ARTICLE

Identification of novel drug targets for Alzheimer's disease by integrating genetics and proteomes from brain and blood

Ya-Nan Ou[1,7], Yu-Xiang Yang[2,7], Yue-Ting Deng[2], Can Zhang[3], Hao Hu[1], Bang-Sheng Wu[2], Yi Liu[1], Yan-Jiang Wang[4], Ying Zhu[5], John Suckling[6], Lan Tan[1] and Jin-Tai Yu[2]✉

Genome-wide association studies (GWASs) have discovered numerous risk genes for Alzheimer's disease (AD), but how these genes confer AD risk is challenging to decipher. To efficiently transform genetic associations into drug targets for AD, we employed an integrative analytical pipeline using proteomes in the brain and blood by systematically applying proteome-wide association study (PWAS), Mendelian randomization (MR) and Bayesian colocalization. Collectively, we identified the brain protein abundance of 7 genes (*ACE*, *ICA1L*, *TOM1L2*, *SNX32*, *EPHX2*, *CTSH*, and *RTFDC1*) are causal in AD ($P < 0.05$/proteins identified for PWAS and MR; PPH4 >80% for Bayesian colocalization). The proteins encoded by these genes were mainly expressed on the surface of glutamatergic neurons and astrocytes. Of them, *ACE* with its protein abundance was also identified in significant association with AD on the blood-based studies and showed significance at the transcriptomic level. *SNX32* was also found to be associated with AD at the blood transcriptomic level. Collectively, our current study results on genetic, proteomic, and transcriptomic approaches has identified compelling genes, which may provide important leads to design future functional studies and potential drug targets for AD.

Molecular Psychiatry; https://doi.org/10.1038/s41380-021-01251-6

INTRODUCTION

Alzheimer's disease (AD) is a worldwide public health threat, currently affecting 6.2 million Americans nowadays; the number is expected to grow to 13.8 million by 2060 [1]. However, there is no effective disease-modifying treatment for AD [2]. Over the past decade, the widespread application of large-scale genome-wide association studies (GWASs) has drastically advanced the discovery of genetic variants associated with AD [3, 4]. Nevertheless, it is still challenging to decipher the underlying biological mechanisms responsible for the vast majority of these genetic effects, which has hindered translating these genetic findings to drug development of AD by targeting these candidate genes [5].

Before understanding the genetic architecture of a disease, we first need to characterize the complex profiles and associated genetic architecture of downstream proteins [6]. In addition to GWAS, proteome-wide association studies (PWASs) have been recently developed that can establish associations between proteome abundance and phenotypic variations [7], helping providing a bridge that can translate the discovery of genetic architectures and elucidate the mechanisms of downstream proteins associated with human pathophysiology [6]. Publicly accessible GWAS summary statistics provide a rich resource for integrative analyses of disease pathogeneses: e.g., Mendelian randomization (MR) and Bayesian colocalization. MR uses genetic variants to estimate the effects of an exposure on an outcome, on the presumption that alleles are randomly allocated to gametes, removing the effects of confounding bias and reverse causation [8]. Statistical colocalization explores whether two associated signals share common causal variants [9]. Recently, frameworks of integrative analysis that combine MR with colocalization have been widely deployed to identify biological mediators modulating genes and clinical outcomes [10, 11]. Moreover, protein levels and functions, in combination with gene expression and epigenetic regulation, vary greatly dependent on tissue types [12]. While investigating the brain tissues that is most closely related with cerebral dysfunction is the most direct approach, it is often difficult to sample, and other body tissues, such as blood, are more tractable for individual patient diagnosis and treatment.

Accordingly, we sought to discover promising drug targets for AD by combining high-throughput proteomics in the brain and blood with genetic data to determine the genomic architecture-associated protein levels in AD. We systematically link protein biomarkers to AD by taking a five-step approach. First, we leveraged protein quantitative trait locus (pQTL) data derived from brain tissue and findings from two large-scale AD GWASs to conduct a PWAS analysis aimed at identifying the candidate protein biomarkers (Step 1). Second, we integrated these data using a MR framework, which harnesses genetic colocalization to

[1]Department of Neurology, Qingdao Municipal Hospital, Qingdao University, Qingdao, China. [2]Department of Neurology and Institute of Neurology, Huashan Hospital, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Shanghai Medical College, Fudan University, Shanghai, China. [3]Genetics and Aging Research Unit, McCance Center for Brain Health, Mass General Institute for Neurodegenerative Diseases (MIND), Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA, USA. [4]Department of Neurology, Daping Hospital, Third Military Medical University, Chongqing, China. [5]State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Institutes of Brain Science, Fudan University, Shanghai, China. [6]Department of Psychiatry, University of Cambridge, Cambridge, UK. [7]These authors contributed equally: Ya-Nan Ou, Yu-Xiang Yang. ✉email: jintai_yu@fudan.edu.cn

Received: 29 May 2021 Revised: 25 July 2021 Accepted: 28 July 2021
Published online: 11 August 2021

highlight genes and AD that are influenced by a shared causal variant (Step 2). Third, by leveraging gene-expression data, we identified the significant genes driving GWAS signals at the transcriptional level (Step 3). Fourth, a specificity analysis was conducted to detect the cell type that targeted genes express on the highest levels (Step 4). Last, we verified the findings by applying them to the proteomic data derived from blood serum to assess the consistency between the two tissues (Step 5). Figure 1 summarizes the overall analysis pipeline applied in this study.

MATERIALS AND METHODS

Brain-derived pQTL and expression quantitative trait locus (eQTL) data

Proteomes and coupled genotyping data were derived from the dorsolateral prefrontal cortex (dPFC) of postmortem brain samples donated by 400 participants of European ancestry from the Religious Orders Study/Memory and Aging Project (ROS/MAP) [13]. Proteomic sequencing was performed using isobaric tandem mass tag peptide labeling; peptides were analyzed by liquid chromatography coupled to mass spectrometry (MS). Proteome Discoverer suite v.2.3 (Thermo Fisher Scientific) and tandem MS spectra were used to search against the canonical UniProtKB human proteome database containing 20,338 total sequences to assign peptide spectral matches. Genotyping was obtained from either whole-genome sequencing or genome-wide genotyping by either the Illumina OmniQuad Express or Affymetrix GeneChip 6.0 platforms [14]. A total of 376 subjects across 8356 proteins with both proteomic and genetic data passed the quality control for the PWAS.

Transcriptomes data were profiled from postmortem brain samples donated by 576 individuals recruited by the ROS/MAP. These transcriptomes were profiled mainly from the dPFC. Picard v.1.83 was used to convert BAM files to FASTQ format; STAR v.2.5.1b was used to align reads to the GRCh38 reference genome and compute gene counts for each sample. Quality control process included removing genes with <1 count per million in at least 50% of the samples and genes with missing gene length and percentage guanine–cytosine content and removing outlier samples. We regressed out effects of batch, sex, postmortem interval, age at death, brain region, and final diagnosis of cognitive status and then estimated the messenger RNA (mRNA) weights. Genome-wide genotyping was generated as described above.

Blood proteomes and eQTL data

Serum protein biomarkers were derived from the largest GWAS analyses for which genome-wide summary statistics were publicly available (AGES Reykjavik project) [15]. The study samples consisted of 5457 predominantly European individuals aged 66–96 years. Serum levels of 4137 human proteins, targeted by 4782 Slow Off-Rate Modified Aptamers (SOMAmers), were determined at SomaLogic Inc. (Boulder, US).

We next obtained whole-blood eQTL data from the GTEx version 8 database of 948 donors from https://gtexportal.org/home/datasets, which is used for the study of human gene expression and their associations with genetic variations [16]. Gene expression in these samples was obtained using paired-end RNA-seq (Illumina TruSeq; Illumina Inc.) and genotype data from whole-genome sequencing. Complete descriptions of the donor enrollment, consent process, bio-specimen procurement methods, sample fixation, and histo-pathological review procedures are available at the GTEx official website (https://www.gtexportal.org).

AD GWAS data

The main analysis utilized the most recent and largest AD GWAS [17]. It is composed of a GWAS from the UK Biobank (UKB) for family history of AD ($N = 53,042$ cases, 355,900 controls) and the Kunkle et al. stage 1 GWAS ($N = 21,982$ cases, 41,944 controls), which was obtained from the International Genomics of Alzheimer's Project (IGAP) GWAS Stage 1 result. The Kunkle et al. stage 1 GWAS is composed of datasets from the Alzheimer Disease Genetics Consortium, Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium, the European Alzheimer's Disease Initiative, and Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium. AD cases were all autopsy-confirmed or clinically confirmed using the published criteria (Supplementary Table 1).

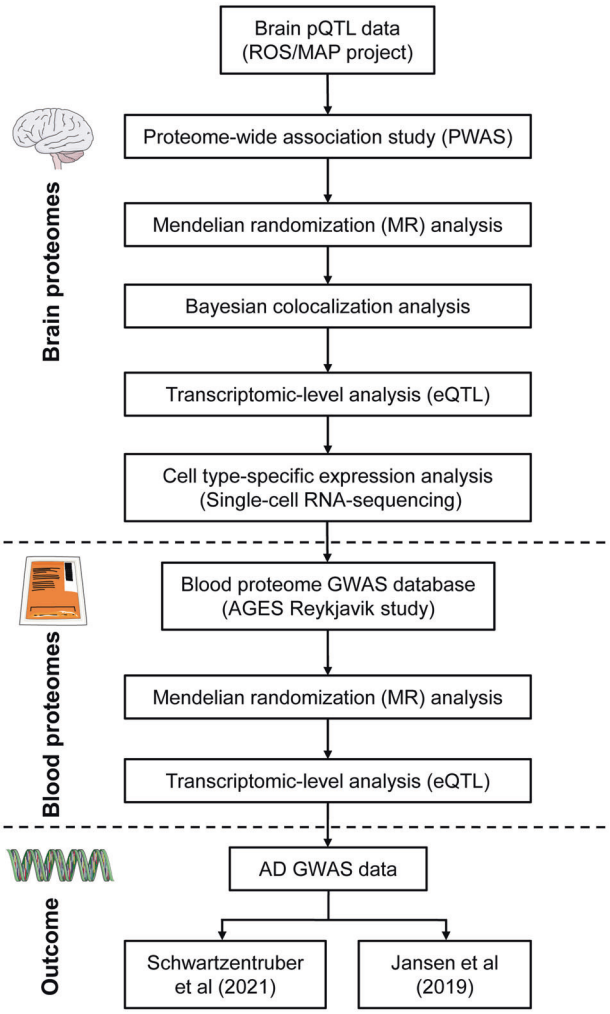

Fig. 1 The analytical pipeline for identifying genetic targets of AD using brain- and blood-based proteomics. First, we leveraged pQTL data derived from the brain tissue and findings from two large-scale AD GWASs to conduct a PWAS analysis (Step 1). Second, we integrated the above data using a MR framework, which harnesses genetic colocalization to highlight genes and AD that are influenced by a shared causal variant (Step 2). Third, by leveraging gene-expression data, we identified the significant genes driving GWAS signals on the transcriptional level (Step 3). Fourth, cell-type specificity analysis was conducted to detect the cell type that targeted genes express on the highest levels (Step 4). Last, we verified the findings by applying the proteomic data derived from blood serum to detect the consistency between the two tissues (Step 5).

We also utilized the GWAS summary statistics from another large AD GWAS [3]. This study included 71,880 AD cases and AD-by-proxy cases based on parental diagnoses (mean age of onset: 64.3 years) and 383,378 controls (mean age at last contact: 58.2 years) of European ancestry from 4 consortia, including the Alzheimer's Disease Working Group of the Psychiatric Genomics Consortium, IGAP, Alzheimer's Disease Sequencing Project and UKB (Supplementary Table 1).

Statistical analysis

Proteome-wide association study. Genome-wide genotyping was dispersed into a linkage disequilibrium (LD) reference panel of ~1,190,321 single-nucleotide polymorphisms (SNPs) to minimize the influence of LD on the estimated test statistics [18]. The SNP-based heritability for each

gene was estimated using both the proteomic and genetic data. For proteins with significant heritability ($P < 0.01$), we used FUSION to compute the effect of SNPs on protein abundance using multiple predictive models, including top1, blup, lasso, enet, and bslmm [18]. We chose the weights of protein calculated from the most predictive model. Weights used in the present analysis were derived from https://www.synapse.org/#!Synapse:syn23627957. 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 × weight for the independent SNPs at the locus to perform the PWAS of AD. A linear regression model was applied to each protein entry. We evaluated the significance level of association with AD as the outcome and normalized protein abundance as the predictor. The Benjamini–Hochberg method was used to control the false discovery rate [19], which was estimated by the p.adjust() function in R. We also computed a Bonferroni-corrected *P* value threshold.

MR analysis. MR relies on three assumptions—the genetic variants of exposure are strongly associated with the exposure of interest, the genetic variants of exposure are independent from confounders of the exposure-outcome relation, and the genetic variants are only linked to the outcome through affecting the exposure of interest [8]. SNPs included for the analysis were strongly and independently ($R^2 < 0.001$) predicted exposures at genome-wide significance (5×10^{-8}). The Wald ratio estimates the log odds change in AD risk per standard deviation change in protein biomarker relative to the risk allele for the instrumenting SNP [20]. Where more than one SNP was available, a weighted mean of the ratio estimates weighted by the inverse variance of the ratio estimates (inverse-variance weighted method) was used [21]. The regression was constrained to pass through the origin, thus leading to a zero intercept. Complementary methods, including MR-Egger, weighted median, simple mode, and weighted mode, were also applied. Estimates were filtered based on a multiple testing threshold of $P < 0.05$/number of proteins analyzed. MR estimates were derived using the "TwoSampleMR" package in R 3.6.0 (github.com/MRCIEU/TwoSampleMR).

Bayesian colocalization analysis. This method assesses whether two associated signals are consistent with shared causal variant(s). Default prior probabilities were used, including $P1 = 1 \times 10^{-4}$, $P2 = 1 \times 10^{-4}$, $P12 = 1 \times 10^{-5}$, where P1 is the probability that a given SNP is associated with AD, P2 is the probability that a given SNP is a significant QTL, and P12 is the probability that a given SNP is both an AD result and an QTL [9]. We extracted the summary association statistics for SNPs from the QTL dataset and then extracted the summary association statistics for the same SNPs from the AD GWAS. This approach has five mutually exclusive hypotheses: (1) there is no causal SNP for either trait (H0); (2) there is one causal SNP for trait 1 only (H1); (3) there is one causal SNP for trait 2 only (H2); (4) there are two distinct causal SNPs, one for each trait (H3); and (5) there is a causal SNP common to both traits (H4). Support for each of the hypotheses is quantified by the posterior probability (PP), denoted by PPH0, PPH1, PPH2, PPH3, and PPH4, respectively [9]. Evidence for colocalization was assessed using the PP for hypothesis 4 (PPH4 >80%) [22, 23]. These PPs were calculated using Bayesian "coloc.abf" function, "coloc" package in R (http://cran.r-project.org/web/packages/coloc).

Cell-type specificity analysis. We further examined the cell type-specific expression of the 7 genes by using human brain single-cell RNA sequencing (RNA-seq) data profiled from the Cell Types database (https://portal.brain-map.org/atlases-and-data/rnaseq). Individual layers of cortex were dissected from tissues covering the middle temporal gyrus (MTG), anterior cingulate gyrus, primary visual cortex, primary motor cortex, primary somatosensory cortex and primary auditory cortex derived from the human brain, and nuclei were dissociated and sorted using the neuronal marker NeuN. Nuclei were sampled from postmortem and neurosurgical (MTG only) donor brains and expression was profiled with SMART-Seq v4 or 10× Genomics Chromium Single Cell 3′ v3 RNA-seq. Specificity values of the gene expression were generated using the CELLEX (CELL-type EXpression-specificity) [24], which is a tool for computing cell-type Expression Specificity (ES) profiles. It employed a "wisdom of the crowd" approach by integrating multiple ES metrics, thus combining complementary cell-type ES profiles, to capture multiple aspects of ES and obtain improved robustness (https://github.com/perslab/CELLEX).

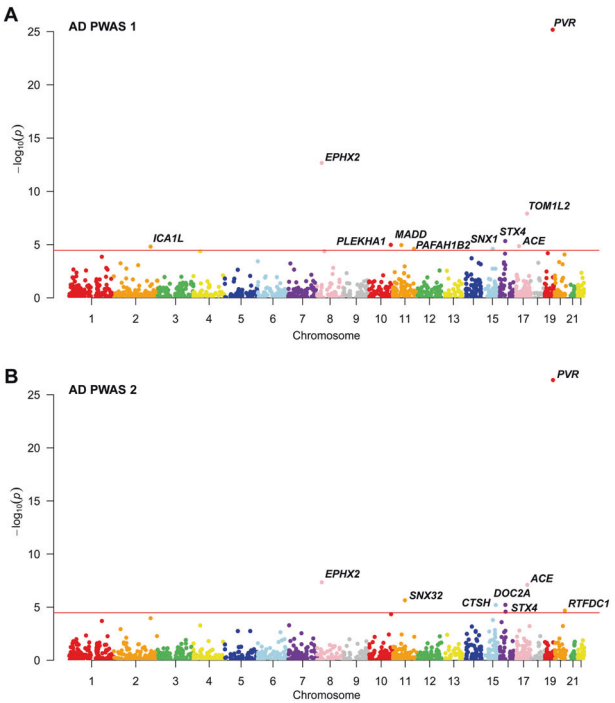

Fig. 2 Manhattan plots of PWAS by integrating brain pQTL and AD GWAS. The PWAS identified ten genes whose brain protein abundance was associated with Schwartzentruber AD GWAS (**A**) and eight genes associated with Jansen AD GWAS (**B**). Each point represents a single test of association between a gene and AD ordered by genomic position on the *X* axis and the association strength on the *Y* axis as the $-\log_{10}(P)$ of a *z*-score test. The red horizontal line reflects the significant threshold of the Bonferroni-corrected $P < 0.05/1468$ (3.41×10^{-5}) for Schwartzentruber AD GWAS and $P < 0.05/1476 = 3.39 \times 10^{-5}$ for Jansen AD GWAS.

RESULTS

PWAS identifies 14 candidate genes associated with AD using brain pQTL

We performed a PWAS by integrating AD GWAS results of 376 human brain proteomes profiled from the dPFC [25]. The PWAS conducted in the Schwartzentruber AD GWAS [17] identified 10 genes [*angiotensin-converting enzyme* (*ACE*), *epoxide hydroxylase* (*EPHX2*), *Islet Cell Autoantigen 1 Like* (*ICA1L*), *MAP Kinase Activating Death Domain* (*MADD*), *Platelet Activating Factor Acetylhydrolase 1b Catalytic Subunit 2* (*PAFAH1B2*), *Pleckstrin Homology Domain Containing A1* (*PLEKHA1*), *PVR Cell Adhesion Molecule* (*PVR*), *Sorting Nexin 1* (*SNX1*), *Syntaxin 4* (*STX4*), and *Target Of Myb1 Like 2 Membrane Trafficking Protein (TOM1L2)*] whose brain protein levels were associated with AD at a Bonferroni-corrected *P* value threshold of 3.41×10^{-5} (0.05/1468; Fig. 2A and Supplementary Table 2). Four of these 10 proteins (ACE, EPHX2, PVR, and STX4) were replicated in another PWAS using Jansen AD GWAS [3] (Fig. 2B and Supplementary Table 3). Another four proteins [Cathepsin H (CTSH), Double C2 Domain Alpha (DOC2A), Replication Termination Factor 2 Domain Containing 1 (RTFDC1), and Sorting Nexin 32 (SNX32)] were further identified in the Jansen AD GWAS. In summary, a total of 14 candidate genes associated with AD were identified using brain pQTL. Detailed information can be found in Table 1.

MR reveals 9 genes associated with AD using brain pQTL

Most of the analyzed proteins could only be instrumented using a single SNP, thus MR estimates were mainly based on the Wald ratio method. After corrections for multiple testing, we identified 6 protein

Table 1. PWAS identified 14 candidate genes in brain proteomes associated with AD.

	PWAS 1				PWAS 2			
	Gene	**CHR**	***z*-score**	***P* value**	**Gene**	**CHR**	***z*-score**	***P* value**
1	*ACE*	17	−5.695	1.23×10^{-8}	*ACE*	17	−5.370	8.04×10^{-8}
2	*EPHX2*	8	7.339	2.16×10^{-13}	*CTSH*	15	4.516	6.30×10^{-6}
3	*ICA1L*	2	−4.319	1.57×10^{-5}	*DOC2A*	16	−4.520	6.13×10^{-6}
4	*MADD*	11	−4.395	1.11×10^{-5}	*EPHX2*	8	5.467	4.58×10^{-8}
5	*PAFAH1B2*	11	−4.215	2.50×10^{-5}	*PVR*	19	−10.800	4.13×10^{-27}
6	*PLEKHA1*	10	4.405	1.06×10^{-5}	*RTFDC1*	20	4.252	2.12×10^{-5}
7	*PVR*	19	−10.525	6.66×10^{-26}	*SNX32*	11	−4.730	2.27×10^{-6}
8	*SNX1*	15	−4.210	2.55×10^{-5}	*STX4*	16	4.200	2.67×10^{-5}
9	*STX4*	16	4.581	4.62×10^{-6}	—	—	—	—
10	*TOM1L2*	17	4.342	1.41×10^{-5}	—	—	—	—

This Table shows all significant genes identified in the AD PWASs well as their *z*-scores and corresponding *P* values.
PWAS 1 describes the results associated with the Schwartzentruber AD GWAS (Bonferroni-corrected $P = 0.05/1468 = 3.41 \times 10^{-5}$), whereas PWAS 2 represents the results in the Jansen AD GWAS (Bonferroni-corrected $P = 0.05/1476 = 3.39 \times 10^{-5}$).

biomarkers that provided strong evidence of an association [$P < 8.18 \times 10^{-5}$ (0.05/611)] in the first AD GWAS (Table 2, Fig. 3A, and Supplementary Table 4). Associations between lower ACE, ICA1L, and Solute Carrier Family 20 Member 2 (SLC20A2) levels and higher AD risk (odds ratio (OR) and 95% confidence interval (CI) = 0.551 (0.412–0.736), $P = 5.76 \times 10^{-5}$ for ACE; OR 95% CI = 0.379 (0.244–0.589), $P = 1.57 \times 10^{-5}$ for ICA1L; OR and 95% CI = 0.454 (0.311–0.662), $P = 4.17 \times 10^{-5}$ for SLC20A2) were identified, as well as associations between higher Microtubule Associated Protein 1S (MAP1S), TOM1L2, and EPHX2 levels and higher AD risk [OR and 95% CI = 2.365 (1.577–3.547), $P = 3.17 \times 10^{-5}$ for MAP1S; OR and 95% CI = 3.855 (2.091–7.105), $P = 1.53 \times 10^{-5}$ for TOM1L2; OR and 95% CI = 1.422 (1.294–1.563), $P = 2.72 \times 10^{-13}$ for EPHX2].

We then evaluated 609 protein biomarkers with AD risk in the Jansen GWAS dataset (Table 2, Fig. 3B, and Supplementary Table 5). Consistent positive associations were observed for EPHX2 level with AD risk (OR and 95% CI = 1.064 (1.042–1.087), $P = 1.09 \times 10^{-8}$). In addition, genetically determined increased levels of CTSH and RTFDC1 increased the risk of AD [OR and 95% CI = 1.052 (1.029–1.070), $P = 6.29 \times 10^{-6}$ for CTSH; OR and 95% CI = 1.138 (1.073–1.208), $P = 1.76 \times 10^{-5}$ for RTFDC1], whereas SNX32 [OR and 95% CI = 0.900 (0.862–0.940), $P = 2.26 \times 10^{-6}$] decreased the risk of AD.

Colocalization between AD risk genes and pQTL in brain

Statistical colocalization analysis reported for each gene, the probability that the GWAS and pQTL share a causal variant, referred to as hypothesis 4 (PPH4). This analysis identified 5 of the 6 genes that provided evidence of genetic colocalization based on a PPH4 >80%. Results indicated that *ACE*, *ICA1L*, *MAP1S*, *SLC20A2*, and *TOM1L2* play a role in AD risk (PPH4 = 96.9, 97.2, 93.2, 89.9, and 87.6%, respectively; Table 2, Fig. 3A, and Supplementary Table 6). All of the 4 genes (*CTSH*, *EPHX2*, *RTFDC1*, and *SNX32*) also passed the PPH4 >80% criterion in the analysis of the Jansen GWAS dataset (Table 2, Fig. 3B, and Supplementary Table 7).

Gene expression in brain and altered AD risk

Considering that DNA is transcribed into mRNA, which is translated into protein, we asked whether genes with evidence for being causal in AD at the protein level had similar evidence at the transcriptional level by conducting MR and colocalization analyses. *ACE* was associated with a negative Wald ratio ($P = 8.85 \times 10^{-8}$), indicating a relationship between decreased gene expression and increased AD risk (Supplementary Tables 8 and 9). *ACE* gene was also associated with AD in the shared genetic effect between gene expression (eQTL) and AD risk (PPH4 = 99.2%; Supplementary Tables 8 and 11). Furthermore, the expression of *SNX32* in brain was associated with reduced AD risk in the Jansen GWAS dataset ($P = 3.12 \times 10^{-6}$, PPH4 = 99.2%; Supplementary Tables 8, 10, and 12).

Cell-type specificity analysis in the brain

Collectively, seven genes (*ACE*, *ICA1L*, *TOM1L2*, *SNX32*, *EPHX2*, *CTSH*, and *RTFDC1*) showed evidence consistent with being causal for AD by applying the discovered brain proteomes analysis (the right two columns in Fig. 4A, B). Using human single-cell RNA-seq data from the Cell Types database (https://portal.brain-map.org/atlases-and-data/rnaseq), we found cell type-specific enrichment for expression of the seven causal genes (Supplementary Table 13). *CTSH*, *EPHX2*, and *TOM1L2* were enriched in astrocytes, whereas *ICA1L* and *RTFDC1* showed enrichment in glutamatergic neurons. *ACE* showed enrichment in endothelial cells, and *SNX32* was enriched in glutamatergic neurons.

Six genes in blood proteomics were associated with AD risk

Applying our MR framework using serum proteomes data, 6 genetically predicted effects in the blood proteomics survived corrections for multiple testing ($P < 7.80 \times 10^{-5}$, based on 641 genes; Table 3, Fig. 3C, and Supplementary Table 14). Of these, the concentrations of 2 proteins were positively associated with AD risk [Apolipoprotein E (APOE), OR and 95% CI = 1.729 (1.652–1.809), $P = 4.00 \times 10^{-123}$; CD33 Molecule (CD33), OR and 95% CI = 1.052 (1.029–1.075), $P = 4.88 \times 10^{-6}$], while the other 4 were inversely associated with AD risk [ACE, OR and 95% CI = 0.926 (0.896–0.955), $P = 1.92 \times 10^{-6}$; Creatine Kinase, M-Type (CKM), OR and 95% CI = 0.713 (0.628–0.810), $P = 1.98 \times 10^{-7}$; Transmembrane Protein 106B (TMEM106B), OR and 95% CI = 0.857 (0.799–0.918), $P = 1.34 \times 10^{-5}$; Triggering Receptor Expressed On Myeloid Cells 2 (TREM2), OR and 95% CI = 0.560 (0.478–0.657), $P = 1.05 \times 10^{-12}$]. We next repeated the analysis using the Jansen GWAS meta-analysis, which verified three AD risk genes: *ACE*, *APOE*, and *CD33* (Table 3, Fig. 3D, and Supplementary Table 15).

Gene expression in blood and altered AD risk

A total of 26 and 15 Wald ratio effects showed evidence of MR at the multiple comparison corrected thresholds in the two AD GWASs, respectively (Supplementary Tables 16–18). When colocalization method was applied to these genes, 6 showed strong evidence for colocalization (PPH4 >80%; Supplementary Tables 16, 19, and 20). Analysis in the Jansen GWAS found that the *SNX32* gene passed the MR *P*-threshold with a value of 4.22×10^{-6} (Supplementary Tables 16 and 18).

Table 2. Candidate genes identified by MR and colocalization using brain pQTL.

	AD GWAS 1					AD GWAS 2				
	Gene	**Beta**	**SE**	***P* value**	**PPH4**	**Gene**	**Beta**	**SE**	***P* value**	**PPH4**
1	*ACE*	−0.600	0.148	5.76×10^{-5}	96.9%	*CTSH*	0.051	0.011	6.29×10^{-6}	100%
2	*ICA1L*	−0.970	0.225	1.57×10^{-5}	97.2%	*EPHX2*	0.062	0.011	1.09×10^{-8}	100%
3	*MAP1S*	0.861	0.207	3.17×10^{-5}	93.2%	*RTFDC1*	0.130	0.030	1.76×10^{-5}	100%
4	*SLC20A2*	−0.790	0.193	4.17×10^{-5}	89.9%	*SNX32*	−0.105	0.022	2.26×10^{-6}	99.7%
5	*TOM1L2*	1.349	0.312	1.53×10^{-5}	87.6%	—	—	—	—	—
6	*EPHX2*	0.352	0.048	2.72×10^{-13}	2.06×10^{-9}	—	—	—	—	—

This Table shows the beta, SE, and *P* values for the MR, as well as the result of PPH4 for colocalization analysis.
AD GWAS 1 describes the results analyzed in the Schwartzentruber AD GWAS (Bonferroni-corrected $P = 0.05/611 = 8.18 \times 10^{-5}$), whereas GWAS 2 represents the results in the Jansen AD GWAS (Bonferroni-corrected $P = 0.05/609 = 8.21 \times 10^{-5}$).

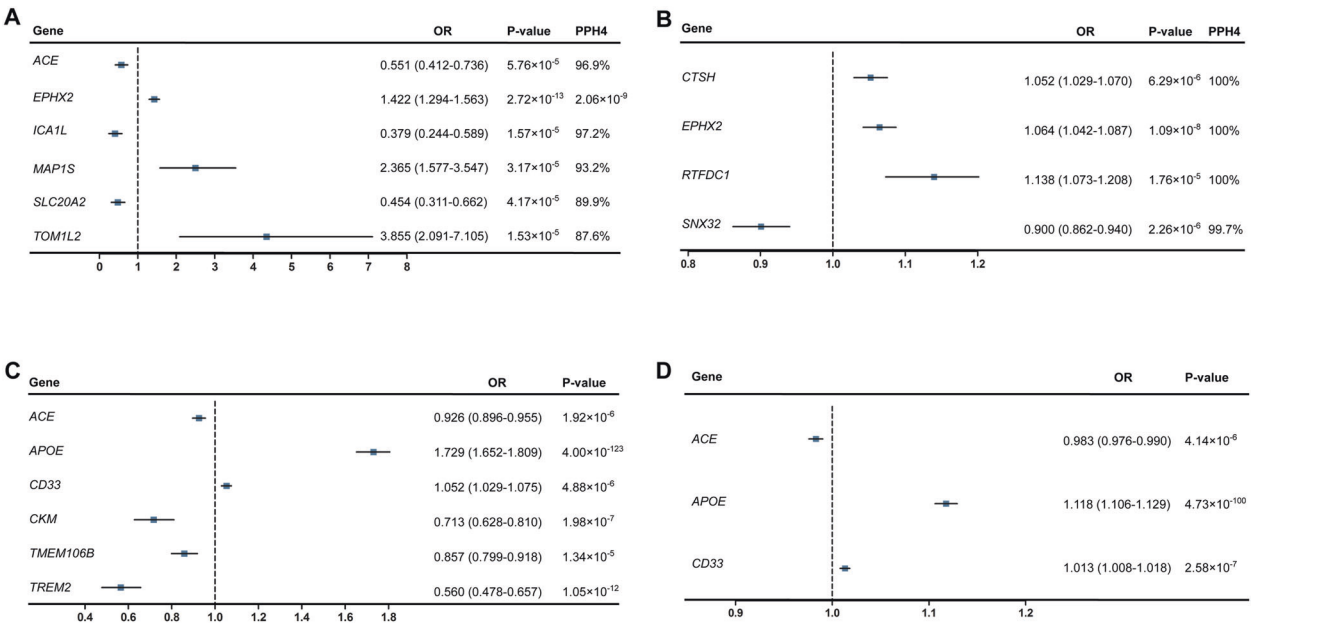

Fig. 3 MR and Bayesian colocalization analysis identified AD-associated genes. A, **B** are identified genes in brain tissue in AD GWAS (**A** Schwartzentruber GWAS; **B** Jansen GWAS), whereas **C**, **D** denote genes in serum in AD GWAS (**C** Schwartzentruber GWAS; **D** Jansen GWAS).

Summary findings

By PWAS, MR, and Bayesian colocalization analyses, *ACE*, *ICA1L*, *TOM1L2*, *SNX32*, *EPHX2*, *CTSH*, and *RTFDC1* showed evidence of being causal in AD brain tissue. Furthermore, *ACE* was supported by MR of blood proteomics and also showed significance at the transcriptional level. *SNX32* was also associated with AD at the blood transcriptional level. Summary plots are shown in Fig. 4A–D.

DISCUSSION

In the present study, we proposed a pipeline of analytical techniques that investigate the functional associations between multiple protein biomarkers in the brain and blood with AD risk. Collectively, we identified seven genes (*ACE*, *ICA1L*, *TOM1L2*, *SNX32*, *EPHX2*, *CTSH*, and *RTFDC1*) from a comprehensive analyses including brain PWAS, MR and colocalization, as well as *ACE* with its protein abundance in significant association with AD on the blood-based studies and *SNX32* being supported at the blood transcriptional level. Furthermore, our results suggested importance of glutamatergic neurons and astrocytes in AD because most of the genes discovered in our studies mainly expressed on the surface of these cells.

Identifying therapeutic targets for diseases is a key goal of human genetics research and is particularly vital for neurodegenerative diseases, including AD, for which no disease-modifying therapies currently exist [2]. Our analysis implicated genes that have been widely investigated in AD, such as *ACE*, *TOM1L2*, *EPHX2*, and *CTSH*, as well as new candidates, including *ICA1L*, *RTFDC1*, and *SNX32*. *TOM1L2* encodes a protein putatively involved in intracellular recruitment of clathrin onto endosomes and has a proportionately high level of connectivity with known AD genes in the 17p LD block [26]. *EPHX2*, cosited at a locus associated with AD, encodes a key enzyme for epoxyeicosatrienoic acid signaling named soluble epoxide hydrolase [27]. *CTSH*, a linked AD gene, is expressed in the temporal cortices of late-onset Alzheimer patients and was shown to be altered [28]. *ICA1L* and *RTFDC1* were new candidate genes, which were also identified as being associated with AD in a recent published integrative study [25]. *ICA1L* was revealed to be related to cerebro- and cardio-vascular diseases in recent GWASs [29, 30]. The removal of Replication termination factor 2 is a key determinant for the ability of cells to manage replication stress and maintain genome integrity [31]. SNX32 is the member of Sorting nexins (SNXs) family, which have been implicated in regulating membrane trafficking in the endocytic

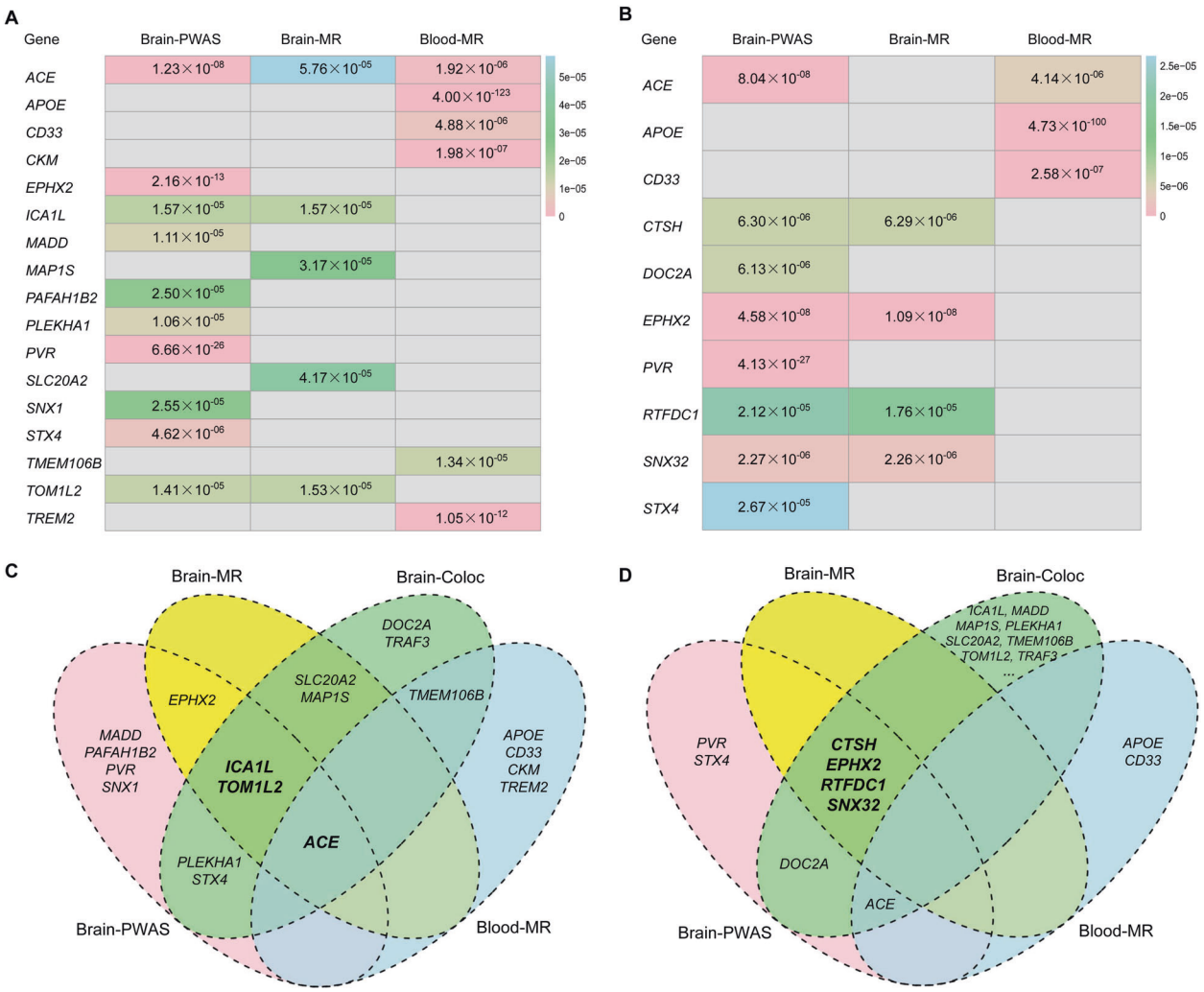

Fig. 4 Summary of the AD significant genes with evidence for being consistent among the four approaches. A, **B** are heatmap plots denoting the genetic correlations between candidate genes and AD. The rows denote three methods for identifying candidate genes associated with AD. The columns are the gene names. The color of the block represents the size of *P* values of each genetic correlation. Genes passed both of the Brain-MR and colocalization analyses were depicted in the Brain-MR column. The missing *P* values (gray background) represent the analyses that were not performed because of the lack of this part of data. **C**, **D** are Venn plots of the AD significant genes with evidence for being consistent among the four approaches. Collectively, **A**, **C** were results in Schwartzentruber GWAS; **B**, **D** were in Jansen GWAS.

Table 3. Candidate genes identified by MR using blood proteomes.

	AD GWAS 1				AD GWAS 2			
	Gene	**Beta**	**SE**	***P* value**	**Gene**	**Beta**	**SE**	***P* value**
1	*ACE*	−0.077	0.016	1.92×10^{-6}	*ACE*	−0.017	0.004	4.14×10^{-6}
2	*APOE*	0.547	0.023	4.00×10^{-123}	*APOE*	0.111	0.005	4.73×10^{-100}
3	*CD33*	0.051	0.011	4.88×10^{-6}	*CD33*	0.013	0.003	2.58×10^{-7}
4	*CKM*	−0.338	0.065	1.98×10^{-7}	—	—	—	—
5	*TMEM106B*	−0.155	0.036	1.34×10^{-5}	—	—	—	—
6	*TREM2*	−0.579	0.081	1.05×10^{-12}	—	—	—	—

This Table shows the beta, SE, and *P* values for the MR analysis of blood proteomes.
AD GWAS 1 describes the results analyzed in the Schwartzentruber AD GWAS (Bonferroni-corrected $P = 0.05/641 = 7.80 \times 10^{-5}$), whereas GWAS 2 represents the results in the Jansen AD GWAS (Bonferroni-corrected $P = 0.05/632 = 7.91 \times 10^{-5}$).

pathway [32]. *SNX32* has been found to be associated with increased risk of AD, however, it influences other complex traits in the opposite direction, which may make *SNX32* a less attractive therapeutic target [22]. The findings suggest that an efficient integration of the pQTL and GWAS signals are helpful for deciphering disease biology. In general, genetic data derived from brain tissue that captures biological effects relevant to this disease are pertinent to characterizing genes associated with AD.

AD is a multifactorial disease affecting both the brain and the periphery. A correlation of $R^2 = 0.70$ between comparisons of QTL derived from brain tissue and whole blood associated with the same gene target was previously reported [33], suggesting that blood is broadly a valid proxy for brain tissue. Blood-based profiles might provide an accessible and effective strategy for evaluating the complicated interplay between brain tissue and the periphery in the pathogenesis of AD. In the present analysis, only one of the above seven protein's abundance (ACE) was verified to be associated with AD risk in serum. We recognize that there may be multiple reasons where this is the case: (1) it may be partly due to potential methodological differences in tissue collection, material extraction, and analytical platforms; (2) a decoupling of protein markers in brain and blood due to their different tissue type and expression profiling; that is, genes that are highly expressed in brain tissues are not necessarily expressed at the same level in peripheral blood; and (3) proteomic analysis in blood has a limited number of proteins and does not fully capture all protein indicators. Previous candidate gene studies and large GWASs have associated *ACE* variants with AD risk. Both ACE expression in AD brain tissue and cerebrospinal fluid (CSF) levels of ACE are associated with Aβ load [34] and AD severity [35]. Other studies have shown that ACE can inhibit Aβ toxicity and aggregation [36]. A large-sample longitudinal study indicated that ACE inhibitor use was associated with reduced risk of AD dementia [37]. Researchers found the therapeutic roles of *ACE*-overexpressing macrophages in preserving synapses and cognition, attenuating neuropathology and neuroinflammation, and enhancing resistance to pathognomonic amyloid-β forms in AD^{+} mice [38]. Moreover, angiotensin II, a product of ACE function, was reported to mediate various neuropathological processes and is now an intervention target in phase II clinical trials of AD [39]. Future high-quality trails targeted on *ACE* might be a promising strategy. In general, our findings along with the extant literature may help shed light on the causal pathways between these candidate genes and AD risk.

The majority of known trait-associated variants reside in non-protein-coding regions of the human genome, with previous research implicating them with transcriptional regulatory mechanisms [40, 41]. As such, our analysis applying eQTL develops our mechanistic regulatory understanding that underlies associations from GWAS. However, only two (ACE and SNX32) of the seven identified proteins provided evidence of gene expression (Supplementary Tables 7–11). There could be several reasons for this lack of agreement. First, though the exact correlation between eQTLs and pQTLs has not been fully determined [42], mRNA expression is known to be poorly correlated with protein levels for many genes [43], which is in part due to many post-transcriptional factors, such as sequence features implicated in protein translation and degradation [44]. Next, technical artifacts in assays and differences in data analysis approaches might play an important role. Compared with pQTL, eQTL studies employ more stringent criteria to identify distant regulatory variants as eQTL is more likely to be related to local regulatory elements [45]. As such, transcriptomic and proteomic analyses provide different perspectives of AD risk genes, and a higher proportion of transcriptomic and proteomic signatures may colocalize as sample sizes of molecular datasets grow.

In general, leveraging three completely different but mutually reinforcing methods (PWAS, MR, and Bayesian colocalization), we identified seven genetically predicted effects across the genome. Our results suggest that a shared genetic architecture between AD and a subset of proteins does exist. PWAS can help establish associations between proteome abundance and phenotypic variations [7]. MR uses genetic variants as a proxy for exposures to infer causality among correlated traits [8]. Colocalization analysis is a Bayesian approach to estimate the probability that two observed association signals in a given locus are consistent with a shared causal variant [9, 46]. Together, these three methods identified candidate proteins implicated in AD pathogenesis and novel biomarkers for AD that will be useful for developing therapeutic strategies [17]. One recent integrative analysis identified 11 proteins in brain tissue being causal in AD, among which 7 passed PWAS criteria, summary-based MR, and colocalization analysis [25]. Our findings are partly consistent, with *ICA1L*, *SNX32*, *CTSH*, and *RTFDC1* identified. However, our present research employed two large AD GWAS databases [3, 17], including one published very recently [17]. Moreover, we integrated data derived from brain tissue and verified the results in serum. Thus, our present results are more novel and comprehensive, though further downstream studies are needed to confirm the exact roles of these genes. Another proteomic analysis employing MR analysis in multiple tissues (brain, CSF, and plasma) found that *CD33*, *TREM2*, and *Membrane Spanning 4-domains A6A* (*MS4A6A*) genes were associated with AD, partly consistent with our results [47]. All of these findings provide new evidence for potential therapeutic targets by linking genetic factors to disease via specific proteins.

The present study has several limitations. First, pQTL and eQTL mapping are unable to resolve all GWAS signals. The role of genes involving AD biological progression is difficult to decipher clearly in a distinct level: here, the protein level. Fully mapping the biological mechanisms of AD requires further focus on epigenetics, such as mQTL, single-cell sequencing, and whole-genome sequencing [48]. Second, the lack of comprehensive coverage of all proteomics due to detection limited to specific proteomes targeted by the SOMAmers method. Studies using other detection technologies and biological samples might reveal further associations. Third, undue focus on single-SNP MR approaches brings a susceptibility to other pitfalls, including the inability to examine potential heterogeneity. Fourth, variations in GWAS samples and statistical power limited our ability to identify certain associations with smaller effects. As such, ongoing expansion in the scale and diversity of GWASs can help with more precise estimates and enable its broader application. Fifth, the diagnostic criteria of AD in the GWASs vary a lot, thus introducing heterogeneity of the results. The future use of AD-related biomarkers (amyloid, tau, and neuroimaging) may be more helpful in identifying potential target proteins. Last, the present analysis was restricted to non-Latino Caucasians, and thus findings should be replicated in cohorts of other ethnicities.

In conclusion, we found strong evidence supporting seven AD-associated causal relationships of protein biomarkers in the brain, of which *ACE* was further verified in blood samples. These findings illuminate the causal pathways underlying AD on both genetic and functional levels, which thus prioritize candidate targets for therapeutic intervention. Future studies leveraging broader large-scale molecular datasets derived from AD-relevant tissues are warranted, which may further aid in not only developing novel insights into the genetic and functional mechanisms but also providing potential druggable targets enabling new interventions of AD.

DATA AVAILABILITY

Data of brain pQTL and eQTL from the ROS/MAP study are available through https://doi.org/10.7303/syn23627957 and https://www.synapse.org/#!Synapse:syn3219045, respectively. Proteomic and transcriptomic data used in this manuscript are available via the AD Knowledge Portal (https://adknowledgeportal.org). Data are available for general research use according to the following requirements for data access and data attribution (https://adknowledgeportal.org/DataAccess/Instructions). Data from AGES Reykjavik study can be accessed at www.sciencemag.org/cgi/content/full/science.aaq1327/DC1. Data from the AGES Reykjavik study are available through collaboration (AGES_data_request@hjarta.is) under a data usage agreement with the IHA. GTEx can be accessed at https://gtexportal.org/home/datasets (GTEx Analysis V8). Summary statistics for the Schwartzentruber's meta-analysis are available through the National Human Genome Research Institute-European Bioinformatics Institute GWAS catalog under accession nos. GCST90012877 and GCST90012878

(https://www.ebi.ac.uk/gwas/downloads/summary-statistics). Summary statistics for the Jansen's GWAS can be made available for download upon publication (https://ctg.cncr.nl/software/summary_statistics). Cell-type specificity data are available at https://portal.brain-map.org/atlases-and-data/rnaseq.

CODE AVAILABILITY

Codes associated with the current submission can be requested by contacting the corresponding author.

REFERENCES

1. 2021 Alzheimer's disease facts and figures. Alzheimers Dement. 2021;17:327–406.
2. Arvanitakis Z, Shah RC, Bennett DA. Diagnosis and management of dementia: review. JAMA. 2019;322:1589.
3. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet. 2019;51:404–13.
4. 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 Aβ, tau, immunity and lipid processing. Nat. Genet. 2019;51:414–30.
5. Cannon ME, Mohlke KL. Deciphering the emerging complexities of molecular mechanisms at GWAS loci. Am J Hum Genet. 2018;103:637–53.
6. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558:73–9.
7. Brandes N, Linial N, Linial M. PWAS: proteome-wide association study—linking genes and phenotypes by functional variation in proteins. Genome Biol. 2020;21:173.
8. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.
9. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations. Wellcome Open Res. 2019;4:186.
10. McGowan LM, Davey Smith G, Gaunt TR, Richardson TG. Integrating Mendelian randomization and multiple-trait colocalization to uncover cell-specific inflammatory drivers of autoimmune and atopic disease. Hum Mol Genet. 2019;28:3293–300.
11. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016;48:481–7.
12. Richardson TG, Hemani G, Gaunt TR, Relton CL, Smith GD. A transcriptome-wide Mendelian randomization study to uncover tissue-dependent regulatory mechanisms across the human phenome. Nat Commun. 2020;11:185.
13. Wang M, Beckmann ND, Roussos P, Wang E, Zhou X, Wang Q, et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. Sci Data. 2018;5:180185.
14. De Jager PL, Ma Y, McCabe C, Xu J, Vardarajan BN, Felsky D, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. Sci Data. 2018;5:180142.
15. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, et al. Co-regulatory networks of human serum proteins link genetics to disease. Science. 2018;361:769–73.
16. Yang J, Yu X, Zhu G, Wang R, Lou S, Zhu W, et al. Integrating GWAS and eQTL to predict genes and pathways for non-syndromic cleft lip with or without palate. Oral Dis. 2020. https://doi.org/10.1111/odi.13699.
17. Schwartzentruber J, Cooper S, Liu JZ, Barrio-Hernandez I, Bello E, Kumasaka N, et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. Nat Genet. 2021;53:392–402.
18. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48:245–52.
19. Wingo AP, Dammer EB, Breen MS, Logsdon BA, Duong DM, Troncosco JC, et al. Large-scale proteomic analysis of human brain identifies proteins associated with cognitive trajectory in advanced age. Nat Commun. 2019;10:1619.
20. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23:R89–98.
21. Rasooly D, Patel CJ. Conducting a reproducible Mendelian randomization analysis using the R analytic statistical environment. Curr Protoc Hum Genet. 2019;101:e82.
22. Kibinge NK, Relton CL, Gaunt TR, Richardson TG. Characterizing the causal pathway for genetic variants associated with neurological phenotypes using human brain-derived proteome data. Am J Hum Genet. 2020;106:885–92.
23. Taylor K, Davey Smith G, Relton CL, Gaunt TR, Richardson TG. Prioritizing putative influential genes in cardiovascular disease susceptibility by applying tissue-specific Mendelian randomization. Genome Med. 2019;11:6.
24. Timshel PN, Thompson JJ, Pers TH. Genetic mapping of etiologic brain cell types for obesity. Elife. 2020;9:e55851.
25. Wingo AP, Liu Y, Gerasimov ES, Gockley J, Logsdon BA, Duong DM, et al. Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. Nat Genet. 2021;53:143–6.
26. Reynolds CA, Hong M-G, Eriksson UK, Blennow K, Wiklund F, Johansson B, et al. Analysis of lipid pathway genes indicates association of sequence variation near SREBF1/TOM1L2/ATPAF2 with dementia risk. Hum Mol Genet. 2010;19:2068–78.
27. Padhy B, Hayat B, Nanda GG, Mohanty PP, Alone DP. Pseudoexfoliation and Alzheimer's associated CLU risk variant, rs2279590, lies within an enhancer element and regulates CLU, EPHX2 and PTK2B gene expression. Hum Mol Genet. 2017;26:4519–29.
28. Chen W, Wang M, Zhu M, Xiong W, Qin X, Zhu X. 14,15-Epoxyeicosatrienoic acid alleviates pathology in a mouse model of Alzheimer's disease. J Neurosci. 2020;40:8188–203.
29. Wang X, Mo X, Zhang H, Zhang Y, Shen Y. Identification of phosphorylation associated SNPs for blood pressure, coronary artery disease and stroke from genome-wide association studies. Curr Mol Med. 2019;19:731–8.
30. Chung J, Marini S, Pera J, Norrving B, Jimenez-Conde J, Roquer J, et al. Genome-wide association study of cerebral small vessel disease reveals established and novel loci. Brain. 2019;142:3176–89.
31. Kottemann MC, Conti BA, Lach FP, Smogorzewska A. Removal of RTF2 from stalled replisomes promotes maintenance of genome integrity. Mol Cell. 2018;69:24.e5–35.e5.
32. Seet LF, Hong W. The Phox (PX) domain proteins and membrane traffic. Biochim Biophys Acta. 2006;1761:878–96.
33. Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nat Commun. 2018;9;2282.
34. Jochemsen HadassaM, Teunissen CharlotteE, Ashby EmmaL, van der Flier WiesjeM, Jones RuthE, Mirjam I Geerlings, et al. The association of angiotensin-converting enzyme with biomarkers for Alzheimer's disease. Alzheimers Res Ther. 2014;6:27.
35. Miners Scott, Ashby Emma, Baig Shabnam, Harrison Rachel, Tayler Hannah, Speedy Elizabeth, et al. Angiotensin-converting enzyme levels and activity in Alzheimer's disease: differences in brain and CSF ACE and association with ACE1 genotypes. Am J Transl Res. 2009;1:163–77.
36. Baranello RJ, Bharani KL, Padmaraju V, Chopra N, Lahiri DK, Greig NH, et al. Amyloid-beta protein clearance and degradation (ABCD) pathways and their role in Alzheimer's disease. Curr Alzheimer Res. 2015;12:32–46.
37. Yasar S, Xia J, Yao W, Furberg CD, Xue QL, Mercado CI, et al. Antihypertensive drugs decrease risk of Alzheimer disease: Ginkgo Evaluation of Memory Study. Neurology. 2013;81:896–903.
38. Koronyo-Hamaoui M, Sheyn J, Hayden EY, Li S, Fuchs DT, Regis GC, et al. Peripherally derived angiotensin converting enzyme-enhanced macrophages alleviate Alzheimer-related disease. Brain. 2020;143:336–58.
39. Kehoe PG, Perry G, Avila J, Tabaton M, Zhu X. The coming of age of the angiotensin hypothesis in Alzheimer's disease: progress toward disease prevention and treatment? J Alzheimers Dis. 2018;62:1443–66.
40. Huo Y, Li S, Liu J, Li X, Luo X-J. Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. Nat Commun 2019;10:670.
41. Gusev A, Mancuso N, Won H, Kousi M, Finucane HK, Reshef Y, et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. Nat Genet. 2018;50:538–48.
42. Nounu A, Greenhough A, Heesom KJ, Richmond RC, Zheng J, Weinstein SJ, et al. A combined proteomics and Mendelian randomization approach to investigate the effects of aspirin-targeted proteins on colorectal cancer. Cancer Epidemiol Biomarkers Prev. 2020;30:564–75.
43. Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, et al. Genomic variation. Impact of regulatory variation from RNA to protein. Science. 2015;347:664–7.
44. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012;13:227–32.
45. GTEx Consortium. Genetic effects on gene expression across human tissues. Nature. 2017;550:204–13.
46. Giambartolomei C, Zhenli Liu J, Zhang W, Hauberg M, Shi H, Boocock J, et al. A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics. 2017;34:2538–45.
47. Yang C, Farias FHG, Ibanez L, Suhy A, Sadler B, Fernandez MV, et al. Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. Nat Neurosci. 2021.
48. Prokopenko D, Morgan SL, Mullin K, Hofmann O, Chapman B, Kirchner R, et al. Whole-genome sequencing reveals new Alzheimer's disease–associated rare variants in loci related to synaptic function and neuronal development. Alzheimers Dement. 2021.

ACKNOWLEDGEMENTS

This study was supported by grants from the National Natural Science Foundation of China (82071201, 91849126), Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01) and ZHANGJIANG LAB, Tianqiao and Chrissy Chen Institute, and the State Key Laboratory of Neurobiology and Frontiers Center for Brain Science of Ministry of Education, Fudan University. This work was made possible by the generous sharing of statistics from the public databases. We thank the participants of the ROS and MAP for their time and effort. Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. We thank the AGES Reykjavik study, GTEx project, and Cell Types database for their kind dedication. We thank the Schwartzentruber's GWAS meta-analysis and Jansen's GWAS. Statistics were made possible by their generous sharing of GWAS summary statistics. Access to data is shown below. The full list of acknowledgements can be found in the Supplementary Text.

AUTHOR CONTRIBUTIONS

J-TY conceptualized the study and revised the manuscript. Y-NO, Y-XY, B-SW, Y-TD, and YL analyzed and interpreted the data. Y-NO prepared all the figures and tables. Y-NO and Y-XY drafted and revised the manuscript. HH, LT, JS, CZ, YZ, and Y-JW revised the manuscript. All authors contributed to the writing and revisions of the paper and approved the final version.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-021-01251-6.

Correspondence and requests for materials should be addressed to J.-T.Y.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.