

## **Brain and Blood Transcriptome-Wide Association Studies Identify Five Novel Genes Associated with Alzheimer's Disease**

Makaela A. Mews<sup>1</sup>, Adam C. Naj<sup>2,3</sup>, Anthony J. Griswold<sup>4,5</sup>, Alzheimer's Disease Genetics Consortium, Jennifer E. Below<sup>6</sup>, William S. Bush<sup>7</sup>

1) System Biology and Bioinformatics, Department of Nutrition, Case Western Reserve University School of Medicine, Cleveland, OH, 44106, USA

2) Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104, USA

3) Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104, USA

4) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, 33136, USA

5) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, 33136, USA

6) Vanderbilt Genetics Institute and Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, 37232, USA

7) Cleveland Institute for Computational Biology, Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, 44106, USA

## ABSTRACT

**INTRODUCTION:** Transcriptome-wide Association Studies (TWAS) extend genome-wide association studies (GWAS) by integrating genetically-regulated gene expression models. We performed the most powerful AD-TWAS to date, using summary statistics from *cis*-eQTL meta-analyses and the largest clinically-adjudicated Alzheimer’s Disease (AD) GWAS.

**METHODS:** We implemented the OTTERS TWAS pipeline, leveraging *cis*-eQTL data from cortical brain tissue (MetaBrain; N=2,683) and blood (eQTLGen; N=31,684) to predict gene expression, then applied these models to AD-GWAS data (Cases=21,982; Controls=44,944).

**RESULTS:** We identified and validated five novel gene associations in cortical brain tissue (*PRKAG1*, *C3orf62*, *LYSMD4*, *ZNF439*, *SLC11A2*) and six genes proximal to known AD-related GWAS loci (Blood: *MYBPC3*; Brain: *MTCH2*, *CYB561*, *MADD*, *PSMA5*, *ANXA11*). Further, using causal eQTL fine-mapping, we generated sparse models that retained the strength of the AD-TWAS association for *MTCH2*, *MADD*, *ZNF439*, *CYB561*, and *MYBPC3*.

**DISCUSSION:** Our comprehensive AD-TWAS discovered new gene associations and provided insights into the functional relevance of previously associated variants.

## 1 BACKGROUND

Alzheimer's disease (AD) has a strong genetic component, with heritability ( $h^2$ ) estimates ranging from  $h^2 \sim 60-80\%$  based on twin studies [1,2]. However, known AD risk variants discovered through genome-wide association studies (GWAS) explain only  $\sim 30\%$  of the genetic variance in disease risk while the remaining  $\sim 70\%$  is attributed to undiscovered AD variants [3]. Multiple recent large-scale GWAS [4,5] have been conducted with varying stringency in phenotyping criteria, with some studies using AD-by-proxy (GWAX) phenotypes [6] that may increase study heterogeneity [7] but also dramatically increase sample size. These GWAS have been successful at identifying new AD risk variants, though nearly all of these variants fall in non-coding regions making their roles in disease risk difficult to interpret. A key component to the translation of these findings into drug targets is understanding how these non-coding variants influence gene expression [8–10]. Transcriptomic integration with GWAS has the potential to both improve statistical power for discovery of novel genetic associations and accelerate associated drug development by mapping variants to their functional outcomes.

Mirroring the advancement of GWAS, studies of expression quantitative trait loci (eQTLs) -- which were initially small in scale -- have now been aggregated via meta-analysis for extremely well-powered studies [11,12]. While eQTLs generally have larger effects and are detectable in smaller sample sizes than variants associated with disease [13], prior work examining the genetic architecture of *cis*-regulatory variation [14] demonstrates that weaker eQTLs form polygenic components that complement strong, sparse eQTL effects [12]. While TWAS methods were originally developed to use

individual-level data [15], recent methodological developments have extended these approaches to allow for the use of summary statistics in both the creation of expression prediction models and for association analysis [16–25].

Previous transcriptome-wide association studies (TWAS) for AD have utilized smaller-scale eQTL reference datasets applied to both clinical AD and AD-by-proxy individual-level and summary data. Methodological extensions to the original TWAS framework like BGW-TWAS [26], T-GEN [27], VC-TWAS [28], UTMOST [29,30], InTACT [31], and MR-JTI [32,33] have been developed and applied to AD, leading to improvements in gene discovery by incorporating *trans*-eQTLs, epigenetic annotations, random *cis*-eQTL effect modeling, and multi-tissue modeling [34] that leverages shared tissue expression profiles to boost power. While these studies have advanced our understanding of how heritable gene expression is associated with AD risk, they have identified varying numbers of significant AD TWAS associations, from 8 [35] to 415 genes [32]. However, there has been limited functional validation or fine mapping of the results. Despite methodological improvements, most of these studies have relied on Genotype-Tissue Expression (GTEx) [36] data for model training. While GTEx provides genetic data across 54 diverse tissue types, representing unparalleled tissue diversity, the sample size of only 838 donors restricts generalizability.

In this study, we leverage the largest available AD GWAS summary statistics comprising over 65,000 individuals along with the largest blood and brain *cis*-eQTL summary statistics from published meta-analyses to identify novel AD-TWAS associations. Using the OTTERS approach [17], which combines diverse modeling techniques to account for variations in the sparseness or polygenicity of gene regulation, we generated

candidate TWAS associations. These candidates were then rigorously filtered and validated to distill the most robust TWAS signals. Additionally, we fine-mapped causal eQTLs and performed conditional analyses to provide deeper insights into the genetic architecture underlying the identified TWAS associations.

## 2 METHODS

### 2.1 Data Resources

#### 2.1.1 Sources of Summary Statistics

We leveraged the largest available *cis*-eQTL meta-analysis summary statistics for two key AD-associated tissues:

- Cortical brain tissue (N=2,683) from the MetaBrain study [11]
- Blood (N=31,684) from the eQTLGen study [12]

We performed an inverse variance-weighted fixed-effects meta-analysis across the MetaBrain study-specific results, excluding the GTEx study, which we used for model performance evaluation. Similarly, we obtained a version of the eQTLGen meta-analysis summary statistics that excluded GTEx.

We used the largest clinically diagnosed AD meta-analysis from Kunkle et al. 2019 (Stage I: AD Cases=21,982; Controls=44,944) to perform our primary analysis so our generated results would be most reflective of an AD phenotype (**Supplementary Table 1**) [37]. However, we also performed all analyses using AD-related Dementia (ADRD) summary statistics from Bellenguez et al. 2022 as this is the largest GWAS study to date related to AD [5] and include these analyses in **Supplementary Table 2**.

We limited our analysis to European-descent populations, as all four meta-analyses were based on this population.

### *2.1.2 Linkage Disequilibrium Reference Dataset*

Because we used summary statistics, an individual-level reference dataset was needed to estimate linkage disequilibrium (LD) for model building in OTTERS and fine-mapping. We used 503 unrelated EUR samples from the 1000 Genomes high coverage release [38] (Link:

[https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data\\_collections/1000G\\_2504\\_high\\_coverage/working/20220422\\_3202\\_phased\\_SNV\\_INDEL\\_SV/](https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/20220422_3202_phased_SNV_INDEL_SV/)) as a reference and filtered the dataset to only contain variants common to both the 1000 Genomes and GTEX (v8) (Link: <https://www.gtexportal.org/home/downloads/adult-gtex/overview>) .

### *2.1.3 Standardization of Summary Statistics*

Both MetaBrain [11] and ADRD summary statistics [5] were available in build 38. We lifted eQTLGen [12] and AD summary statistics [37] from GRCh37/hg19 to GRCh38/hg38 using UCSC liftOver (<https://hgdownload.soe.ucsc.edu/downloads.html>). To ensure consistency in effect direction/coded allele, we aligned all variants to GRCh38, removing variants with a reference or alternative allele that did not match the GRCh38 reference and flipping direction of the effect as appropriate. After our alignment procedure, we also wanted to ensure that the SNPs used in expression prediction models were available in all other datasets, so we restricted our analysis to only include variants consistently available within eQTL summary statistics, AD summary statistics, 1000 Genomes, and GTEx. Using these criteria, we examined a total of 7,859,831

SNPs for the brain cortex analysis and 7,653,537 SNPs for the whole blood analysis.

When using the ADRD summary statistics, we explored a total of 7,952,002 SNPs for the analysis of brain cortex analysis and 7,638,319 SNPs for the analysis of whole blood analysis.

## 2.2 OTTERS Implementation

We extended the recently published OTTERS framework [17] in several ways. In *Stage I: Training*, we used the standard five polygenic risk score (PRS) methods for summary statistic data implemented in OTTERS: p-value thresholding with LD clumping (P+T) at two thresholds ( $p < 0.05$  &  $p < 0.001$ ) [25]; elastic net regression implemented in the method *lassosum* [19]; Bayesian regression with a continuous shrinkage (CS) prior on SNP effect sizes implemented in the method PRS-CS [23]; and finally a non-parametric Bayesian multiple Dirichlet process regression method SDPR [24] (**Figure 1**). In *Stage II: Testing*, the p-values produced from each of these modeling approaches were subsequently aggregated using the aggregated Cauchy association test (ACAT) to generate an omnibus p-value (ACAT-O) [39] (**Figure 1**). An overview of our processing and results filtering workflow is shown in **Figure 1**.

While the original implementation of OTTERS (<https://github.com/daiqile96/OTTERS>) used the *pseudovvalidate* option within the *lassosum* (v.0.4.5) method to select the optimal tuning parameter between Ridge and Lasso regression, we instead used the *validate* option to generate our models with GTEX (v8) EUR normalized expression, covariate, and genetic data from Brain\_Cortex and Whole\_Blood (Link: <https://www.gtexportal.org/home/downloads/adult-gtex>) as validation sets. This modification increased model sparsity and had a higher predictive performance relative

to the *pseudovvalidate* option (**Supplemental Figure 1**). As we used GTEx v8 expression data to evaluate the performance of all imputation models, we prevented overfitting of the *lassosum* model by only including a randomly generated 50% split of the GTEx data (Brain: N=92/183; Blood: N=279/558) to serve as a validation testing dataset during model training and the remaining 50% served as an independent dataset for model evaluation. We also expanded the shrinkage and mixing parameters (balance between Ridge and Lasso regression) of the OTTERS *lassosum* implementation by also testing shrinkage penalties of 0.05, 0.1, 0.25, 0.33, 0.75, and 0.95 and mixing parameters of 0.05, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8, 0.9, and 0.95.

### 2.2.1 Post-processing of findings from OTTERS

In the original OTTERS implementation, invalid gene models are filtered based on a  $R^2 > 0.01$  and after models are adjusted for genomic control. We improve upon the existing approach by instead requiring an  $R > 0.1$  between the predicted and actual normalized expression of the gene based on GTEx v8 data as using only the  $R^2$  value masks genes models with negative correlations that are biologically nonsensical. In *Stage II: Testing*, we remove invalid models prior to adjusting for genomic control to keep them from skewing the adjustment (**Figure 1**). Overall, we used valid z-scores to adjust p-values for genomic control and subsequently ran the initial ACAT-O test. Multiple-testing was controlled by a Bonferroni-correction of the ACAT-O p-values identified at the end of *Stage II: Testing* (**Figure 1**). In *Stage III: Filtering*, we tested if z-score direction was concordant among tested methods and replaced gene models with NA values if there was any discordance across the methods, effectively requiring all methods to agree on the direction of effect (**Figure 1**). After we applied this



concordance restriction, we applied genomic control to the concordant z-scores, and re-ran that ACAT-O test to yield ACAT-O<sub>CON</sub> (concordance-restricted ACAT-O) p-values. By applying genomic control at this stage of the analysis, we reduced the overall ACAT inflation (Brain:  $\lambda = 1.58 > \lambda_{\text{CON}} = 1.17$ ; Blood:  $\lambda = 1.27 > \lambda_{\text{CON}} = 1.04$ ) (**Supplementary Figures 2-5**).

While each of the tested methods has advantages and disadvantages that lead to differing SNP selection, ultimately all methods receive the same set of input eQTL summary statistics. By enforcing concordance among the tested methods, we can enrich our results for higher confidence TWAS associations. We subsequently annotated whether a gene fell within a known AD GWAS association based on whether a 1Mb window around the gene's transcription start and end site overlapped with a 1Mb window around *APOE* or whether a 1Mb window around the gene fell within a 1Mb window around a known AD/ADRD GWAS association. We used a combined list of lifted-over AD GWAS associations [37] and ADRD GWAS associations [5] for this annotation of the results. Finally, we restricted our associations to genes where there was at least a nominally significant p-value ( $p < 0.01$ ) for a minimum of two out of five methods tested (in addition to a Bonferroni-corrected significant ACAT-O<sub>CON</sub> p-value) (**Figure 1**). We implemented these criteria to minimize the potential for false-positive associations and prevent any single method from biasing our results while still allowing for the detection of multiple different model types.

### 2.2.2 Independent Validation of OTTERS-identified Genes

In *Stage IV: Validation*, results meeting our post-processing criteria were further validated in an independent analysis of differential gene expression among AD cases and controls (**Figure 1**). We performed one-tailed T-tests (using the direction of effect predicted in our results) within a non-Hispanic white (NHW) whole-blood dataset [40] (AD Cases=119; Controls=117) and a partially independent brain RNA microarray dataset, KRONOSII [41], that was curated for high AD pathology and low secondary pathology (AD Cases=168; Controls=177). More specifically, we utilized the *RNA\_eQTL\_residual-corrected-data\_KRONOSII\_Brainome.txt* for the validation of brain cortex results (Link: [https://drive.google.com/drive/folders/15WmJsBXIH3Feib9Vu\\_bU2VpYQ2pDc8WX](https://drive.google.com/drive/folders/15WmJsBXIH3Feib9Vu_bU2VpYQ2pDc8WX)). For the whole blood validation, we used residualized normalized expression data after covariate adjustment for sex, age-at-exam, and principal components 1-12 (PC1-12) to capture population substructure estimated analysis using *flashpcaR* [42]. Because we used summary statistics, it was difficult to rule out the possibility that some samples from the KRONOSII study were included in the MetaBrain study or the AD summary statistics, however we assume based on documentation of the KRONOSII study that any overlap would be somewhat limited [41]. We used a Bonferroni-corrected p-value based on the number of genes tested to determine if a TWAS hit was validated in the actual expression data. In the microarray expression data, multiple probes mapped to a single gene in some instances, and we required at least one probe to be significantly differentially expressed between AD Cases and Controls for that the TWAS gene model to be validated.

### *2.2.3 Fine-mapping Validated Associations*

In *Stage V: Fine-mapping*, we identify the specific eQTLs driving the significant TWAS results by applying the SuSiE (Sum of Single Effects) [43,44] approach (*susieR* package v.0.12.35) to the original summary statistics while also using the same LD reference panel used in the OTTERS analysis. Credible sets of causal eQTLs were generated for eight of eleven validated TWAS associations, though seven of these eight SuSiE analyses reported convergence issues in the model fitting, which could indicate a mismatch between the eQTL summary statistics and the 1000 Genomes EUR LD matrix. For consistency however, we used the same LD matrix for all analyses. We extracted the variants falling within identified credible sets and subsequently re-ran the modified OTTERS pipeline using only the SuSiE identified SNPs to assess how the sparse fine-mapped eQTLs influence TWAS model results. We only re-ran our model for our validated TWAS associations from blood and brain. As we only performed analyses on a small set of genes, we did not adjust for genomic control and instead focused on the overall ACAT-O p-values. We annotated the credible SNPs included in the models using the Functional genomics repository (FILER) [45].

### *2.2.4 GWAS Conditional Analysis*

For our validated AD-TWAS associations that were proximal to known AD/ADRD GWAS loci, we sought to determine whether the genetically-regulated expression of the identified genes was driving the original GWAS association. To this end, we leveraged individual-level data from the Alzheimer's Disease Genetics Consortium (ADGC), encompassing 29,681 NHW participants across 35 cohorts. We conducted a series of logistic regression models that included both the predicted genetically-regulated

expression of the AD-TWAS gene and the GWAS SNP dosage, along with covariates such as sex, age, cohort, and population substructure (PC1-3). These analyses closely mirrored the approach used by Kunkle et al. 2019 [37]. When the logistic regression results warranted further investigation, we also performed linear regression models using either the GWAS SNP dosage or the genetically-regulated gene expression as the outcome variable, in order to disentangle the significant predictors. For this conditional analysis, we tested the top two most significantly associated modeling methods that produced valid gene models, with the exception of *MTCH2*, where we tested all sparse eQTL models as this was the only gene where every approach yielded a significant AD-TWAS association (**Supplementary Table 4**). Additionally, we used PLINK1.9 to perform linkage disequilibrium calculations ( $-r^2$  square) within this ADGC cohort to assess the strength of LD between AD/ADRD GWAS SNP and the causal eQTLs contained within the sparse genetically-regulated gene expression models.

### 3 RESULTS

#### 3.1 OTTERS Results and Quality Filtering

For brain-cortex, 993 models out of 9,282 were significantly associated with AD (Bonferroni-corrected ACAT-O  $p < 5.39 \times 10^{-6}$ ) (**Figure 1: Stage II**). We applied additional stringent filtering criteria to these prediction models, with 505 models showing concordant effect among all valid model z-scores (**Figure 1: Stage III**) and of those models 80 have nominal associations among at least two methods excluding five of these genes that fall within a 1Mb window of *APOE* (**Figure 1: Stage III**). For whole blood, 503 models out of 7,214 were significantly associated with AD (Bonferroni-corrected ACAT-O  $p < 6.93 \times 10^{-6}$ ) (**Figure 1: Stage II**). After filtering, 222 models show

concordant effect and of those models 42 have nominal associations amongst at least two individual methods excluding two of these genes that fall within a 1Mb window of *APOE* (**Figure 1: Stage III**).

### 3.2 Annotation of Significant TWAS Associations

We assumed significant TWAS associations from our OTTERS approach were driven by either 1) re-weighting and aggregating multiple sub-significant GWAS associations into a single TWAS gene association, or 2) identification of a TWAS-significant gene by either directly selecting a GWAS-associated variant in the expression prediction model or indirectly tagging a GWAS-associated variant through linkage disequilibrium. Thus, we annotated our TWAS results as either novel or GWAS-proximal, respectively, by creating 1Mb regions around variants identified as significant GWAS associations from Kunkle et al. 2019 and Bellenguez et al. 2022, and assessing whether this window overlaps with a 1Mb window around each TWAS gene [5,37]. Using this criterion, we identified 80 significant genes (excluding five genes falling within the *APOE* window) of which 50 were novel genes and 30 were GWAS association-proximal genes within brain cortex. Similarly, we identified 42 significant genes within blood (excluding two genes falling within *APOE* window) of which 18 were novel genes and 24 were GWAS association-proximal genes. Comparing results between blood and brain, there were no novel genes common between the two tissues. For genes that are proximal to known GWAS loci, we identified four genes that were identified in both whole blood and brain cortex, and which also had concordant effect directions in both tissues (*KANSL1*, *ARL17A*, *LRRC37A2*, and *C1QTNF4*) (**Supplementary Table 3**).

### 3.3 Validation in Expression Datasets

To validate TWAS associations from whole blood, we examined differences in gene expression within an independent whole-blood RNA-seq dataset of NHW, clinically-diagnosed AD cases (N=119) and age-and-sex-matched controls (N=117). Based on the direction of effect from each TWAS association, we tested if measured gene expression differed significantly between AD cases and controls using a one-tailed t-test. We were able to test 40/41 (98%) of our novel and GWAS-Proximal genes.

From this analysis, we validated one gene (*MYBPC3*) that was proximal to the known AD/ADRD GWAS locus *SPI1* (**Table 1**). Specifically, we found a positive association between *MYBPC3* (myosin-binding protein C) expression and AD risk (ACAT-O<sub>CON</sub> p =  $4.43 \times 10^{-8}$ ; t-test p =  $6.5 \times 10^{-4}$ ).

Brain cortex associations were validated using a brain microarray-based expression dataset of neuropathologically curated cases that were enriched for Alzheimer's disease pathology and minimized co-occurring neurodegeneration markers such as Lewy bodies, and controls were neuropathologically confirmed as having minimal pathology loads [41]. Based on the microarray data, we were able to assess 40/80 of TWAS associations identified using the AD GWAS summary statistics. We validated ten associations. Five of 23 novel AD TWAS associations were statistically significant (Bonferroni-corrected t-test p-value < 0.001; *C3orf62*, *LYSMD4*, *PRKAG1*, *ZNF439*, *SLC11A2*). In addition, we validated five AD TWAS associations (5/17) that were located proximal to a known AD/ADRD GWAS loci (Bonferroni-corrected t-test p-value < 0.001 *PSMA5*, *ANXA11*, *MTCH2*, *MADD*, *CYB561*).

### 3.4 Fine-mapping TWAS Associations

We next explored whether a sparse genetic model based on fine-mapped causal eQTLs for each *cis*-regulatory region would produce a similar strength TWAS association with improved interpretability. To test this assumption, we analyzed eQTL meta-analysis summary statistics using SuSiE and subsequently re-ran our OTTERS pipeline using only this subset of fine-mapped eQTLs (*Stage V: Fine-mapping, Figure 1*). Credible sets of causal eQTLs were generated for eight of eleven validated TWAS associations. Of the eight validated associations with credible sets, five genes (*MTCH2*, *MADD*, *CYB561*, *ZNF439*, and *MYBPC3*) had equivalent or improved AD-TWAS association statistics when the OTTERS analysis was restricted to the fine-mapped causal eQTLs within these credible sets (**Table 2**). In contrast, the remaining three genes (*C3orf62*, *LYSMD4*, and *ANXA11*) were no longer significant, indicating that their effects rely on a more polygenic genetic architecture than captured by the SuSiE fine-mapping (which was limited to 10 credible sets).

#### 3.4.1 Brain: *MTCH2* & *MADD*

The *MTCH2* OTTERS association was primarily driven by the PRS-CS approach modeling 1,336 SNPs (ACAT-O<sub>CON</sub>  $p = 3.12 \times 10^{-7}$ ;  $p_{\text{PRS-CS\_CON}} = 1.79 \times 10^{-7}$ ;  $Z_{\text{PRS-CS}} = -3.57$ ;  $R^2_{\text{PRS-CS}} = 5.7\%$ ). This model was reduced to seven credible sets with eight causal eQTLs by SuSiE fine-mapping, which produced significant AD-TWAS associations using all strategies and each method selected between one to three causal eQTLs. The most parsimonious model was generated by the lassosum approach and contained a single eQTL that falls with active enhancer/promoter regions based on chromatin accessibility and histone modification data (**Supplemental text**). This model

demonstrated slightly improved model performance compared to the original (5.7% vs. 6.3% variance in gene expression explained) and maintained a strong association to AD ( $p_{\text{Lassosum}} = 3.53 \times 10^{-8}$ ;  $z_{\text{Lassosum}} = -5.51$ ). This variant is located approximately 235kb downstream from the established *SPI1* AD locus. In our analysis, a valid model for *SPI1* within brain was generated only using SDPR and was not significantly associated with AD (ACAT- $O_{\text{CON}}$   $p = 0.47$ ;  $z_{\text{SDPR}} = 0.59$ ;  $R^2_{\text{SDPR}} = 1.0\%$ ). We ran a conditional analysis using a series of logistic regression models within a subset of cohorts analyzed by Kunkle et al. 2019 and confirmed that there are individually significant associations between the *SPI1* locus and AD status ( $z = -3.2$ ,  $p = 1.35 \times 10^{-3}$ ) and the genetically-regulated expression of *MTCH2* and AD status ( $z = -4.6$ ,  $p = 4.08 \times 10^{-6}$ ). However, when both variables are included within the model only the genetically-regulated expression of *MTCH2* remains significant ( $z = -3.3$ ,  $p = 9.09 \times 10^{-4}$ ) whereas the *SPI1* association is no longer significant ( $z = -0.18$ ,  $p = 0.86$ ). These results suggest that the genetically-regulated expression of *MTCH2* appears to be a driving factor behind the *SPI1* AD GWAS locus association.

However, *MADD* is located approximately 90kb upstream from the *SPI1* GWAS association [37] and 350kb upstream from *MTCH2*. The *MADD* OTTERS association was primarily driven by the P+T(0.001) approach modeling 307 SNPs (ACAT- $O_{\text{CON}}$   $p = 6.64 \times 10^{-8}$ ;  $p_{\text{P+T}_0.001_{\text{CON}}} = 2.22 \times 10^{-8}$ ;  $z_{\text{P+T}_0.001} = -1.65$ ;  $R^2_{\text{P+T}_0.001} = 4.9\%$ ). This model was reduced to six credible sets with 14 potential causal eQTLs by SuSiE fine-mapping, which produced a six SNP SDPR model demonstrating slightly improved model performance to the original (4.9% vs. 6.3% variance in gene expression explained). This model also maintained a strong negative association to AD ( $p_{\text{SDPR}} = 3.01 \times 10^{-8}$ ;  $z_{\text{SDPR}} = -$



5.54). Based on our earlier results with *MTCH2*, we assessed in a series of conditional models whether *MADD* contributes to the *SPI1* locus association and found that *MADD* genetically-regulated expression was not associated with AD in either the individual ( $z = -1.2$ ,  $p=0.23$ ) or combined models ( $z = -0.21$ ,  $p = 0.83$ ). Of note, we were able to impute the genetically-regulated expression of *MADD* using only five out of six SNPs within the model as the other SNP was unable to be imputed in the ADGC cohort, which may have slightly reduced the strength of the initial *MADD* association to AD. However, we confirmed using a linear model with imputed *MTCH2* expression as the outcome that the genetically-regulated expression of *MADD* ( $t_{MADD} = 10.05$ ) and the *SPI1* dosage ( $t_{SPI1} = 143.35$ ) were both significantly associated predictor variables ( $p < 2 \times 10^{-16}$ ). Moreover, the *SPI1* locus is in moderate LD with the single causal eQTL within the *MTCH2* expression model ( $r^2 = 0.45$ ) whereas the *MADD* SNPs are in weak LD with the *MTCH2* eQTL ( $r^2 < 0.09$ ) which is reflected in the strength of their effect sizes on *MTCH2* expression. Overall, these results suggest that both the *MADD* AD-TWAS association and *SPI1* locus GWAS association are driven by the *cis*-regulatory region surrounding *MTCH2*.

### 3.4.2 Brain: *CYB561*

The *CYB561* OTTERS association was driven by the SDPR approach modeling 1,616 SNPs (ACAT- $O_{CON}$   $p = 1.99 \times 10^{-33}$ ;  $p_{SDPR\_CON} = 3.98 \times 10^{-34}$ ;  $z_{SDPR} = -9.83$ ;  $R^2_{SDPR} = 28.1\%$ ). Fine-mapping identified seven credible sets each containing a single causal eQTL SNP. OTTERS analysis of these SNPs identified a six SNP SDPR model with nearly identical performance ( $R^2_{SDPR} = 24\%$ ) and produced a TWAS result of similar significance ( $p_{SDPR} = 1.8 \times 10^{-20}$ ;  $z_{SDPR} = -9.3$ ). *CYB561* is located 34kb upstream to the

identified GWAS associations located near *ACE* identified in Bellenguez et al [5], and only 24kb upstream of the *ACE* association in Kunkle et al. [37]. In our analysis, all modeling approaches produced valid results for *ACE* within the brain and overall the *ACE* gene model was nominally negatively associated with AD risk (ACAT- $O_{CON}$   $p = 0.001$ ;  $z = [-0.5, -2.6]$ ;  $R^2_{SDPR} = [6\%, 10\%]$ ).

As the *ACE* GWAS locus is located nearly in the middle of the promoter elements of *ACE* (~16kb) and *CYB561* (~24kb), we assessed in a series of logistic regression models if the genetically-regulated expression of *CYB561* is driven by its close proximity to the *ACE* GWAS locus. We found that while there was a significant association between the *ACE* locus and AD status ( $z = 2.9$ ,  $p = 3.49 \times 10^{-3}$ ), there was not a significant association between the genetically-regulated expression of *CYB561* and AD status ( $z = -0.757$ ,  $p = 0.449$ ). Moreover, we confirmed using a linear model with the dosage of the *ACE* GWAS locus as the outcome variable that the genetically-regulated expression of *CYB561* was the strongest significantly associated predictor variable ( $t = -18.9$ ,  $p < 2 \times 10^{-16}$ ) despite being in weak LD with the *ACE* GWAS locus ( $r^2 < 0.02$ ). These results suggest that the regulatory architecture underlying this genomic region is more complex and requires further exploration in larger sample sizes to clarify the association between *CYB561* genetically-regulated expression and AD.

### 3.4.3 Brain: *ZNF439*

The OTTERS association for *ZNF439* was driven by the lassosum approach, which modeled 1,336 SNPs. This model yielded a highly significant association to AD (ACAT- $O_{CON}$   $p = 2.21 \times 10^{-52}$ ;  $p_{Lassosum\_CON} = 1.1 \times 10^{-52}$ ;  $z_{Lassosum} = 9.94$ ;  $R^2_{Lassosum} = 5.7\%$ ).

Though causal eQTL fine-mapping, we identified five credible sets each containing a

single causal eQTL and subsequent application of these SNPs to the OTTERS approach led to the generation of a four SNP lassosum model demonstrating slightly reduced model performance in terms of variance explained ( $R^2_{\text{Lassosum}} = 5.0\%$ ) but further strengthened the association to AD ( $p_{\text{Lassosum}} = 3.79 \times 10^{-97}$ ;  $z_{\text{Lassosum}} = 20.92$ ). Functional annotation of these causal eQTLs revealed that the effect size direction of each SNP within the model corresponded to histone markers of active transcription or repressive chromatin, underscoring the additional biological complexity underlying the functional roles of these genetic variants (**Supplemental Text**).

### 3.4.3 Blood: *MYBPC3*

The original OTTERS association to *MYBPC3* was driven most strongly by the P+T 0.001 approach modeling 1,081 SNPs (ACAT- $O_{\text{CON}}$   $p = 4.43 \times 10^{-8}$ ;  $p_{\text{P+T}_{0.001_{\text{CON}}}} = 1.75 \times 10^{-8}$ ;  $z_{\text{P+T}_{0.001}} = 1.41$ ;  $R^2_{\text{P+T}_{0.001}} = 3.6\%$ ). Fine-mapping of *MYBPC3* yielded four credible sets each containing a single causal eQTL. Restricting the OTTERS method to the credible causal eQTLs, the SDPR, lassosum, and PRS-CS modeling approaches were able to explain only a negligible proportion of gene expression, falling short of our  $R^2 > 1\%$  filtering threshold. In contrast, the P+T methods used a two-SNP model and maintained a statistically significant positive association with AD risk ( $p_{\text{P+T}_{0.001}} = 6.48 \times 10^{-44}$ ,  $z_{\text{P+T}_{0.001}} = 13.9$ ,  $R^2_{\text{P+T}_{0.001}} = 1\%$ ).

As *MYBPC3* is approximately 90kb upstream the *SPI1* AD GWAS locus, we also ran a conditional analysis using a series of logistic regression models to determine whether the genetically-regulated expression of *MYBPC3* within blood may be contributing to the *SPI1* locus signal. We confirmed that the *SPI1* locus was individually associated with AD status ( $z = -3.2$ ,  $p = 1.35 \times 10^{-3}$ ) and that there was a marginal association between

the genetically-regulated expression of *MYBPC3* and AD status ( $z = 1.8$ ,  $p = 0.07$ ). However, when both variables were included within the model, only the *SPI1* locus association remained significant ( $z = -2.73$ ,  $p = 0.006$ ), while the *MYBPC3* genetically-regulated expression association was no longer marginally significant ( $z = -0.66$ ,  $p = 0.51$ ). Of note, we were only able to impute the genetically-regulated expression of *MYBPC3* using a single-SNP, as the other SNP could not be imputed in the ADGC cohort, which may have reduced the strength of the initial *MYBPC3* association to AD. Additionally, the single SNP within the *MYBPC3* expression model is in moderate LD with the *SPI1* GWAS variant ( $r^2 = 0.49$ ), which may explain the previously found marginal association to AD. Overall, these results suggest that the *MYBPC3* AD-TWAS association in blood does not contribute to the AD-GWAS signal at the *SPI1* locus.

### 3.5 Conditional Analyses in polygenic AD-TWAS Associations

We identified two genes (Brain: *PSMA5* and *ANXA11*) that were located proximal to known GWAS loci and potentially driven by polygenic eQTL architectures that were unable to be fine-mapped.

#### 3.5.1 Brain: *PSMA5*

Our identified *PSMA5* OTTERS association was driven most strongly by the SDPR approach, which modeled 2,592 SNPs (ACAT- $O_{CON}$   $p = 5.25 \times 10^{-8}$ ;  $p_{SDPR\_CON} = 1.31 \times 10^{-8}$ ;  $z_{SDPR} = -4.59$ ;  $R^2_{SDPR} = 3.7\%$ ), followed closely by the PRS-CS approach also modeling 2,592 SNPs ( $p_{PRS-CS\_CON} = 2.81 \times 10^{-4}$ ;  $z_{PRS-CS} = -2.48$ ;  $R^2_{PRS-CS} = 5.9\%$ ). However, we were unable to fine-map any causal eQTLs within *PSMA5* using SuSiE. Given that the *SORT1* ADRD GWAS locus [5] is located approximately 66kb from the

transcription start site of *PSMA5*, we assessed in a series of logistic regression models whether the *PSMA5*AD-TWAS association was driven by its proximity to the *SORT1* locus. Importantly, in our brain analyses we did not generate a valid model to predict *SORT1* expression.

We found a significant association between the *SORT1* locus and AD status ( $z = 6.44$ ,  $p = 2.62 \times 10^{-3}$ ) in the ADGC NHW cohort. While the SDPR model of *PSMA5* genetically-regulated expression was not significantly associated with AD ( $z = 1.40$ ,  $p = 0.16$ ), the PRS-CS model was significantly associated with AD ( $z = 2.25$ ,  $p = 0.024$ ). Notably, there was minimal LD between the *SORT1* locus and the SNPs contained with the *PSMA5* model ( $r^2 < 0.04$ ). In the combined logistic regression model, both the *SORT1* locus and PRS-CS genetically-regulated expression of *PSMA5* retained their significant associations to AD ( $p_{\text{SORT1}} = 2.88 \times 10^{-3}$ ;  $p_{\text{PRS-CS}_{\text{PSMA5}}} = 0.027$ ), consistent with the slight LD identified between these genetic signals. Collectively, our results suggest that the identified TWAS association between *PSMA5* and AD represents a distinct genetic signal, independent of the previously reported GWAS association at the nearby *SORT1* ADRD locus.

### 3.5.2 Brain: *ANXA11*

The *ANXA11* TWAS association was driven mostly strongly by the SDPR approach, which modeled 3,157 SNPs (ACAT- $O_{\text{CON}}$   $p = 3.37 \times 10^{-12}$ ;  $p_{\text{SDPR}_{\text{CON}}} = 8.42 \times 10^{-13}$ ;  $z_{\text{SDPR}} = 5.77$ ;  $R^2_{\text{SDPR}} = 1.8\%$ ). Through fine-mapping, we were able to identify six credible sets containing seven causal eQTLs. However, when we re-ran the OTTERS approach restricted to these causal eQTLs, the AD-TWAS association for *ANXA11* diminished to nominal significance ( $p = [0.034, 0.044]$ ), despite the genetically-regulated expression

explained by the model increasing from 1.8% to 4.9%. Given that *ANXA11* is located ~343kb upstream of the *TSPAN14* ADRD GWAS locus, we assessed whether *TSPAN14* locus might be driving the *ANXA11* AD-TWAS association. Using a series of logistic regression models, we found that neither the *TSPAN14* locus nor the genetically-regulated expression of *ANXA11* was associated with AD status in this cohort. Interestingly, we did identify several SNPs within the 3,157 SNP SDPR model that were within high LD ( $r^2 > 0.9$ ) with the *TSPAN14* locus. Additionally, *ANXA11* was also identified as a significant ADRD-TWAS association (**Supplementary Table 2**), suggesting that the initial AD-TWAS association may be driven instead by co-occurring forms of dementia (e.g., Fronto-temporal dementia) and/or misdiagnosis. Furthermore, our findings suggest that the *TSPAN14* ADRD GWAS locus may be associated with a broader ADRD phenotype rather than a clinical AD phenotype.

#### 4 DISCUSSION

In this study, we leveraged the largest available eQTL meta-analysis summary statistics from both cortical brain tissue and blood, and applied them to the largest clinically adjudicated AD case-control GWAS dataset through a summary-statistics based TWAS framework. This approach utilized five different methods to capture a range of sparse to polygenic genetic architectures underlying gene expression regulation. Using this TWAS framework, we identified and validated five novel genes within cortical brain tissue (*PRKAG1*, *C3orf62*, *LYSMD4*, *ZNF439*, *SLC11A2*) where the genetically-regulated expression of these genes was significantly associated with AD status. Additionally, we identified and validated six genes that were proximal to known AD/ADRD GWAS associations (Blood: *MYBPC3*; Brain: *MTCH2*, *CYB561*, *MADD*, *PSMA5*, *ANXA11*).

Finally, we fine-mapped causal eQTLs and obtained similar or improved TWAS power for five genes (*MTCH2*, *MADD*, *CYB561*, *ZNF439* and *MYBPC3*) when using sparse eQTL prediction models.

Our identified novel validated associations span a wide spectrum of functional understanding, with genes having limited information on function (*LYSMD4*, *C3orf62*, *ZNF439*) to genes with recognized roles in energy metabolism (*PRKAG1*) and iron homeostasis (*SLC11A2*).

*PRKAG1* (Protein Kinase AMP-activated Non-Catalytic Subunit Gamma 1) encodes a component of AMPK and is involved in sensing cellular energy. During metabolic stress, AMPK inhibits macromolecule biosynthesis and cell growth while activating energy-producing pathways. In our study, we found that low levels of *PRKAG1* within cortical brain tissue were associated with an increased AD risk (ACAT-O<sub>CON</sub>  $p = 1.73 \times 10^{-7}$ ). Interestingly, a 2012 study using healthy adult human cortical slices exposed to sublethal doses of amyloid-beta soluble oligomers found *PRKAG1* to be upregulated after exposure [46]. This work suggests our association may be driven by lack of proper biological compensation during high AD pathology and further work is needed to clarify the role of *PRKAG1*. While *PRKAG1* has limited associations with Alzheimer's disease, activators of AMPK are recognized for their therapeutic potential in treating AD such as by promoting autophagy and reducing insulin resistance [47]. For instance, researchers have found that moderate aerobic exercise can counteract amyloid-beta-induced learning and memory impairment in animal models through in part AMPK restoration [48].

*SLC11A2* (Solute carrier family 11 member 2) encodes a proton-coupled divalent metal ion transporter that plays a critical role in iron homeostasis and has been previously associated with amyotrophic lateral sclerosis (ALS) and Parkinson's disease. In our study, we found reduced expression of *SLC11A2* within the brain was associated with an increased AD risk (ACAT-O<sub>CON</sub>  $p = 4.39 \times 10^{-14}$ ). While there are limited studies specifically exploring the connection between *SLC11A2* and AD, one prior study used 216 AD cases and 323 controls found a nominally significant ( $p = 0.08$ ) association between a variant (rs407135) within *SLC11A2* and AD [49]. Interestingly, in a prion mouse model *Slc11a2* expression within the hippocampus was found to be sustainably reduced during disease progression which is concordant with our findings [50].

Our TWAS associations identified as proximal to known AD/ADRD loci include *PSMA5* (component of 20S core proteasome complex), *ANXA11* (calcium-dependent phospholipid-binding protein), *MTCH2* (mitochondrial insertase and regulator of apoptosis and lipid homeostasis), *MADD* (adaptor protein regulating apoptosis through activation of mitogen-activated protein kinase), and *CYB561* (Transmembrane Ascorbate-Dependent Reductase).

Several previous AD TWAS studies have identified *MTCH2* as being associated with AD status, and our work further supports the notion that the genetically-regulated expression of *MTCH2* appears to drive the GWAS association at the *SPI1* locus [34,35]. Notably, we were able to retain the strength of this association using only a single causal eQTL for *MTCH2* genetically-regulated expression. This result builds upon prior work by Gockley et al. 2021 as they also identified *MTCH2* genetically-regulated expression as significantly negatively associated with AD using six neo-cortical brain



tissues (N\_RNA-seq = 888) [35]. Through joint-conditional probability analysis and colocalization analysis, these researchers identified that *MTCH2* likely shares a single causal variant with the *SPI1* locus [35]. However, their best performing models for *MTCH2* expression contained 260 SNPs, none of which overlapped with the SNPs identified as causal eQTLs through fine-mapping and/or retained in our OTTERS models, which highlights the importance of fine-mapping causal eQTLs using the largest available eQTL resource (**Supplementary Text**). *MADD* is also located near *SPI1* and has been found to be negatively associated AD status and our results support that this association may be driven by *MTCH2*. Critically, we cannot exclude the possibility that other nearby genes contribute to AD risk at this locus, and *SPI1* has been functionally validated with extensive modeling within myeloid cells, especially microglia, for its role in AD risk [51].

Interestingly, *ANXA11* has been found in prior work to be associated with amyotrophic lateral sclerosis (ALS) with or without frontotemporal dementia (FTD) and our work further supports the role of *ANXA11* in neurodegeneration [52]. *PSMA5* has also been associated with AD based on several *in vivo* animal models that have identified *PSMA5* downregulation is involved in APP-induced inhibition of cell proliferation, and our work further motivates additional studies to clarify how *PSMA5* expression modulation impacts AD pathogenesis [53–55]

There are several considerations to our study. First, we examined only *cis*-regulatory variants, as the OTTERS framework was designed for *cis*-eQTLs, and prior work has found that *trans*-eQTL effects tend to be relatively weaker [56]. While we used a standard 1 Mb window around the transcription start and end sites, we acknowledge

that regions outside of this window may also harbor relevant eQTLs due to the complex 3D architecture of chromatin interactions.

Additionally, we conducted our analysis on cortical brain tissue, as it had the largest brain sample size available, in order to maximize statistical power for detecting tissue-specific effects, which can be challenging to disentangle using multi-tissue modeling. Prior work has suggested that approximately 8,000 samples are optimal for imputation model training in TWAS, while 56,000 samples are needed in the GWAS summary statistics to achieve maximal power [57]. Given these recommendations, we recognize that while our study represents the most well-powered AD TWAS to date through the integration of brain eQTL meta-analyses and multiple modeling approaches, larger brain sample sizes will be critical for fully elucidating the underlying genetic mechanisms of AD risk.

Although we performed fine-mapping of our TWAS associations, our results suggest only a small portion of TWAS associations are explained by fine-mapped causal eQTLs. While we propose that models not amenable to fine-mapping are likely driven by more polygenic gene expression models, it is also possible that these are explained by other effects unmeasured in our analysis, like long-range linkage disequilibrium, haplotype effects, or the tagging of structural variants or other unobserved genetic variants.

Additionally, using our stringent validation criteria, we may have missed TWAS associations that are driven by gene expression changes that do not persist into later stages of the disease captured by our differential expression association analyses. However, as these results represent the best independent evaluation of our TWAS effects, we have chosen to specifically highlight genes meeting this validation criterion.

Finally, our study was limited to individuals of European descent, as the eQTL meta-analysis summary statistics were generated predominantly using European-descent samples. Ongoing studies capturing greater genomic diversity will hopefully soon produce resources for similar analyses in African American and Hispanic/Latino populations.

Overall, by combining multiple large scale eQTL summary statistics with AD GWAS results, we have performed a TWAS that identified several new gene associations to AD and provided functional insights for several previously associated GWAS loci.

## REFERENCES

- [1] Karlsson IK, Escott-Price V, Gatz M, Hardy J, Pedersen NL, Shoai M, et al. Measuring heritable contributions to Alzheimer's disease: polygenic risk score analysis with twins. *Brain Commun* 2022;4:fcab308. <https://doi.org/10.1093/braincomms/fcab308>.
- [2] Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* 2006;63:168–74. <https://doi.org/10.1001/archpsyc.63.2.168>.
- [3] Ridge PG, Hoyt KB, Boehme K, Mukherjee S, Crane PK, Haines JL, et al. Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiol Aging* 2016;41:200.e13-200.e20. <https://doi.org/10.1016/j.neurobiolaging.2016.02.024>.
- [4] Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for

Alzheimer's disease. *Nat Genet* 2021;53:1276–82. <https://doi.org/10.1038/s41588-021-00921-z>.

[5] Bellenguez C, Küçükali F, Jansen IE, Kleindan L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* 2022;54:412–36. <https://doi.org/10.1038/s41588-022-01024-z>.

[6] Liu JZ, Erlich Y, Pickrell JK. Case-control association mapping by proxy using family history of disease. *Nat Genet* 2017;49:325–31. <https://doi.org/10.1038/ng.3766>.

[7] Escott-Price V, Hardy J. Genome-wide association studies for Alzheimer's disease: bigger is not always better. *Brain Commun* 2022;4:fcac125. <https://doi.org/10.1093/braincomms/fcac125>.

[8] Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet* 2015;47:856–60. <https://doi.org/10.1038/ng.3314>.

[9] Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. *Am J Hum Genet* 2018;102:717–30. <https://doi.org/10.1016/j.ajhg.2018.04.002>.

[10] King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet* 2019;15:e1008489. <https://doi.org/10.1371/journal.pgen.1008489>.

- [11] de Klein N, Tsai EA, Vochteloo M, Baird D, Huang Y, Chen C-Y, et al. Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. *Nat Genet* 2023;55:377–88. <https://doi.org/10.1038/s41588-023-01300-6>.
- [12] Vösa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, 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;53:1300–10. <https://doi.org/10.1038/s41588-021-00913-z>.
- [13] Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci* 2013;368:20120362. <https://doi.org/10.1098/rstb.2012.0362>.
- [14] Wheeler HE, Shah KP, Brenner J, Garcia T, Aquino-Michaels K, Consortium Gte, et al. Survey of the Heritability and Sparse Architecture of Gene Expression Traits across Human Tissues. *PLOS Genetics* 2016;12:e1006423. <https://doi.org/10.1371/journal.pgen.1006423>.
- [15] Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* 2015;47:1091–8. <https://doi.org/10.1038/ng.3367>.
- [16] Melton HJ, Zhang Z, Wu C. SUMMIT-FA: a new resource for improved transcriptome imputation using functional annotations. *Hum Mol Genet* 2023:ddad205. <https://doi.org/10.1093/hmg/ddad205>.

- [17] Dai Q, Zhou G, Zhao H, Vösa U, Franke L, Battle A, et al. OTTERS: a powerful TWAS framework leveraging summary-level reference data. *Nat Commun* 2023;14:1271. <https://doi.org/10.1038/s41467-023-36862-w>.
- [18] Zhang Z, Bae YE, Bradley JR, Wu L, Wu C. SUMMIT: An integrative approach for better transcriptomic data imputation improves causal gene identification. *Nat Commun* 2022;13:6336. <https://doi.org/10.1038/s41467-022-34016-y>.
- [19] Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. *Genet Epidemiol* 2017;41:469–80. <https://doi.org/10.1002/gepi.22050>.
- [20] Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun* 2019;10:5086. <https://doi.org/10.1038/s41467-019-12653-0>.
- [21] Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *The American Journal of Human Genetics* 2015;97:576–92. <https://doi.org/10.1016/j.ajhg.2015.09.001>.
- [22] Privé F, Arbel J, Vilhjálmsson BJ. LDpred2: better, faster, stronger. *Bioinformatics* 2021;36:5424–31. <https://doi.org/10.1093/bioinformatics/btaa1029>.
- [23] Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019;10:1776. <https://doi.org/10.1038/s41467-019-09718-5>.

[24] Zhou G, Zhao H. A fast and robust Bayesian nonparametric method for prediction of complex traits using summary statistics. *PLOS Genetics* 2021;17:e1009697.

<https://doi.org/10.1371/journal.pgen.1009697>.

[25] International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009;460:748–52.

<https://doi.org/10.1038/nature08185>.

[26] Luningham JM, Chen J, Tang S, De Jager PL, Bennett DA, Buchman AS, et al. Bayesian Genome-wide TWAS Method to Leverage both cis- and trans-eQTL Information through Summary Statistics. *Am J Hum Genet* 2020;107:714–26.

<https://doi.org/10.1016/j.ajhg.2020.08.022>.

[27] Liu W, Li M, Zhang W, Zhou G, Wu X, Wang J, et al. Leveraging functional annotation to identify genes associated with complex diseases. *PLoS Comput Biol* 2020;16:e1008315. <https://doi.org/10.1371/journal.pcbi.1008315>.

[28] Tang S, Buchman AS, De Jager PL, Bennett DA, Epstein MP, Yang J. Novel Variance-Component TWAS method for studying complex human diseases with applications to Alzheimer's dementia. *PLoS Genet* 2021;17:e1009482.

<https://doi.org/10.1371/journal.pgen.1009482>.

[29] Sun Y, Zhu J, Zhou D, Canchi S, Wu C, Cox NJ, 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;13:141.

<https://doi.org/10.1186/s13073-021-00959-y>.

- [30] Hu Y, Li M, Lu Q, Weng H, Wang J, Zekavat SM, et al. A statistical framework for cross-tissue transcriptome-wide association analysis. *Nat Genet* 2019;51:568–76. <https://doi.org/10.1038/s41588-019-0345-7>.
- [31] Bae YE, Wu L, Wu C. InTACT: An adaptive and powerful framework for joint-tissue transcriptome-wide association studies. *Genet Epidemiol* 2021;45:848–59. <https://doi.org/10.1002/gepi.22425>.
- [32] Wang YH, Luo PP, Geng AY, Li X, Liu T-H, He YJ, et al. Identification of highly reliable risk genes for Alzheimer’s disease through joint-tissue integrative analysis. *Frontiers in Aging Neuroscience* 2023;15.
- [33] Zhou D, Jiang Y, Zhong X, Cox NJ, Liu C, Gamazon ER. A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. *Nat Genet* 2020;52:1239–46. <https://doi.org/10.1038/s41588-020-0706-2>.
- [34] Chen H-H, Petty LE, Sha J, Zhao Y, Kuzma A, Valladares O, et al. Genetically regulated expression in late-onset Alzheimer’s disease implicates risk genes within known and novel loci. *Transl Psychiatry* 2021;11:618. <https://doi.org/10.1038/s41398-021-01677-0>.
- [35] Gockley J, Montgomery KS, Poehlman WL, Wiley JC, Liu Y, Gerasimov E, et al. Multi-tissue neocortical transcriptome-wide association study implicates 8 genes across 6 genomic loci in Alzheimer’s disease. *Genome Med* 2021;13:76. <https://doi.org/10.1186/s13073-021-00890-2>.



[36] THE GTEX CONSORTIUM. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 2020;369:1318–30.

<https://doi.org/10.1126/science.aaz1776>.

[37] Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A $\beta$ , tau, immunity and lipid processing. *Nat Genet* 2019;51:414–30.

<https://doi.org/10.1038/s41588-019-0358-2>.

[38] Byrska-Bishop M, Evani US, Zhao X, Basile AO, Abel HJ, Regier AA, et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* 2022;185:3426-3440.e19.

<https://doi.org/10.1016/j.cell.2022.08.004>.

[39] Liu Y, Chen S, Li Z, Morrison AC, Boerwinkle E, Lin X. ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. *Am J Hum Genet* 2019;104:410–21. <https://doi.org/10.1016/j.ajhg.2019.01.002>.

[40] Griswold AJ, Sivasankaran SK, Van Booven D, Gardner OK, Rajabli F, Whitehead PL, et al. Immune and Inflammatory Pathways Implicated by Whole Blood Transcriptomic Analysis in a Diverse Ancestry Alzheimer's Disease Cohort. *J Alzheimers Dis* 2020;76:1047–60. <https://doi.org/10.3233/JAD-190855>.

[41] Petyuk VA, Chang R, Ramirez-Restrepo M, Beckmann ND, Henrion MYR, Piehowski PD, et al. The human brainome: network analysis identifies HSPA2 as a novel Alzheimer's disease target. *Brain* 2018;141:2721–39.

<https://doi.org/10.1093/brain/awy215>.

- [42] Abraham G, Qiu Y, Inouye M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* 2017;33:2776–8. <https://doi.org/10.1093/bioinformatics/btx299>.
- [43] Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the “Sum of Single Effects” model. *PLoS Genet* 2022;18:e1010299. <https://doi.org/10.1371/journal.pgen.1010299>.
- [44] Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J R Stat Soc Series B Stat Methodol* 2020;82:1273–300. <https://doi.org/10.1111/rssb.12388>.
- [45] Kuksa PP, Leung YY, Gangadharan P, Katanic Z, Kleidermacher L, Amlie-Wolf A, et al. FILER: a framework for harmonizing and querying large-scale functional genomics knowledge. *NAR Genom Bioinform* 2022;4:lqab123. <https://doi.org/10.1093/nargab/lqab123>.
- [46] Sebollela A, Freitas-Correa L, Oliveira FF, Paula-Lima AC, Saraiva LM, Martins SM, et al. Amyloid- $\beta$  oligomers induce differential gene expression in adult human brain slices. *J Biol Chem* 2012;287:7436–45. <https://doi.org/10.1074/jbc.M111.298471>.
- [47] Yang L, Jiang Y, Shi L, Zhong D, Li Y, Li J, et al. AMPK: Potential Therapeutic Target for Alzheimer’s Disease. *Curr Protein Pept Sci* 2020;21:66–77. <https://doi.org/10.2174/1389203720666190819142746>.
- [48] Azimi M, Gharakhanlou R, Naghdi N, Khodadadi D, Heysieattalab S. Moderate treadmill exercise ameliorates amyloid- $\beta$ -induced learning and memory impairment,

possibly via increasing AMPK activity and up-regulation of the PGC-1 $\alpha$ /FNDC5/BDNF pathway. *Peptides* 2018;102:78–88. <https://doi.org/10.1016/j.peptides.2017.12.027>.

[49] Jamieson SE, White JK, Howson JMM, Pask R, Smith AN, Brayne C, et al. Candidate gene association study of solute carrier family 11a members 1 (SLC11A1) and 2 (SLC11A2) genes in Alzheimer's disease. *Neurosci Lett* 2005;374:124–8. <https://doi.org/10.1016/j.neulet.2004.10.038>.

[50] Spiers JG, Cortina Chen H-J, Barry TL, Bourgoignon J-M, Steinert JR. Redox stress and metal dys-homeostasis appear as hallmarks of early prion disease pathogenesis in mice. *Free Radic Biol Med* 2022;192:182–90. <https://doi.org/10.1016/j.freeradbiomed.2022.09.025>.

[51] Huang K-L, Marcora E, Pimenova AA, Di Narzo AF, Kapoor M, Jin SC, et al. A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nat Neurosci* 2017;20:1052–61. <https://doi.org/10.1038/nn.4587>.

[52] Wang Y, Duan X, Zhou X, Wang R, Zhang X, Cao Z, et al. ANXA11 mutations are associated with amyotrophic lateral sclerosis–frontotemporal dementia. *Front Neurol* 2022;13:886887. <https://doi.org/10.3389/fneur.2022.886887>.

[53] Wu Y, Zhang S, Xu Q, Zou H, Zhou W, Cai F, et al. Regulation of global gene expression and cell proliferation by APP. *Sci Rep* 2016;6:22460. <https://doi.org/10.1038/srep22460>.

[54] Mirzaei M, Pushpitha K, Deng L, Chitranshi N, Gupta V, Rajput R, et al. Upregulation of Proteolytic Pathways and Altered Protein Biosynthesis Underlie Retinal

Pathology in a Mouse Model of Alzheimer's Disease. *Mol Neurobiol* 2019;56:6017–34.

<https://doi.org/10.1007/s12035-019-1479-4>.

[55] Chocron ES, Munkácsy E, Kim HS, Karpowicz P, Jiang N, Van Skike CE, et al.

Genetic and pharmacologic proteasome augmentation ameliorates Alzheimer's-like pathology in mouse and fly APP overexpression models. *Sci Adv* n.d.;8:eabk2252.

<https://doi.org/10.1126/sciadv.abk2252>.

[56] Liu X, Li YI, Pritchard JK. *Trans* Effects on Gene Expression Can Drive

Omnigenic Inheritance. *Cell* 2019;177:1022-1034.e6.

<https://doi.org/10.1016/j.cell.2019.04.014>.

[57] He R, Xue H, Pan W. Statistical power of transcriptome-wide association studies.

*Genet Epidemiol* 2022;46:572–88. <https://doi.org/10.1002/gepi.22491>.

## ACKNOWLEDGEMENTS

We would like all the authors who contributed to the summary statistics from the eQTL datasets (MetaBrain, eQTLGen) and AD/ADRD GWAS as these resources were invaluable to our work. We are also grateful to the authors of the OTTERS pipeline for their clear published methodology. The conditional analyses utilized data from the Alzheimer's Disease Genetics Consortium (ADGC) (<https://www.adgenetics.org/>). The researchers associated with the ADGC (**Collaborators Appendix**) contributed to the planning and execution of the ADGC, and/or provided data, but were not involved in the

analysis or writing of this article. The full acknowledgement statement for the Alzheimer's Disease Genetics Consortium can be found here:

<https://www.adgenetics.org/content/acknowledgements>. We are grateful to the contributors who gathered the samples utilized in this study, as well as the patients and their families, whose assistance and participation enabled this work to be carried out.

### **CONFLICT OF INTEREST STATEMENT**

All authors declare that the research was conducted in the absence of any financial or commercial relationships that could be construed as a potential conflict of interest.

### **SOURCES OF FUNDING**

This work was funded by the National Institute on Aging (NIA) (Grant: RF1AG061351; PIs: Adam C. Naj, PhD, Jennifer E. Below, PhD, and William S. Bush, PhD & Grant: RF1AG070935; PIs: Anthony J. Griswold, PhD and William S. Bush, PhD). Our funding sources had no role in the preparation or submission of this manuscript.

### **CONSENT STATEMENT**

All study procedures were approved by the institutional review boards at each corresponding study center and written informed consent was obtained either from the study participant or legal guardian.

**Keywords:** TWAS, eQTL, AD, GWAS, gene expression

### **Collaborators Appendix**

The collaborators associated with the Alzheimer's Disease Genetics Consortium (ADGC) include the following individuals: Erin Abner, PhD; Perrie M. Adams, PhD;

Alyssa Aguirre, LCSW; Marilyn S. Albert, PhD; Roger L. Albin, MD; Mariet Allen, PhD; Lisa Alvarez, Howard Andrews, PhD; Liana G. Apostolova, MD; Steven E. Arnold, MD; Sanjay Asthana, MD; Craig S. Atwood, PhD; Gayle Ayres, DO; Robert C. Barber, PhD; Lisa L. Barnes, PhD; Sandra Barral, PhD; Jackie Bartlett, PhD; Thomas G. Beach, MD PhD; James T. Becker, PhD; Gary W. Beecham, PhD; Penelope Benchek, PhD; David A. Bennett, MD; John Bertelson, MD; Sarah A. Biber, PhD; Thomas D. Bird, MD; Deborah Blacker, MD; Bradley F. Boeve, MD; James D. Bowen, MD; Adam Boxer, MD, PhD; James B. Brewer, MD; James R. Burke, MD PhD; Jeffrey M. Burns, MD MS; William S. Bush, PhD; Joseph D. Buxbaum, PhD; Goldie Byrd, PhD; Laura B. Cantwell, MPH; Chuanhai Cao, PhD; Cynthia M. Carlsson, MD; Minerva M. Carrasquillo, PhD; Kwun C. Chan, PhD; Scott Chasse, PhD; Yen-Chi Chen, PhD; Marie-Francoise Chesselet, PhD; Nathaniel A. Chin, MD; Helena C. Chui, MD; Jaeyoon Chung, PhD; Suzanne Craft, PhD; Paul K. Crane, MD MPH; Marissa Cranney, BS; Carlos Cruchaga, PhD; Michael L. Cuccaro, PhD; Jessica Culhane, PhD; C. Munro Cullum, PhD; Eveleen Darby, MA MS; Barbara Davis, MA; Philip L. De Jager, MD PhD; Charles DeCarli, MD; John C. DeToledo, MD; Dennis W. Dickson, MD; Nic Dobbins, PhD; Ranjan Duara, MD; Nilufer Ertekin-Taner, MD PhD; Denis A. Evans, MD; Kelley M. Faber, MS; Thomas J. Fairchild, PhD; Daniele Fallin, PhD; Kenneth B. Fallon, MD; David W. Fardo, PhD; Martin R. Farlow, MD; John Farrell, PhD; Lindsay A. Farrer, PhD; Victoria Fernandez-Hernandez, Tatiana M. Foroud, PhD; Matthew P. Frosch, MD PhD; Douglas R. Galasko, MD; Adriana Gamboa, BS; Kathryn M. Gauthreaux, PhD; Tamar Gefen, PhD; Daniel H. Geschwind, MD PhD; Bernardino Ghetti, MD; John R. Gilbert, PhD; Alison M. Goate, D.Phil; Thomas Grabowski, MD; Neill R. Graff-Radford, MD; Anthony R. Griswold, PhD;

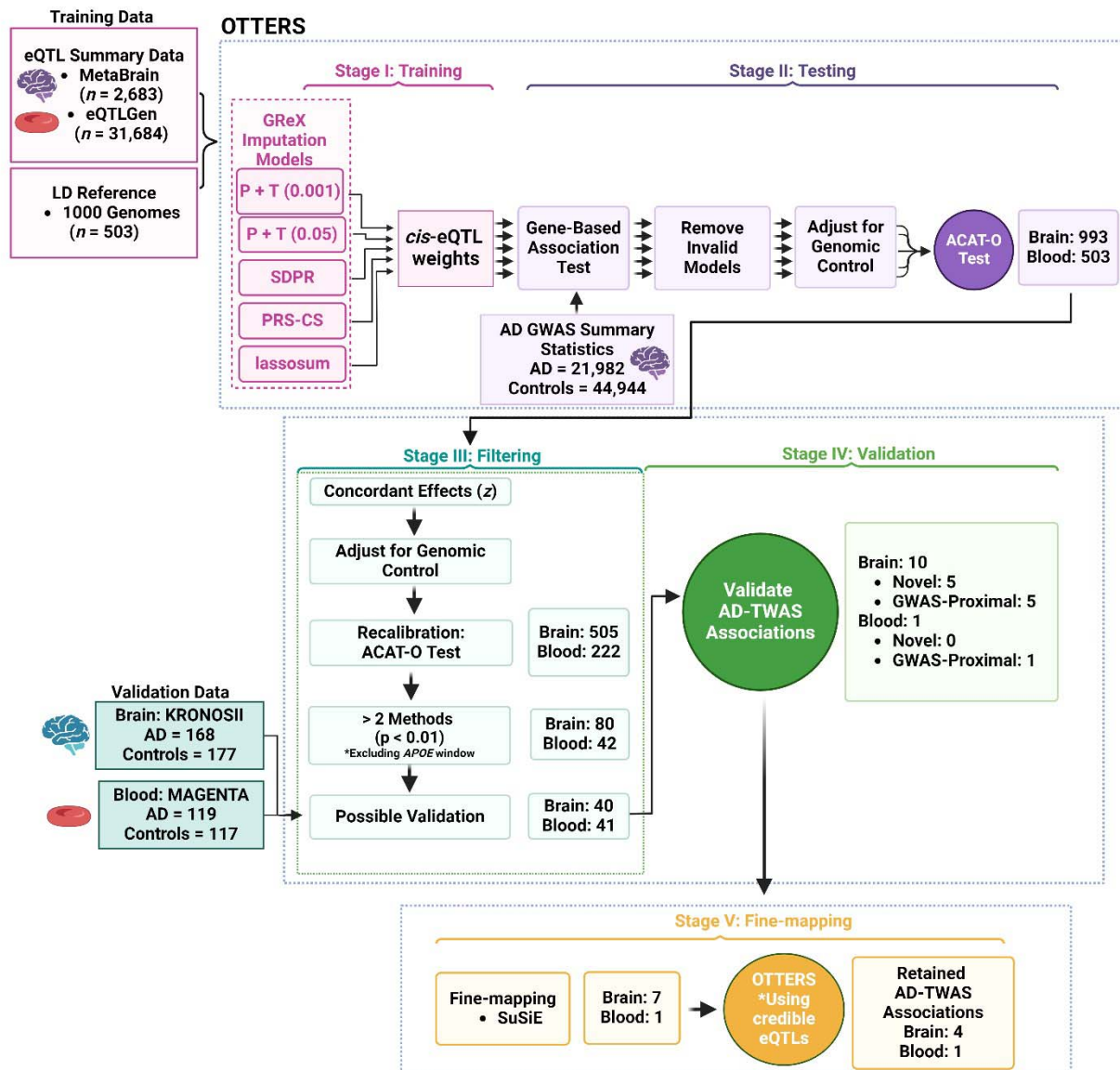
Jonathan L. Haines, PhD; Hakon Hakonarson, MD PhD; Kathleen Hall, PhD; James R. Hall, PhD; Ronald L. Hamilton, MD; Kara L. Hamilton-Nelson, MPH; Xudong Han, PhD; Oscar Harari, PhD; John Hardy, PhD; Lindy E. Harrell, MD PhD; Elizabeth Head, PhD; Victor Henderson, MD MS; Michelle Hernandez, BS; Lawrence S. Honig, MD PhD; Ryan M. Huebinger, PhD; Matthew J. Huentelman, PhD; Christine M. Hulette, MD; Bradley T. Hyman, MD PhD; Linda Hynan, PhD; Laura Ibanez, BS; Gail P. Jarvik, MD PhD; Suman Jayadev, MD; Lee-Way Jin, MD PhD; Kimberly Johnson, MSW PhD; Leigh Johnson, PhD; Bruce Jones, PhD; Gyungah Jun, PhD; M. Ilyas Kamboh, PhD; Moon Il Kang, PhD; Anna Karydas, BA; Mindy J. Katz, MPH; John S.K. Kauwe, PhD; Jeffrey A. Kaye, MD; C. Dirk Keene, MD PhD; Benjamin Keller, PhD; Aisha Khaleeq, MD; Ronald Kim, MD; Janice Knebl, DO; Neil W. Kowall, MD; Joel H. Kramer, PsyD; Walter A. Kukull, PhD; Brian W. Kunkle, PHD MPH; Amanda P. Kuzma, MS; Frank M. LaFerla, PhD; James J. Lah, MD PhD; Eric B. Larson, MD MPH; Melissa Lerch PhD; Alan J. Lerner MD; Yuk Ye Leung, PhD; James B. Leverenz, MD; Allan I. Levey, MD PhD; Andrew P. Lieberman, MD PhD; Richard B. Lipton, MD; Oscar L. Lopez, MD; Kathryn L. Lunetta, PhD; Constantine G. Lyketsos, MD MHS; Douglas Mains, DrPH; Jennifer Manly, PhD; Logue Mark, PhD; David Marquez, PhD; Daniel C. Marson, JD PhD; Eden R. Martin, PhD; Eliezer Masliah, MD; Paul Massman, PhD; Arjun V. Masurkar, MD PhD; Richard Mayeux, MD; Wayne C. McCormick, MD MPH; Susan M. McCurry, PhD; Stefan McDonough, PhD; Ann C. McKee, MD; Marsel Mesulam, MD; Jesse Mez, PhD; Bruce L. Miller, MD; Carol A. Miller, MD; Charles Mock, PhD; Abhay Moghekar, MD; Thomas J. Montine, MD PhD; Edwin Monuki, Sean D. Mooney, PhD; John C. Morris, MD; Shubhabrata Mukherjee, PhD; Amanda J. Myers, PhD; Adam C. Naj, PhD; Trung

Nguyen, PhD; James Noble, PhD; Kelley Nudelman, PhD; Sid E. O'Bryant, PhD; Kyle Ormsby, PhD; Marcia Ory, PhD MPH; Raymond Palmer, PhD; Joseph E. Parisi, MD; Henry L. Paulson, MD PhD; Valory Pavlik, PhD; David Paydarfar, MD; Victoria Perez, BS; Margaret A. Pericak-Vance, PhD; Ronald C. Petersen, MD PhD; Marsha Polk, BS; Liming Qu, MS; Mary Quiceno, MD; Joseph F. Quinn, MD; Ashok Raj, MD; Farid Rajabli, PhD; Vijay Ramanan, PhD; Eric M. Reiman, MD; Joan S. Reisch, PhD; Christiane Reitz, MD PhD; John M. Ringman, MD; Erik D. Roberson, MD PhD; Monica Rodriguear, MA; Ekaterina Rogaeva, PhD; Howard J. Rosen, MD; Roger N. Rosenberg, MD; Donald R. Royall, MD; Mary Sano, PhD; Andrew J. Saykin, PsyD; Gerard D. Schellenberg, PhD; Julie A. Schneider, MD; Lon S. Schneider, MD; William W. Seeley, MD; Richard M. Sherva, PhD; Dean K. Shibata, PhD; Scott Small, MD; Amanda G. Smith, MD; Janet Smith, BS; Yeunjoo Song, PhD; Salvatore Spina, MD; Peter St George-Hyslop, MD FRCP; Robert A. Stern, PhD; Alan Stevens, PhD; Stephen Strittmatter, MD PhD; David Sultzer, BS; Russell H. Swerdlow, MD; Andrew Teich, PhD; Jeffrey Tilson, PhD; Giuseppe Tosto, MD; John Q. Trojanowski, MD PhD; Juan C. Troncoso, MD; Debby W. Tsuang, MD; Otto Valladares, MS; Vivianna M. Van Deerlin, MD PhD; Christopher Van Dyck, MD; Linda J. Van Eldik, PhD; Jeffery M. Vance, MD PhD; Badri N. Vardarajan, MS; Robert Vassar, PhD; Harry V. Vinters, MD; Li-San Wang, PhD; Sandra Weintraub, PhD; Kathleen A. Welsh-Bohmer, PhD; Nick Wheeler, PhD; Ellen Wijsman, PhD; Kirk C. Wilhelmsen, MD PhD; Benjamin Williams, MD; Jennifer Williamson, MS; Henrick Wilms, MD; Thomas S. Wingo, MD; Thomas Wisniewski, MD; Randall L. Woltjer, MD PhD; Martin Woon, PhD; Steven G. Younkin, MD PhD; Lei Yu, PhD; Yi Zhao, MS; Xiongwei Zhou, PhD; Congcong Zhu, PhD.









**FIGURE 1.** OTTERS Workflow. OTTERS uses eQTL summary data and a LD reference panel to train five imputation models (P+T 0.001, P+T 0.05, lassosum, SDPR, PRS-CS) (Stage I). These models are subsequently applied to AD GWAS summary data and the p-values of valid models are combined using ACAT-O (Stage II). We filter the results based on concordant z scores and validate the AD-TWAS associations using RNA-seq data (Stage III & IV). Finally, we further explore the AD-TWAS association using fine-

mapping and re-running of the OTTERS workflow using only credible eQTLs to assess whether a sparse or polygenic architecture is needed for the TWAS association (Stage V). Created in BioRender.com. Abbreviations: LD, Linkage Disequilibrium; GReX, genetically-regulated gene expression; TWAS, Transcriptome-wide association study; GWAS, Genome-wide association study; AD, Alzheimer's Disease; eQTL, expression quantitative trait loci; OTTERS, Omnibus Transcriptome Test using Expression Reference Summary data; SuSiE, Sum of Single Effects model

OTTERS Association					OTTERS Components					Validation
Chr	Gene	Tissue	Type*	ACAT-O <sub>CON</sub> p <sup>†</sup>	lassosum: z (p)	SDPR: z (p)	PRS-CS: z (p)	P0.001: z (p)	P0.05: z (p)	T-test: t-value (p) <sup>‡</sup>
11	<i>MYBPC3</i>	Blood	GWAS-Proximal	<b>4.43e-08</b>	1.19 (0.0286)	<b>14.62 (5.1e-06)</b>	1.94 (0.0493)	<b>1.41 (1.75e-08)</b>	<b>1.41 (1.8e-08)</b>	<b>3.25 (6.54e-04)</b>
3	<i>C3orf62</i>	Brain	Novel	<b>2.25e-59</b>	<b>10.63 (4.5e-60)</b>	<b>11.5 (4.13e-46)</b>	1.12 (0.103)	0.49 (0.0986)	0.5 (0.0951)	<b>3.41 (3.7e-04)</b>
12	<i>SLC11A2</i>	Brain	Novel	<b>4.39e-14</b>	<b>-4.97 (2.19e-14)</b>	-	-	-1.21 (4.17e-05)	-	<b>-3.75 (1.06e-04)</b>
12	<i>PRKAG1</i>	Brain	Novel	<b>1.73e-07</b>	-	<b>-3.32 (3.84e-05)</b>	-0.65 (0.341)	<b>-1.61 (4.34e-08)</b>	-0.59 (0.0506)	<b>-6.12 (1.36e-09)</b>
15	<i>LYSMD4</i>	Brain	Novel	<b>5.58e-26</b>	-2.28 (4.48e-04)	<b>-8.63 (1.12e-26)</b>	-2.23 (1.11e-03)	-0.06 (0.84)	-0.06 (0.829)	<b>-4.19 (1.79e-05)</b>
19	<i>ZNF439</i>	Brain	Novel	<b>2.21e-52</b>	<b>9.94 (1.1e-52)</b>	-	2.45 (3.47e-04)	-	-	<b>3.28 (5.73e-04)</b>
1	<i>PSMA5</i>	Brain	GWAS-Proximal	<b>5.25e-08</b>	-	<b>-4.59 (1.31e-08)</b>	-2.48 (2.81e-04)	-0.97 (9.54e-04)	-0.45 (0.132)	<b>-7.19 (2.16e-12)</b>
10	<i>ANXA11</i>	Brain	GWAS-Proximal	<b>3.37e-12</b>	-	<b>5.77 (8.42e-13)</b>	2.13 (1.8e-03)	0.37 (0.214)	0.31 (0.305)	<b>3.54 (2.31e-04)</b>
11	<i>MTCH2</i>	Brain	GWAS-Proximal	<b>3.12e-07</b>	-	-	<b>-3.57 (1.79e-07)</b>	<b>-1.51 (2.71e-07)</b>	<b>-1.41 (2.98e-06)</b>	<b>-5.37 (7.48e-08)</b>
11	<i>MADD</i>	Brain	GWAS-Proximal	<b>6.64e-08</b>	-	-	-2.68 (8.89e-05)	<b>-1.65 (2.22e-08)</b>	-1.31 (1.27e-05)	<b>-6.07 (1.77e-09)</b>
17	<i>CYB561</i>	Brain	GWAS-Proximal	<b>1.99e-33</b>	-0.51 (0.429)	<b>-9.83 (3.98e-34)</b>	-2.95 (1.62e-05)	-0.05 (0.863)	-0.07 (0.808)	<b>-5.15 (2.26e-07)</b>

**TABLE 1.** Validated TWAS associations using AD GWAS summary statistics.

\* Whether an association is deemed novel or GWAS-Proximal is based on whether a 1 Mb window around the transcription start and end site overlaps with a 1 Mb window around a known AD/ADRD GWAS loci.

† ACAT-O<sub>CON</sub> p derived from p-values generated during Stage III (Brain Cortex: Bonferroni-corrected  $p < 5.39 \times 10^{-6}$ ; Blood: Bonferroni-corrected  $p < 6.93 \times 10^{-6}$ ) (see Figure 1)

‡ One-tailed t-test statistics and p-values, direction of effect determined based on direction of AD-TWAS association

Bold text indicates p-values significant after Bonferroni correction

Abbreviations: TWAS, Transcriptome-wide association study; GWAS, Genome-wide association study; AD, Alzheimer's Disease; ADRD, Alzheimer's Disease and Related Dementias; OTTERS, Omnibus Transcriptome Test using Expression Reference Summary data



OTTERS Association					OTTERS Components				
Chr	Gene	Tissue	Type*	ACAT-O p <sup>†</sup>	lassosum: z (p)	SDPR: z (p)	PRS-CS: z (p)	P0.001: z (p)	P0.05: z (p)
11	<i>MYBPC3</i>	Blood	GWAS-Proximal	<b>6.48e-44</b>	-	-	-	<b>13.90 (6.48e-44)</b>	<b>13.90 (6.48e-44)</b>
19	<i>ZNF439</i>	Brain	Novel	<b>3.79e-97</b>	<b>20.92 (3.79e-97)</b>	-	-	-	-
11	<i>MTCH2</i>	Brain	GWAS-Proximal	<b>1.81e-13</b>	<b>-5.51 (3.53e-08)</b>	<b>-7.57 (3.62e-14)</b>	<b>-6.54 (6.04e-11)</b>	<b>-5.54 (3.05e-08)</b>	<b>-6.24 (4.43e-10)</b>
11	<i>MADD</i>	Brain	GWAS-Proximal	<b>1.51e-07</b>	-0.72 (0.474)	<b>-5.54 (3.01e-08)</b>	-3.09 (1.99e-03)	-0.16 (0.875)	-0.36 (0.716)
17	<i>CYB561</i>	Brain	GWAS-Proximal	<b>8.81e-20</b>	-2.05 (0.040)	<b>-9.28 (1.76e-20)</b>	-1.90 (0.057)	-0.66 (0.511)	-0.66 (0.509)

**TABLE 2.** Validated TWAS associations using fine-mapped causal eQTLs applied to AD GWAS summary statistics.

\* Whether an association is deemed novel or GWAS-Proximal is based on whether a 1 Mb window around the transcription start and end site overlaps with a 1 Mb window around a known AD/ADRD GWAS loci.

† ACAT-O p derived from p-values generated during Stage V: Fine-mapping (Brain Cortex: Bonferroni-corrected  $p < 5.39 \times 10^{-6}$ ; Blood: Bonferroni-corrected  $p < 6.93 \times 10^{-6}$ ) (see Figure 1)

Bold text indicates p-values significant after Bonferroni correction

Abbreviations: TWAS, Transcriptome-wide association study; GWAS, Genome-wide association study; AD, Alzheimer's Disease; ADRD, Alzheimer's Disease and Related Dementias; eQTL, expression quantitative trait loci; OTTERS, Omnibus Transcriptome Test using Expression Reference Summary data

