


RESEARCH ARTICLE

Open Access



# Genome-wide association study of brain biochemical phenotypes reveals distinct genetic architecture of Alzheimer's disease related proteins

Stephanie R. Oatman<sup>1†</sup>, Joseph S. Reddy<sup>2†</sup>, Zachary Quicksall<sup>2</sup>, Minerva M. Carrasquillo<sup>1</sup>, Xue Wang<sup>2</sup>, Chia-Chen Liu<sup>1</sup>, Yu Yamazaki<sup>1</sup>, Thuy T. Nguyen<sup>1</sup>, Kimberly Malphrus<sup>1</sup>, Michael Heckman<sup>2</sup>, Kristi Biswas<sup>1</sup>, Kwangsik Nho<sup>3,4,5</sup>, Matthew Baker<sup>1</sup>, Yuka A. Martens<sup>1</sup>, Na Zhao<sup>1</sup>, Jun Pyo Kim<sup>4</sup>, Shannon L. Risacher<sup>3,4</sup>, Rosa Rademakers<sup>1,6</sup>, Andrew J. Saykin<sup>3,7</sup>, Michael DeTure<sup>1</sup>, Melissa E. Murray<sup>1</sup>, Takahisa Kanekiyo<sup>1</sup>, for the Alzheimer's Disease Neuroimaging Initiative, Dennis W. Dickson<sup>1</sup>, Guojun Bu<sup>1</sup>, Mariet Allen<sup>1</sup> and Nilüfer Ertekin-Taner<sup>1,8\*</sup> 

## Abstract

**Background** Alzheimer's disease (AD) is neuropathologically characterized by amyloid-beta (A $\beta$ ) plaques and neurofibrillary tangles. The main protein components of these hallmarks include A $\beta$ 40, A $\beta$ 42, tau, phosphor-tau, and APOE. We hypothesize that genetic variants influence the levels and solubility of these AD-related proteins in the brain; identifying these may provide key insights into disease pathogenesis.

**Methods** Genome-wide genotypes were collected from 441 AD cases, imputed to the haplotype reference consortium (HRC) panel, and filtered for quality and frequency. Temporal cortex levels of five AD-related proteins from three fractions, buffer-soluble (TBS), detergent-soluble (Triton-X = TX), and insoluble (Formic acid = FA), were available for these same individuals. Variants were tested for association with each quantitative biochemical measure using linear regression, and GSA-SNP2 was used to identify enriched Gene Ontology (GO) terms. Implicated variants and genes were further assessed for association with other relevant variables.

**Results** We identified genome-wide significant associations at seven novel loci and the *APOE* locus. Genes and variants at these loci also associate with multiple AD-related measures, regulate gene expression, have cell-type specific

<sup>†</sup>Stephanie R. Oatman and Joseph S. Reddy contributed equally to this work.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

\*Correspondence:

Nilüfer Ertekin-Taner

Taner.Nilufer@mayo.edu; taner.nilufer@mayo.edu

Full list of author information is available at the end of the article



enrichment, and roles in brain health and other neuropsychiatric diseases. Pathway analysis identified significant enrichment of shared and distinct biological pathways.

**Conclusions** Although all biochemical measures tested reflect proteins core to AD pathology, our results strongly suggest that each have unique genetic architecture and biological pathways that influence their specific biochemical states in the brain. Our novel approach of deep brain biochemical endophenotype GWAS has implications for pathophysiology of proteostasis in AD that can guide therapeutic discovery efforts focused on these proteins.

**Keywords** Alzheimer's, Genetics, APOE, Amyloid, Tau, Association, Brain, GWAS, Neuroscience, Diseases, Biochemistry

## Background

Alzheimer's disease (AD) is a progressive neurodegenerative disorder, neuropathologically characterized by the accumulation of amyloid beta (A $\beta$ ) plaques and neurofibrillary tangles (NFT) in the brain [1, 2]. While AD neuropathology broadly follows characteristic patterns, heterogeneity in the composition, location and burden of the two primary lesions has been reported across post-mortem datasets [3–8]. The main component of insoluble amyloid plaques is A $\beta$ 42, while A $\beta$ 40 is often found deposited in the brain cerebrovasculature called cerebral amyloid angiopathy (CAA) [8]. A $\beta$  is generated by the normal cleavage of the amyloid-beta precursor protein (APP), which then can oligomerize and form extracellular deposits [9, 10]. Some mutations in the *APP* gene cause a familial early-onset form of AD through modification of APP cleavage resulting in an increase in A $\beta$ 42 production [11]. Increased tau levels are also observed in AD, along with abnormal hyperphosphorylation leading to aggregation into insoluble NFT within the cell body [2, 12]. Under normal conditions in the brain the soluble tau protein is found relatively un-phosphorylated and bound to microtubules for stabilization [13–15]. Previous genetic studies of late-onset AD (LOAD) have found variants associated with the risk of developing AD; the most significant of which is the well-established *APOE*- $\epsilon$ 4 allele [16–18]. *APOE* encodes apolipoprotein E (APOE) which functions mainly in lipid transport, but is also known to play a role in A $\beta$  metabolism and its insoluble forms are often found co-deposited with A $\beta$  plaques [19]. Beyond insoluble deposits of amyloid and tau species, soluble and membrane-associated biochemical states of these proteins have also been associated with AD-related phenotypes. In the temporal cortex, soluble levels of A $\beta$ 40 and A $\beta$ 42 are significantly elevated in AD compared to controls and A $\beta$ 40 levels positively correlate with disease duration [20, 21]. Membrane-associated forms of A $\beta$  show a significant positive correlation with A $\beta$  positron emission tomography (PET) imaging in AD, while cortical A $\beta$ 42 levels have been reported to correlate with worse clinical severity and increased rate of cognitive decline [20, 21]. Moreover, it has been shown that when tau interacts with the plasma membrane, the propensity

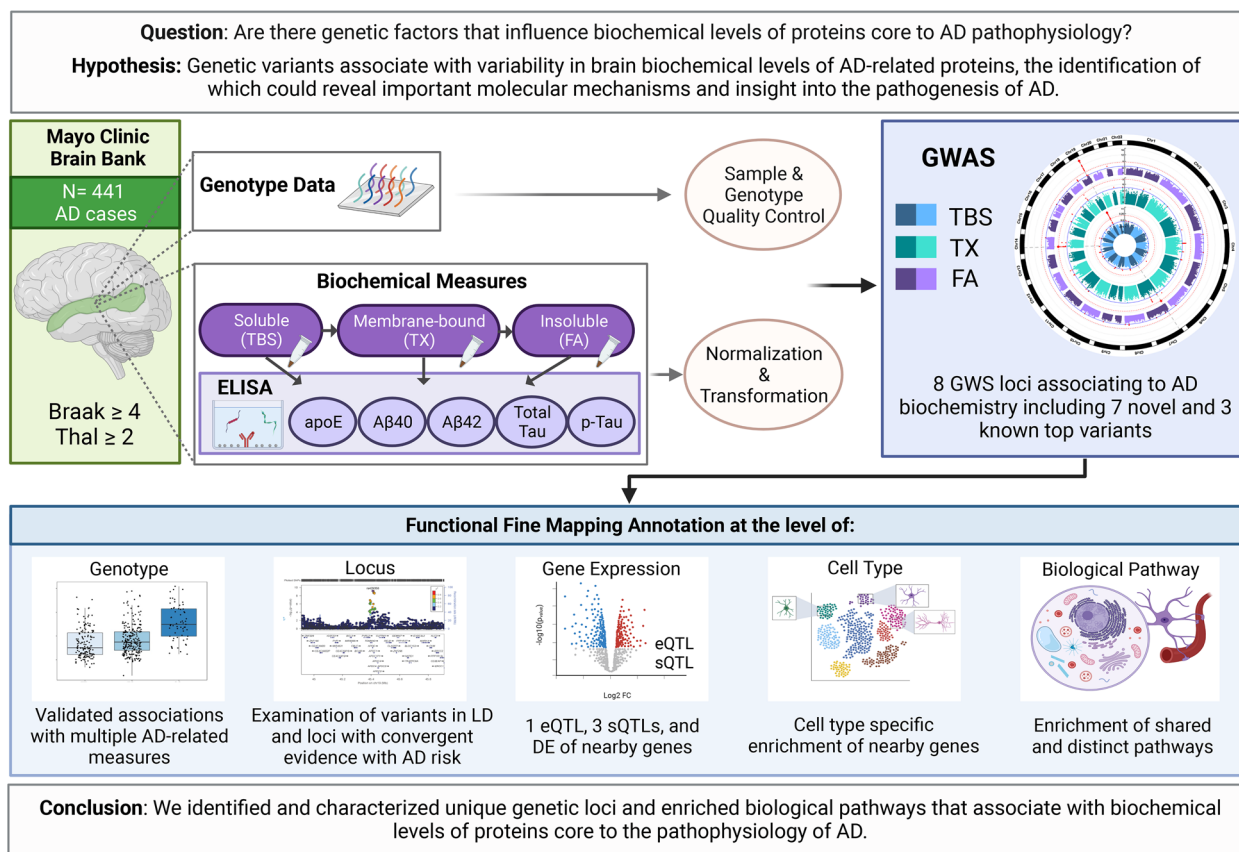
for fibrillization increases, and within the context of AD, variability in soluble tau has been shown to occur in the presence of A $\beta$  pathology but before significant NFT pathology [22–31]. Apart from *APOE*- $\epsilon$ 4, genetic risk factors associated with different brain biochemical states of distinct proteins core to AD pathology have yet to be identified and characterized [32].

We hypothesize that important insights into the pathogenesis of AD may be gained by identifying genetic variants associated with variability in brain levels of AD-related protein endophenotypes including A $\beta$ 40, A $\beta$ 42, tau, phosphorylated tau (p-Tau) and APOE. Furthermore, different biochemical states (soluble, membrane, and insoluble) of AD-related proteins may have distinct genetic variants that influence their levels within the brain. Such findings may provide key insights into production or clearance pathways for these disease-associated proteins, leading to novel therapeutic targets or biomarkers. To investigate this, we utilized genetic and biochemical measures collected from the temporal cortex of 441 post-mortem AD cases. We performed a genome-wide association study (GWAS) for levels of all five proteins, collected from three biochemical states in the brain (Fig. 1). Our findings reveal novel genetic loci and highlight the unique genetic architecture for specific biochemical states of AD-related protein endophenotypes. This study establishes deep brain biochemical endophenotype GWAS as a novel approach to dissect the biochemical heterogeneity of AD proteins which is essential to fine-tune therapeutic efforts targeting these proteins.

## Methods

### Brain samples

Post-mortem temporal cortex samples included in this study were a part of the Mayo Clinic AD-CAA (MC-CAA) study on the AD-knowledge portal (<https://adknowledgeportal.synapse.org>, see data sharing). All samples had a confirmed AD neuropathological diagnosis, a Braak stage  $\geq$  four, Thal phase  $\geq$  three, and were from non-Hispanic White decedents of Northern European descent. In total, 441 samples had both genome-wide genotyping data and biochemical measures available for



**Fig. 1** Graphical Abstract. Graphical depiction of this study: AD = Alzheimer’s Disease, N = Number, Aβ = amyloid beta, p-Tau = phosphorylated tau, TBS = Tris Buffered Saline, TX = Triton-X, FA = Formic Acid, GWS = Genome-wide significant, LD = Linkage Disequilibrium, eQTL = expression quantitative trait locus, sQTL = splicing quantitative trait locus, DE = Differential expression. Created with BioRender.com

analyses. This study was approved by the appropriate Mayo Clinic Institutional Review Board.

**Neuropathology**

Braak stage, Thal phase, and CAA scores were measured by the Mayo Clinic Brain Bank using previously established protocols [32–36]. Intermediate Braak stages were grouped with the next lowest stage as follows, stage 3.5 is 3, 4.5 is 4, and 5.5 is 5 as detailed previously [32, 37].

**Biochemical measures**

Biochemical measures from 441 of the 469 superior temporal cortex brain samples previously described [32] were utilized for this study based on the availability of genome-wide genotypes. Biochemical measures include five AD-related proteins (APOE, Aβ40, Aβ42, tau, and phospho-tau (Thr231)) from three tissue fractions. Briefly, supernatant fractions were collected after three sequential buffer treatments of tissue homogenate and resulting pellets: first with tris-buffered saline buffer (TBS), second with detergent (TBS/1% Triton X) buffer

(TX), and finally with formic acid buffer (FA), representing soluble, lipid-membrane and insoluble biochemical fractions. AD-related biochemical measures were quantified in each fraction via ELISA and normalized against total protein quantities. All biochemical measures were transformed by either the natural log or square root to achieve an approximately normal distribution including the Aβ40/42 ratio (Fig. S1). In the subset of 441 samples analyzed, we evaluated the association of the AD-related proteins within and among these tissue fractions through pairwise correlation and found similar results as reported previously [32] (Data not shown).

**Genotyping**

DNA was isolated from brain tissue using the AutoGen245T instrument according to manufacturer’s protocols, incubated with two μl (4 mg.ml) RNaseA solution (Qiagen, Germany) and stored at -80°C until use. Genome-wide genotypes from 477 samples were previously collected [37] using the Infinium Omni2.5 Exome8 v1.3 genotyping array and results were exported using

the Illumina GenomeStudio software v1.9.4. Data was formatted into PLINK (v1.9) files [38, 39] (lgen, fam, and map) and quality control of the samples and genotypes was performed, as described in detail elsewhere [37]. Four hundred sixty samples passed quality control (QC), of which 441 also had biochemical measures [32]. Variants passing quality control ( $N = 1,383,987$ ) were imputed to the haplotype reference consortium (HRC) panel [40, 41] and those with an imputation quality  $R^2 \geq 0.7$  and  $MAF \geq 2\%$  were kept, yielding a total of 6,726,078 variants for analysis. Genotype dosages were converted to hard calls when needed with uncertainty  $> 0.1$  set to missing. Minor allele frequencies and Hardy-Weinberg  $p$ -values were calculated for all reported variants using dosages in PLINK [38, 39].

Genotypes for key variants, or their proxies ( $r^2 = 1$ ,  $D' = 1$  in 1000 Genomes EUR), were validated by Taqman genotyping or Sanger sequencing following the manufacturer's protocols. These assays were also used to collect genotypes from an additional 1564 Mayo Clinic Brain Bank (MCBB) samples with available DNA to enable assessment with AD-related neuropathology measures of Braak stage, Thal phase, neuropathological diagnosis of AD, and age at death (combined  $N = 2005$ , Table S7). These combined 2005 AD samples are collectively referred to as the Mayo Clinic Brain Bank Expansion Cohort and are non-overlapping with the AMP-AD Mayo Clinic cohort described below. Taqman genotyping assays were performed using 10 ng of dried-down DNA and the QuantStudio 7 Flex system (Thermo Fisher Scientific, USA) for 7 variants (Table S14). Genotypes for *APOE*-rs429358 were previously collected using Taqman assays and queried from a database. One variant, *STRN4*-rs34805055, failed genotyping assay design and had no viable proxies, so Sanger sequencing was performed to validate all minor allele carrier samples. Sequencing was done on the ABI 3730 Genetic Analyzer instrument (Thermo Fisher Scientific, USA) following PCR amplification with the following primer pair: forward (5'- GGA AAGCAGCTCTGATAC) and reverse (5'- CGCATTCTG AGTCTCTG) (Integrated DNA Technologies, USA).

#### AMP-AD datasets

The Mayo RNAseq study [42], The Mount Sinai Brain Bank (MSBB) study [43] and The Religious Orders Study and Memory and Aging Project (ROSMAP) Study [44], were obtained from the AD-knowledge portal (<https://adknowledgeportal.synapse.org>). Available brain tissue RNAseq data, whole-genome genotypes, and neuropathological variables collected from these three studies were downloaded and used for fine-mapping of GWS loci and association analyses. The inclusion of these large, well-characterized, and harmonized datasets from the

AMP-AD consortia with complementary multi-omics datasets allow us to investigate and characterize GWS variants and loci across multiple regions of the brain for associations with AD-related phenotypes and brain gene expression levels in these cohorts. The RNA-seq data consists of seven datasets, two from Mayo Clinic (TCX and CER), four from MSBB (BM10, BM22, BM36, and BM44), and one from ROS-MAP (DLPFC) and previously underwent consensus reprocessing (AMP-AD, RNAseq Harmonization Study) [45]. Additional QC and diagnosis harmonization of these datasets based on neuropathological measures retrieved from individual metadata files are described in detail elsewhere [37].

In all cohorts, diagnosis was determined primarily by neuropathology made by experienced neuropathologists. The following criteria were used for diagnoses: AMP-AD Mayo dataset AD patients had a Braak stage  $\geq 4$  while nonADs had a Braak stage  $\leq 3$ . AMP-AD MSBB dataset AD patients had a Braak stage  $\geq 4$  and CERAD score  $\geq 2$  while nonADs had Braak stage  $\leq 3$  and CERAD score  $\leq 1$ . AMP-AD ROS-MAP dataset AD patients had a Braak stage  $\geq 4$  and CERAD score  $\leq 2$  while nonADs had Braak stage  $\leq 3$  and CERAD score  $\geq 3$ . Of note, MSBB and ROS-MAP used different CERAD definitions. In ROS-MAP, CERAD score (1-4) was based on semiquantitative estimates of neuritic plaque density in one or more neocortical regions following recommendations by the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) protocol [46]. In MSBB, a CERAD 1 = Normal, 2 = Definite AD, 3 = probable AD, and 4 = possible AD. In ROSMAP, 1 = Definite AD, 2 = probable AD, 3 = Possible AD, and 4 = No AD.

Whole genome sequencing (WGS) data from each AMP-AD cohort was processed separately using an automated pipeline at the New York Genome Center. 150 bp paired-end reads were aligned to GRCh37 human reference genome using Burrows-Wheeler Aligner [47] (BWA-MEM v0.7.08). After marking duplicates with Picard tools [48] (v1.83) and local read alignment around indels, base quality score recalibration (BQSR) was performed using Genome Analysis Toolkit [49] (GATK v3.4.0). Variant calling and joint genotyping were performed using GATK's HaplotypeCaller (GATK v3.4.0) and GenotypeGVCFs (GATK v3.5), respectively, to generate a multisample VCF file for each dataset. Variant quality was assessed using GATK's variant quality score recalibration (VQSR) tool. After obtaining multi-sample VQSR-ed VCFs for each individual study from the AD knowledge portal (see data sharing), genotypes were imported into PLINK [38, 39] (v1.9) for additional sample and variant QC using an in-house next-generation sequencing QC pipeline. Bi-allelic autosomal variants that pass VQSR FILTER, having a genotyping rate  $\geq 98\%$

and a minor allele frequency  $\geq 2\%$ , and a Bonferroni adjusted HWE  $p$ -value in controls  $>0.05$  were retained for downstream analysis. Variants within high variability regions of the genome that can lead to spurious associations were excluded. Samples with a call rate  $\geq 98\%$ , sex concordant with clinical information as evaluated using the inbreeding coefficient of the X-chromosome (males  $\geq 0.7$ , females  $\leq 0.3$ ) and a heterozygosity estimate within 3 standard deviations (SD) of mean were retained. Relatedness among samples within each cohort was evaluated using KING [50] robust and only one sample from each pair or family of samples related to the third degree (kinship estimate  $\geq 0.0442$ ) was retained. Population substructure was evaluated using Eigenstrat [51, 52] and outliers beyond 6 SD of the top 10 principal components were removed over five iterations while refitting PCs after each iteration. After performing sample and variant QC within each cohort, data from all three datasets was merged and relatedness and population substructure were re-evaluated to exclude related samples and population outliers across all three datasets. In summary, unrelated samples of relatively homogeneous non-Hispanic White ancestry that met the aforementioned sample and variant QC metrics were retained for downstream analyses.

#### Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset

The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD [53]. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease [54]. Inclusion and exclusion criteria, clinical and neuroimaging protocols, and other information about ADNI can be found at [www.adni-info.org](http://www.adni-info.org). Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). Demographic information, apolipoprotein E (APOE) and genome, Pre-processed [ $^{18}\text{F}$ ] Florbetapir PET scans, plasma total A $\beta$ 42, and plasma tau phosphorylated at the threonine 181 (p-tau) data, and clinical information are available and were downloaded from the ADNI data repository ([www.loni.usc.edu/ADNI/](http://www.loni.usc.edu/ADNI/)). [ $^{18}\text{F}$ ] Florbetapir PET scans were intensity-normalized using a whole cerebellum reference region to create SUVR images [55]. CSF biomarkers (Amyloid- $\beta$  1-42 peptide (A $\beta$ 1-42), total tau (t-tau), and p-tau) were generated by the validated and highly automated Roche Elecsys<sup>®</sup> electrochemiluminescence immunoassays and

the same reagent lot for each of these three biomarkers [56]. The ADNI participants were genotyped using several Illumina genotyping platforms. After quality control procedures for participants and SNPs, un-genotyped SNPs for non-Hispanic participants of European ancestry were imputed separately in each platform using Markov Chain Haplotyping with the Haplotype Reference Consortium data as a reference panel [57].

#### Statistical analysis

Power calculations were performed in R (v4.0.2) with the `genpwr` package. For a sample size of 441, and an  $\alpha = 5\text{E-}08$ , we have 80% power to detect effect sizes of 0.42 and 0.96 when the minor allele frequency (MAF) is 0.5 and 0.05, respectively. Principal component analysis with automatic outlier exclusion was performed using Eigenstrat [51, 52], no population outliers were identified. PLINK was utilized to perform PCA without outlier exclusion to examine samples in this cohort relative to 1000G superpopulations. PCA plots were generated using the `plot_ly()` package in R (v3.6) (Fig. S5).

PLINK (v2.00a2LM) was used to perform genome-wide association tests for variant dosage associations with each biochemical measure adjusting for age, sex and the first three population principal components (PCs). When specified, the *APOE*- $\epsilon$ 2 and - $\epsilon$ 4 alleles as well as *APOE* diplotypes 23, 24, 34, and 44 (33 was set as the reference) were included as covariates. There were no *APOE* 22 samples. Genomic inflation values ( $\lambda$ ) were calculated in R (v3.6.2) for each biochemical measure with and without adjustment of *APOE*- $\epsilon$ 2 and - $\epsilon$ 4 alleles (Fig. S6); there was no evidence for genomic inflation ( $0.97 < \lambda < 1.02$ ). QQ plots were generated in R (v3.6) with the `ggplot` package (Fig. S6). A genome-wide significance (GWS) threshold was calculated based on the number of independent variants in our dataset which also accounts for the inclusion of low frequency variants [58]. Quality control of the imputed variants to the HRC reference panel yielded a total of 6,726,078 variants. To calculate the number of independent variants, we used the '--indep-pairwise' flag in PLINK with the window size set to 50 kb, step size set to 5, and  $R^2$  threshold set to 0.8. After this pruning, we had 1,679,420 independent variants which we used to calculate our study GWS threshold of  $2.98 \times 10^{-8}$  ( $0.05/1,679,420$ ). To determine if GWS associations were independent from the effect of the *APOE*- $\epsilon$ 4 allele, conditional analyses were run in PLINK 2.0 implementing the '--condition' command in a linear regression model conditioning on the *APOE*- $\epsilon$ 4 tagging variant rs429358 and adjusting for age, sex, and PCs1-3. LD analysis was performed using PLINK ( $\pm 1\text{ Mb}$ ,  $D' \geq 0.8$  and  $r^2 \geq 0.2$ ). The estimated proportion of biochemical measure variance explained by the GWS index SNPs was based on the

$R^2$  calculated through linear regression models regressing appropriate index SNPs on each biochemical measure.

Variants were tested for association with AD neuropathology and other related measures in the AMP-AD and expanded MCBB datasets using multi-variable regression analysis. Braak stage and Thal phase were assessed with ordinal regression in R (v4.0.2), diagnosis with logistic regression in PLINK, and age at death with linear regression in PLINK (v2.00a2LM). Samples with age at death greater than 90 years in the MCBB were redacted to 90 to parallel protocols of the AMP-AD datasets. All models included sex as a covariate, and age at death, *APOE-ε2*, and *APOE-ε4* when appropriate or specified. Meta-analysis was performed in R (v4.0.2) with the *meta* function for both fixed and random effects models.

The AMP-AD datasets were used to assess of each locus with brain gene expression. Differential expression analysis between diagnosis (AD case or control) and normalized gene expression levels was performed using linear regression implemented in R (v3.5.2) adjusting for age at death, sex, RNA integrity number (RIN), and sequencing batch. eQTL analysis was performed by testing variant association with CQN gene expression levels in a linear mixed model using the *lme4* package in R (v3.5.2) adjusting for diagnosis, sex, age at death, RIN, tissue source, and the first three PCs, with the flow cell added as a random effects variable.

Variants were investigated in the ADNI dataset by performing a linear regression association analysis of variants with AD endophenotypes using additive genetic models adjusted for age and sex with or without *APOEε4* carrier status inclusion as an additional covariate.

Colocalization analyses for GWS loci with evidence of a significant QTL was performed. For loci with eQTL evidence from AMP-AD datasets, we used the *coloc*() R package [59, 60] with summary GWAS and QTL statistics as inputs  $\pm 500$  Kb from the index variant and linkage disequilibrium estimates from the 1000 Genomes EUR dataset. Single causal variant colocalization was investigated with the *coloc.abf* function while multiple causal variants were investigated with the *runsusie* and *susie.coloc* functions [61] in *coloc* which uses the Sum of Single Effects (SuSiE) regression framework [60]. The suggested decision rule of hypothesis 4 (H4) having a posterior probability (PP) value  $> 0.9$  was used to indicate colocalization. For loci with QTL evidence from GTEx, the ezQTL webserver tool made available through the NIH Division of Cancer Epidemiology and Genetics at the National Cancer Institute ([analysisstools.cancer.gov/ezqtl/#/home](https://analysisstools.cancer.gov/ezqtl/#/home)) [62] was utilized to investigate single causal variant colocalization with HyPrColoc [63] and multiple causal variant colocalization with eCAVIAR [64]. GWAS summary statistics were lifted from hg19

to hg38 using the UCSC chain file (hg19ToHg38, downloaded from [genome.ucsc.edu/](https://genome.ucsc.edu/) on 11/04/2022) and the R package *rtracklayer::liftOver*(). GTEx v8 summary data and 1000 Genomes EUR LD data were precompiled by ezQTL.

#### Pathway analysis

Gene set enrichment analysis was performed for each GWAS result with GSA-SNP2 software [65] against the MSigDb c5.all.v5.2 database [66, 67]. Options selected include European race, GRCh37(hg19) padding build, and pathway size window of 10-200. GSA-SNP2 results were matched with Gene Ontology (GO) [68, 69] term IDs using an in-house script. Significant GO terms and *p*-values were input into REViGO [70] to summarize significantly enriched pathways. REViGO settings were as follows: medium (0.7) allowed similarity, *Homo sapiens* (Gene Ontology Jan 2017) database, and SimRel semantic similarity measure. Summary bar charts were created in R (v3.6.2) with ggplot by taking reduced pathway groups from the REViGO outputs and the most significant *p*-value of that group for each biochemical measure.

#### Variant annotations

We queried existing data and results from multiple resources to further annotate key variants and investigate the implicated loci. These additional datasets represent either the largest, most comprehensive, or most applicable dataset available to characterize the loci of interest for their associations with AD-related phenotypes, regulatory potential or other human diseases and phenotypes. Associations of GWS variant dosage with sqrt (CAA) were performed previously [37] and results were queried for key variants identified in this study. Cell-specific differential gene expression analysis was queried from Mathys et al. 2019 [71] between AD pathology and no pathology samples in six cell types (excitatory neurons, inhibitory neurons, microglia, oligodendrocytes, astrocytes, and oligodendrocyte precursor cells), downloaded from the supplemental material (Table S2) on June 10, 2019, and limited to genes  $\pm 1$  Mb from the GWS variants in the Ensembl hg19 build (release 103) [72]. Only genes that passed study-level significance (FDR corrected *p*-value  $\leq 0.01$  and a fold change  $\geq 0.25$ ) in at least one cell type were included in our evaluation.

Summary statistics for the LOAD GWAS (Kunkle et al. 2019 [16]- NG00075, Lambert et al. 2013 [17]- NG00036) and CSF GWAS (Cruchaga et al. 2013 [73]- NG00049) were downloaded from NIAGADS. International Genomics of Alzheimer's Project (IGAP) is a large three-stage study based upon genome-wide association studies (GWAS) on individuals of European ancestry. In stage 1, IGAP used genotyped and imputed data on 11,480,632

single nucleotide polymorphisms (SNPs) to meta-analyze GWAS datasets consisting of 21,982 Alzheimer's disease cases and 41,944 cognitively normal controls from four consortia: The Alzheimer Disease Genetics Consortium (ADGC); The European Alzheimer's disease Initiative (EADI); The Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE); and The Genetic and Environmental Risk in AD Consortium Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES). In stage 2, 11,632 SNPs were genotyped and tested for association in an independent set of 8362 Alzheimer's disease cases and 10,483 controls. Meta-analysis of variants selected for analysis in stage 3A ( $n = 11,666$ ) or stage 3B ( $n = 30,511$ ) samples brought the final sample to 35,274 clinical and autopsy-documented Alzheimer's disease cases and 59,163 controls.

The 1000 genomes phase\_3 (GBR) dataset was queried for variants in LD ( $\pm 50\text{kb}$ ,  $r^2 \geq 0.8$ ,  $D' \geq 0.8$ ) through Ensembl and NCBI LDlink (<https://ldlink.nci.nih.gov/>) [74]. The Genotype-Tissue Expression (GTEx) Project v8 (<https://gtexportal.org/>) [75, 76] was queried for significant eQTLs and sQTLs between August and December 2020.

A graphical description of the datasets integrated in each analysis is outlined in Fig. S7.

## Results

### Genome-wide association study identifies seven novel loci associated with AD brain biochemical endophenotypes

We utilized a cohort of 441 autopsy-confirmed AD cases from the Mayo Clinic Brain Bank with genome-wide genotypes and temporal cortex (TCX) biochemical measures of AD-related protein endophenotypes including APOE, A $\beta$ 40, A $\beta$ 42, total tau, and p-Tau from soluble (TBS), membrane (TX), and insoluble (FA) tissue fractions. Demographics including neuropathology scores are outlined in Table S1. Quantitative brain biochemical measures were previously collected and transformed to approximate a normal distribution (Table S2, Fig. S1) [32]. To identify genetic associations with brain levels of AD-related proteins, genome-wide association studies were performed for each normalized biochemical fraction as well as the normalized ratio of A $\beta$ 40/42, adjusting for age, sex, and the first three population principal components (PCs) in the primary model. When specified, we also adjusted for APOE genotypes as follows: conditional analysis on the APOE- $\epsilon$ 4 tagging variant (rs429358) imputed dosages, including as covariates APOE diploypes, the APOE- $\epsilon$ 2 tagging variant (rs7412) dose, and/or APOE- $\epsilon$ 4 dose, as well as stratified by APOE diplotype. Altogether, we identified genome-wide significant (GWS,

$P < 2.98 \times 10^{-8}$ ) SNP-endophenotype associations at 8 unique loci: 6 unique loci for A $\beta$ 40, 3 for APOE, and 1 for the ratio of A $\beta$ 40/A $\beta$ 42 (Table 1, Figs. 2, 3 and 4, Table S3).

Seven of these loci involve novel intronic variants that have not been previously implicated in genetic association studies of AD or related endophenotypes: rs116580059 near *SCIN* for APOE in TX fraction (rs116580059-APOE<sub>TX</sub> (*SCIN*)), rs11845003-APOE<sub>TX</sub> (*NPAS3*), rs116726862-A $\beta$ 40<sub>TX</sub> (*SLC9A9*), rs148028977-A $\beta$ 40<sub>TX</sub> (*RFX7*), rs34805055-A $\beta$ 40<sub>TX</sub> (*STRN4*), rs77785770-A $\beta$ 40<sub>TX</sub> (*KCNN2*) and rs9890231-A $\beta$ 40<sub>TBS</sub> (*ITGB4*). Assessment of the index SNPs at each locus across all biochemical measures determined that each is nominally ( $P < 0.05$ ) associated with additional biochemical measures (Fig. 3). The estimated proportion of biochemical measure variance explained by the index SNPs based on the  $R^2$  of linear regression models ranged from 7 to 27%, with APOE- $\epsilon$ 4 (rs429358) alone explaining between 6.6 to 14% (Table S4).

More broadly, we detected 1813 variants with a  $P < 1 \times 10^{-5}$  ranging from 26 variants for total tau<sub>TBS</sub> to 341 for APOE<sub>TX</sub> (Table S3). While not reaching GWS, the most significant SNPs for the remaining traits include, A $\beta$ 42: TBS-rs147370282, TX-rs10219590, FA-rs461939; A $\beta$ 40/42: TBS-rs9890231, TX-rs483082; total tau: TBS-rs34678552, TX-rs76878089, FA-rs1634993; and p-Tau: TBS-rs2294557, TX-rs10987782, and FA-rs11651012 (Fig. S2, Table S3). Interestingly we found associations that approach GWS for total tau<sub>TBS</sub> ( $P = 8.03\text{E-}07$ ) and total tau<sub>TX</sub> ( $P = 9.48\text{E-}06$ ) with rs117691004 which is located in the *PRKN* gene known to play a role in Parkinson's Disease [77] (Table S3).

### Multiple variants at the APOE locus associate with brain biochemical measures of AD-related proteins

Presence of the APOE- $\epsilon$ 4 allele has previously been reported to associate with biochemical measure levels in this dataset [32], however, using the GWAS data we can explore the effects of APOE $\epsilon$ 4 dose and additional genetic variation at this locus. We found a total of 30 unique GWS variants at or proximal to the APOE gene with at least one associated with six of the biochemical measures (Table S3). The known exonic AD risk APOE- $\epsilon$ 4 tagging variant (rs429358) was the most significant SNP for four of the traits: A $\beta$ 40<sub>TX</sub>, A $\beta$ 40<sub>FA</sub>, A $\beta$ 40/42<sub>FA</sub>, APOE<sub>FA</sub>. The proximal *NECTIN2* intronic variant (rs283815) was the top SNP for APOE<sub>TBS</sub> (Figs. 4 and 5), and an intergenic SNP between APOE and *APOC1* (rs483082) was the most significant for A $\beta$ 40/42<sub>TX</sub>. These two SNPs are in linkage disequilibrium (LD) with APOE- $\epsilon$ 4 in our dataset (rs283815:  $r^2 = 0.73$ ,  $D' = 0.87$ ; rs483082:  $r^2 = 0.92$ ,  $D' = 0.99$ ) which is likewise associated with the same biochemical traits (Fig. 3, Table S3). In

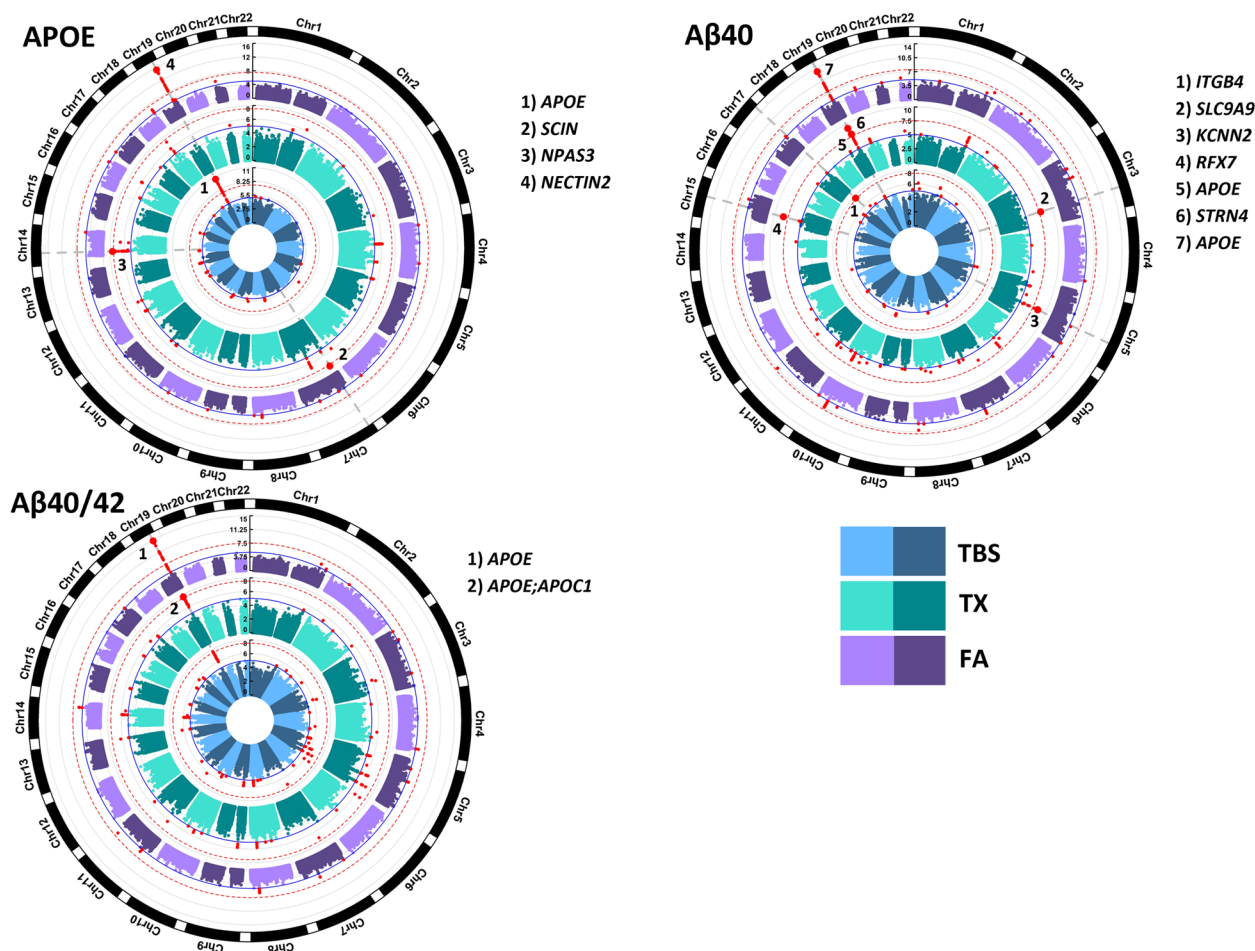
**Table 1** Description of genome wide significant SNPs

Protein	Buffer	Transf	SNP	Chr	Locus	Position Annotation	Minor Allele	N	MAF	MAF Bin	HWE-P	Additive Model		Conditional APOE-ε4 model		CADD Score	Regulome Rank			
												Beta	(CI 95%)	Beta	(CI 95%)					
Novel	Aβ40	TBS	In	rs9890231	17	ITGB4	Intron ITGB4	G	439	0.04	LF	1.00	-1.88	(-2.52--1.23)	2.09E-08	-1.84	(-2.47--1.21)	2.06E-08	3.371	5
	Aβ40	TX	In	rs77785770	5	KCNM2	Intron KCNM2	G	441	0.05	LF	0.60	0.79	(0.53--1.05)	3.12E-09	0.72	(0.47--0.97)	2.15E-08	0.977	5
	Aβ40	TX	In	rs148028977	15	RFX7	Intron RFX7	T	441	0.02	LF	1.00	1.24	(0.83--1.64)	4.47E-09	1.1	(0.71--1.49)	6.98E-08	12.64	5
	Aβ40	TX	In	rs116726862	3	SLC9A9	Intron SLC9A9	T	441	0.03	LF	1.00	0.95	(0.63--1.27)	1.51E-08	0.57	(0.38--0.77)	2.78E-08	3.507	3a
	Aβ40	TX	In	rs34805055	19	STRN4	Intron STRN4	T	441	0.08	MF	1.00	0.61	(0.41--0.82)	1.15E-08	0.85	(0.54--1.16)	1.53E-07	6.379	4
	APOE	TX	sqrt	rs116580059	7	SCIN	Intron SCIN	C	439	0.03	LF	1.00	3.34	(2.20--4.48)	1.82E-08	3.38	(2.24--4.51)	1.08E-08	1.254	3a
	APOE	TX	sqrt	rs11845003	14	NPAS3	Intron NPAS3	T	439	0.02	LF	1.00	4.05	(2.66--5.44)	2.05E-08	4.04	(2.65--5.42)	1.99E-08	7.418	4
Known	Aβ40	TX	In	rs429358	19	APOE	Exon APOE	C	441	0.4	HF	0.07	0.37	(0.25--0.48)	6.88E-10	-	-	-	0.007	4
	Aβ40	FA	In	rs429358	19	APOE	Exon APOE	C	441	0.4	HF	0.07	0.91	(0.68--1.15)	6.75E-14	-	-	-	0.007	4
	Aβ40/42	TX	In	rs483082	19	APOE/APOC1	Intergenic	T	441	0.42	HF	0.49	0.33	(0.22--0.44)	2.11E-08	0.29	(-0.08--0.66)	0.12	10.25	4
	Aβ40/42	FA	In	rs429358	19	APOE	Exon APOE	C	441	0.4	HF	0.07	0.94	(0.72--1.16)	1.48E-15	-	-	-	0.007	4
	APOE	FA	In	rs429358	19	APOE	Exon APOE	C	441	0.4	HF	0.07	0.49	(0.38--0.60)	4.53E-16	-	-	-	0.007	4
	APOE	TBS	sqrt	rs283815	19	APOE/NECTIN2	Intron NECTIN2	G	441	0.42	HF	0.02	-2.41	(-3.10--1.72)	2.60E-11	-1.51	(-2.74--0.27)	1.75E-02	1.827	7

Descriptions of top genome-wide significant (GWS) SNPs associated with each biochemical measure in N = 441 AD cases. Seven are novel and three are known. Models include an additive model adjusted for age, sex, and PC1-3, and a conditional model conditioning on APOE-ε4 (rs429358) adjusted for age, sex, and PC1-3

Transf: Transformation, N Number, MAF Minor Allele Frequency, HWE-P Hardy Weinberg Equilibrium P-value, CI Confidence Interval, P P-value





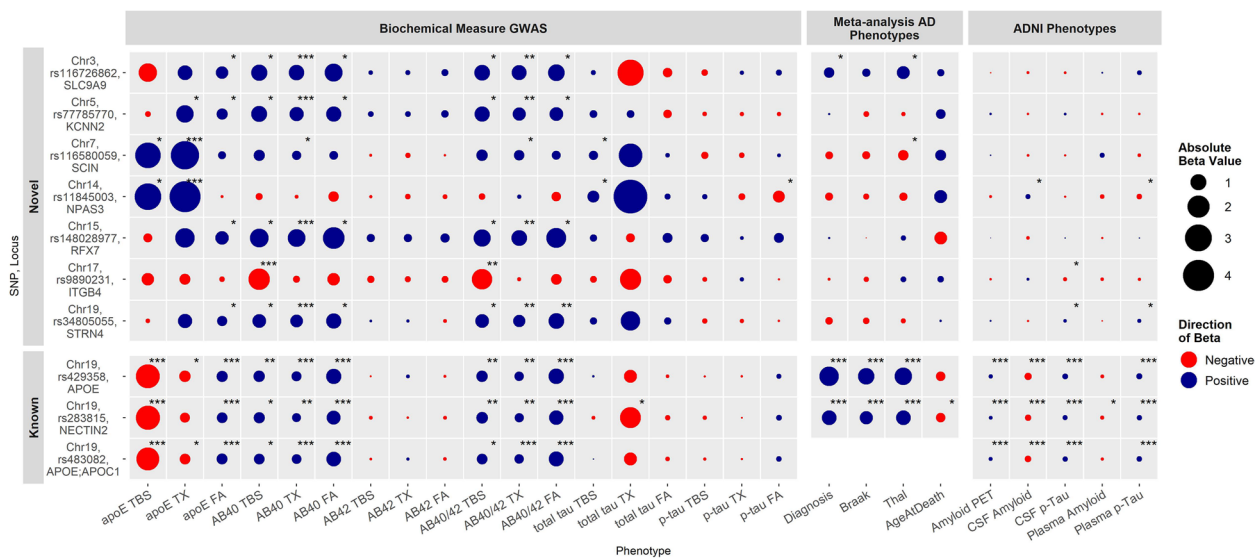
**Fig. 2** Circular Manhattan Plots of brain AD-related protein GWAS. Circular Manhattan plots for each protein measured in three biochemical fractions. Plots for proteins with SNPs that reach genome-wide significance (GWS) are shown. Red dotted line marks GWS threshold of  $p$ -value =  $2.98E-08$ , solid blue line marks  $p$ -value =  $1E-05$ . Top SNPs at GWS loci have dots increased in size and labeled with the closest gene name. SNPs with a  $p$ -value  $< 1E-05$  are colored red. Radial axes measure  $-\log_{10}(P$ -value). Inner most blue circle is the soluble TBS fraction, middle green circle is the membrane TX fraction and outer most purple circle is the insoluble FA fraction

an *APOE-ε4* conditional analysis, only rs283815 remains nominally significant (Table 1). We further examined the rs283815- $APOE_{TBS}$  association after adjusting for *APOE-ε2* and *APOE-ε4* dose, for *APOE* diplotypes, and in the *APOE-33* only sample subset ( $N = 141$ ,  $\beta = -3.79$ ,  $P = 2.16E-03$ ) finding that this association remained nominally significant (Table S5). Taken together, this suggests that the effects of rs483082 and *APOE-ε4* on  $A\beta_{40/42_{TX}}$  levels likely represent the same signal. However, because the rs283815- $APOE_{TBS}$  association is present even in the *APOE-33* only sample subset, as well as after adjusting for combinations of *APOE* diplotypes, our results suggest that more than one genetic variant at the *APOE* locus beyond that of the *APOE-ε4* signal likely contribute to soluble APOE levels in the TCX. The rs283815 variant has previously been implicated in AD risk in males [78] and imaging of cerebral amyloid deposition [79], although it did not survive adjustment for

*APOE-ε4* in the latter. These variants at the *APOE* region (rs429358, rs283815, and rs483082) represent the most significant associations across all fractions of  $A\beta_{40}$ ,  $A\beta_{40/42}$ , and APOE, but no fractions of  $A\beta_{42}$ , total tau, or p-Tau ( $P > 0.031$ ) (Table S6); indicating that they likely impact disease risk through effects on APOE and  $A\beta_{40}$ , but not tau. Furthermore, the direction of association for APOE fractions indicates a shift in biochemical state with minor allele carriers having lower soluble APOE ( $APOE_{TBS}$ ) and higher insoluble APOE ( $APOE_{FA}$ ) (Fig. 3), suggesting a role in promoting aggregation of APOE rather than overall levels.

#### AD brain biochemical endophenotype GWS variants also associate with disease risk, age at death, and AD-related neuropathology and biomarkers

To further characterize the GWS variants with respect to other AD-related phenotypes, we evaluated their



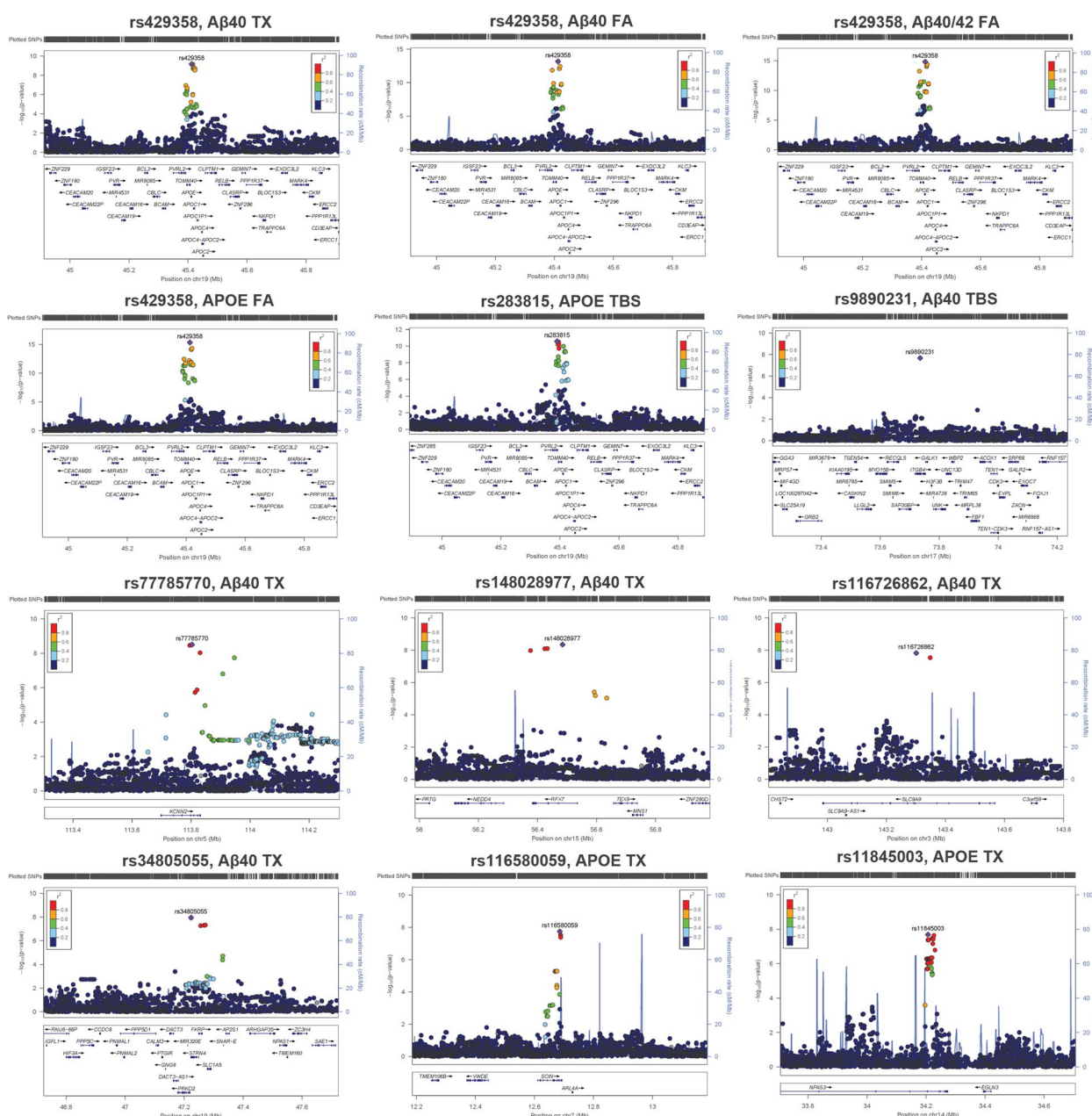
**Fig. 3** Association of Genome-wide significant SNPs across all Biochemical Measures, Meta-Analysis, and ADNI dataset with AD-related Phenotypes. Associations of novel (top) and known (bottom) GWAS Index SNPs across all 18 biochemical measures in the Mayo Clinic Brain Bank cohort ( $n = 441$ ) and results from meta-analysis with AD-related phenotypes in up to 4 cohorts ( $n = 3707$ ). Meta-analysis was conducted using data from four independent autopsy datasets, namely the AMP-AD Mayo Clinic ( $n = 344$ ), Mount Sinai Brain Bank (MSBB,  $n = 267$ ), Rush (ROSMAP,  $n = 1091$ ) and the Mayo expanded brain bank dataset ( $n = 2005$ ). Mayo AMP-AD and expanded brain bank datasets were non-overlapping, and latter also included the 441 AD donors from the brain biochemical measures GWAS. Meta-analysis results are fixed effects models adjusted for sex and age at death when appropriate. Rs483082 was not significant after conditioning on rs429358 (*APOE-ε4*) and so was not carried forward for meta-analysis. Proxy SNPs genotypes were used for rs148028977 and rs116580059 in the Mayo expanded dataset. Note, Thal measures were only available from the expanded Mayo dataset and the AMP-AD Mayo dataset. Rs34805055 was not genotyped in the expanded Mayo dataset, therefore meta-analysis excluded this cohort for this SNP. ADNI associations include amyloid PET ( $N = 784$ ), CSF amyloid ( $N = 1154$ ), CSF p-Tau ( $N = 1151$ ), plasma amyloid ( $N = 262$ ), and plasma p-Tau ( $N = 787$ ) adjusted for age and sex. Dot color indicates direction of beta value (blue = positive, red = negative), size of dot indicates absolute beta value. Associations with a  $p$ -value  $\leq 2.98E-8$  indicated by (\*\*\*),  $1E-05 \leq p$ -value  $< 2.98E-8$  indicated by (\*\*), and  $0.05 \leq p$ -value  $< 1E-05$  indicated by (\*)

association with AD risk, AD-related neuropathological variables (Braak stage, Thal phase, cerebral amyloid angiopathy = CAA), AD-related biomarkers (amyloid [<sup>18</sup>F] Flortetapir PET scans, plasma total Aβ42, and phosphorylated tau (pTau), CSF Aβ42, total tau (t-tau) and pTau) and age at death in additional samples.

We expanded the cohort size and validated the GWAS genotypes by genotyping these variants or their proxies in the GWAS study samples ( $N = 441$ ) and additional Mayo Clinic Brain Bank (MCBB) participants ( $N = 1564$ ) using TaqMan assays or Sanger sequencing (Table S7). We refer to these 2005 participants as the Mayo Clinic Expanded Cohort. A high level of concordance, > 98%, was observed between the array-based genotyped or imputed alleles and those collected by TaqMan and sequencing (Table S8). Association results using the TaqMan genotypes in place of the array-based genotypes showed only minor variations in effect size and significance, demonstrating consistency of the results (Table S5). Genotypes for the index GWS variants were also extracted from three independent whole genome sequence (WGS) datasets

available from the Accelerating Medicines Partnership AD (AMP-AD) study through the AD knowledge portal ([www.synapse.org](http://www.synapse.org)) which includes the Mayo Clinic RNAseq (Mayo,  $n = 344$ ) [42], Mount Sinai Brain Bank (MSBB,  $n = 267$ ) [43], and Rush Religious Orders Study and Memory and Aging Project (ROSMAP,  $n = 1091$ ) [44] studies (Table S7). We note that the AMP-AD Mayo and Mayo Clinic Expanded Cohort are non-overlapping. Meta-analyses for available genotypes and common AD-related phenotypes were conducted across these 4 independent datasets for each SNP using fixed and random effects models (Fig. 3, Fig. S3, Table S9).

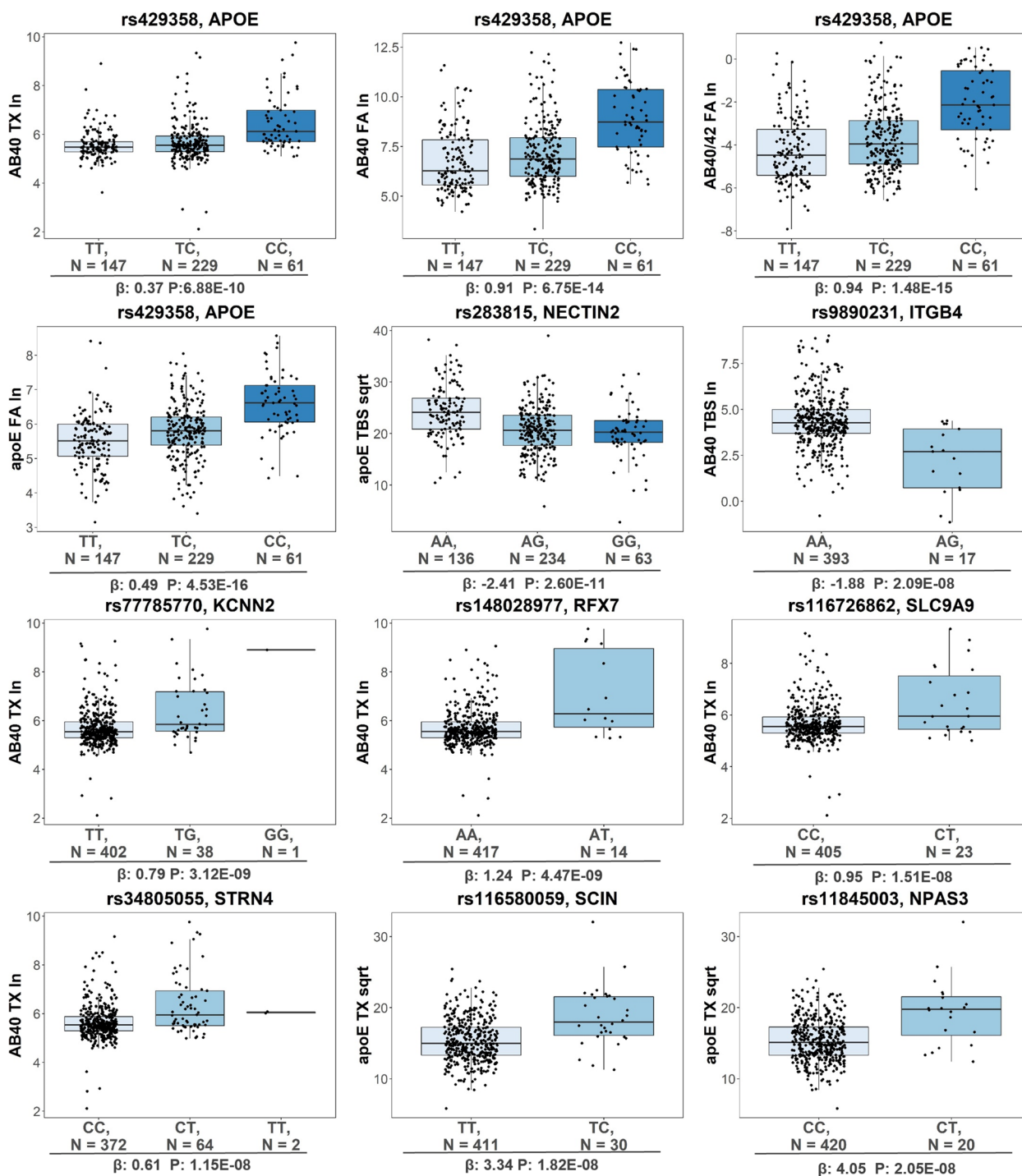
In addition, we investigated the association of GWS variants with other AD-related endophenotypes that were not available in the aforementioned datasets. We queried a previous GWAS of CAA [37] to determine the association of the GWS variants in the current study with this vascular AD pathology. We also evaluated the association of the GWS variants with AD-related biomarkers including amyloid PET measures, CSF and plasma Aβ and tau in the ADNI dataset.



**Fig. 4** Locus Zoom Plots. Locus Zoom plots (locuszoom.org) of GWS SNPs showing associations +/- 500 kb from variant of interest (labeled). Right Y-axis shows the *p*-value, left Y-axis shows rate of recombination, and X- axis shows position on chromosome and nearby gene positions. Each plot point represents a variant in the dataset color coded by (*r*<sup>2</sup>) value

As expected, the *APOE-ε4* variant (rs429358) and the proximal *NECTIN2* variant (rs283815) significantly associated with AD risk, Thal phase, Braak stage, amyloid PET, CSF amyloid and p-Tau, plasma p-Tau and as reported previously for *APOE-ε4*, also CAA [32]. Rs283815 was also associated with age at death and plasma amyloid, although the rs283815 associations were no longer significant after adjustment for

*APOE-ε2* or *APOE-ε4* (Fig. 3, Tables S9 and S10). Of the novel variants, we found rs116726862 (*SLC9A9* intron, increased Aβ<sub>40-TX</sub>) was associated with higher Thal phase (*P* = 7.50E-03), increased CAA (*P* = 4.70E-02), and a trend for increased AD risk (*P* = 0.07), higher Braak stage (*P* = 0.07), and higher plasma p-Tau levels (*P* = 0.08). The *SCIN* intronic SNP rs116580059 (increased APOE<sub>TX</sub>) was associated with lower Thal



**Fig. 5** Box Plots of Genome-wide Significant SNP Genotypes. Box plots of hard-call genotypes for each genome-wide significant SNP from each biochemical measure GWAS. Each dot represents an individual sample, N=Number. Variant rs number and gene closest to GWS SNP listed at the top, beta and  $p$ -value listed at the bottom, biochemical measures on the y-axis. Genotype: Light blue- Homozygous major, Medium blue- Heterozygotes, Dark blue- Homozygous minors

phase ( $P = 0.015$ ) and a trend ( $P = 0.06$ ) for lower Braak stage. We found rs11845003 (*NPAS3* intron, increased  $APOE_{TX}$ ) associated with increased CSF amyloid levels

( $P = 0.02$ ), decreased plasma p-Tau levels ( $P = 0.04$ ), and trends for decreased amyloid PET ( $P = 0.09$ ). Rs9890231 (intron *ITGB4*, decreased  $A\beta_{40_{TBS}}$ )

associated with decreased CSF p-Tau ( $P = 0.01$ ) and a trend for lower amyloid PET ( $P = 0.07$ ). The intronic *STRN4* variant rs34805055 (increased  $A\beta_{40_{TX}}$ ) associated with increased CSF p-Tau ( $P = 0.03$ ) and plasma tau levels ( $P = 0.04$ ) (Fig. 3, Tables S9 and S10).

In all cases, we find these additional AD-related phenotype associations are entirely consistent with what we would expect based on prior knowledge. Specifically, higher amyloid PET, CSF pTau, plasma pTau, lower CSF and plasma  $A\beta$  have congruent associations with the AD risk variant at the *APOE* locus and other variants that behave like *APOE* with respect to brain biochemistry associations (i.e. *SLC9A9*, *STRN4* variants). In contrast, variants that have a pattern of brain biochemistry associations opposite to that of *APOE* also have opposite direction of associations for the other AD-phenotypes as well (i.e. *NPAS3* and *ITGB4* variants). Although not all associations would survive Bonferroni correction, these biologically congruent associations facilitate identification of potential molecular mechanisms connecting genetic variants, biochemical measures, and other AD phenotypes. Taken together, these results validate the array-based genotype calls by an independent assay. Importantly, they also implicate at least two of the novel variants more broadly in AD risk and neuropathology (variants near *SLC9A9* and *SCIN*), and four with additional AD-related endophenotypes (near *SLC9A9*, *NPAS3*, *ITGB4*, and *STRN4*) in a direction that is consistent with the brain biochemical findings and known associations with the well-established *APOE* risk variant.

#### Brain transcriptome analyses implicate expression dysregulation at some of the novel AD brain biochemical endophenotype loci

We hypothesized that some of the variants might function through their influence on the expression or splicing of nearby genes. We performed *cis*-expression quantitative trait locus (*cis*-eQTL) analysis (SNP  $\pm 1$  Mb) in three AMP-AD transcriptome datasets collected from seven brain regions of AD cases and controls available through the AD knowledge portal and also queried independent results from the GTEx portal ([www.gtexportal.org/home/](http://www.gtexportal.org/home/), queried 08/2020) [75, 76]. We found that rs34805055- $A\beta_{40_{TX}}$  (intron *STRN4*) was significantly associated with the downregulation of *PRKD2* gene expression in the Mayo Clinic TCX dataset ( $\beta = -0.29$ ,  $q$ -value =  $2.6E-03$ ) and *RN7SL364P* in the ROS-MAP dorsolateral prefrontal cortex (DLPFC) dataset ( $\beta = -0.38$ ,  $q$ -value =  $3.0E-02$ ). The *PRKD2* gene is located approximately 3 kb downstream of rs34805055 while *RN7SL364P* is a pseudogene located within an intron of *PRKD2*. In the GTEx dataset, *PRKD2*

expression and splicing QTLs were found for this variant in healthy brain cortex tissue (Normalized Effect Size (NES) =  $-0.46$ ,  $P = 4.1E-05$ ), and other tissue types. The rs283815- $APOE_{TBS}$  (*NECTIN2* intron) variant significantly associated with *TOMM40* splicing in healthy cerebellum tissue (NES =  $-0.94$ ,  $P = 3.80E-16$ ), whilst rs9890231- $A\beta_{40_{TBS}}$  (intron *ITGB4*) associated with *ITGB4* splicing in several tissues including healthy tibial nerve tissue (NES =  $-1$ ,  $P = 6.50E-09$ ). Altogether, these results indicate that the rs34805055- $A\beta_{40_{TX}}$  variant may influence gene regulation and splicing of the *PRKD2* gene rather than the index gene *STRN4*, while rs9890231- $A\beta_{40_{TBS}}$  and rs283815- $APOE_{TBS}$  may influence splicing of *ITGB4* and *TOMM40*, respectively. The remaining novel index variants do not appear to influence gene regulation in the CNS at the bulk tissue level.

We next investigated the colocalization of the significant GWAS and QTL signals at each locus by testing for colocalization under single and multiple causal variant assumptions. We found that none of the traits (GWAS and QTL pairs) showed evidence of colocalization under the assumption that a locus has a single causal variant with both traits. Relaxing this assumption to allow for multiple causal variants at a locus to associate with a trait, we found the rs34805055- $A\beta_{40_{TX}}$  (*STRN4*) trait and Mayo TCX *PRKD2* eQTL showed evidence of colocalization (H4 posterior probability = 0.999). It should be noted, however, there was only one credible set identified in this analysis with rs34805055 and rs62134781 being the top signals in the GWAS and eQTL datasets, respectively. A single credible set suggests a single causal variant at the locus for both traits, however, our original analysis under this assumption did not show evidence of colocalization and these two variants are not in LD ( $R^2 = 0.1347$ ,  $D' = 1$ ). No other traits showed colocalization with multiple causal variants.

We also examined each implicated locus (variant  $\pm 1$  Mb, hg19) to determine if there were differentially expressed (DE) genes between AD cases and controls. We used the AMP-AD RNAseq datasets [42–44] to assess 267 expressed genes. We found that while all loci harbored DE genes in at least 1 brain region, 3 index genes and 1 gene implicated through the above QTL analysis were consistently DE in 2 or more brain regions. *KCNN2* was downregulated in AD for 3 datasets while *RFX7*, *SLC9A9*, and *PRKD2* were upregulated in AD for 2 (Table S11). Bulk tissue profiling captures changes across multiple cell types and may miss cell-specific molecular changes so we also queried results from a published single-cell RNAseq (scRNA-seq) dataset [71]. We found 67 genes at the implicated loci that were DE in at least one cell type between samples with AD pathology and those with no AD pathology, 90%

of which were dysregulated in neurons (Table S12). Of the index genes, *ITGB4* was upregulated in astrocytes, *NECTIN2* was downregulated in neurons, while *APOE* was upregulated in neurons and microglia, and downregulated in astrocytes in AD. Interestingly, although we focus on the AD pathology vs no pathology analyses, Mathys et al. [71] also reported DE genes between early and late AD pathology in which we see upregulation of *PRKD2* in neurons (fold change = 0.51) in late AD. Collectively these results implicate dysregulation of *PRKD2* and *ITGB4* at the novel loci identified by this study and *APOE*, *NECTIN2* and *TOMM40* at the established Chr19q13 AD locus (Fig. S4). Larger datasets with genetic and single-cell data will be needed to further explore whether these variants influence cell-specific gene expression changes.

#### Genes at the AD brain biochemical endophenotype GWAS loci are implicated in neuronal health and disease

We hypothesized that variants either in LD with the GWS index variants or others at the AD brain biochemical endophenotype GWAS loci may also associate with function or disease(s) of the central nervous system (CNS).

First, we identified variants in LD with the GWS index variants by querying the AMP-AD WGS Mayo TCX and CER datasets as well as the 1000 Genomes GBR dataset. We then searched these LD variants in the GTEx database as well as the GWAS catalog for significant associations beyond those already reported for the index variants (queried 09/28/2022). Eight of the nine GWS index variants had variants in LD with them. All but one of these were in noncoding regions, the exception being rs157581 in LD with rs283815, a synonymous missense variant in *TOMM40* (Table S9). At this same locus, we found multiple additional variants in LD with rs283815 at *NECTIN2*-*APOE*<sub>TBS</sub> locus with previous associations to AD, CAA, cognition, and AD biomarkers including CSF levels of A $\beta$ 42 and tau [78–84]. Interestingly, we also identified a complete proxy of the index variant rs116726862 at *SLC9A9*-A $\beta$ 40<sub>TX</sub> locus, rs115134872 ( $D' = 1$ ,  $r^2 = 1$ ), previously associated with survival in amyotrophic lateral sclerosis [85].

Next, we investigated all variants present in the index genes through a gene search in the GWAS catalog (<https://www.ebi.ac.uk/gwas/>, queried 06/07/2021) [86]. Five of the seven novel index genes have variants that associate ( $P < 1E-05$ ) with AD-related phenotypes (Table S9). These include the *SLC9A9* locus associated with brain A $\beta$ 40<sub>TX</sub> (i.e., *SLC9A9*-A $\beta$ 40<sub>TX</sub> locus) in this study and working memory [87], response to cholinesterase inhibitors in AD [88] and epistatic interactions with tau measurements [89]. *SLC9A9*-A $\beta$ 40<sub>TX</sub> locus was also

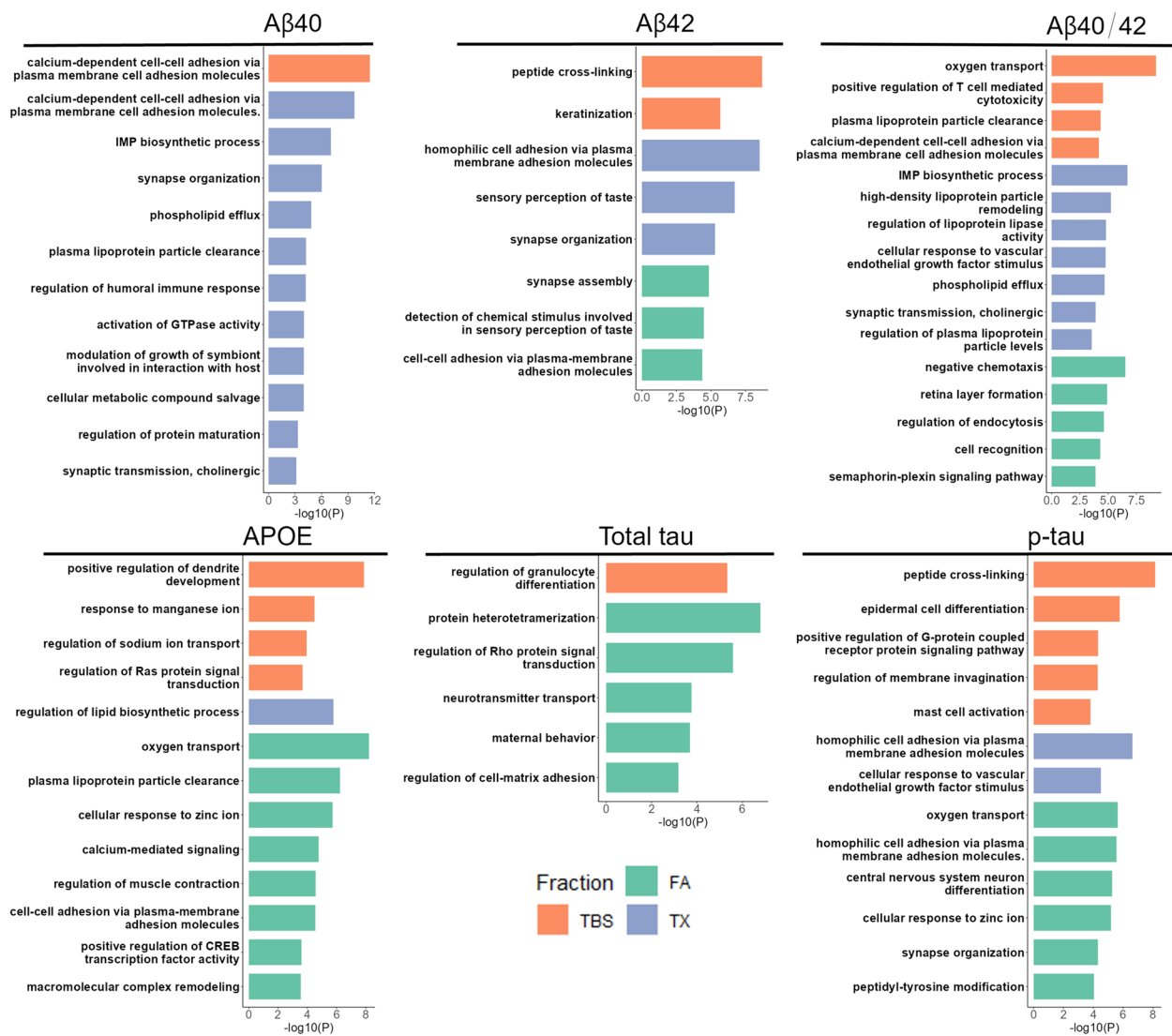
implicated in the neuropsychiatric disorders of autism [90, 91] and attention deficit hyperactivity disorder [92].

Another brain biochemical endophenotype GWAS locus with AD and other neuropsychiatric disease-related associations was *NPAS3*-*APOE*<sub>TX</sub> which is also associated with neuritic and diffuse plaque measurements [93], CSF levels of soluble TREM2 [94], epistatic interactions with tau measurements [89], schizophrenia [95, 96] and bipolar [95, 97, 98] disorder. The *KCNN2*-A $\beta$ 40<sub>TX</sub> locus was also associated with age of onset for AD [99], epistatic interactions with amyloid [89], schizophrenia [95, 96], bipolar [95, 97, 98] disorder and hippocampal sclerosis [100]. Additionally, *RFX7*-A $\beta$ 40<sub>TX</sub> and *SCIN*-*APOE*<sub>TX</sub> loci have associations with regional brain volume [101] and inflammation markers [102], respectively.

This convergent evidence supports important roles for most of the AD brain biochemical endophenotype GWAS genes and loci in neuronal health and disease.

#### Gene set enrichment analysis identifies shared and distinct pathways among GWAS genes for different AD brain biochemical measures

To identify pathways that are enriched for each AD brain biochemical fraction GWAS we performed gene set enrichment analysis using GSA-SNP2 [65] with Gene Ontology (GO) terms [65]. We identified both shared and distinct significantly enriched pathway terms for each biochemical measure GWAS (Fig. 6). Shared biological pathway terms have known roles in AD such as synapse organization, cell-to-cell adhesion, and immune-related processes. Distinct enrichment for different biochemical fractions (TBS, TX, FA) was related to known functions for each protein, indicating that discrete molecular mechanisms likely influence specific biochemical states of each AD-related protein. For example, *APOE* is known to function in lipid metabolism and we found enrichment for lipoprotein clearance pathways. Notably, for soluble A $\beta$ 42<sub>TBS</sub> and p-Tau<sub>TBS</sub> we found enrichment in peptide cross-linking pathways indicating a genetic influence on systems that may play a role in the transition from soluble to aggregated forms of these proteins. Further, we see the enrichment of synapse organization and central nervous system neuron differentiation pathways for p-Tau<sub>FA</sub>. This is consistent with the well-established knowledge that NFTs comprising hyper-phosphorylated tau correlate with neuronal loss and severity in AD [103]. Interestingly, we found distinct enrichment in endocytosis regulation pathways for A $\beta$ 40/42<sub>FA</sub> and sensory perception of taste pathways for A $\beta$ 42<sub>TX</sub> and A $\beta$ 42<sub>FA</sub>. These results implicate variants proximal to genes involved in both known and novel pathways that may play a role in AD pathogenesis through impacts on specific or multiple AD-related proteins and their distinct biochemical states in the brain.



**Fig. 6** Gene set enrichment analysis in all biochemical measures. Gene set enrichment was performed via GSA-SNP2 with the MiSigDB (c5.all.v5.2). Significantly enriched pathways ( $q$ -value  $< 0.05$ ) were matched with Gene Ontology (GO) IDs and input in REViGO for reduction of redundant pathways and summarization. Significant pathway groups from REViGO and the highest  $p$ -value of that group are plotted in bar plots for each protein. X-axis =  $-\log_{10}(P)$ -value, Orange = soluble (TBS) fraction, Blue = membrane (TX) fraction, Green = insoluble (FA) fraction

### Some AD risk GWAS variants also influence distinct biochemical measures

Finally, we wanted to determine if there are shared genetic risk factors for AD risk and specific brain AD-related biochemical measures. We queried large-scale AD risk GWAS [16, 17] for the novel variants identified in this study but did not find a significant association ( $P > 0.05$ ) for those outside the *APOE* locus. This would suggest that while the GWS variants may influence these brain biochemical endophenotypes, they do not have a statistically significant impact on the more heterogeneous phenotype of AD risk. We also investigated 28

GWS late-onset AD (LOAD) established risk variants for association with each biochemical measure [16]. Of the 26 variants present in our dataset, only associations with the *APOE* $\epsilon 4$  tagging variant survived Bonferroni correction for 26 tests ( $P < 1.93E-03$ ), however; 9 others were nominally significant ( $P < 0.05$ ) for at least one biochemical fraction with six having directions of effect consistent with AD risk, based on prior knowledge, and the direction of effect for the *APOE* $\epsilon 4$  tagging variant (Table S13). These include *APOE* $\epsilon 4$ -rs429358, discussed previously, and the risk allele for rs9331896 (C) at the *CLU* locus which associates with higher brain  $APOE_{TBS}$ ,

APOE<sub>TX</sub> and total tau<sub>TBS</sub>, and lower p-Tau<sub>TX</sub>. The risk allele for rs73223431 (T) at the *PTK2B* locus associates with increased p-Tau<sub>TX</sub> and p-Tau<sub>FA</sub>, but not APOE or Aβ measures. The protective allele of rs1080826 (A) at the *EPHA1* locus associated with lower levels of Aβ42<sub>TX</sub>, Aβ42<sub>FA</sub> and total tau<sub>FA</sub> (Table S13). While associations outside the *APOE* locus would not survive Bonferroni correction, these results using unique biochemical measures from brain tissue can still indicate the underlying pathological mechanisms by which these established AD-risk variants might influence disease.

## Discussion

Genetic, model system, and neuropathology studies have clearly established Aβ, tau and APOE as disease hallmarks, biomarkers and therapeutic targets in AD and other neurodegenerative diseases [104–106]. Even though there are several strategies targeting these molecules for therapeutic benefit in AD, critical knowledge gaps hinder progress. All three molecules undergo complex processing and exist in heterogeneous biochemical states in the human brain. Discovering genetic and other factors contributing to this molecular complexity and biochemical heterogeneity can yield novel therapeutic avenues. Further, the relationship of the various biochemical states of these molecules and their genetic determinants with other AD-related outcomes can help clarify the beneficial vs. detrimental mechanism of action for targeted therapies. Finally, a comprehensive genetic map of the various biochemical states of key AD proteins can help pave the way for personalized medicine targeting specific perturbed pathways in an individualized fashion.

In this study we sought to identify the genetic determinants contributing to variability in the brain biochemical states of key AD-related proteins. We report the identification of eight independent GWS loci that associate with brain levels of five hallmark AD-related proteins isolated from three tissue fractions. Seven loci are novel and associated with Aβ40, APOE, and Aβ40/Aβ42 biochemical levels. Aside from these novel loci, we also observe significant associations across the biochemical measures for the *APOE-ε4* tagging variant (rs429358). Notably, we detected a signal within *NECTIN2* nearby the *APOE* locus that cannot be entirely explained by *APOE-ε4*. These results demonstrate the contribution of genetic factors besides *APOE-ε4* at the *APOE* locus to the variability of brain biochemical states of AD-related proteins.

Our study also provides insights into the pathological mechanisms through which the *APOE* and novel loci may act through to influence the brain biochemical measures. Although the *APOE-ε4* variant (rs429358) and the

*NECTIN2* variant (rs283815) were associated with measures of APOE, Aβ40, and Aβ40/42 ratio, these variants did not show significant associations with Aβ42, total tau, and p-Tau, consistent with previous findings in this dataset [32]. These results suggest that the *APOE* locus may influence mechanisms associated with levels of APOE and Aβ40 proteins rather than Aβ42 or tau in the brain. This contrasts with some previous studies that suggest the *APOE* locus does affect tau levels independent of Aβ [73, 107, 108], others showing brain region specific effects of *APOE-ε4* [109–111], while another study showed an interaction effect [112]. In our current study, while we cannot rule out the possibility there may be associations at the *APOE* locus that affect Aβ42 or tau brain biochemical levels, based on the 95% confidence intervals of these associations (Table S6), these effects would likely be small. Nonetheless, it will be important for future studies to investigate this further in additional datasets, particularly those with tissue fraction specific data such as ours, case-control cohorts, and longitudinally.

Importantly, this study identified additional, biologically congruent associations for five of the seven novel loci including *SLC9A9*, *SCIN*, *NPAS3*, *ITGB4*, and *STRN4*. Based on the associations of the *APOE* locus, we would expect increased Aβ40<sub>TX</sub> (*SLC9A9* and *STRN4*) to correlate with increased AD risk while increases in APOE<sub>TX</sub> (*SCIN* and *NPAS3*) and Aβ40<sub>TBS</sub> (*ITGB4*) would correlate with decreased AD risk. Remarkably, we see these relationships recapitulated in the independent datasets we evaluated. The likely detrimental variants at loci *SLC9A9* and *STRN4* associate with increased AD risk, Braak, Thal, CSF p-Tau, and/or plasma p-Tau levels, while the likely beneficial variants at loci *SCIN*, *NPAS3*, and *ITGB4* associate with decreased Thal and CSF p-Tau levels, increased CSF amyloid levels, and/or trending with decreased amyloid PET measures. These biologically congruent associations in independent datasets provide support that the effects of these novel loci on brain biochemical levels may have roles in the broader pathophysiology of AD.

Of the novel loci, *SLC9A9* which significantly associates with Aβ40<sub>TX</sub> also has nominal associations and trends with higher levels of CAA, Braak, Thal, plasma p-Tau, and AD risk. In a previous study, Aβ40<sub>TX</sub> levels were shown to positively correlate with CAA scores [32]. Our findings suggest that the *SLC9A9* locus may influence AD neuropathologies, including CAA, by mediating brain Aβ40<sub>TX</sub> levels. *SLC9A9* encodes a sodium hydrogen exchanger with multiple functions in regulating the endosome, an organelle critical for the processing of amyloid [113]. Notably, the *SLC9A9* locus has associations with other AD-related phenotypes such as working memory [87], response to cholinesterase inhibitors in



AD [88], and other neuropsychiatric diseases [114–116]. Based on our findings, we postulate that fundamental functions of SLC9A9 in the endosome, including amyloid processing, may underlie its influence on AD and other neuropsychiatric disease-related outcomes.

Many of the other index genes discovered in our brain biochemical endophenotype GWAS have established roles relevant to AD pathology, neurological disorders, and brain function. The *KCNN2*, *RFX7*, *STRN4* and *ITGB4* loci significantly associate with brain A $\beta$ 40 levels, the first three for the membrane-bound (TX) and the last for the soluble (TBS) fraction. *KCNN2*, which encodes a calcium-activated potassium channel, resides at a locus with many other AD-related and neuropsychiatric associations (Table S9). In a transcriptional network analysis, *KCNN2* was the top-ranked network driver gene for classifying AD cases vs. controls [117] and has been shown to have alternative splicing in AD [118]. *STRN4*, like *SLC9A9* and *KCNN2*, encodes a membrane-bound protein. *STRN4* was reported to be a key binding partner and possible regulator of MAP4K [119], the inhibition of which was shown to be neuroprotective [120, 121]. MAP4K is an upstream regulator of YAP [119], the deficiency of which by A $\beta$  sequestering led to neuronal necrosis in the early stages of AD [122]. These findings imply that *STRN4* may be a potential regulator of a molecular cascade, including MAP4K and YAP involved in neuronal health and A $\beta$  metabolism. We also found evidence through transcriptome studies that another gene at the *STRN4* locus, *PRKD2*-a protein kinase- may be the index gene. Future studies are needed to distinguish the actual functional gene at this locus.

Of the four A $\beta$ 40<sub>TX</sub> associated loci genes, *RFX7* is the only transcription factor. Gene-based rare variant analysis of *RFX7* in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort showed trending significance with entorhinal cortex thickness [123]. *ITGB4*, the index gene at the A $\beta$ 40<sub>TBS</sub> locus, encodes a transmembrane integrin involved in cell-to-cell adhesion, is differentially expressed in AD [124–127] with potential roles in the blood-brain barrier [128, 129], schizophrenia and bipolar disorder [130].

Though most of the novel associations were with brain A $\beta$ 40 biochemical fractions, our study also discovered two significant loci, *NPAS3* and *SCIN*, for brain APOE<sub>TX</sub> levels. *NPAS3*, encoding neuronal PAS domain protein, has been implicated in neurogenesis [131], general cognitive function [132, 133], psychiatric disorders [134] like schizophrenia [135], and has protein aggregation potential [136]. *NPAS3* is also known to regulate transcriptional levels of *VGF* [137–139] which is a key regulator in protection against AD pathogenesis in 5xFAD mice models [117] and a top target identified by the AMP-AD

consortium ([agora.adknowledgeportal.org/](http://agora.adknowledgeportal.org/)). Finally, *SCIN* which encodes an actin-binding protein has variants that associate with inflammation markers [102] and rate of cognitive decline in ADs [140]. The *SCIN* locus variant associated with higher APOE<sub>TX</sub> is also associated with lower Thal phase in the same cohort, suggesting that higher membrane-bound levels of APOE might have a protective role in AD. Notably, upregulated expression of *SCIN* was identified as part of a pan-neurodegenerative gene signature across AD, Lewy Body disease, and ALS-FTD [141].

In summary, investigating the functions and other genetic associations of the index genes near biochemical endophenotype GWAS loci support the conclusion that these genes and loci have functional consequences on brain health and neuropsychiatric disease. While such genetic localizations do not definitively prove the involvement of the index genes in the tested phenotypes, they nevertheless provide new testable hypotheses. Moreover, these findings underscore the potential of brain biochemical endophenotypes in the discovery of novel AD-related genes and pathways. Indeed, using the genetic association findings from our GWAS, we detected shared and distinct enrichment in GO pathways known to be important in AD risk as well as novel pathways. In general, these shared biological pathways highlight known broad biological processes in AD such as synapse organization or immune functions. In contrast, distinct pathways pinpoint processes that may relate to specific functions or biochemical states of these proteins, such as lipid metabolism for APOE and peptide cross-linking for soluble A $\beta$ 42<sub>TBS</sub> and p-Tau<sub>TBS</sub>, respectively. This suggests that genetic influences affect not only total levels but also specific biochemical states of these AD-related proteins in the brain. These findings have implications for identifying therapeutic targets that may play a role in the transition of these proteins into pathogenic biochemical states rather than their overall levels.

In this study, we also investigated the functional mechanisms of the GWS loci by expression and splicing QTL, as well as bulk and single-cell transcriptome analyses of our and other published data. QTL analyses suggested that *STRN4*-A $\beta$ 40<sub>TX</sub> rs34805055, *ITGB4*-A $\beta$ 40<sub>TBS</sub> rs9890231, and *APOE/NECTIN2*-APOE<sub>TBS</sub> rs283815 may modulate brain biochemical levels by impacting the expression or splicing of nearby genes, *PRKD2*, *ITGB4*, and *TOMM40*, respectively. Differential expression of genes at each locus revealed significant bulk or cell-specific transcriptional changes in AD for *PRKD2*, *ITGB4* and *APOE/NECTIN2/TOMM40*. Even so, colocalization analyses suggested that these signals are not a result of a single causal variant except for rs34805055-A $\beta$ 40<sub>TX</sub> (*STRN4*)-*PRKD2* eQTL in the Mayo TCX dataset which had inconsistent evidence

of colocalization. It should be noted, however, that many of these colocalization analyses included small datasets from healthy individuals which may not have enough power to reject the null hypothesis. Additional transcriptome studies, particularly cell type specific QTLs, are needed to further characterize the putative regulatory function of these variants.

Notably, our study identified novel GWS loci for brain biochemical levels of AD proteins that were missed in prior studies investigating CSF levels of these proteins [73] or overall neuropathology [100]. This is likely because these studies capture more global brain changes or represent combined biochemical states of these proteins. With the exception of *APOE-ε4*, none of the GWS variants identified here associated with all three biochemical fractions of a protein, suggesting that these loci likely reflect the genetic determinants for specific biochemical states of these proteins. Our findings highlight the potential for deep biochemical phenotyping and demonstrate that this approach can dissect the genetic loci and pathways involved in the specific biochemical states of AD-related proteins, which in turn has implications for understanding disease mechanisms and therapeutic development.

Our study has many strengths including extensive biochemical measures of key AD proteins Aβ, tau and APOE from three brain tissue fractions in a large sample of neuropathologically diagnosed AD patients. We validated and annotated the significant GWAS loci genes and variants by leveraging additional large-scale WGS, RNA-seq, and scRNA-seq datasets with Braak, Thal, CAA, age at death, and additional AD-related biomarker measures. We demonstrated enriched pathways that are both shared as well as those that are AD-protein and biochemical state-specific.

Nevertheless, this study has several limitations including the biochemical measures being available from a single neuropathologically-diagnosed AD cohort of 441 individuals; to our knowledge other autopsy AD cohorts with such deep brain biochemical phenotyping are lacking. Although we have leveraged other AD-related outcome associations from these and other independent samples to validate and annotate our findings, future studies in additional samples with brain biochemical measures are needed for replication as well as increased power. It should also be noted that since we used a more liberal minor allele frequency threshold of 2% and thus five of the novel GWS variants are low frequency (<5%), we validated all novel index GWS variants with secondary confirmation methods and demonstrated high concordance with and reliability of our GWAS imputed genotypes. Nonetheless, the low frequency of these variants in combination with the relatively small GWAS sample

size may influence the precision of our effect estimates. Additional studies to expand this work into larger datasets should be performed in the future to confirm these effects. Additionally, although we perform a multiple testing correction of the GWS threshold for the increased number of independent variants, we do not perform an additional correction for the number of biochemical measures tested, as these measures are not independent of each other [32]. We also note that our AD samples have other co-pathologies, as is commonly observed in neuropathologic AD. Thus, there is a possibility that these co-pathologies may have reduced the power of this study by introducing further heterogeneity. Despite this potential confounding, we were still able to achieve GWS for *APOE* and 7 novel loci. We note that of the biochemical endophenotypes analyzed, APOE and Aβ40 had GWS associations, but not tau or Aβ42. This may be because certain proteins in specific biochemical states may be under stronger genetic control vs. may have more precise measurements reflecting their true biological variability vs. a combination of these factors. Larger scale studies with increasing measurement precision may reveal genetic factors governing the other biochemical measures. Finally, this study was conducted on non-Hispanic white individuals of Northern European descent, making it necessary to expand it to individuals of non-European ancestry.

## Conclusions

Our results strongly suggest that, although the biochemical measures tested reflect proteins core to the pathology of AD, there are unique genetic loci associated with and enriched biological pathways for specific brain biochemical states of these proteins. These findings are expected to dissect the pathophysiology of the biochemical state of AD and finesse therapeutic target discovery efforts focused on these proteins. More broadly, this study presents a new approach that will be applicable to other neurodegenerative diseases to uncover novel mechanisms of proteostasis.

## Abbreviations

A1	Tested Allele
Aβ	Amyloid-β
AD	Alzheimer's disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
AMP-AD	Accelerating Medicines Partnership- Alzheimer's Disease
APOE	Apolipoprotein E
APP	Amyloid Precursor Protein
β	Beta value
BIC	Bioinformatics core
BM	Brodman Area
CAA	Cerebral amyloid angiopathy
CER	Cerebellum
CI	Confidence Interval
CNS	Central nervous system
DE	Differential expression
DEG	Differentially expressed genes
DLPFC	Dorsal lateral prefrontal cortex

eQTL	expression quantitative trait locus
FA	Formic acid soluble tissue fraction
FDR	False Discovery Rate
GAC	Genome analysis core
GBR	British in England and Scotland population
GO	Gene ontology
GWAS	Genome-wide association study
GWS	Genome-wide significant
HRC	Haplotype reference consortium
HWE	Hardy Weinberg Equilibrium
LD	Linkage disequilibrium
LOAD	Late-onset Alzheimer's Disease
MAF	Minor Allele Frequency
MCI	Mild cognitive impairment
MRI	Magnetic Resonance Imaging
MSBB	Mount Sinai Brain Bank
N	Number
NES	Normalized Effect Size
NFT	Neurofibrillary tangles
P	P-value
PC	Principal Component
PET	Positron emission tomography
PHF	Paired Helical Filaments
pTau	Phosphorylated tau
QQ-plot	Quantile-quantile plot
SD	Standard Deviation
SE	Standard Error
sQTL	Splicing quantitative trait locus
ROSMAP	Religious Orders Study and Rush Memory Aging Project
scRNA-seq	Single-cell RNA sequencing
sqrt (CAA)	Square root transformed CAA scores
TBS	Tris-buffered saline soluble tissue fraction
TCX	Temporal cortex
Transf	Transformation
TX	Detergent (1% Triton-X) soluble tissue fraction
WGS	Whole genome sequencing

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13024-022-00592-2>.

**Additional file 1: Table S1.** Description of Dataset. **Table S2.** Description of Biochemical Measures. **Table S3.** Summary Statistics for SNPs with a  $p$ -value <  $1 \times 10^{-5}$  in all biochemical measures. **Table S4.** Estimated proportion of biochemical variance explained by GWS index SNPs. **Table S5.** Additional association model results for the GWS loci. **Table S6.** GWAS Summary statistics for all *APOE* index variants across phenotypes. **Table S7.** Description of Independent AMP-AD and Mayo Expansion Datasets. **Table S8.** Concordance between Imputed GWAS genotypes and Taqman/Sequencing genotypes. **Table S9.** Additional Annotations of GWS loci. **Table S10.** Analysis of GWS variants in the ADNI cohort. **Table S11.** Differentially expressed genes in two or more brain regions proximal to GWS variants in the AMP-AD RNAseq datasets. **Table S12.** Cell type specific differential expression of genes proximal to GWS variants. **Table S13.** Analysis of Significant AD GWAS SNPs from Kunkel et al 2019. **Table S14.** TaqMan Assays used for genotyping key variants. **Figure S1.** Histograms of biochemical measures. **Figure S2.** Circular Manhattan Plots with no GWS SNPs. **Figure S3.** Forest Plots. **Figure S4.** Summary of gene regulatory annotations for each GWS locus. **Figure S5.** Population Substructure via Eigenstrat Analysis. **Figure S6.** Quantile-Quantile (QQ) Plots of biochemical measure GWAS. **Figure S7.** Datasets used for manuscript analytics.

## Acknowledgements

We thank the patients and families for their participation, without whom these studies would not have been possible. We thank our colleagues at the Mayo Clinic Genome Analysis Core (GAC) and Bioinformatics Core (BC) for their collaboration.

**AD Knowledge Portal: AMP-AD datasets:** The results published here are in whole or in part based on data obtained from the AMP-AD Knowledge Portal ([doi:https://doi.org/10.7303/syn2580853](https://doi.org/10.7303/syn2580853)). **Mayo Clinic:** The Mayo RNAseq study data was led by Dr. Nilüfer Ertekin-Taner, Mayo Clinic, Jacksonville, FL as part of the multi-PI U01 AG046139 (MPIs Golde, Ertekin-Taner, Younkin, Price). Samples were provided from the following sources: The Mayo Clinic Brain Bank and Banner Sun Health Research Institute. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data includes samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinsons Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimers Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimers Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinsons Research.

**MSBB:** These data were generated from postmortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr. Eric Schadt from Mount Sinai School of Medicine. **ROSMAP:** Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161 (ROS), R01AG15819 (ROSMAP; genomics and RNAseq), R01AG17917 (MAP), R01AG30146, R01AG36042 (5hC methylation, ATACseq), RC2AG036547 (H3K9Ac), R01AG36836 (RNAseq), R01AG48015 (monocyte RNAseq) RF1AG57473 (single nucleus RNAseq), U01AG32984 (genomic and whole exome sequencing), U01AG46152 (ROSMAP AMP-AD, targeted proteomics), U01AG46161 (TMT proteomics), U01AG61356 (whole genome sequencing, targeted proteomics, ROSMAP AMP-AD), the Illinois Department of Public Health (ROSMAP), and the Translational Genomics Research Institute (genomic). Additional phenotypic data can be requested at [www.radc.rush.edu](http://www.radc.rush.edu). **AGORA:** The results published here are in whole or in part based on data obtained from Agora, a platform initially developed by the NIA-funded AMP-AD consortium that shares evidence in support of AD target discovery ([agora.adknowledgeportal.org/](http://agora.adknowledgeportal.org/)). **Mayo Clinic AD-CAA:** The Mayo Clinic AD-CAA study was led by Dr. Guojun Bu and Dr. Nilufer Ertekin-Taner at Mayo Clinic, Jacksonville, FL as part of the multi-PI RF1AG051504 (MPIs Bu and Ertekin-Taner) using samples from the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50AG016574, R37AG027924, Cure Alzheimer's Fund, and support from Mayo Foundation.

**ADNI dataset:** Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

**GTEx:** The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for

the analyses described in this manuscript were obtained from the GTEx Portal between 07/2020-10/2020.

**NIAGADS:** Data for this study (Kunkle et al. 2019 [16]- NG00075, Lambert et al. 2013 [17]- NG00036, and Cruchaga et al. 2013 [73]- NG00049) were prepared, archived, and distributed by the National Institute on Aging Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania (U24-AG041689), funded by the National Institute on Aging.

**IGAP:** We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The i-Select chips was funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD/PERADES was supported by the Medical Research Council (Grant n° 503480), Alzheimer's Research UK (Grant n° 503176), the Wellcome Trust (Grant n° 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant n° 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01-AG-12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC-10-196728.

#### Authors' contributions

SRO, MA, JSR, and NET wrote the manuscript; NET, SRO, MA, and JSR designed the study; SRO, JSR, MA, ZSQ, MMC, XW, KN, JK, SLR, and AJS performed and consulted on data analysis; MH consulted on statistical methods; CL, YY, YAM, NZ, TK, and GB collected biochemical measures; DWD, MD, and MEM provided neuropathological data and tissue samples; TTN and KGM isolated DNA from tissue samples; SRO, KB, MB, RR performed and consulted on targeted genotyping and sequencing; ADNI collected and made available ADNI data. NET oversaw the study and provided direction, funding, and resources. All authors reviewed and contributed to the manuscript. The author(s) read and approved the final manuscript.

#### Funding

This work was supported by National Institute on Aging [RF1 AG051504 to N.E.T and G. B, U01 AG046139 to N.E.T, and R01 AG061796 to N.E.T]. Data collection was also supported through funding by NIA grants P50AG016574, R37AG027924, Cure Alzheimer's Fund, and support from Mayo Foundation. KN, JK, SLR and AJS were supported in part by P30 AG010133, P30 AG072976, R01 AG019771, R01 AG057739, U01 AG024904, R01 LM013463, R01 AG068193, T32 AG071444, U01 AG068057, U01 AG072177, R01 LM012535, and R01 AG069901.

#### Availability of data and materials

The data in this manuscript can be accessed via the AD Knowledge Portal, MC-CAA study. For access to content described in this manuscript see the following <https://doi.org/10.7303/syn35564295.1>. The AD Knowledge Portal is a platform for accessing data, analyses and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target Discovery Program and other National Institute on Aging (NIA)-supported programs to enable open-science practices and accelerate translational learning. The data, analyses and tools are shared early in the research cycle without a publication embargo on secondary use. Data is available for general research use according to the following requirements for data access and data attribution (<https://adknowledgeportal.synapse.org/DataAccess/Instructions>).

#### Declarations

##### Ethics approval and consent to participate

This study was approved by the appropriate Mayo Clinic Institutional Review Board.

##### Consent for publication

Not applicable.

#### Competing interests

None to report.

#### Author details

<sup>1</sup>Department of Neuroscience, Mayo Clinic, 4500 San Pablo Road, Jacksonville, FL 32224, USA. <sup>2</sup>Department of Quantitative Health Sciences, Mayo Clinic, Jacksonville, FL, USA. <sup>3</sup>Indiana Alzheimer Disease Center, Indiana University School of Medicine, Indianapolis, IN, USA. <sup>4</sup>Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA. <sup>5</sup>School of Informatics and Computing, Indiana University School of Medicine, Indianapolis, IN, USA. <sup>6</sup>VIB-UA Center for Molecular Neurology, VIB, University of Antwerp, Antwerp, Belgium. <sup>7</sup>Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA. <sup>8</sup>Department of Neurology, Mayo Clinic, 4500 San Pablo Road, Birdsall 3, Jacksonville, FL 32224, USA.

Received: 19 July 2022 Accepted: 19 December 2022

Published online: 07 January 2023

#### References

- 2017 Alzheimer's disease facts and figures. *Alzheimer's Dement.* 2017;13:325–73. <https://doi.org/10.1016/j.jalz.2017.02.001>.
- DeTure MA, Dickson DW. The neuropathological diagnosis of Alzheimer's disease. *Mol Neurodegener.* 2019;14:32. <https://doi.org/10.1186/s13024-019-0333-5>.
- Lam B, Masellis M, Freedman M, Stuss DT, Black SE. Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. *Alzheimers Res Ther.* 2013;5:1–1. <https://doi.org/10.1186/alzrt155>.
- Yasuhara O, Kawamata T, Aimi Y, McGeer EG, McGeer PL. Two types of dystrophic neurites in senile plaques of Alzheimer disease and elderly non-demented cases. *Neurosci Lett.* 1994;171:73–6. [https://doi.org/10.1016/0304-3940\(94\)90608-4](https://doi.org/10.1016/0304-3940(94)90608-4).
- Janocko NJ, et al. Neuropathologically defined subtypes of Alzheimer's disease differ significantly from neurofibrillary tangle-predominant dementia. *Acta Neuropathol.* 2012;124:681–92. <https://doi.org/10.1007/s00401-012-1044-y>.
- Murray ME, et al. Differential clinicopathologic and genetic features of late-onset amnesic dementias. *Acta Neuropathol.* 2014;128:411–21. <https://doi.org/10.1007/s00401-014-1302-2>.
- Mehta RI, Schneider JA. What is Alzheimer's disease? The neuropathological heterogeneity of clinically defined Alzheimer's dementia. *Curr Opin Neurol.* 2021;34:237–45. <https://doi.org/10.1097/wco.0000000000000912>.
- Lau HHC, Ingelsson M, Watts JC. The existence of Abeta strains and their potential for driving phenotypic heterogeneity in Alzheimer's disease. *Acta Neuropathol.* 2020. <https://doi.org/10.1007/s00401-020-02201-2>.
- Golde TE, Eckman CB, Younkin SG. Biochemical detection of Abeta isoforms: implications for pathogenesis, diagnosis, and treatment of Alzheimer's disease. *Biochim Biophys Acta.* 2000;1502:172–87. [https://doi.org/10.1016/s0925-4439\(00\)00043-0](https://doi.org/10.1016/s0925-4439(00)00043-0).
- Masters CL, et al. Alzheimer's disease. *Nat Rev Dis Primers.* 2015;1:15056. <https://doi.org/10.1038/nrdp.2015.56>.
- Bi C, Bi S, Li B. Processing of mutant  $\beta$ -amyloid precursor protein and the clinicopathological features of familial Alzheimer's disease. *Aging Dis.* 2019;10:383–403. <https://doi.org/10.14336/AD.2018.0425>.
- Iqbal K, Liu F, Gong CX, Grundke-Iqbal I. Tau in Alzheimer disease and related tauopathies. *Curr Alzheimer Res.* 2010;7:656–64. <https://doi.org/10.2174/156720510793611592>.
- Mandelkow EM, et al. Tau domains, phosphorylation, and interactions with microtubules. *Neurobiol Aging.* 1995;16:355–62. [https://doi.org/10.1016/0197-4580\(95\)00025-A](https://doi.org/10.1016/0197-4580(95)00025-A).
- Pooler AM, Noble W, Hanger DP. A role for tau at the synapse in Alzheimer's disease pathogenesis. *Neuropharmacology.* 2014;76:1–8. <https://doi.org/10.1016/j.neuropharm.2013.09.018>.
- Guo T, Noble W, Hanger DP. Roles of tau protein in health and disease. *Acta Neuropathol.* 2017;133:665–704. <https://doi.org/10.1007/s00401-017-1707-9>.
- Kunkle BW, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A $\beta$ , tau, immunity and lipid processing. *Nat Genet.* 2019;51:414–30. <https://doi.org/10.1038/s41588-019-0358-2>.

17. Lambert JC, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet.* 2013;45:1452–8. <https://doi.org/10.1038/ng.2802>.
18. Jansen IE, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet.* 2019;51:404–13. <https://doi.org/10.1038/s41588-018-0311-9>.
19. Kanekiyo T, Xu H, Bu G. ApoE and A $\beta$  in Alzheimer's disease: accidental encounters or partners? *Neuron.* 2014;81:740–54. <https://doi.org/10.1016/j.neuron.2014.01.045>.
20. Steinerman JR, et al. Distinct pools of beta-amyloid in Alzheimer disease-affected brain: a clinicopathologic study. *Arch Neurol.* 2008;65:906–12. <https://doi.org/10.1001/archneur.65.7.906>.
21. Roberts BR, et al. Biochemically-defined pools of amyloid- $\beta$  in sporadic Alzheimer's disease: correlation with amyloid PET. *Brain.* 2017;140:1486–98. <https://doi.org/10.1093/brain/awx057>.
22. Ait-Bouziad N, et al. Discovery and characterization of stable and toxic tau/phospholipid oligomeric complexes. *Nat Commun.* 2017;8:1678. <https://doi.org/10.1038/s41467-017-01575-4>.
23. Gray EG, Paula-Barbosa M, Roher A. Alzheimer's disease: paired helical filaments and cytomembranes. *Neuropathol Appl Neurobiol.* 1987;13:91–110. <https://doi.org/10.1111/j.1365-2990.1987.tb00174.x>.
24. Ekinci FJ, Shea TB. Phosphorylation of tau alters its association with the plasma membrane. *Cell Mol Neurobiol.* 2000;20:497–508. <https://doi.org/10.1023/a:1007075115574>.
25. Shea TB. Phospholipids alter tau conformation, phosphorylation, proteolysis, and association with microtubules: implication for tau function under normal and degenerative conditions. *J Neurosci Res.* 1997;50:114–22. [https://doi.org/10.1002/\(sici\)1097-4547\(19971001\)50:1<114::Aid-jnr12>3.0.Co;2-b](https://doi.org/10.1002/(sici)1097-4547(19971001)50:1<114::Aid-jnr12>3.0.Co;2-b).
26. Jones EM, et al. Interaction of tau protein with model lipid membranes induces tau structural compaction and membrane disruption. *Biochemistry.* 2012;51:2539–50. <https://doi.org/10.1021/bi201857v>.
27. Elbaum-Garfinkle S, Ramlall T, Rhoades E. The role of the lipid bilayer in tau aggregation. *Biophys J.* 2010;98:2722–30. <https://doi.org/10.1016/j.bpj.2010.03.013>.
28. Brunello CA, Merezko M, Uronen R-L, Huttunen HJ. Mechanisms of secretion and spreading of pathological tau protein. *Cell Mol Life Sci.* 2020;77:1721–44. <https://doi.org/10.1007/s00018-019-03349-1>.
29. Alafuzoff I, et al. Staging of neurofibrillary pathology in Alzheimer's disease: a study of the BrainNet Europe consortium. *Brain Pathol.* 2008;18:484–96. <https://doi.org/10.1111/j.1750-3639.2008.00147.x>.
30. Koss DJ, et al. Soluble pre-fibrillar tau and  $\beta$ -amyloid species emerge in early human Alzheimer's disease and track disease progression and cognitive decline. *Acta Neuropathol.* 2016;132:875–95. <https://doi.org/10.1007/s00401-016-1632-3>.
31. Mattsson-Carlgrén N, et al. A $\beta$  deposition is associated with increases in soluble and phosphorylated tau that precede a positive tau PET in Alzheimer's disease. *Sci Adv.* 2020;6:eaz2387. <https://doi.org/10.1126/sciadv.aaz2387>.
32. Liu C-C, et al. Tau and apolipoprotein E modulate cerebrovascular tight junction integrity independent of cerebral amyloid angiopathy in Alzheimer's disease. *Alzheimers Dement.* 2020;16:1372–83. <https://doi.org/10.1002/alz.12104>.
33. Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 1991;82:239–59. <https://doi.org/10.1007/bf00308809>.
34. Thal DR, Rüb U, Orantes M, Braak H. Phases of a beta-deposition in the human brain and its relevance for the development of AD. *Neurology.* 2002;58:1791–800. <https://doi.org/10.1212/wnl.58.12.1791>.
35. Murray ME, et al. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol.* 2011;10:785–96. [https://doi.org/10.1016/S1474-4422\(11\)70156-9](https://doi.org/10.1016/S1474-4422(11)70156-9).
36. Murray ME, et al. Clinicopathologic and 11C-Pittsburgh compound B implications of Thal amyloid phase across the Alzheimer's disease spectrum. *Brain.* 2015;138:1370–81. <https://doi.org/10.1093/brain/awv050>.
37. Reddy JS, et al. Genome-wide analysis identifies a novel LINC-PINT splice variant associated with vascular amyloid pathology in Alzheimer's disease. *Acta Neuropathol Commun.* 2021;9:93. <https://doi.org/10.1186/s40478-021-01199-2>.
38. Chang CC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience.* 2015;4. <https://doi.org/10.1186/s13742-015-0047-8>.
39. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75. <https://doi.org/10.1086/519795>.
40. Loh P-R, et al. Reference-based phasing using the haplotype reference consortium panel. *Nat Genet.* 2016;48:1443–8. <https://doi.org/10.1038/ng.3679>.
41. McCarthy S, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48:1279–83. <https://doi.org/10.1038/ng.3643>.
42. Allen M, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data.* 2016;3:160089. <https://doi.org/10.1038/sdata.2016.89>.
43. Wang M, et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci Data.* 2018;5:180185. <https://doi.org/10.1038/sdata.2018.185>.
44. De Jager PL, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data.* 2018;5:180142. <https://doi.org/10.1038/sdata.2018.142>.
45. Wan YW, et al. Meta-analysis of the Alzheimer's disease human brain transcriptome and functional dissection in mouse models. *Cell Rep.* 2020;32:107908. <https://doi.org/10.1016/j.celrep.2020.107908>.
46. Mirra SS, et al. The consortium to establish a registry for Alzheimer's disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology.* 1991;41:479–86. <https://doi.org/10.1212/wnl.41.4.479>.
47. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM; 2013. <https://arxiv.org/abs/1303.3997>
48. Picard Toolkit. Broad Institute, GitHub Repository. Broad Institute; 2019. Picard version 1.83. <https://broadinstitute.github.io/picard/>.
49. Van der Auwera GA, et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013;43(1110):11–33. <https://doi.org/10.1002/0471250953.bi1110s43>.
50. Manichaikul A, et al. Robust relationship inference in genome-wide association studies. *Bioinformatics.* 2010;26:2867–73. <https://doi.org/10.1093/bioinformatics/btq559>.
51. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2:e190. <https://doi.org/10.1371/journal.pgen.0020190>.
52. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38:904–9. <https://doi.org/10.1038/ng1847>.
53. Weiner MW, et al. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimers Dement.* 2012;8:S1–68. <https://doi.org/10.1016/j.jalz.2011.09.172>.
54. Saykin AJ, et al. Genetic studies of quantitative MCI and AD phenotypes in ADNI: progress, opportunities, and plans. *Alzheimers Dement.* 2015;11:792–814. <https://doi.org/10.1016/j.jalz.2015.05.009>.
55. Nho K, et al. Serum metabolites associated with brain amyloid beta deposition, cognition and dementia progression. *Brain Commun.* 2021;3:fcab139. <https://doi.org/10.1093/braincomms/fcab139>.
56. Hansson O, et al. CSF biomarkers of Alzheimer's disease concord with amyloid- $\beta$  PET and predict clinical progression: a study of fully automated immunoassays in BioFINDER and ADNI cohorts. *Alzheimers Dement.* 2018;14:1470–81. <https://doi.org/10.1016/j.jalz.2018.01.010>.
57. Kim JP, et al. BMI1 is associated with CSF amyloid- $\beta$  and rates of cognitive decline in Alzheimer's disease. *Alzheimers Res Ther.* 2021;13:164. <https://doi.org/10.1186/s13195-021-00906-4>.
58. Fadista J, Manning AK, Florez JC, Groop L. The (in) famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet.* 2016;24:1202–5. <https://doi.org/10.1038/ejhg.2015.269>.
59. Giambartolomei C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10:e1004383. <https://doi.org/10.1371/journal.pgen.1004383>.
60. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J R Stat Soc Ser B Stat Methodol.* 2020;82:1273–300. <https://doi.org/10.1111/rssb.12388>.

61. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* 2021;17:e1009440. <https://doi.org/10.1371/journal.pgen.1009440>.
62. Zhang T, Klein A, Sang J, Choi J, Brown KM. ezQTL: a web platform for interactive visualization and colocalization of quantitative trait loci and GWAS. *Genom Proteom Bioinform.* 2022. <https://doi.org/10.1016/j.gpb.2022.05.004>.
63. Foley CN, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun.* 2021;12:764. <https://doi.org/10.1038/s41467-020-20885-8>.
64. Hormozdiari F, et al. Colocalization of GWAS and eQTL signals detects target genes. *Am J Hum Genet.* 2016;99:1245–60. <https://doi.org/10.1016/j.ajhg.2016.10.003>.
65. Yoon S, et al. Efficient pathway enrichment and network analysis of GWAS summary data using GSA-SNP2. *Nucleic Acids Res.* 2018;46:e60. <https://doi.org/10.1093/nar/gky175>.
66. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci.* 2005;102:15545–50. <https://doi.org/10.1073/pnas.0506580102>.
67. Liberzon A, et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1:417–25. <https://doi.org/10.1016/j.cels.2015.12.004>.
68. Ashburner M, et al. Gene ontology: tool for the unification of biology. *Nat Genet.* 2000;25:25–9. <https://doi.org/10.1038/75556>.
69. The Gene Ontology Consortium. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 2018;47:D330–8. <https://doi.org/10.1093/nar/gky1055>.
70. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One.* 2011;6:e21800. <https://doi.org/10.1371/journal.pone.0021800>.
71. Mathys H, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature.* 2019;570:332–7. <https://doi.org/10.1038/s41586-019-1195-2>.
72. Howe KL, et al. Ensembl 2021. *Nucleic Acids Res.* 2020;49:D884–91. <https://doi.org/10.1093/nar/gkaa942>.
73. Cruchaga C, et al. GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron.* 2013;78:256–68. <https://doi.org/10.1016/j.neuron.2013.02.026>.
74. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics.* 2015;31:3555–7. <https://doi.org/10.1093/bioinformatics/btv402>.
75. Aguet F, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017;550:204–13. <https://doi.org/10.1038/nature24277>.
76. GTEx Consortium, L A, Aguet F, et al. A novel approach to high-quality postmortem tissue procurement: the GTEx project. *Biopreserv Biobank.* 2015;13:311–9. <https://doi.org/10.1089/bio.2015.0032>.
77. Dawson TM, Dawson VL. The role of parkin in familial and sporadic Parkinson's disease. *Mov Disord.* 2010;25(Suppl 1):S32–9. <https://doi.org/10.1002/mds.22798>.
78. Nazarian A, Yashin AI, Kulminski AM. Genome-wide analysis of genetic predisposition to Alzheimer's disease and related sex disparities. *Alzheimers Res Ther.* 2019;11:5. <https://doi.org/10.1186/s13195-018-0458-8>.
79. Yan Q, et al. Genome-wide association study of brain amyloid deposition as measured by Pittsburgh compound-B (PiB)-PET imaging. *Mol Psychiatry.* 2021;26:309–21. <https://doi.org/10.1038/s41380-018-0246-7>.
80. Arpawong TE, et al. Genetic variants specific to aging-related verbal memory: insights from GWASs in a population-based cohort. *PLoS One.* 2017;12:e0182448. <https://doi.org/10.1371/journal.pone.0182448>.
81. Li J, et al. Genome-wide association and interaction studies of CSF T-tau/A $\beta$ (42) ratio in ADNI cohort. *Neurobiol Aging.* 2017;57(247):e241–247.e248. <https://doi.org/10.1016/j.neurobiolaging.2017.05.007>.
82. Li J, et al. Genetic interactions explain variance in cingulate amyloid burden: an AV-45 PET genome-wide association and interaction study in the ADNI cohort. *Biomed Res Int.* 2015;2015:647389. <https://doi.org/10.1155/2015/647389>.
83. Liu C, et al. Genome-wide association and mechanistic studies indicate that immune response contributes to Alzheimer's disease development. *Front Genet.* 2018;9:410. <https://doi.org/10.3389/fgene.2018.00410>.
84. Ramanan VK, et al. APOE and BCHE as modulators of cerebral amyloid deposition: a florbetapir PET genome-wide association study. *Mol Psychiatry.* 2014;19:351–7. <https://doi.org/10.1038/mp.2013.19>.
85. Fogh I, et al. Association of a Locus in the CAMTA1 gene with survival in patients with sporadic amyotrophic lateral sclerosis. *JAMA Neurol.* 2016;73:812–20. <https://doi.org/10.1001/jamaneurol.2016.1114>.
86. Buniello A, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47:D1005–d1012. <https://doi.org/10.1093/nar/gky1120>.
87. Donati G, Dumontheil I, Meaburn EL. Genome-wide association study of latent cognitive measures in adolescence: genetic overlap with intelligence and education. *Mind Brain Educ.* 2019;13:224–33. <https://doi.org/10.1111/mbe.12198>.
88. Martinelli-Boneschi F, et al. Pharmacogenomics in Alzheimer's disease: a genome-wide association study of response to cholinesterase inhibitors. *Neurobiol Aging.* 2013;34:1711.e1717–3. <https://doi.org/10.1016/j.neurobiolaging.2012.12.008>.
89. Wang H, et al. Genome-wide interaction analysis of pathological hallmarks in Alzheimer's disease. *Neurobiol Aging.* 2020;93:61–8. <https://doi.org/10.1016/j.neurobiolaging.2020.04.025>.
90. Grove J, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet.* 2019;51:431–44. <https://doi.org/10.1038/s41588-019-0344-8>.
91. Matoba N, et al. Common genetic risk variants identified in the SPARK cohort support DDHD2 as a candidate risk gene for autism. *Transl Psychiatry.* 2020;10:265. <https://doi.org/10.1038/s41398-020-00953-9>.
92. Mick E, et al. Family-based genome-wide association scan of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry.* 2010;49:898–905.e893. <https://doi.org/10.1016/j.jaac.2010.02.014>.
93. Chibnik LB, et al. Susceptibility to neurofibrillary tangles: role of the PTPRD locus and limited pleiotropy with other neuropathologies. *Mol Psychiatry.* 2018;23:1521–9. <https://doi.org/10.1038/mp.2017.20>.
94. Liu C, Yu J. Genome-wide association studies for cerebrospinal fluid soluble TREM2 in Alzheimer's disease. *Front Aging Neurosci.* 2019;11:297. <https://doi.org/10.3389/fgene.2019.00297>.
95. Huang J, et al. Cross-disorder genomewide analysis of schizophrenia, bipolar disorder, and depression. *Am J Psychiatry.* 2010;167:1254–63. <https://doi.org/10.1176/appi.ajp.2010.09091335>.
96. Goes FS, et al. Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am J Med Genet B Neuropsychiatr Genet.* 2015;168:649–59. <https://doi.org/10.1002/ajmg.b.32349>.
97. Stahl EA, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet.* 2019;51:793–803. <https://doi.org/10.1038/s41588-019-0397-8>.
98. Ferreira MA, et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet.* 2008;40:1056–8. <https://doi.org/10.1038/ng.209>.
99. Herold C, et al. Family-based association analyses of imputed genotypes reveal genome-wide significant association of Alzheimer's disease with OSBPL6, PTPRG, and PDC13. *Mol Psychiatry.* 2016;21:1608–12. <https://doi.org/10.1038/mp.2015.218>.
100. Beecham GW, et al. Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet.* 2014;10:e1004606. <https://doi.org/10.1371/journal.pgen.1004606>.
101. Alliey-Rodriguez N, et al. NRXN1 is associated with enlargement of the temporal horns of the lateral ventricles in psychosis. *Transl Psychiatry.* 2019;9:230. <https://doi.org/10.1038/s41398-019-0564-9>.
102. Ahola-Olli AV, et al. Genome-wide association study identifies 27 loci influencing concentrations of circulating cytokines and growth factors. *Am J Hum Genet.* 2017;100:40–50. <https://doi.org/10.1016/j.ajhg.2016.11.007>.
103. Nelson PT, et al. Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature. *J Neuropathol Exp Neurol.* 2012;71:362–81. <https://doi.org/10.1097/NEN.0b013e31825018f7>.
104. Martens YA, et al. ApoE cascade hypothesis in the pathogenesis of Alzheimer's disease and related dementias. *Neuron.* 2022;110:1304–17. <https://doi.org/10.1016/j.neuron.2022.03.004>.

105. Tan CC, Zhang XY, Tan L, Yu JT. Tauopathies: mechanisms and therapeutic strategies. *J Alzheimers Dis.* 2018;61:487–508. <https://doi.org/10.3233/JAD-170187>.
106. Hampel H, et al. The amyloid-beta pathway in Alzheimer's disease. *Mol Psychiatry.* 2021;26:5481–503. <https://doi.org/10.1038/s41380-021-01249-0>.
107. Deming Y, et al. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol.* 2017;133:839–56. <https://doi.org/10.1007/s00401-017-1685-y>.
108. Shi Y, et al. ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature.* 2017;549:523–7. <https://doi.org/10.1038/nature24016>.
109. Ramanan VK, et al. Association of Apolipoprotein E  $\epsilon$ 4, educational level, and sex with tau deposition and tau-mediated metabolic dysfunction in older adults. *JAMA Netw Open.* 2019;2:e1913909. <https://doi.org/10.1001/jamanetworkopen.2019.13909>.
110. Salvadó G, et al. Differential associations of APOE- $\epsilon$ 2 and APOE- $\epsilon$ 4 alleles with PET-measured amyloid- $\beta$  and tau deposition in older individuals without dementia. *Eur J Nucl Med Mol Imaging.* 2021;48:2212–24. <https://doi.org/10.1007/s00259-021-05192-8>.
111. Therriault J, et al. Association of Apolipoprotein E  $\epsilon$ 4 with medial temporal tau independent of amyloid- $\beta$ . *JAMA Neurol.* 2020;77:470–9. <https://doi.org/10.1001/jamaneuro.2019.4421>.
112. Therriault J, et al. APOE $\epsilon$ 4 potentiates the relationship between amyloid- $\beta$  and tau pathologies. *Mol Psychiatry.* 2021;26:5977–88. <https://doi.org/10.1038/s41380-020-0688-6>.
113. Nixon RA. Amyloid precursor protein and endosomal-lysosomal dysfunction in Alzheimer's disease: inseparable partners in a multifactorial disease. *FASEB J.* 2017;31:2729–43. <https://doi.org/10.1096/fj.20170359>.
114. Patak J, Faraone SV, Zhang-James Y. Sodium hydrogen exchanger 9 NHE9 (SLC9A9) and its emerging roles in neuropsychiatric comorbidity. *Am J Med Genet B Neuropsychiatr Genet.* 2020;183:289–305. <https://doi.org/10.1002/ajmg.b.32787>.
115. Markunas CA, et al. Genetic variants in SLC9A9 are associated with measures of attention-deficit/hyperactivity disorder symptoms in families. *Psychiatr Genet.* 2010;20:73–81. <https://doi.org/10.1097/YPG.0b013e3283351209>.
116. Liu G, et al. Genetic variants and multiple sclerosis risk gene SLC9A9 expression in distinct human brain regions. *Mol Neurobiol.* 2017;54:6820–6. <https://doi.org/10.1007/s12035-016-0208-5>.
117. Beckmann ND, et al. Multiscale causal networks identify VGF as a key regulator of Alzheimer's disease. *Nat Commun.* 2020;11:3942. <https://doi.org/10.1038/s41467-020-17405-z>.
118. Heinzen EL, et al. Alternative ion channel splicing in mesial temporal lobe epilepsy and Alzheimer's disease. *Genome Biol.* 2007;8:R32. <https://doi.org/10.1186/gb-2007-8-3-r32>.
119. Seo G, et al. MAP 4K Interactome reveals STRN4 as a key STRIPAK complex component in hippo pathway regulation. *Cell Rep.* 2020;32:107860. <https://doi.org/10.1016/j.celrep.2020.107860>.
120. Bos PH, et al. Development of MAP 4 kinase inhibitors as motor neuron-protecting agents. *Cell Chem Biol.* 2019;26:1703–1715.e1737. <https://doi.org/10.1016/j.chembiol.2019.10.005>.
121. Wu C, Watts ME, Rubin LL. MAP 4K4 activation mediates motor neuron degeneration in amyotrophic lateral sclerosis. *Cell Rep.* 2019;26:1143–1156.e1145. <https://doi.org/10.1016/j.celrep.2019.01.019>.
122. Tanaka H, et al. YAP-dependent necrosis occurs in early stages of Alzheimer's disease and regulates mouse model pathology. *Nat Commun.* 2020;11:507. <https://doi.org/10.1038/s41467-020-14353-6>.
123. Kim D, et al. Knowledge-driven binning approach for rare variant association analysis: application to neuroimaging biomarkers in Alzheimer's disease. *BMC Med Inform Decis Mak.* 2017;17:61. <https://doi.org/10.1186/s12911-017-0454-0>.
124. Chen W-T, et al. Spatial Transcriptomics and in situ sequencing to study Alzheimer's disease. *Cell.* 2020;182:976–991.e919. <https://doi.org/10.1016/j.cell.2020.06.038>.
125. Wang X-L, Li L. Cell type-specific potential pathogenic genes and functional pathways in Alzheimer's disease. *BMC Neurol.* 2021;21:381. <https://doi.org/10.1186/s12883-021-02407-1>.
126. Jing Q, et al. A comprehensive analysis identified hub genes and associated drugs in Alzheimer's disease. *Biomed Res Int.* 2021;2021:8893553. <https://doi.org/10.1155/2021/8893553>.
127. Chowdhury UN, Islam MB, Ahmad S, Moni MA. Systems biology and bioinformatics approach to identify gene signatures, pathways and therapeutic targets of Alzheimer's disease. *Inform Med Unlocked.* 2020;21:100439. <https://doi.org/10.1016/j.imu.2020.100439>.
128. Milner R, Campbell IL. Increased expression of the  $\beta$ 4 and  $\alpha$ 5 integrin subunits in cerebral blood vessels of transgenic mice chronically producing the pro-inflammatory cytokines IL-6 or IFN- $\alpha$  in the central nervous system. *Mol Cell Neurosci.* 2006;33:429–40. <https://doi.org/10.1016/j.mcn.2006.09.004>.
129. Verkerke M, Hol EM, Middeldorp J. Physiological and pathological ageing of astrocytes in the human brain. *Neurochem Res.* 2021;46:2662–75. <https://doi.org/10.1007/s11064-021-03256-7>.
130. O'Brien NL, et al. Rare variant analysis in multiply affected families, association studies and functional analysis suggest a role for the ITGB4 gene in schizophrenia and bipolar disorder. *Schizophr Res.* 2018;199:181–8. <https://doi.org/10.1016/j.schres.2018.03.001>.
131. Kamm GB, Pisciotto F, Klinger R, Franchini LF. The developmental brain gene NPAS3 contains the largest number of accelerated regulatory sequences in the human genome. *Mol Biol Evol.* 2013;30:1088–102. <https://doi.org/10.1093/molbev/mst023>.
132. Davies G, et al. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). *Mol Psychiatry.* 2015;20:183–92. <https://doi.org/10.1038/mp.2014.188>.
133. Trampush JW, et al. GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Mol Psychiatry.* 2017;22:336–45. <https://doi.org/10.1038/mp.2016.244>.
134. Wong J, et al. Expression of NPAS3 in the human cortex and evidence of its posttranscriptional regulation by miR-17 during development, with implications for schizophrenia. *Schizophr Bull.* 2013;39:396–406. <https://doi.org/10.1093/schbul/sbr177>.
135. Macintyre G, et al. Association of NPAS3 exonic variation with schizophrenia. *Schizophr Res.* 2010;120:143–9. <https://doi.org/10.1016/j.schres.2010.04.002>.
136. Nucifora LG, et al. A mutation in NPAS3 that segregates with schizophrenia in a small family leads to protein aggregation. *Complex Psychiatry.* 2016;2:133–44. <https://doi.org/10.1159/000447358>.
137. Sha L, et al. Transcriptional regulation of neurodevelopmental and metabolic pathways by NPAS3. *Mol Psychiatry.* 2012;17:267–79. <https://doi.org/10.1038/mp.2011.73>.
138. Luoma LM, Berry FB. Molecular analysis of NPAS3 functional domains and variants. *BMC Mol Biol.* 2018;19:14. <https://doi.org/10.1186/s12867-018-0117-4>.
139. Yang D, et al. NPAS3 regulates transcription and expression of VGF: implications for neurogenesis and psychiatric disorders. *Front Mol Neurosci.* 2016;9. <https://doi.org/10.3389/fnmol.2016.00109>.
140. Sherva R, et al. Genome-wide association study of rate of cognitive decline in Alzheimer's disease patients identifies novel genes and pathways. *Alzheimers Dement.* 2020;16:1134–45. <https://doi.org/10.1002/alz.12106>.
141. Noori A, Mezlini AM, Hyman BT, Serrano-Pozo A, Das S. Systematic review and meta-analysis of human transcriptomics reveals neuroinflammation, deficient energy metabolism, and proteostasis failure across neurodegeneration. *Neurobiol Dis.* 2021;149:105225. <https://doi.org/10.1016/j.nbd.2020.105225>.

## Publisher's Note

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