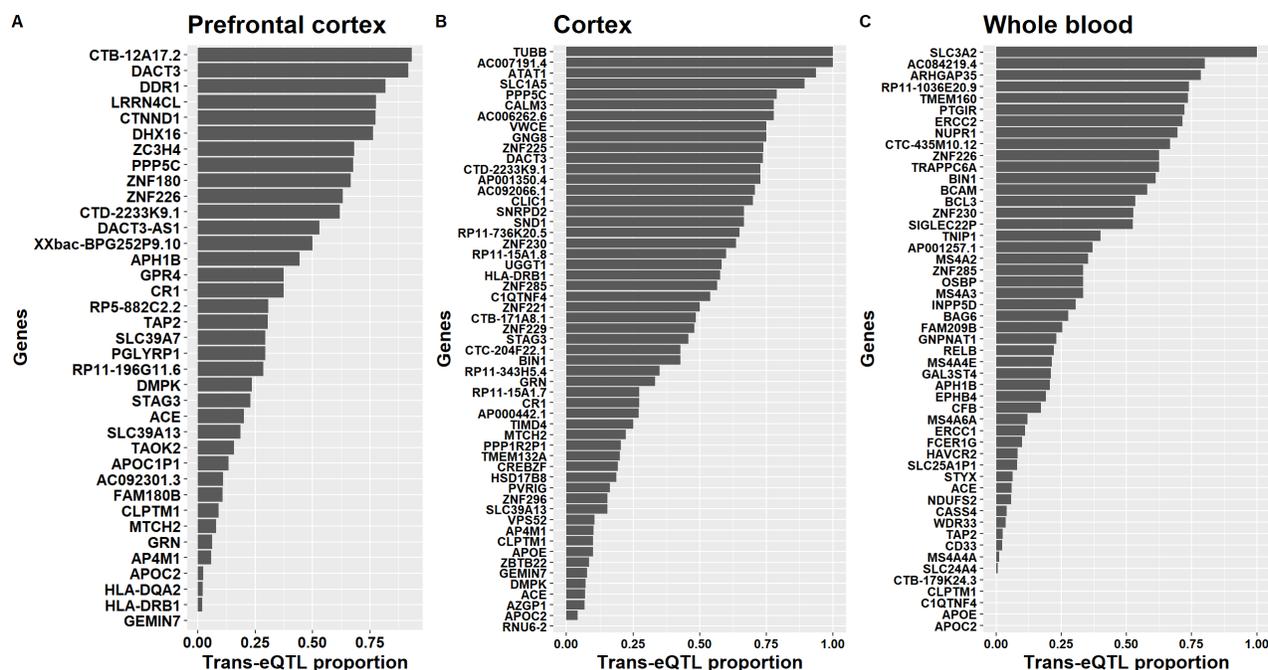


Figure 3. Bar graphs of *trans*-eQTL proportion of significant genes in the three tissues. Graph A, B, C are the bar graphs of *trans*-eQTL proportion of significant genes in the prefrontal cortex, cortex, and whole blood, respectively. For each bar graph, x-axis represents the proportion of *trans*-eQTLs, ranging from 0-1; y-axis represents all the genes that have a significant p -value in the tissue.

3.3 ACAT-O results of AD dementia across three tissues

After combining the TWAS p -values of genes across three tissues using the ACAT-O method, we obtained a total of 17,468 genes with an ACAT-O combined p -value. Among these, 93 genes had a significant p -value. We created a Manhattan plot to display the significance levels of genes with an ACAT-O combined p -value across all 22 chromosomes (Figure 4). As illustrated by the Manhattan plot, the majority of significant genes were located on chromosome 19, which is consistent with prior research findings. Additionally, we observed clusters of significant genes on chromosomes 6, 7, and 11.

Furthermore, a number of significant genes were also identified on chromosomes 1, 2, 5, 14, 15, 16, 17, and 20.

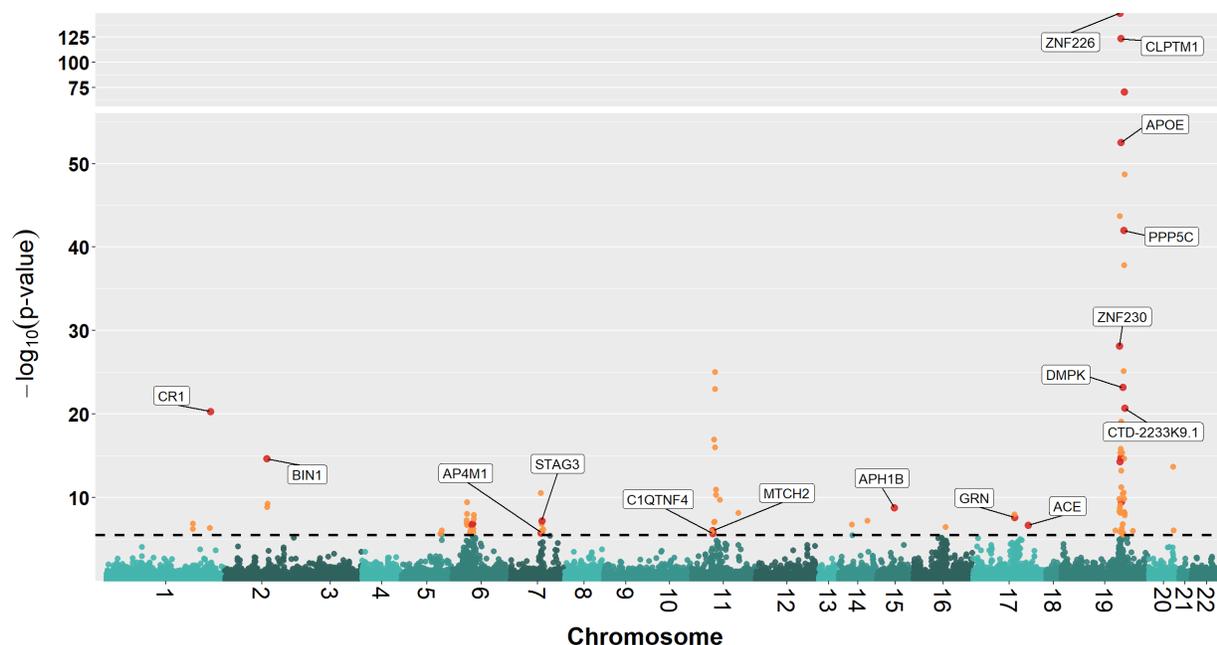


Figure 4. Manhattan plot of TWAS ACAT-O combined results. The horizontal dashed line represents the significance threshold of the p -value, which is determined by dividing 0.05 by the total number of genes. Orange markers indicate genes that were significant in only one tissue type, while red markers with labels highlight genes that were significant in at least two tissue types.

To examine the influence of utilizing *trans*-eQTLs, we conducted a separate TWAS analysis that excluded *trans*-eQTLs and only utilized *cis*-eQTLs. This analysis showed that 64 of the previously identified 93 significant genes retained their significance, while 29 genes were no longer significant, indicating that the significance of these genes was primarily derived from *trans*-eQTLs. We compiled a summary of part of the 93 genes significantly associated with AD dementia as displayed in Table 1, categorizing them based on whether their significance resulted from *cis*-eQTLs or *trans*-eQTLs.

Genes significant due to <i>cis</i> -eQTLs			Genes significant due to <i>trans</i> -eQTLs		
Gene	CHR	P value	Gene	CHR	P value
CR1 ^{ab}	1	4.99E-21	RP11-343H5.4 ^b	1	4.80E-07
BIN1 ^{bc}	2	2.22E-15	DDR1 ^a	6	2.08E-07
HLA-DRB1 ^{ab}	6	1.74E-07	BAG6 ^c	6	1.69E-07
TAP2 ^{ac}	6	1.28E-07	ATAT1 ^b	6	1.43E-07
AP4M1 ^{ab}	7	1.67E-06	DHX16 ^a	6	5.46E-08
STAG3 ^{ab}	7	5.91E-08	XXbac-BPG252P9.10 ^a	6	9.65E-09
C1QTNF4 ^{bc}	11	2.00E-06	TUBB ^b	6	3.88E-10
SLC39A13 ^{ab}	11	1.19E-06	CTNND1 ^a	11	9.02E-08
MTCH2 ^{ab}	11	8.79E-07	SLC3A2 ^c	11	2.01E-10
MS4A3 ^c	11	9.82E-17	VWCE ^b	11	4.95E-11
MS4A4A ^c	11	9.99E-24	AP001350.4 ^b	11	1.22E-17
MS4A6A ^c	11	1.03E-25	RP5-882C2.2 ^a	17	1.03E-08
APH1B ^{ac}	15	1.56E-09	PGLYRP1 ^a	19	2.82E-06
ACE ^{abc}	17	2.00E-07	CTC-435M10.12 ^c	19	8.44E-07
GRN ^{ab}	17	2.37E-08	SNRPD2 ^b	19	1.52E-07
GEMIN7 ^{ab}	19	3.09E-10	ARHGAP35 ^c	19	1.45E-08
CTB-179K24.3 ^c	19	6.47E-14	TMEM160 ^c	19	6.75E-09
ZNF285 ^{bc}	19	5.11E-15	CTB-12A17.2 ^a	19	5.35E-09
APOC2 ^{abc}	19	1.78E-15	SLC1A5 ^b	19	1.49E-10
BCAM ^c	19	4.44E-16	AC007191.4 ^b	19	3.59E-11
GPR4 ^a	19	4.44E-16	CALM3 ^b	19	2.44E-11
RELB ^c	19	1.51E-16	GNG8 ^b	19	2.22E-15
DMPK ^{ab}	19	6.52E-24	TRAPPC6A ^c	19	8.13E-20
ZNF230 ^{bc}	19	7.13E-29	CTD-2233K9.1 ^{ab}	19	2.00E-21
CTC-204F22.1 ^b	19	1.91E-44	PTGIR ^c	19	7.70E-26
APOE ^{bc}	19	3.03E-53	DACT3-AS1 ^a	19	1.61E-38
CLPTM1 ^{abc}	19	2.29E-124	PPP5C ^{ab}	19	9.86E-43
ZNF226 ^{ac}	19	1.14E-149	ZC3H4 ^a	19	2.04E-49
CASS4 ^c	20	2.15E-14	DACT3 ^{ab}	19	1.19E-71

a: Genes significant in the prefrontal cortex b: Genes significant in the cortex
c: Genes significant in the whole blood

Table 1. Part of significant genes of TWAS ACAT-O combined results. The right column of the table includes all the genes significant primarily due to *trans*-eQTLs, while part of the significant genes primarily driven by *cis*-eQTLs are included in the left column of the table.

3.4 Known risk genes for AD dementia

We next sought to quantify the extent of overlapping results between our BGW-TWAS findings and TWAS findings by TIGAR [27]. Comparing our ACAT-O combined results with TWAS findings by TIGAR with reference transcriptomic data of prefrontal cortex tissue [28], 9 out of 93 genes were identified by both methods. These include *HLA-DRB1* on chromosome 6, *OSBP* on chromosome 11, *ACE* on chromosome 17, as well as *BCL3*, *CLPTM1*, *DMPK*, *GEMIN7*, *ZNF230*, and *ZNF296* on chromosome 19. Additionally, 34 out of the 93 genes we identified were also detected in other TWAS or GWAS studies as being related to AD. Particular attention should be paid to *ACE*, *APOC2*, and *CLPTM1*, as they showed significance across all three tissue types.

Our study identified several well-known genes related to AD dementia. For example, according to the burden Z-score direction, we observed that the GR_{EX} of *APOE* in the cortex was positively associated with risk of AD dementia, while a significant negative association was observed in the whole blood. The GR_{EX} of another well-known AD risk gene on chromosome 9, *APOC2*, was found to be positively associated with risk of AD dementia in the prefrontal cortex and the cortex, and negatively associated with risk of AD dementia in the whole blood. We also identified that the GR_{EX} of *ACE* and *CLPTM1* was positively associated with risk of AD dementia in all the three tissues we tested.

In addition to well-known genes, we also identified some genes that have been detected only in recent TWAS studies for AD. For instance, a TWAS conducted with hippocampus tissue in 2021 [29] showed that the expression of *DACT3*, *SNRPD2* and *DMPK* in the hippocampus affects the risk of AD, while we also identified these genes as significant in our study. Moreover, a recent study [30] that integrated eQTL data and GWASs of late-onset AD (LOAD) by a Bayesian statistical method identified risk loci on gene *ZNF226*, which is consistent with our findings. Although *TNIP1* has recently been identified to having eQTLs of AD within blood tissue in a study [18], it was not detected as related to AD in another

study [31]. Our TWAS identified *TNIP1* as a risk gene of AD dementia in whole blood, which add evidence to support the previous statement.

3.5 Novel findings of risk genes for AD dementia

Among the ACAT-O results, we found that 50 out of the 93 significant genes identified by BGW-TWAS were novel findings that have not been reported in previous TWAS nor GWAS studies. Notably, the majority of genes found significant due to *trans*-eQTLs were not accounted by prior TWAS, with the exception of *TRAPPC6A*, *DMPK*, and *DACT3* on chromosome 19.

There is evidence from previous biological studies indicating that some of the novel genes we found by BGW-TWAS have associations with AD. For example, *UGGT1*, located on chromosome 2, is a gene responsible for creating N-glycosylation-related proteins in the endoplasmic reticulum. A 2022 study [32] revealed that *UGGT1* protein expression was upregulated in AD brain capillaries. In contrast, our study showed that GReX of *UGGT1* in the cortex is negatively related to AD dementia.

Additionally, a 2022 study [33] suggested that *BAG6* could prevent the accumulation and aggregation of misfolded proteins with exposed hydrophobic regions, which is a potential cause of AD dementia. This study is in accordance with our finding that GReX of *BAG6* is positively associated with AD dementia in the whole blood.

We identified *DHX16* as a novel risk gene for AD dementia, which was also recognized as a candidate risk gene for early onset AD in the study by Victoria Fernandez et al [34]. According to this study, the RNA helicase Dhx16 is involved in transcription alterations and DNA methylation changes that play a role in memory-related neurological and neuropsychiatric disorders.

Our BGW-TWAS results demonstrated that *TUBB* was a novel AD risk gene in the cortex. *TUBB* protein is a principal component of microtubules, which are formed by the polymerization of α -tubulin and β -tubulin dimers that bind to Guanosine-5'-triphosphate (GTP). It has been reported that higher levels of β -tubulin can be associated with aberrant hyper-phosphorylated tau aggregates, which play a crucial role in etiology of AD [35]. This finding indicates the potential impact of TUBB on AD development.

3.6 Protein-protein association networks and phenotype enrichment analysis by STRING

To further understand the underlying biological mechanism of our identified TWAS risk genes, we generated protein-protein association networks and performed phenotype enrichment analysis of the significant genes using the STRING tool (Figure 5). STRING is a powerful online database and software platform designed to provide comprehensive insights into protein-protein interactions and functional associations. We utilized STRING to construct protein-protein association networks for the significant genes identified in ACAT-O results. These networks are essential for understanding the complex interplay among proteins and their functional relationships within cellular pathways and systems. After the network analysis, STRING also generated phenotype enrichment analysis results. This analysis helps elucidate the functional implications of these genes in the context of AD dementia and other related disorders.

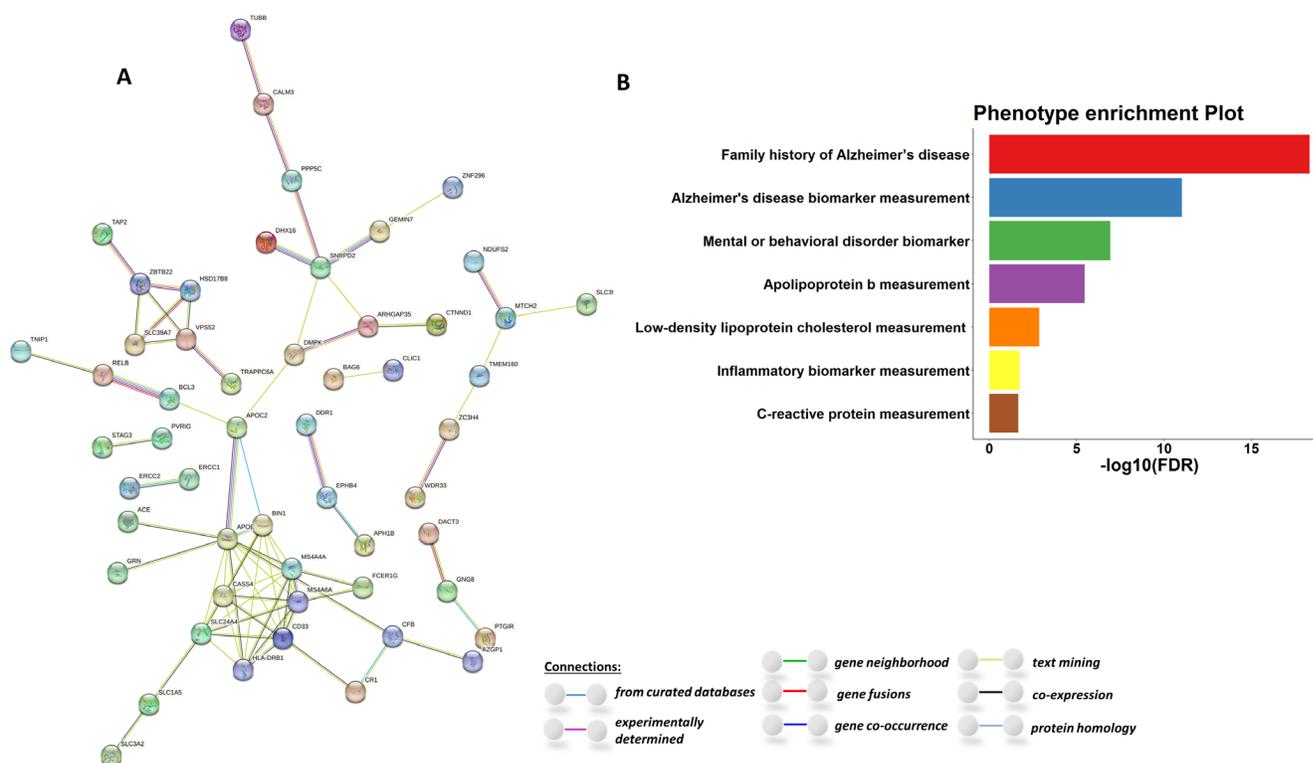


Figure 5. Protein-protein association networks and phenotype enrichment plot. Graph A is the protein-protein association network of the 93 significant genes generated by STRING. The colors of different connections have different meanings as shown below the plot. Graph B is the phenotype enrichment plot of the significant genes, with x-axis being $-\log_{10}(\text{FDR})$ of the phenotype, and y-axis showing different phenotypes enriched in the genes.

3.6.1 Protein-protein association networks

The networks identified by STRING highlight a major functional cluster composed of known AD risk genes *APOE*, *BIN1*, *CASS4*, *MS4A4A*, *MS4A6A*, *SLC24A4*, *CD33*, and *HLA-DRB1*. Three potential pathways were found to be enriched within these gene networks. Two of them, the CD20-like family/membrane-spanning 4-domains subfamily and MHC class I protein-binding genes, involve immune system processes. These pathways include B-cell surface antigen CD20 or MS4A and MHC class I proteins, which present antigens to T cells. It is noticeable that the CD20-like family / MS4A genes *MS4A4A* and *MS4A6A* were reported as AD risk genes [36], and MHC class I proteins are proved to be critical for maintaining neuronal structural complexity in aging brains [37]. The third pathway involves simultanagnosia and

sodium/potassium/calcium exchanger 4, including *SLC24A4* and related genes such as *CD33*, *CASS4*, and *MS4A4A*. There has been evidence supporting the association between *SLC24A4* and AD, with previous studies showing that AD risk variants in *SLC24A4* correlate with increased expression in blood and brain regions [38]. Our study also found that the GRex of *SLC24A4* in the whole blood positively correlated with risk of AD dementia. Apart from the known genes within the main cluster, several novel genes are connected to it. For example, *SLC1A5* and *SLC3A2* are connected to *SLC24A34*, while *CFB* and *AZGP1* are connected to *APOE*.

In addition to the main cluster discussed above, two other clusters are associated with *APOC2*. The first cluster consists of *BCL3*, *RELB*, and *TNIP1*, while the second cluster includes *DMPK*, *ARHGAP35*, *CTNND1*, *SNRPD2*, *DHX16*, *GEMIN7*, *ZNF296*, *PPP5C*, *CALM3*, and *TUBB*. All genes in the first cluster are known AD risk genes. Both *BCL3* and *RELB* have AD risk variants near the *APOE* region [39], and AD-related eQTLs have been detected within *TNIP1* [18]. Among the genes in the second cluster, *DMPK*, *SNRPD2*, *GEMIN7* and *ZNF296* are known AD risk genes, with the remaining being novel findings. Among the known genes, *DMPK* and *SNRPD2* were identified as AD risk genes in the hippocampus and putamen by a previous TWAS [29]. Moreover, *GEMIN7* and *ZNF296* has been identified as AD risk genes in a TWAS conducted on cortex and the amygdala region of the brain [40], respectively.

Another network worth paying attention to is the cluster composed of *TAP2*, *ZBTB22*, *HSD17B8*, *SLC39A7*, *VPS52* and *TRAPPC6A*. Except for *TRAPPC6A*, these genes are all novel findings associated with AD risk on chromosome 6. Among these genes, *ZBTB22* and *SLC39A7* have been shown to be associated with zinc finger in some capacity: *ZBTB22* is a gene that encodes a transcription factor containing both zinc finger and *BTB* domains, and *SLC39A7* is a gene that encodes a zinc transporter protein. Although we know that some zinc finger proteins influence the accumulation of tau proteins to affect the neurofibrillary tangles[41], which is a confirmed pathogenic factor of AD, we are not clear how *ZBTB22* and *SLC39A7* may related to this process. In addition, *TAP2* acts as a molecular scaffold for the final

stage of MHC class I folding, which was detected as a potential AD-related function in the first cluster. As for the gene *TRAPPC6A*, it has been previously implicated in association with AD [42].

Another cluster, which is composed of *WDR33*, *ZC3H4*, *TMEM160*, *MTCH2*, *NDUFS2*, and *SLC39A13*, is combined with known AD risk genes *MTCH2* and *NDUFS2* and novel AD risk genes. *MTCH2* is a gene that plays a crucial role in mitochondrial metabolism, and a recent study has shown that decrease of *MTCH2* level in the forebrain can impair cognitive functions related to the hippocampus and may eventually lead to AD [43]. Similarly, *NDUFS2* is a gene related to the oxidative phosphorylation process of mitochondrial metabolism which affects the risk of AD [14].

Within the network clusters discussed above, 19 novel genes are identified, and 17 of them lack known functions related to AD dementia. Although our understanding of these novel genes is limited, they are involved in networks with known AD risk genes. Further research is needed to investigate the potential roles of these novel genes in AD dementia and their interactions with neighboring genes in the cluster.

3.6.2 Phenotype enrichment analysis

The phenotype enrichment analysis showed that certain human phenotypes, such as family history of Alzheimer's disease (FDR = 4.57e-19) and mental or behavioral disorders (FDR = 1.18e-07), were enriched with our detected BGW-TWAS significant genes. Genes associated with family history of Alzheimer's disease included *APOE*, *BIN1*, *CR1*, etc., while genes associated with mental or behavioral disorder biomarkers included *HLA-DRB1*, *MS4A6A*, *GRN*, etc. Other enriched phenotypes included biomarkers for Alzheimer's disease (FDR = 9.52e-12), Apolipoprotein b (FDR = 3.44e-06), low-density lipoprotein cholesterol (FDR = 1.40e-03), inflammatory biomarkers (FDR = 1.79e-02), and C-reactive protein (FDR = 2.16e-02). Specific genes associated with these phenotypes included *MS4A4A*, *SLC24A4*, *APOC2*, and *ZNF226*.

Apart from phenotypes directly related to AD, our analysis identified significant genes associated with other phenotypes, such as Apolipoprotein B (ApoB), low-density lipoprotein cholesterol (LDL-C), inflammatory biomarkers, and C-reactive protein (CRP). These phenotypes have been reported to have potential associations with AD in previous studies. Elevated LDL-C levels have been implicated in AD development, as they can contribute to lipid-rich deposits in the brain, potentially leading to the formation of amyloid-beta plaques and neurofibrillary tangles, which are two major AD pathological features [44]. In addition, as a major protein component of LDL-C, several studies have suggested a link between elevated ApoB levels and a higher risk of AD dementia [45-48]. Chronic inflammation has been proposed as a significant factor to AD pathogenesis [49]. Increased levels of inflammatory biomarkers, such as cytokines and chemokines, have been observed in brains of AD patients [50]. CRP is an acute-phase protein produced in response to inflammation. Elevated levels of CRP have been associated with an increased risk of AD dementia [51]. Investigating these connections further may offer valuable insights into the risk factors of AD dementia and its development.

3.7 eQTLs of the significant genes

To investigate how eQTL contributed to the association significance by TWAS, we looked into the eQTL weights that were estimated by BGW-TWAS and used as variant weights for gene-based association tests in the second stage of TWAS. Particularly, these eQTL weights were given by the product of the posterior causal probability and estimated eQTL effect size. In Figure 6, we showed scatter plots of non-zero eQTL weights of a couple of example genes.

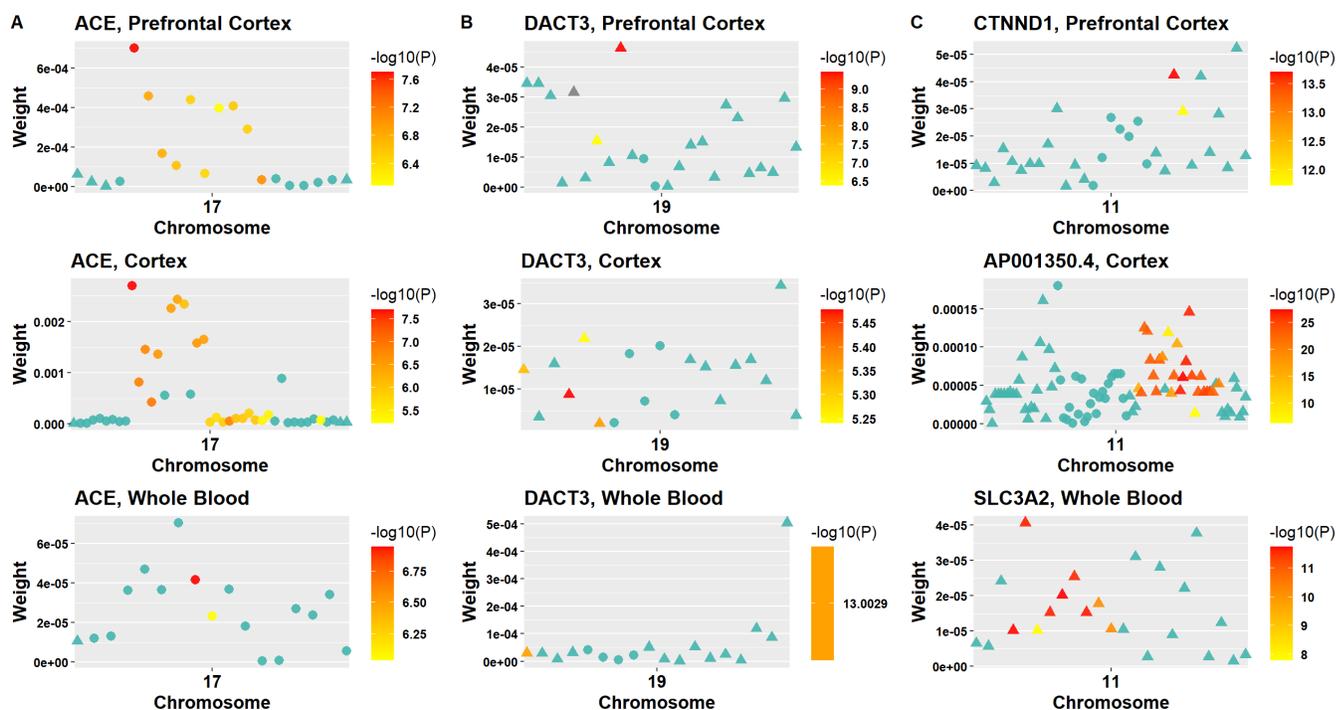


Figure 6. scatter plots of eQTL weights. In this set of scatter plots, the y-axis depicts the absolute values of eQTL weights, and x-axis shows the genome location of the eQTLs. For each eQTL depicted in the plots, circles denote *cis*-eQTLs, and triangles refer to *trans*-eQTLs. The scale of color shows the degree of the $-\log(p\text{-value})$ of the eQTL.

Across the three tissues, column A in Figure 6 shows the eQTLs of gene *ACE* whose significance is primarily due to *cis*-eQTLs, while column B shows the eQTLs of gene *DACT3* whose significance is primarily due to *trans*-eQTLs. Column C of the figure shows the eQTLs of three genes — *CTNND1*, *AP001350.4*, and *SLC3A2* — in the prefrontal cortex, cortex, and whole blood, respectively. In column A, it is evident that the *cis*-eQTLs within the 1 MB region of *ACE* on chromosome 17 primarily contribute to its significance. Conversely, for *DACT3*, the *trans*-eQTLs outside its region on chromosome 19 account for most of its significance. Column C further demonstrates examples of how *trans*-eQTLs contribute to the significance of genes. It is noticeable that all the *trans*-eQTLs of the example genes (and actually for most of the genes we studied) are still on the same chromosome where the gene located. Overall, these

plots in Figure 4 demonstrated that *trans*-eQTLs played a crucial role in identifying risk genes of AD dementia.

4. Discussion

In this study, we performed a BGW-TWAS analysis across three tissues - prefrontal cortex, cortex, and whole blood – to identify genes associated with AD dementia, and we identified 37, 55, and 51 genes with significant *p*-values in the three tissues respectively. After combining the TWAS *p*-values of genes across three tissues using ACAT-O, we obtained 93 genes with significant *p*-values. We further conducted a separate TWAS analysis that excluded *trans*-eQTLs and only utilized *cis*-eQTLs, and found that 64 of the previously identified 93 significant genes retained their significance, while 29 genes were no longer significant, indicating that the significance of these genes was primarily driven by *trans*-eQTLs.

Among the 93 genes identified as significant in our study, we observed that 9 of them were shared with a previous TWAS by TIGAR on the prefrontal cortex. Furthermore, 34 of these genes were also identified as AD risk genes in other TWAS or GWAS studies. Alongside the well-known genes, we also discovered 50 novel genes that have not been previously reported in either TWAS or GWAS investigations.

Importantly, most of the genes exhibiting significance due to *trans*-eQTLs were not detected in previous studies, indicating the potential of *trans*-eQTLs in uncovering novel disease-associated genes.

Through protein-protein association network analysis using STRING, we identified multiple network clusters containing both known and novel AD risk genes. Our findings are consistent with previous studies highlighting the critical involvement of *APOE* on chromosome 19 in AD pathology, as evidenced by its extensive connectivity with other known genes within the network plot. The analysis also underlines the important function of MHC class I protein binding genes (*MS4A4A* and *MS4A6A* in our

study) in the protein network related to AD dementia. Nevertheless, we also identified several clusters mainly comprised of novel AD risk genes that are not directly associated with well-established gene functions and networks of AD dementia. One such cluster included unknown risk genes *ZBTB22* and *SLC39A7*, which are associated with zinc finger generation and may potentially play a role in AD progression. Another cluster featured known AD risk genes *MTCH2* and *NDUFS2*, which are related to the oxidative phosphorylation process of mitochondrial metabolism. Despite these findings, the specific mechanisms by which certain genes contribute to AD pathogenesis are still largely unknown, particularly for novel genes. As such, further research is necessary to explore these novel network cluster findings and unravel their biological roles in AD pathology.

Through the phenotype enrichment analysis, we detected phenotypes enriched in the significant genes. Apart from the phenotypes directly related to AD dementia, some other phenotypes included Apolipoprotein B, low-density lipoprotein cholesterol, inflammatory biomarkers, and C-reactive protein. These phenotypes have been reported to have potential associations with AD dementia in various studies, suggesting a complex interplay between genetic, metabolic, and inflammatory factors in the pathogenesis of the disease. Further studies are needed to elucidate the precise mechanisms underlying these associations and to explore their implications for prevention and treatment of AD dementia.

Our study has several limitations that should be noted. First, due to computation burden of running BGW-TWAS, we only applied our BGW-TWAS analysis to three tissues – the prefrontal cortex, cortex, and whole blood - using the GTEx dataset to investigate AD dementia. However, other tissues, such as hippocampus, muscle, and spinal cord, are also known to play crucial roles in AD dementia [52-54]. Considering only three of these tissues may not capture the full spectrum of gene expression changes that contribute to pathogenesis of AD dementia. Future studies expanding the BGW-TWAS analysis to additional tissues could provide a more comprehensive understanding of the genetic basis of AD dementia and help identify novel therapeutic targets.

Second, although the summary-level GWAS data of AD dementia were meta-analysis results of multiple studies with diverse population, the GTEx reference transcriptomic data used in this study are mainly consisted with individuals of European descent. Especially, eQTL effect sizes can differ across populations due to factors such as population-specific linkage disequilibrium patterns and allele frequencies. Consequently, the gene-trait associations and potential risk loci identified in our study may not fully capture the genetic architecture of AD dementia in non-European populations. Future studies should include more reference transcriptomic populations in the analysis to obtain a more comprehensive understanding of genetics of AD dementia.

Third, the BVS model employed in the BGW-TWAS method inherently assumes a sparse representation of eQTLs contributing to gene expression. This sparsity assumption implies that only a small number of eQTLs have a substantial effect on gene expression, while the majority of eQTLs exert little or no impact. Although this assumption can be computationally advantageous, it may not always accurately represent the underlying genetic architecture of complex traits. Future research could explore alternative modeling approaches that relax the sparsity assumption, allowing for a more flexible representation of eQTL-gene relationships.

Overall, our study highlights the importance of considering *trans*-eQTLs in TWAS analysis as it can help identify significant risk genes that would have been missed in *cis*-eQTL-only TWAS. We identified several well-known AD-related genes, as well as novel genes that have potential associations with AD dementia. This study provides further insights into the genetic architecture of AD dementia and could help in the identification of potential therapeutic targets. To our knowledge, our study is the first to conduct a genome-wide TWAS of AD dementia utilizing both *cis*- and *trans*-eQTLs in the three tissue types, and our approach may offer a promising avenue for identifying hitherto unknown disease-related genes. Our findings could also serve as a valuable reference for future TWAS on AD dementia.

References

1. 2022 Alzheimer's disease facts and figures. *Alzheimers Dement*, 2022. **18**(4): p. 700-789.
2. Long, J.M. and D.M. Holtzman, *Alzheimer Disease: An Update on Pathobiology and Treatment Strategies*. *Cell*, 2019. **179**(2): p. 312-339.
3. Gatz, M., et al., *Role of genes and environments for explaining Alzheimer disease*. *Arch Gen Psychiatry*, 2006. **63**(2): p. 168-74.
4. Jansen, I.E., et al., *Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk*. *Nat Genet*, 2019. **51**(3): p. 404-413.
5. Gamazon, E.R., et al., *A gene-based association method for mapping traits using reference transcriptome data*. *Nat Genet*, 2015. **47**(9): p. 1091-8.
6. Zhao, B.X., et al., *Transcriptome-wide association analysis of brain structures yields insights into pleiotropy with complex neuropsychiatric traits*. *Nature Communications*, 2021. **12**(1).
7. Gusev, A., et al., *Integrative approaches for large-scale transcriptome-wide association studies*. *Nat Genet*, 2016. **48**(3): p. 245-52.
8. Nagpal, S., et al., *TIGAR: An Improved Bayesian Tool for Transcriptomic Data Imputation Enhances Gene Mapping of Complex Traits*. *Am J Hum Genet*, 2019. **105**(2): p. 258-266.
9. Vosa, U., et al., *Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression*. *Nat Genet*, 2021. **53**(9): p. 1300-1310.
10. Westra, H.J., et al., *Systematic identification of trans eQTLs as putative drivers of known disease associations*. *Nat Genet*, 2013. **45**(10): p. 1238-1243.
11. Luningham, J.M., et al., *Bayesian Genome-wide TWAS Method to Leverage both cis- and trans-eQTL Information through Summary Statistics*. *Am J Hum Genet*, 2020. **107**(4): p. 714-726.
12. Consortium, G.T., *The GTEx Consortium atlas of genetic regulatory effects across human tissues*. *Science*, 2020. **369**(6509): p. 1318-1330.
13. Serrano-Pozo, A., et al., *Neuropathological alterations in Alzheimer disease*. *Cold Spring Harb Perspect Med*, 2011. **1**(1): p. a006189.
14. Lunnon, K., et al., *Mitochondrial genes are altered in blood early in Alzheimer's disease*. *Neurobiol Aging*, 2017. **53**: p. 36-47.
15. Souza, V.C., et al., *Whole-Blood Levels of MicroRNA-9 Are Decreased in Patients With Late-Onset Alzheimer Disease*. *Am J Alzheimers Dis Other Demen*, 2020. **35**: p. 1533317520911573.
16. Sullivan, P.F., C. Fan, and C.M. Perou, *Evaluating the comparability of gene expression in blood and brain*. *Am J Med Genet B Neuropsychiatr Genet*, 2006. **141B**(3): p. 261-8.
17. Barbeira, A.N., et al., *Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics*. *Nat Commun*, 2018. **9**(1): p. 1825.
18. Wightman, D.P., et al., *A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease*. *Nat Genet*, 2021. **53**(9): p. 1276-1282.
19. Liu, Y., et al., *ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies*. *Am J Hum Genet*, 2019. **104**(3): p. 410-421.
20. Yongtao Guan, M.S., *Bayesian variable selection regression for genome-wide association studies and other large-scale problems*. *Ann. Appl. Stat.*, 2011. **5**(3): p. 1780-1815.
21. McCulloch, E.I.G.a.R.E., *Variable Selection Via Gibbs Sampling*. *Journal of the American Statistical Association*, 1993. **88**(423): p. 881-889.
22. Zhou, X., P. Carbonetto, and M. Stephens, *Polygenic modeling with bayesian sparse linear mixed models*. *PLoS Genet*, 2013. **9**(2): p. e1003264.
23. Yang, J., et al., *A Scalable Bayesian Method for Integrating Functional Information in Genome-wide Association Studies*. *Am J Hum Genet*, 2017. **101**(3): p. 404-416.

24. Chen, J., et al., *A scalable Bayesian functional GWAS method accounting for multivariate quantitative functional annotations with applications for studying Alzheimer disease*. HGG Adv, 2022. **3**(4): p. 100143.
25. A. P. Dempster, N.M.L.a.D.B.R., *Maximum Likelihood from Incomplete Data via the EM Algorithm*. Journal of the Royal Statistical Society. Series B (Methodological), 1977. **39**(1): p. 1-38.
26. Szklarczyk, D., et al., *The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets*. Nucleic Acids Res, 2021. **49**(D1): p. D605-D612.
27. Hu, T., et al., *Omnibus proteome-wide association study (PWAS-O) identified 43 risk genes for Alzheimer's disease dementia*. medRxiv, 2022: p. 2022.12.25.22283936.
28. De Jager, P.L., et al., *A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research*. Sci Data, 2018. **5**: p. 180142.
29. Liu, N., et al., *Hippocampal transcriptome-wide association study and neurobiological pathway analysis for Alzheimer's disease*. PLoS Genet, 2021. **17**(2): p. e1009363.
30. Rao, S., et al., *An APOE-independent cis-eSNP on chromosome 19q13.32 influences tau levels and late-onset Alzheimer's disease risk*. Neurobiol Aging, 2018. **66**: p. 178 e1-178 e8.
31. Katsumata, Y., et al., *Multiple gene variants linked to Alzheimer's-type clinical dementia via GWAS are also associated with non-Alzheimer's neuropathologic entities*. Neurobiol Dis, 2022. **174**: p. 105880.
32. Suzuki, M., et al., *Upregulation of ribosome complexes at the blood-brain barrier in Alzheimer's disease patients*. J Cereb Blood Flow Metab, 2022. **42**(11): p. 2134-2150.
33. Kasu, Y.A.T., et al., *BAG6 prevents the aggregation of neurodegeneration-associated fragments of TDP43*. iScience, 2022. **25**(5): p. 104273.
34. al, V.F.e., *An enrichment of rare variants and the lysosomal pathways are important contributors to early onset Alzheimer disease*. Alzheimer's & Dementia, 2021. **17**(S3).
35. Puig, B., et al., *BetaII-tubulin and phospho-tau aggregates in Alzheimer's disease and Pick's disease*. J Alzheimers Dis, 2005. **7**(3): p. 213-20; discussion 255-62.
36. Antunez, C., et al., *The membrane-spanning 4-domains, subfamily A (MS4A) gene cluster contains a common variant associated with Alzheimer's disease*. Genome Med, 2011. **3**(5): p. 33.
37. Lazarczyk, M.J., et al., *Major Histocompatibility Complex class I proteins are critical for maintaining neuronal structural complexity in the aging brain*. Sci Rep, 2016. **6**: p. 26199.
38. Tan, M.S., et al., *Associations of Alzheimer's disease risk variants with gene expression, amyloidosis, tauopathy, and neurodegeneration*. Alzheimers Res Ther, 2021. **13**(1): p. 15.
39. Nho, K., et al., *Association analysis of rare variants near the APOE region with CSF and neuroimaging biomarkers of Alzheimer's disease*. BMC Med Genomics, 2017. **10**(Suppl 1): p. 29.
40. Sun, Y., et al., *A transcriptome-wide association study of Alzheimer's disease using prediction models of relevant tissues identifies novel candidate susceptibility genes*. Genome Med, 2021. **13**(1): p. 141.
41. Bu, S., et al., *Zinc Finger Proteins in Neuro-Related Diseases Progression*. Front Neurosci, 2021. **15**: p. 760567.
42. Hamilton, G., et al., *Alzheimer's disease genes are associated with measures of cognitive ageing in the lothian birth cohorts of 1921 and 1936*. Int J Alzheimers Dis, 2011. **2011**: p. 505984.
43. Ruggiero, A., et al., *Loss of forebrain MTCH2 decreases mitochondria motility and calcium handling and impairs hippocampal-dependent cognitive functions*. Sci Rep, 2017. **7**: p. 44401.
44. Di Paolo, G. and T.W. Kim, *Linking lipids to Alzheimer's disease: cholesterol and beyond*. Nat Rev Neurosci, 2011. **12**(5): p. 284-96.

45. Wingo, T.S., et al., *Association of Early-Onset Alzheimer Disease With Elevated Low-Density Lipoprotein Cholesterol Levels and Rare Genetic Coding Variants of APOB*. JAMA Neurol, 2019. **76**(7): p. 809-817.
46. Hu, H., et al., *Association of serum Apolipoprotein B with cerebrospinal fluid biomarkers of Alzheimer's pathology*. Ann Clin Transl Neurol, 2020. **7**(10): p. 1766-1778.
47. Button, E.B., et al., *Vasoprotective Functions of High-Density Lipoproteins Relevant to Alzheimer's Disease Are Partially Conserved in Apolipoprotein B-Depleted Plasma*. Int J Mol Sci, 2019. **20**(3).
48. Picard, C., et al., *Apolipoprotein B is a novel marker for early tau pathology in Alzheimer's disease*. Alzheimers Dement, 2022. **18**(5): p. 875-887.
49. Heneka, M.T., et al., *Neuroinflammation in Alzheimer's disease*. Lancet Neurol, 2015. **14**(4): p. 388-405.
50. Holmes, C., et al., *Systemic inflammation and disease progression in Alzheimer disease*. Neurology, 2009. **73**(10): p. 768-74.
51. Locascio, J.J., et al., *Plasma amyloid beta-protein and C-reactive protein in relation to the rate of progression of Alzheimer disease*. Arch Neurol, 2008. **65**(6): p. 776-85.
52. Braak, H. and E. Braak, *Neuropathological staging of Alzheimer-related changes*. Acta Neuropathol, 1991. **82**(4): p. 239-59.
53. Manczak, M., M.J. Calkins, and P.H. Reddy, *Impaired mitochondrial dynamics and abnormal interaction of amyloid beta with mitochondrial protein Drp1 in neurons from patients with Alzheimer's disease: implications for neuronal damage*. Hum Mol Genet, 2011. **20**(13): p. 2495-509.
54. Pechlivanidou, M., et al., *Glial Gap Junction Pathology in the Spinal Cord of the 5xFAD Mouse Model of Early-Onset Alzheimer's Disease*. Int J Mol Sci, 2022. **23**(24).