HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2021 July 28.

Published in final edited form as:

Nat Genet. 2021 February ; 53(2): 143–146. doi:10.1038/s41588-020-00773-z.

Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis

Aliza P. Wingo[1,2,#], **Yue Liu**[3], **Ekaterina S. Gerasimov**[3], **Jake Gockley**[4], **Benjamin A. Logsdon**[4], **Duc M. Duong**[5], **Eric B. Dammer**[5], **Chloe Robins**[3], **Thomas G. Beach**[6], **Eric M. Reiman**[7], **Michael P. Epstein**[8], **Philip L. De Jager**[9], **James J. Lah**[3], **David A. Bennett**[10], **Nicholas T. Seyfried**[5], **Allan I. Levey**[3], **Thomas S. Wingo**[3,8,#]

[1]Division of Mental Health, Atlanta VA Medical Center, Decatur, GA, USA

[2]Department of Psychiatry, Emory University School of Medicine, Atlanta, GA, USA

[3]Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA

[4]Sage Bionetworks, Seattle, WA, USA

[5]Department of Biochemistry, Emory University School of Medicine, Atlanta, GA, USA

[6]Banner Sun Health Research Institute, Sun City, AZ 85351, USA

[7]Banner Alzheimer's Institute, Arizona State University and University of Arizona, Phoenix, AZ 85351, USA

[8]Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA

[9]Center for Translational and Computational Neuroimmunology, Department of Neurology and the Taub Institute for Research for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, NY, USA

[10]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, USA

[#]Corresponding authors aliza.wingo@emory.edu (A.P.W.) and thomas.wingo@emory.edu (T.S.W), Emory University School of Medicine, 615 Michael Street NE, Atlanta, GA 30322-1047.

Author contributions

APW and TSW conceptualized and designed the study. APW, DMD, EBD, TGB, EMR, PLD, JJL, DAB, NTS, AIL, and TSW acquired the data. APW, YL, ESG, JG, BAL, and TSW conducted analyses. APW, YL, ESG, JG, BAL, EBD, CR, MPE, JJL, DAB, NTS, AIL, and TSW interpreted the data. APW and TSW wrote the first draft of the manuscript. All authors critically revised and reviewed the manuscript.

Data availability

Phenotype data from ROS/MAP are available at https://www.radc.rush.edu. Discovery proteomic data are available at https://www.synapse.org/#!Synapse:syn17015098. Confirmation phenotypic and proteomic data are available at https://www.synapse.org/#!Synapse:syn9884314. Protein weights for the discovery and confirmation datasets and the pQTL summary statistics are available at https://www.synapse.org/#!Synapse:syn23191787. Transcript weights and their transcriptomic data sources are available at https://www.synapse.org/#!Synapse:syn20803583.

Competing interests

The authors declare no competing interests.

Introductory paragraph

Genome-wide association studies (GWAS) have identified many risk loci for Alzheimer's disease (AD)[1,2], but how these loci confer AD risk remains unclear. Here, we aimed to identify loci that confer AD risk through their effects on brain protein abundances to provide new insights into AD pathogenesis. To that end, we integrated AD GWAS results with human brain proteomes to perform a proteome-wide association study (PWAS) of AD, followed by Mendelian randomization and colocalization analysis. We identified 11 genes that are consistent with being causal in AD, acting via their *cis*-regulated brain protein abundances. Nine replicated in a confirmatory PWAS and eight represent novel AD risk genes not identified before by AD GWAS. Furthermore, we demonstrated that our results were independent of *APOE E4.* Together, our findings provide new insights into AD pathogenesis and promising targets for further mechanistic and therapeutic studies.

AD affects 35 million people worldwide but there is no effective disease-modifying treatment for it[3]. To support the development of new AD therapeutics, genetic studies of AD, especially GWAS, have identified many risk loci[1,2], but how these risk loci contribute to AD remains unclear. To gain insight into how these loci contribute to AD pathogenesis, we integrated AD GWAS results[1] with human brain proteomes[4] to identify genes that confer AD risk through their effects on brain protein abundance.

In the discovery phase, we performed a PWAS by integrating AD GWAS results (N=455,258)[1] with 376 human brain proteomes profiled from the dorsolateral prefrontal cortex (dPFC; Supplementary Table 1a)[4] using the FUSION pipeline[5]. Before integration, the proteomic profiles underwent quality control and effects of clinical characteristics and technical factors were regressed out before we estimated effects of genetic variants on protein abundance, referred to as protein weights. After quality control, the proteomic profiles included 8356 proteins, of which 1475 were heritable and their protein weights could be estimated for the PWAS. The PWAS identified 13 genes whose *cis*-regulated brain protein levels were associated with AD at false discovery rate (FDR) $p<0.05$ (Figure 1, Table 1, Extended Data Figure 1a, Supplementary Table 2).

A confirmatory PWAS was performed using the same AD GWAS[1] and an independent set of 152 human brain proteomes profiled from the dPFC (Supplementary Table 1b)[6]. After quality control, 8168 proteins remained and 1139 were heritable. Correlation between the protein weights in the discovery and confirmatory datasets was high (median 0.85, interquartile range 0.21; Supplementary table 3). Three of the 13 discovery PWAS-significant proteins could not be tested in the confirmatory PWAS – one protein was not profiled and two were profiled but did not have significant heritability, likely due to the smaller sample size. Ten of these 13 proteins could be tested and all 10 proteins replicated in the confirmatory PWAS (Table 1; Extended Data Figure 1b; Supplementary table 4).

Associations in the PWAS of AD may result when a variant is associated with protein expression (i.e., the variant is a protein quantitative trait locus [pQTL]) and AD simultaneously, or from a coincidental overlap between pQTLs and sites in linkage disequilibrium with AD GWAS sites. The former is interpreted as evidence supporting either

a pleiotropic or causal role for the gene (and will be referred to as consistent with being causal for simplicity) while the latter suggests a non-causal role. We investigated these possibilities using two independent but complementary approaches. First, using a Bayesian colocalization method, COLOC[7], we examined the posterior probability for a shared causal variant between a pQTL and AD for the 13 discovery AD PWAS-significant genes. We found 9 of 13 genes consistent with being causal (Table 2; Supplementary Table 5). Second, we used the summary data-based Mendelian randomization (SMR)[8] and its accompanying heterogeneity in dependent instruments (HEIDI)[8]. SMR results suggest that the *cis*-regulated protein abundance mediates the association between genetic variants and AD for all these 13 genes, but HEIDI results argue against a causal role for 4 genes due to linkage disequilibrium (Table 2; Supplementary Table 6). Thus, 9 of the 13 genes have evidence consistent with a causal role in AD by SMR/HEIDI. In sum, we found 7 genes with consistent results for causality by both COLOC and SMR/HEIDI (*CTSH, DOC2A, ICA1L, LACTB, PLEKHA1, SNX32,* and *STX4*; Table 2), and 4 genes with conflicting results for causality by these two approaches (*ACE, CARHSP1, RTFDC1, and STX6*; Table 2). Results for *EPHX2* and *PVR* argued against causality (Table 2).

Combining evidence for replication and results of causality tests, there were 5 genes (*CTSH, DOC2A, ICA1L, LACTB, and SNX32*) with evidence for both replication and causality (Table 3). There were 4 genes with evidence for replication and mixed results supporting causality (*ACE, CARHSP1, RTFDC1, and STX6;* Table 3). Thus, among the 13 discovery PWAS-significant genes, 11 were consistent with being causal in AD, and 9 of 11 replicated in the confirmatory PWAS (Table 3).

Since the *APOE E4* allele is strongly associated with AD, we investigated whether *APOE E4* influenced our PWAS findings. To that end, we regressed out the effect of *APOE E4* from the proteomes and used the regressed proteomic profiles to perform the PWAS of AD. That analysis found the 13 original PWAS-significant genes and 6 additional significant genes at FDR $p<0.05$ (*ACOT8*, *DDX58, ISLR2*, *PITPNC1, TBC1D1, and TRIM65*; Supplementary table 7). All the 13 genes had the same directions of association as those in the discovery PWAS. Moreover, results from COLOC and SMR/HEIDI tests found the same evidence of causality as the original findings except that *ACE* was now consistent with causality by both COLOC and SMR/HEIDI compared to mixed findings before (Supplementary tables 8-9). The 6 additional genes were not consistent with being causal by COLOC (Supplementary Table 8). These observations suggest that our findings are unlikely to be influenced by *APOE E4*.

To understand the specificity of the AD PWAS results, we performed PWAS for other brain-relevant and biometric traits. We expected the degree of overlap of significant genes to roughly correspond to their genetic correlations. GWAS results from individuals of European descent for clinical AD (N=63,926)[2], amyotrophic lateral sclerosis (ALS; N=80,610)[9], Parkinson's disease (PD; N=1,474,097)[10], neuroticism (N=390,278)[11], height (N=693,529)[12], body mass index (BMI; N=681,275)[12], and waist-to-hip ratio adjusting for BMI (WHRadjBMI; N=694,649)[13] were combined with the discovery proteomic profiles (N=376) to perform a PWAS of each trait. The PWAS of clinical AD identified 4 genes, ALS 7 genes, PD 17 genes, neuroticism 72 genes, height 662 genes, BMI 395 genes, and

WHRadjBMI 244 genes (Supplementary Tables 10-16). Overlap of the significant genes between the discovery AD PWAS and PWAS of other traits was 75% for clinical AD, 0% for ALS, 5.9% for PD, 2.8% for neuroticism, 1.7% for height, 1.5% for BMI, and 0.4% for WHRadjBMI (Extended Data Figure 2). The small overlap with biometric traits is not surprising given their estimates of genetic correlation with AD[1]. These results suggest the specificity of our AD PWAS findings.

Given the central dogma of molecular biology that DNA is transcribed into mRNA, which is translated into protein, we asked whether the identified 11 genes with evidence for being causal in AD at the protein level had similar evidence at the transcript level. We integrated the AD GWAS results[1] with 888 human brain transcriptomes to perform a TWAS of AD using FUSION[5]. The 888 transcriptomes were mainly from the frontal cortex donated by participants of European descent (Supplementary table 1c), and quality control was analogous to that of the proteomes to remove technical and clinical characteristics before estimating the effect of genetic variants on mRNA expression. Among the 13,650 transcripts after quality control, 6870 were heritable. The AD TWAS identified 40 genes whose genetically regulated mRNA expression levels were associated with AD at FDR $p<0.05$ (Extended Data Figure 3; Supplementary table 17). Among the 11 potentially causal genes identified at the protein-level, five genes, *ACE*, *CARHSP1*, *SNX32*, *STX4*, and *STX6,* showed at least nominal significance with similar directions of association with AD as seen at the protein-level (Table 3; Supplementary table 18a).

For the 5 genes with evidence at both the transcript and protein levels, results from SMR test for two molecular traits[14] suggested their protein abundance is mediated by mRNA expression (Supplementary table 18a,b). For the three genes with suggestive evidence for *cis*-regulated mRNA's association with AD (*CTSH*, *LACTB*, and *RTFDC1*), only *CTSH* had evidence to suggest protein expression is mediated by mRNA expression level (Supplementary table 18a,b). In sum, about half (6 of 11) of the genes with evidence consistent with being causal in AD at the protein level were also associated with AD at the transcript level.

We previously identified 31 modules of co-expressed proteins in ROS/MAP reference proteomes using Weight Gene Co-expression Network Analysis[4,15]. We found that 6 of the 11 potential AD causal proteins belonged to one of these modules while 5 did not. For these 6 proteins, each belonged to a different module, which implies that our PWAS findings are not simply the result of correlated protein expression[16].

Using human single-cell RNA-sequencing data profiled from the dPFC[17] we found cell-type specific enrichment for expression of 6 of the 11 causal genes at FDR p-value < 0.05 (adjusted for 17,775 genes). *DOC2A*, *ICA1L*, *PLEKHA1*, and *SNX32* were enriched in excitatory neurons, whereas *CARHSP1* showed enrichment in oligodendrocytes and *CTSH* in astrocytes and microglia (Extended Data Figure 4; Supplementary table 19).

Lastly, 8 of the 11 identified causal genes were not within 1Mb of AD genome-wide significant sites[1] while 3 were (*LACTB*, *RTFDC1*, and *STX4*), implying that these 8 genes

were from novel sites. The 8 genes were in regions with suggestive AD associations in GWAS (p-values of 5.3×10^{-5} to 1.9×10^{-7}), which is in line with other TWAS studies[18–20].

In conclusion, we identified 11 brain proteins that have evidence consistent with being causal in AD for future mechanistic studies to find new treatments for the disease.

Methods

Human Brain Proteomic and Genetic Data in the Discovery PWAS

We generated human brain proteomes from the dorsolateral prefrontal cortex (dPFC) of post-mortem brain samples donated by 400 participants of European descent of the Religious Orders Study and Rush Memory and Aging Project (ROS/MAP)[21]. Participants in the ROS/MAP studies gave informed consent for longitudinal assessments, agreed to an Anatomic Gift Act, and consented to repurposing their data and biospecimens for future studies. The Institutional Review Board of Rush University Medical Center approved the ROS/MAP studies.

We performed proteomic sequencing using isobaric tandem mass tag (TMT) peptide labeling and analyzed these peptides by liquid chromatography coupled to mass spectrometry. Samples were randomized by age, sex, post-mortem interval, cognitive diagnosis, and pathologies into 50 batches prior to TMT labeling to minimize batch effects. Peptides from each individual sample (N=400) and the global internal standard (GIS; N=100) were labeled using the TMT 10-plex kit (ThermoFisher) and high pH fractionation was used to increase peptide depth as previously described[22]. Two of the exact same GIS were included in each batch. We used Proteome Discoverer suite (version 2.3 ThermoFisher Scientific) and MS2 spectra searched against the canonical UniProtKB Human proteome database (February 2019) with 20,338 total sequences to assign peptide spectral matches. Peptide spectral matches (PSM) were filtered using percolator to a false discovery rate (FDR) of less than 1%, and, after spectral assignment, peptides were collated into proteins such that the combined probabilities of their constituent peptides achieved an FDR of 1%. Peptides shared among multiple proteins were assigned based on parsimony. Integration of ion quantification from MS2 or MS3 scans with a tolerance of 20 ppm at the most confident centroid setting was used to quantify reporter ions.

After quantification of the proteins, we identified proteins that were not reliably measured using the two GIS that were run in each batch. Proteins whose measurements fell outside the 95% confidence interval for any batch were removed from further analysis. Proteomic analysis identified 12,691 proteins and after we excluded proteins with missing values in more than 50% of the 400 subjects, 8356 proteins remained. To remove the effects of protein loading differences, we scaled each protein abundance with a sample-specific total protein abundance and $\log_2$ transformed the abundance. Next, we identified and removed poorly performing samples using iterative principal component analysis (PCA) to remove samples with greater than four standard deviations from the mean of either the first or second principal component. Subsequently, regression was used to estimate and remove the effects of proteomic sequencing batch, MS reporter quantification mode, sex, age at death, postmortem interval, study (ROS vs. MAP), and the final clinical diagnosis of cognitive

status from the proteomic profile. Expanded details on the proteomic sequencing and quality control are published here[4].

Genotyping was obtained from either whole genome sequencing (WGS) or genome-wide genotyping by either Illumina OmniQuad Express or Affymetrix GeneChip 6.0 platforms as described here[23]. Quality control of genotyping from either source was performed separately using Plink[24]. WGS data was preferred over array-based genotyping in cases where individuals had genotyping data from both sources. Individuals with overall genotyping missingness >5% were excluded. Variants were excluded if they had evidence of deviation from Hardy Weinberg equilibrium (p-value $< 1\times10^{-8}$), missing genotype rate >5%, minor allele frequency <1%, or are not a single nucleotide polymorphism (SNP). Next, KING[25] was used to remove individuals estimated to be closer than second degree relatives. For array-based data, we imputed genotyping to 1000 Genome Project Phase 3[26] using the Michigan Imputation Server[27] and SNPs with imputation $R^2 > 0.3$ were retained. Principal component analysis was performed to compare genetic ancestry of these individuals to CEU from 1000 Genomes Project (Extended Data Figure 5; Supplementary Table 20). All samples were kept for analyses. All of our analyses used only the 1,190,321 HapMap SNPs present in the 489 individuals of European descent from the 1000 Genomes Project, which was provided by FUSION[5] and commonly referred to as the linkage disequilibrium reference panel. After quality control, there were 376 subjects with both proteomic and genetic data for our discovery PWAS.

Human Brain Proteomic and Genetic Data in the Confirmation PWAS

The confirmation human brain proteomes were profiled from the dPFC of post-mortem brain samples from 198 participants of European descent recruited by the Banner Sun Health Research Institute (Banner). Participants in this study were recruited from the retirement communities in the greater Phoenix, Arizona, USA. All enrolled participants or their legal representatives signed an informed consent and the study was approved by the Institutional Review Board of Banner Sun Health Research Institute. Participants consented to annual standardized medical, neurological, and neuropsychological testing. Research diagnoses were made using approved research guidelines and a final clinicopathological diagnosis was made after review of all clinical, medical records, and neuropathological findings[6]. Only subjects with a final diagnosis of normal cognition or AD were included in the proteomic analysis. Proteomic profiling was performed using the same approach as described above for the discovery proteomes with two differences: only MS2 scans were obtained and MS2 spectra were searched against the UniProtKB human brain proteome database downloaded in April 2015. Due to different databases, exact Uniprot IDs were used when comparing the discovery and confirmation results. In total, there were 11,518 proteins quantified. We applied the same quality control procedure as was done in the discovery proteomic dataset to the confirmation proteomic data. Likewise, we used regression to remove the effects of proteomic sequencing batch, age, sex, post-mortem interval, and final clinical diagnosis of cognitive status from the confirmatory proteomic profiles before estimating the protein weights.

Genotyping was performed using the Affymetrix Precision Medicine Array using DNA extracted from the brain with the Qiagen GenePure kit. We applied the same approach to quality control as described for the discovery dataset, including removing individuals based on data completeness or relatedness, removing sites with evidence of deviation from Hardy Weinberg equilibrium, missingness above 5%, minor allele frequency below 1%, or are not a SNP. Genotyping was imputed to the 1000 Genome Project Phase 3[26] using the Michigan Imputation Server[27]. SNPs with imputation $R^2 > 0.3$ were retained. Finally, only sites included in the linkage disequilibrium reference panel were used in our confirmation PWAS, as recommended by the FUSION pipeline. After quality control, there were 152 subjects with both proteomic and genetic data to include in our confirmation analyses.

Brain Transcriptomic and Genetic Data in the AD TWAS

The brain transcriptomes were profiled from post-mortem brain samples donated by 783 individuals of European descent recruited by ROS/MAP, Mayo, and Mount Sinai Brain Bank studies[23,28,29]. These transcriptomes were profiled mainly from the dPFC and also from frontal cortex, temporal cortex, inferior frontal gyrus, superior temporal gyrus, and perirhinal gyrus. Details on alignment, quality control, and normalization of the RNA-sequencing data have been described previously[30]. Briefly, Picard was used to convert BAM files to FASTQ format and STAR[31] was used to align reads to the GRCh38 reference genome and compute gene counts for each sample. We removed genes with < 1 count per million in at least 50% of the samples and genes with missing gene length and percent GC content. Next, we removed outlier samples. Then, we regressed out effects of batch, sex, post-mortem interval, age at death, brain region, and final diagnosis of cognitive status from the transcriptomic profiles before estimating mRNA weights.

For subjects with transcriptomic data, their genome-wide genotyping was generated as described previously[23,28,29]. Quality control of the genotyping data was performed as described above for the discovery ROS/MAP dataset. After quality control, there were 13,650 mRNAs quantified from 783 individuals using 888 transcriptomes. Genotyping was filtered to include only sites in the linkage disequilibrium reference panel provided by FUSION before estimating mRNA weights as described below.

AD GWAS summary statistics

We used the summary association statistics from the latest GWAS of AD by Jansen et al[1], which had 455,258 Caucasian participants, most of whom were from the UK Biobank with family history of dementia.

Statistical Approach

We used FUSION[5] to estimate protein weights in the discovery and confirmation dataset, separately. For simplicity, we described here the process for the discovery dataset and followed the same steps for the confirmation dataset. As mentioned above, we subset ROS/MAP genome-wide genotyping into a linkage disequilibrium reference panel of 1,190,321 SNPs provided by FUSION to minimize the influence of linkage disequilibrium on the estimated test statistics[5]. Next, the SNP-based heritability for each gene was estimated using the discovery proteomic and genetic data. For proteins with significant

heritability (i.e. heritability p-value <0.01), we used FUSION to compute the effect of SNPs on protein abundance using multiple predictive models - top1, blup, lasso, enet, bslmm[5]. Protein weights from the most predictive model were selected. Subsequently, we used FUSION to combine the genetic effect of AD (AD GWAS Z-score) with the protein weights by calculating the linear sum of $Z_{score \times weight}$ for the independent SNPs at the locus to perform the PWAS of AD[5]. Lambda (λ_{obs})and lambda 1,000 (λ_{1000}), which is a standardized estimate of genomic inflation scaled to a study of 1,000 cases and 1,000 controls[32–34], were calculated for each PWAS. Lambda 1,000 was calculated using the following formula[32–34]:
$\lambda_{1000} = 1 + (\lambda_{obs} - 1) \times \left(1/n_{cases} + 1/n_{controls}\right) \div \left(1/1,000_{cases} + 1/1,000_{controls}\right)$. They were found to be consistent with other studies using FUSION that calculated lambda[34] (Extended Data Figure 1). The slightly higher λ_{obs} in the confirmation PWAS may reflect some difference in the heterogeneity of the datasets.

For the transcriptomic data, we calculated the transcript weights using FUSION with two modifications to accommodate individuals with transcriptomic profiles from more than one brain region. First, the flag *-scale 1* was added to handle pre-scaled expression values. Second, the family ID in the plink FAM file was used to ensure that samples from the same individual were always in the same fold within cross validation, and that no fold differed by more than 5% in size from any other fold. RNA weights were estimated using all five models and the most predictive model was used. Next, we used FUSION to combine the genetic effect of AD (AD GWAS Z-score) with the mRNA expression weights to perform the TWAS of AD.

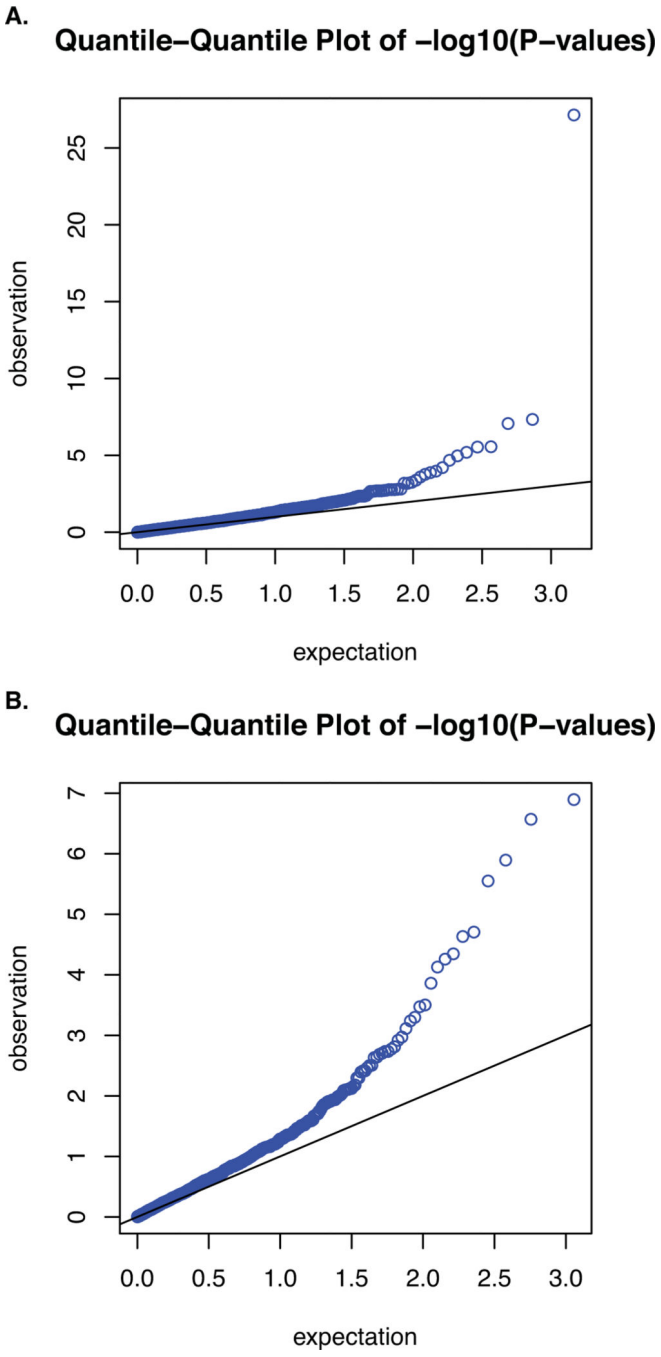
For the colocalization test we used the COLOC software[7] to estimate the posterior probability of the protein and AD sharing a causal variant, as well as the posterior probability of the protein and AD not sharing a causal variant using the marginal association statistics. For summary data-based Mendelian Randomization, SMR software[8] was used to test whether the AD PWAS-significant genes (from the FUSION) were associated with AD via their *cis*-regulated brain protein expression. We used plink[24] to estimate protein quantitative trait loci (pQTL) in the discovery proteomic dataset by linear regression. Then, we applied SMR to the pQTL results and the AD GWAS summary statistics. We used the conservative unadjusted p-value <0.05 from the heterogeneity in dependent instrument (HEIDI) to declare that presence of linkage likely influences the main SMR findings. For genes with both mRNA and protein abundance associated with AD, we applied SMR for two molecular traits[14] to the eQTL summary statistics from Siebert et al[35] and pQTL summary statistics described above to determine if the mRNA mediates the influence of SNP on proteins.

We examined the cell-type specific expression of the 11 genes with evidence for a causal role in AD at the brain protein level using human brain single-cell RNA-sequencing data profiled from the dPFC from Mathys et al[17]. First, we performed data preprocessing and transformation on the raw single-cell RNA-sequencing data using the Seurat package[36]. We removed genes with fewer than 3 counts in a cell and cells with unique feature counts over 2,500 or less than 200. The RNA counts were then normalized and scaled using the

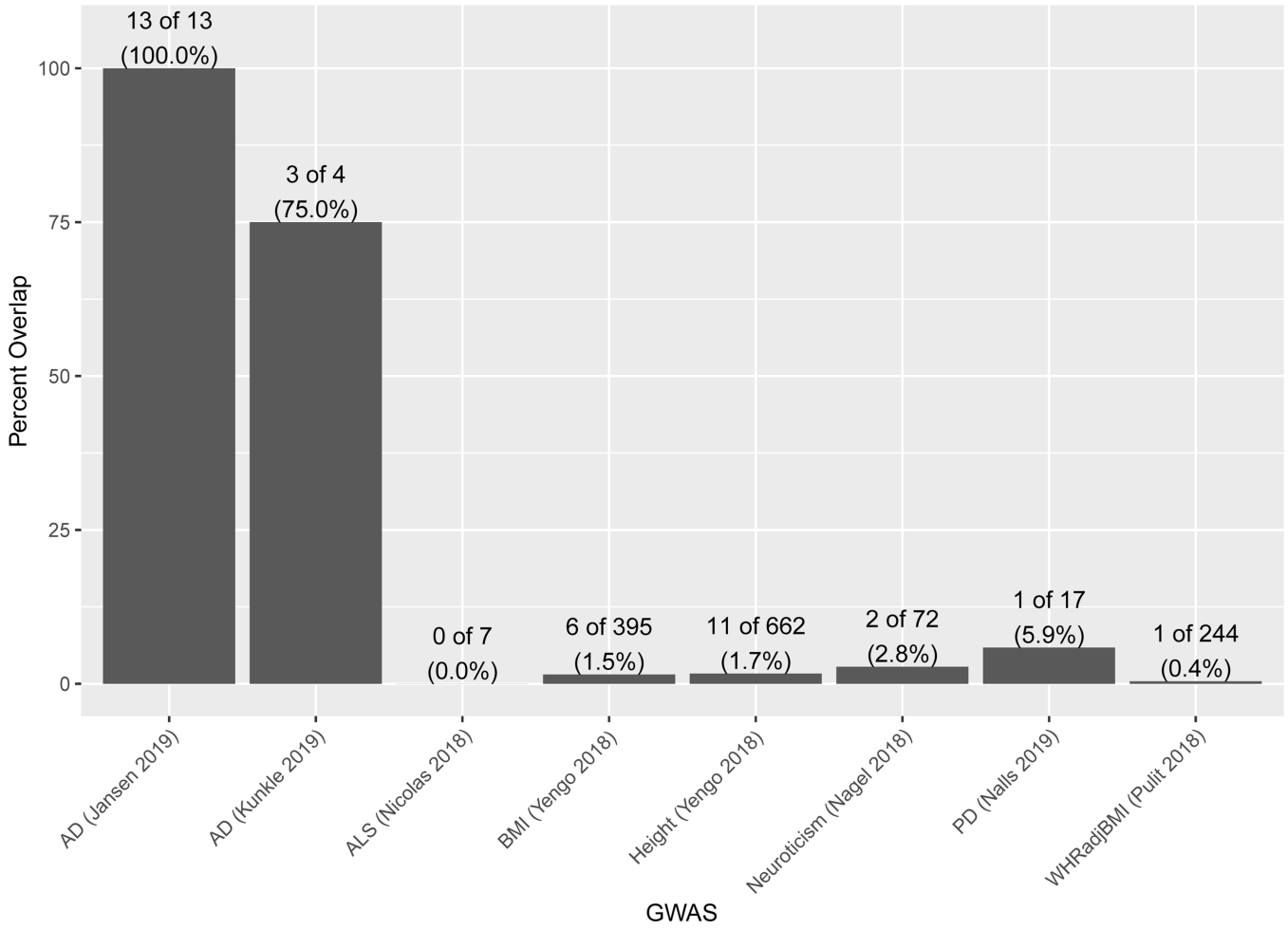
NormalizeData and ScaleData functions. The RNA-sequencing data had 17,926 genes in 70,634 cells before and 17,775 genes in 53,083 cells after quality control and normalization. We focused on the 5 main cell types - excitatory neuron, inhibitory neuron, astrocyte, microglia, and oligodendrocyte. For the 11 potentially AD causal genes, we performed differential expression analysis to compare their expression levels in one cell type versus the rest of the other cell types to determine if they are highly expressed in a particular cell type. Multiple testing correction applied to this analysis was corrected for all 17,775 genes.

To determine the novelty of the genes identified in the discovery PWAS, we asked whether each gene was within 1Mb window of the 2358 significant AD GWAS sites ($p < 5\times10^{-8}$) that correspond to the 29 independent risk loci[1].

Extended Data

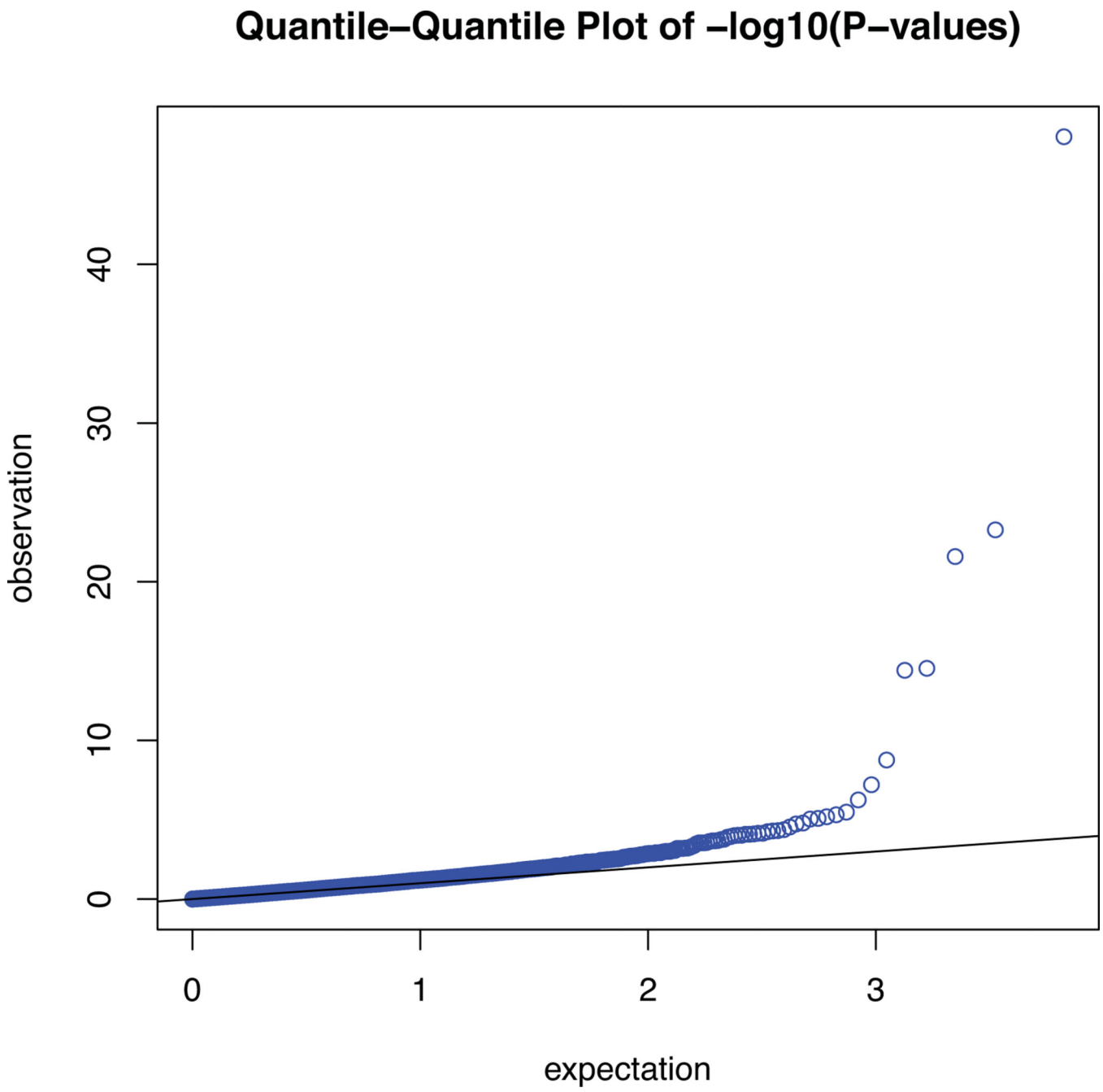


Extended Data Fig. 1. Quantile-quantile plots for the discovery and replication PWAS of AD
Quantile-quantile plot for **A)** the discovery PWAS of AD ($\lambda = 1.36$; $\lambda_{1000} = 1.003$) and **B)** confirmatory PWAS of AD ($\lambda = 1.39$; $\lambda_{1000} = 1.003$).



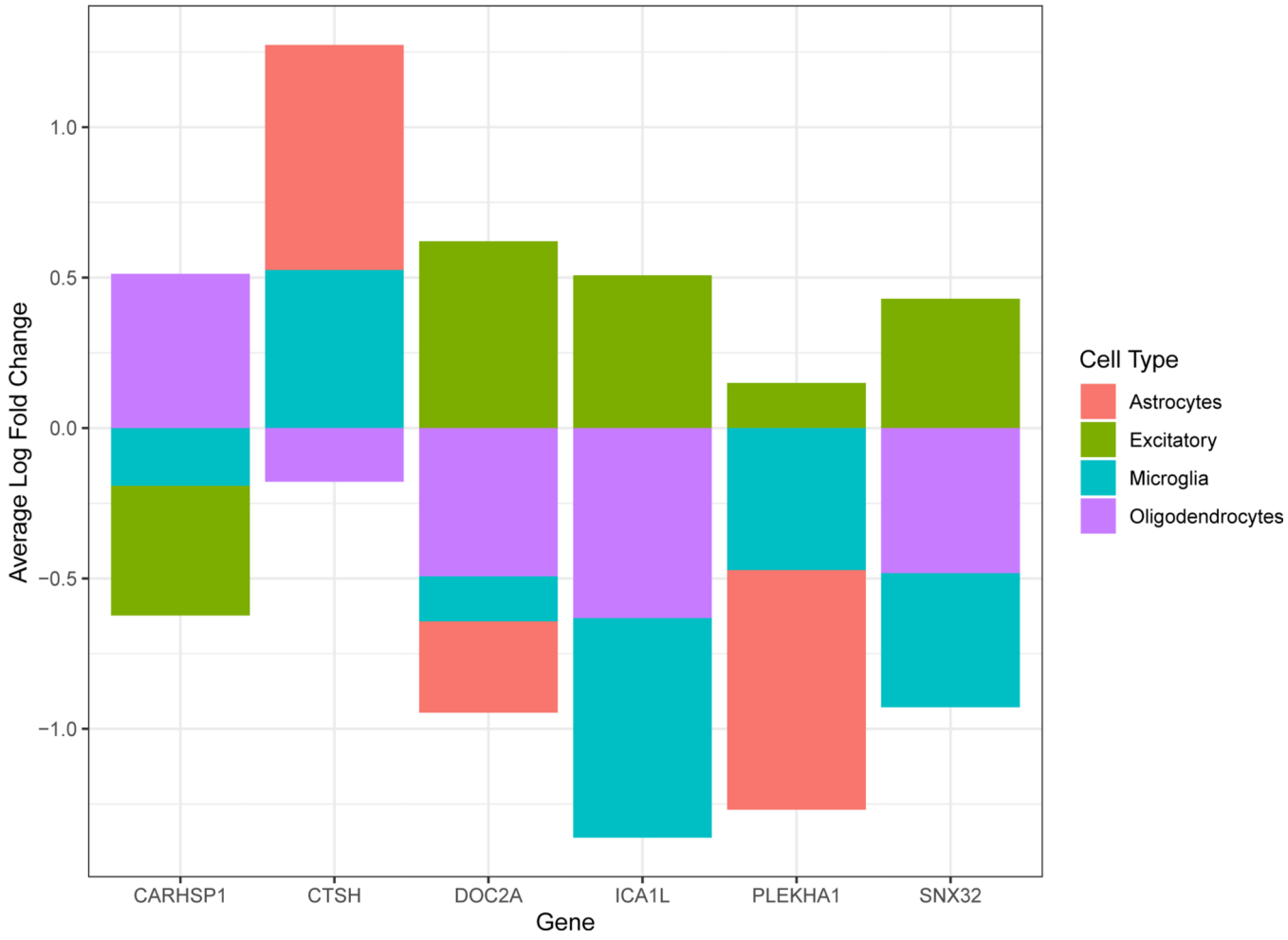
Extended Data Fig. 2. Overlap of significant genes between AD and other traits

Overlap between results of the AD PWAS and PWAS for other traits. All the PWAS used the discovery ROS/MAP proteomic dataset (n=376) and GWAS summary results from Caucasian individuals. The following outcomes were tested: clinical AD GWAS (N=63,926), amyotrophic lateral sclerosis (ALS; N=80,610), body mass index (BMI; N=681,275), height (N=693,529), neuroticism (N=390,278), Parkinson's disease (PD; N=1,474,097), and waist-to-hip ratio adjusting for BMI (WHRadjBMI; N=694,649). Significant genes considered for overlap are those with FDR $p<0.05$.



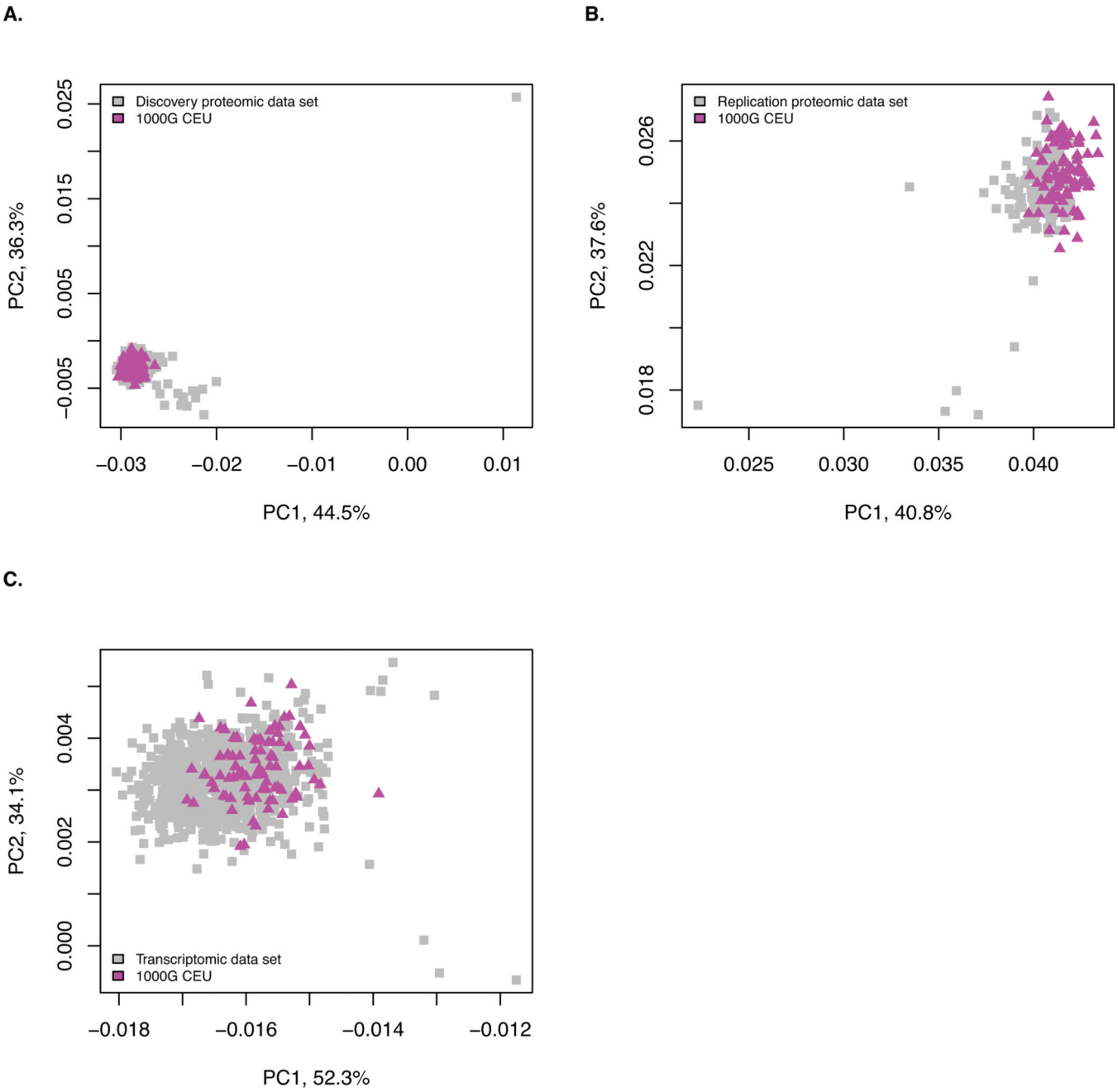
Extended Data Fig. 3. Quantile-quantile plot for the TWAS of AD

Quantile-quantile plot for the TWAS of AD ($\lambda = 1.22$; $\lambda_{1000} = 1.002$).



Extended Data Fig. 4. Single cell-type expression

Single-cell type expression for AD PWAS-significant genes with evidence of causality in AD. Using human brain single-cell RNA-sequencing data profiled from the dPFC, we found that 6 genes (of the 11 genes) had evidence of enrichment in a cell type at FDR $p < 0.05$. Enrichment testing was performed using Wilcoxon rank sum test, as implemented by the Seurat package, and multiple testing was accounted for by FDR adjusted for 17,775 tested genes. *CARHSP1* showed enrichment in oligodendrocytes. *CTSH* showed enrichment in astrocytes and microglia. *DOC2A*, *ICA1L*, *PLEKHA1*, and *SNX32* were enriched in excitatory neurons.



Extended Data Fig. 5. Genetic principal components of genetic ancestry for each dataset

Genetic principal components of genetic ancestry for each dataset. The first two genetic principal components for individuals in each dataset are plotted (grey boxes) with individuals from the 1000G CEU dataset (purple triangles) for **A)** the discovery proteomic dataset, **B)** the replication proteomic dataset, and **C)** the transcriptomic dataset.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to the participants of the ROS, MAP, Mayo, Mount Sinai Brain Bank, and Banner Sun Health Research Institute Brain and Body Donation Program for their time and participation. The following NIH grants supported this work: P30 AG066511 (A.I.L.), P30 AG10161 (D.A.B.), P30 NS055077 (A.I.L.), P50 AG025688 (A.I.L.), R01 AG015819 (D.A.B.), R01 AG017917 (D.A.B.), R01 AG053960 (N.T.S.), R01 AG056533 (T.S.W., A.P.W.) R01 AG057911 (N.T.S.), R01 AG061800 (N.T.S.), R56 AG060757 (T.S.W.), R56 AG062256 (T.S.W.), RC2 AG036547 (D.A.B.), RF1 AG057470 (T.S.W.), U01 AG046152 (P.L.D.), U01 AG046161 (A.I.L.), U01 AG061356 (P.L.D.), U01 AG061357 (A.I.L.), and U01 MH115484 (A.P.W.). NIH grants include those that supported the Accelerating Medicine Partnership for AD, the NINDS Emory Neuroscience Core, and Goizueta Alzheimer's Disease Research Center (ADRC) at Emory University, the Rush University ADRC, and Arizona State University ADRC that made this work possible. The following Veterans Administration grants supported this work: I01 BX003853 (A.P.W.) and IK4 BX005219 (A.P.W.). The Brain and Body Donation Program has been supported by NIH, the Arizona Department of Health Services, the Arizona Biomedical Research Commission and the Michael J. Fox Foundation for Parkinson's Research. Additional support includes grants from the Alzheimer's Association (N.T.S.), Alzheimer's Research UK (N.T.S.), The Michael J. Fox Foundation for Parkinson's Research (N.T.S.), and the Weston Brain Institute Biomarkers Across Neurodegenerative Diseases Grant 11060 (N.T.S.). The views expressed in this work do not necessarily represent the views of the Veterans Administration or the United States Government.

References

1. Jansen IE, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nature genetics 51, 404–413 (2019). [PubMed: 30617256]
2. Kunkle BW, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. Nature genetics 51, 414–430 (2019). [PubMed: 30820047]
3. Ballard C, et al. Alzheimer's disease. Lancet (London, England) 377, 1019–1031 (2011).
4. Wingo AP, et al. Shared proteomic effects of cerebral atherosclerosis and Alzheimer's disease on the human brain. Nature neuroscience 23, 696–700 (2020). [PubMed: 32424284]
5. Gusev A, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nature genetics 48, 245–252 (2016). [PubMed: 26854917]
6. Beach TG, et al. Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program. Neuropathology 35, 354–389 (2015). [PubMed: 25619230]
7. Giambartolomei C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS genetics 10, e1004383 (2014).
8. Zhu Z, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nature genetics 48, 481–487 (2016). [PubMed: 27019110]
9. Nicolas A, et al. Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. Neuron 97, 1268–1283.e1266 (2018). [PubMed: 29566793]
10. Nalls MA, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. The Lancet. Neurology 18, 1091–1102 (2019). [PubMed: 31701892]
11. Nagel M, et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nature genetics 50, 920–927 (2018). [PubMed: 29942085]
12. Yengo L, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Human molecular genetics 27, 3641–3649 (2018). [PubMed: 30124842]
13. Pulit SL, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Human molecular genetics 28, 166–174 (2018).
14. Wu Y, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nat Commun 9, 918 (2018). [PubMed: 29500431]
15. Langfelder P. & Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics 9, 559 (2008). [PubMed: 19114008]

16. Wainberg M, et al. Opportunities and challenges for transcriptome-wide association studies. Nature genetics 51, 592–599 (2019). [PubMed: 30926968]
17. Mathys H, et al. Single-cell transcriptomic analysis of Alzheimer's disease. Nature 570, 332–337 (2019). [PubMed: 31042697]
18. Gusev A, et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. Nature genetics 50, 538–548 (2018). [PubMed: 29632383]
19. Huckins LM, et al. Gene expression imputation across multiple brain regions provides insights into schizophrenia risk. Nature genetics 51, 659–674 (2019). [PubMed: 30911161]
20. Raj T, et al. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. Nature genetics 50, 1584–1592 (2018). [PubMed: 30297968]
21. Bennett DA, et al. Religious Orders Study and Rush Memory and Aging Project. Journal of Alzheimer's disease : JAD 64, S161–s189 (2018). [PubMed: 29865057]
22. Mertins P, et al. Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. Nature protocols 13, 1632–1661 (2018). [PubMed: 29988108]
23. De Jager PL, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. Scientific data 5, 180142 (2018). [PubMed: 30084846]
24. Purcell S, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. American journal of human genetics 81, 559–575 (2007). [PubMed: 17701901]
25. Manichaikul A, et al. Robust relationship inference in genome-wide association studies. Bioinformatics (Oxford, England) 26, 2867–2873 (2010).
26. Abecasis GR, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65 (2012). [PubMed: 23128226]
27. Das S, et al. Next-generation genotype imputation service and methods. Nature genetics 48, 1284–1287 (2016). [PubMed: 27571263]
28. Allen M, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. Scientific data 3, 160089 (2016). [PubMed: 27727239]
29. Wang M, et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. Scientific data 5, 180185 (2018). [PubMed: 30204156]
30. Logsdon BA, et al. Meta-analysis of the human brain transcriptome identifies heterogeneity across human AD coexpression modules robust to sample collection and methodological approach. bioRxiv, 510420 (2019).
31. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England) 29, 15–21 (2013).
32. Devlin B. & Roeder K. Genomic control for association studies. Biometrics 55, 997–1004 (1999). [PubMed: 11315092]
33. Freedman ML, et al. Assessing the impact of population stratification on genetic association studies. Nature genetics 36, 388–393 (2004). [PubMed: 15052270]
34. Wu L, et al. A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. Nature genetics 50, 968–978 (2018). [PubMed: 29915430]
35. Sieberts SK, et al. Large eQTL meta-analysis reveals differing patterns between cerebral cortical and cerebellar brain regions. Scientific data 7, 340 (2020). [PubMed: 33046718]
36. Butler A, Hoffman P, Smibert P, Papalexi E. & Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature biotechnology 36, 411–420 (2018).

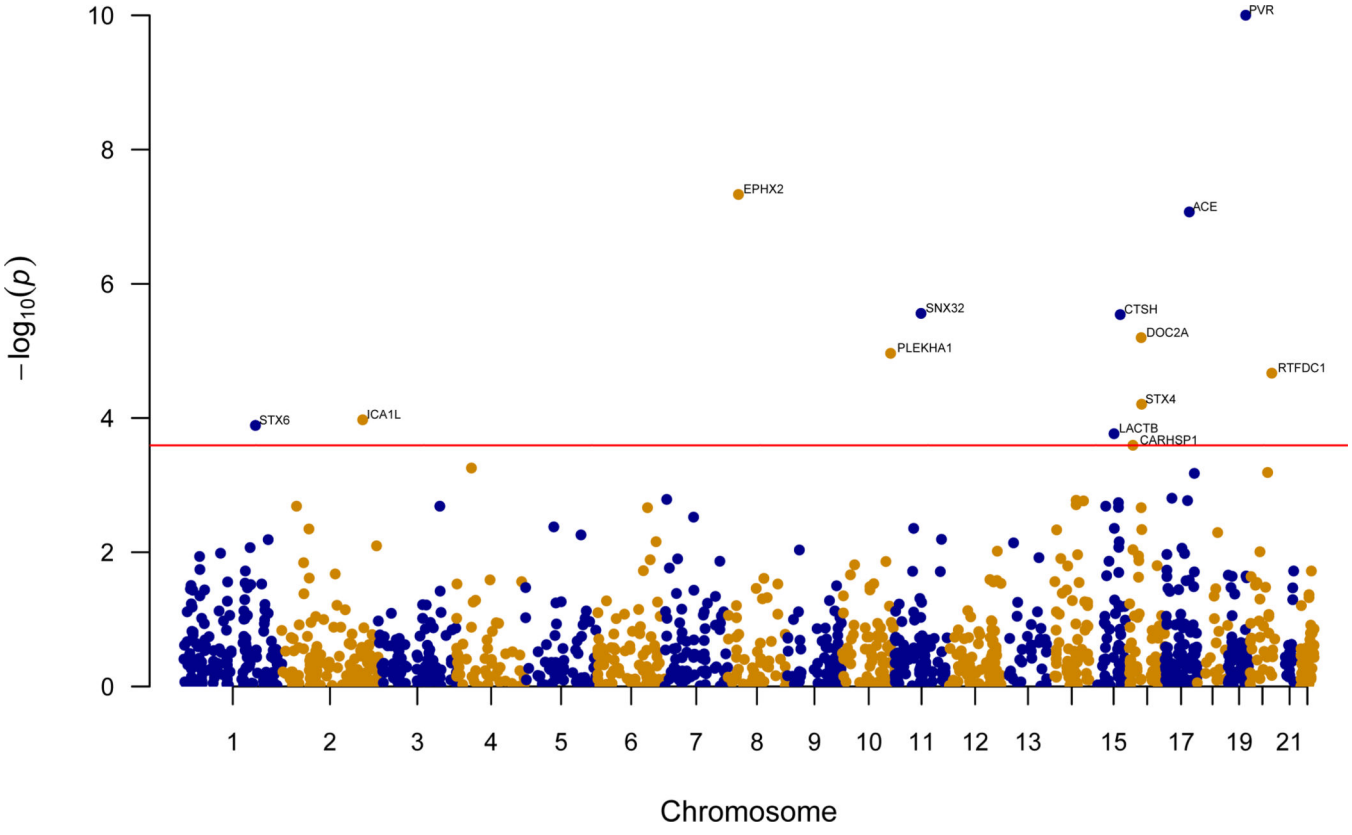


Figure 1:

Manhattan plot for the discovery AD PWAS integrating the AD GWAS (N=455,258) with the discovery ROS/MAP proteomes (N=376). Each point represents a single test of association between a gene and AD ordered by genomic position on the x axis and the association strength on the y axis as the -$\log_{10}$ p value of a z-score test. The discovery PWAS identified 13 genes whose *cis*-regulated brain protein abundances were associated with AD at FDR p< 0.05. The red horizontal line reflects the significant threshold of FDR p <0.05 and is set at the highest unadjusted p value that is below that threshold ($p = 2.6\times10^{-4}$).

Table 1:

The discovery AD PWAS identified 13 significant genes, of which, 10 were found in the confirmatory PWAS and all 10 replicated.

			Discovery PWAS			Confirmatory PWAS		Evidence for Replication
	Gene	**CHR**	**PWAS.Z**	**PWAS.*p***	**PWAS.FDR.*p***	**PWAS.Z**	**PWAS.*p***	
1	ACE	17	−5.36	8.5×10^{-8}	4.2×10^{-5}	−5.28	1.3×10^{-7}	yes
2	EPHX2	8	5.46	4.7×10^{-8}	3.4×10^{-5}	4.68	2.8×10^{-6}	yes
3	SNX32	11	−4.69	2.8×10^{-6}	8.4×10^{-4}	−4.27	2.0×10^{-5}	yes
4	DOC2A	16	−4.51	6.4×10^{-6}	1.6×10^{-3}	−4.23	2.3×10^{-5}	yes
5	LACTB	15	3.76	1.7×10^{-4}	2.1×10^{-2}	4.08	4.5×10^{-5}	yes
6	ICA1L	2	−3.88	1.1×10^{-4}	1.6×10^{-2}	−3.96	7.5×10^{-5}	yes
7	CARHSP1	16	3.66	2.6×10^{-4}	2.9×10^{-2}	3.48	5.1×10^{-4}	yes
8	RTFDC1	20	4.25	2.1×10^{-5}	3.9×10^{-3}	3.10	2.0×10^{-3}	yes
9	STX6	1	3.83	1.3×10^{-4}	1.7×10^{-2}	2.96	3.1×10^{-3}	yes
10	CTSH	15	4.68	2.9×10^{-6}	8.4×10^{-4}	2.36	1.8×10^{-2}	yes
11	PLEKHA1 *	10	4.40	1.1×10^{-5}	2.3×10^{-3}	–	–	–
12	PVR **	19	−10.94	7.1×10^{-28}	1.0×10^{-24}	–	–	–
13	STX4 **	16	4.00	6.2×10^{-5}	1.0×10^{-2}	–	–	–

This table gives the z scores for the AD PWAS associations with their corresponding p-values and FDR-adjusted p-values for all significant genes in the AD discovery PWAS. Confirmatory AD PWAS z scores and their corresponding unadjusted p-values are provided for the significant genes in the discovery AD PWAS.

(*) **Asterisk** indicates protein not profiled in the confirmatory proteomic dataset.

(**) **Double asterisks** denote proteins profiled but did not have significant SNP-based heritability estimates in the confirmatory proteomic dataset.

Table 2:

COLOC and SMR analysis of the 13 significant genes in the discovery AD PWAS. Eleven genes had evidence consistent with a causal role by either COLOC or SMR.

			COLOC		SMR		
	Gene	**Chr**	**H_4**	**Causal variant**	**SMR.p**	**HEIDI.p**	**Causal variant**
1	CTSH	15	0.962	yes	3.1×10^{-5}	0.464	yes
2	DOC2A	16	0.907	yes	1.0×10^{-3}	0.742	yes
3	ICA1L	2	0.672	yes	4.1×10^{-4}	0.977	yes
4	LACTB	15	0.754	yes	3.8×10^{-4}	0.070	yes
5	PLEKHA1 *	10	0.581	yes	3.0×10^{-3}	0.455	yes
6	SNX32	11	0.975	yes	2.7×10^{-5}	0.588	yes
7	STX4 *	16	0.918	yes	5.0×10^{-3}	0.808	yes
8	ACE	17	0.976	yes	4.0×10^{-3}	0.039	no
9	RTFDC1	20	0.643	yes	4.6×10^{-5}	0.034	no
10	CARHSP1	16	0.188	no	1.2×10^{-2}	0.397	yes
11	STX6	1	0.072	no	1.0×10^{-2}	0.748	yes
12	EPHX2	8	0	no	7.1×10^{-7}	0.008	no
13	PVR *	19	0.022	no	1.4×10^{-5}	n/a	n/a

For the 13 FDR-significant genes in the discovery AD PWAS, the result of COLOC H_4, which is the Bayesian posterior probability that a genetic variant is shared by both traits (i.e., gene and AD), and P values for SMR and SMR HEIDI tests are given.

*(*)***Asterisk** denotes genes not found in the confirmatory PWAS. **n/a** (not applicable) indicates undetermined result since the number of pQTL SNPs were too small for HEIDI to test. Genes were sorted by whether they are consistent with being a causal variant.

Table 3:

Summary of the 11 AD PWAS-significant genes with evidence for being consistent with a causal role in AD.

			Discovery	Confirmatory	Evidence for causality		TWAS	Novel
	Gene	**Chr**	**PWAS**	**PWAS**	**COLOC**	**SMR**	**significant**	**gene**
1	CTSH	15	significant	replicated	yes	yes	suggestive	yes
2	DOC2A	16	significant	replicated	yes	yes	n/a	yes
3	ICA1L	2	significant	replicated	yes	yes	no	yes
4	LACTB	15	significant	replicated	yes	yes	suggestive	no
5	SNX32	11	significant	replicated	yes	yes	yes	yes
6	ACE	17	significant	replicated	yes	no	yes	yes
7	RTFDC1	20	significant	replicated	yes	no	suggestive	no
8	CARHSP1	16	significant	replicated	no	yes	yes	yes
9	STX6	1	significant	replicated	no	yes	yes	yes
10	STX4*	16	significant	–	yes	yes	yes	no
11	PLEKHA1*	10	significant	–	yes	yes	n/a	yes

*(*)***Asterisk** denotes proteins not found in the confirmation PWAS. **n/a** refers to genes that did not have significant heritability estimates to be included in the TWAS of AD. Full results for TWAS is in Supplementary tables 17-18. "**suggestive**" in "TWAS significant" column refers to genes with 0.05 <TWAS nominal p <0.1. **Novel gene** refers to genes not within 1Mb window of SNPs with p < 5E-08 identified in Jansen et. al. AD GWAS[1].