Multi-Omics Analysis for Identifying Cell-Type-Specific Druggable Targets in Alzheimer's Disease

Shiwei Liu[1,2], Min Young Cho[1,2,3], Yen-Ning Huang[1,2], Tamina Park[1,2], Soumilee Chaudhuri[1,2], Thea Jacobson Rosewood[1,2], Paula J Bice[1,2], Dongjun Chung[4], David A. Bennett[5], Nilüfer Ertekin-Taner[6,7], Andrew J Saykin[1,2], Kwangsik Nho[1,2,8*]

[1]Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, 550 N. University Blvd. Indianapolis, IN, 46202, USA.

[2]Indiana Alzheimer's Disease Research Center, Indiana University School of Medicine, 355 W. 16th Street, Goodman Hall, Suite 4100, Indianapolis, IN, 46202, USA.

[3]Sungkyunkwan University, Seoul, Republic of Korea

[4]Department of Biomedical Informatics, College of Medicine, The Ohio State University, OH, 43210, USA

[5]Department of Neurological Science, Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, 60612, USA.

[6]Department of Neuroscience, Mayo Clinic, Jacksonville, FL, 32224, USA

[7]Department of Neurology, Mayo Clinic, Jacksonville, FL, 32224, USA

[8]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, 340 West 10th Street, Fairbanks Hall, Suite 6200 Indianapolis, Indiana, 46202, USA

*Corresponding author: Kwangsik Nho

Address: Indiana University School of Medicine, 355 W. 16th Street, Goodman Hall, Suite 4100, Indianapolis, IN, 46202, USA.

Email: knho@iu.edu

Abstract

Background: Analyzing disease-linked genetic variants via expression quantitative trait loci (eQTLs) is important for identifying potential disease-causing genes. Previous research prioritized genes by integrating Genome-Wide Association Study (GWAS) results with tissue-level eQTLs. Recent studies have explored brain cell type-specific eQTLs, but they lack a systematic analysis across various Alzheimer's disease (AD) GWAS datasets, nor did they compare effects between tissue and cell type levels or across different cell type-specific eQTL datasets. In this study, we integrated brain cell type-specific eQTL datasets with AD GWAS datasets to identify potential causal genes at the cell type level.

Methods: To prioritize disease-causing genes, we used Summary Data-Based Mendelian Randomization (SMR) and Bayesian Colocalization (COLOC) to integrate AD GWAS summary statistics with cell-type-specific eQTLs. Combining data from five AD GWAS, three single-cell eQTL datasets, and one bulk tissue eQTL meta-analysis, we identified and confirmed both novel and known candidate causal genes. We investigated gene regulation through enhancer activity using H3K27ac and ATAC-seq data, performed protein-protein interaction and pathway enrichment analyses, and conducted a drug/compound enrichment analysis with the Drug Signatures Database (DSigDB) to support drug repurposing for AD.

Results: We identified 27 candidate causal genes for AD using cell type-specific eQTL datasets, with the highest numbers in microglia, followed by excitatory neurons, astrocytes, inhibitory

neurons, oligodendrocytes, and oligodendrocyte precursor cells (OPCs). *PABPC1* emerged as a novel astrocyte-specific gene. Our analysis revealed protein-protein interaction (PPI) networks for these causal genes in microglia and astrocytes. We found the "regulation of aspartic-type peptidase activity" pathway being the most enriched among all the causal genes. AD-risk variants associated with candidate causal gene *PABPC1* is located near or within enhancers only active in astrocytes. We classified the genes into three drug tiers and identified druggable interactions, with imatinib mesylate emerging as a key candidate. A drug-target gene network was created to explore potential drug targets for AD.

Conclusions: We systematically prioritized AD candidate causal genes based on cell type-specific molecular evidence. The integrative approach enhances our understanding of molecular mechanisms of AD-related genetic variants and facilitates the interpretation of AD GWAS results.

Keywords

Causal genes; eQTL; Alzheimer's disease; GWAS; SNP; genetic variant, Gene expression; cell type; astrocytes; drug repurposing

List of abbreviations

AD: Alzheimer's Disease

eQTLs: Expression Quantitative Trait Loci

GWAS: Genome-Wide Association Study

SMR: Summary Data-Based Mendelian Randomization

COLOC: Bayesian Colocalization

LOAD: Late-Onset Alzheimer's Disease

UKBEC: The UK Brain Expression Consortium

GTEx: Genotype-Tissue Expression Consortium

DLPFC: Dorsolateral Prefrontal Cortex

PFC: Prefrontal Cortex

OPCs: Oligodendrocyte Progenitor Cells

TMM: Trimmed Mean of M-values

CPM: Counts Per Million

PCs: Principal Components

HEIDI: Heterogeneity in Dependent Instruments

PPs: Posterior Probabilities

PPI: Protein–Protein Interaction

LD: Linkage Disequilibrium

DSigDB: The Drug Signatures Database

DEG: Differential Gene Expression

Background

Alzheimer's Disease (AD) is a multifaceted neurodegenerative disorder characterized by progressive cognitive decline and memory loss[1]. AD is broadly categorized into early-onset and late-onset forms, with late-onset AD (LOAD) being the most common[2]. The genetic architecture of AD is complex, involving numerous deleterious variants distributed across various genes[2]. Among these, the *APOE* ε4 allele is recognized as the strongest genetic risk

factor for late-onset AD[3]. Genome-Wide Association Studies (GWAS) have significantly advanced our understanding of the genetic basis of AD[4-9]. Early AD GWAS studies identified key loci like *CLU* and *CR1*[5]. The latest AD GWAS study has significantly expanded our understanding of the genetic basis of Alzheimer's disease, identifying 83 genetic variants across 75 loci, including 42 newly discovered variants in European ancestry populations[4].

However, while GWAS studies are instrumental in identifying genetic variants associated with AD, they fail to elucidate the molecular and cellular mechanisms by which the variants contribute to the disease. Only a small fraction of these variants resides within coding regions, while a significant number of non-coding risk variants remain unexplained. To better understand the underlying mechanisms through which these risk variants act, recent studies have employed Expression Quantitative Trait Loci (eQTL) analyses[10-15] for following up study of GWAS results. The eQTL analyses can reveal how non-coding variants identified by GWAS influence the risk of AD through changes in gene expression[16]. Several public eQTL datasets derived from brain tissue have become available, including the Braineac dataset from the UK Brain Expression Consortium (UKBEC)[17], the Genotype-Tissue Expression (GTEx) consortium[18] and the MetaBrain dataset[10]. These datasets have enhanced the interpretation of GWAS findings by elucidating how risk variants regulate gene expression on the tissue level.

Furthermore, a few recent eQTL studies have demonstrated that these non-coding variants affect gene expression in a cell-type-specific manner, underscoring the complexity of their functional impact[11, 14]. Cell type-specific eQTLs enable researchers to determine the cell types that are most influenced by genetic variants and enable the identification of key cell types and regulatory

networks involved in the disease progression, thereby offering enhanced understanding of the underlying mechanisms of diseases. Moreover, previous research has shown that GWAS-identified risk variants in non-coding regions can influence phenotypic outcomes by perturbing transcriptional gene promoters and enhancers[19]. For instance, a study has shown that the AD-associated genes *BIN1* and *PICALM* are regulated through AD risk variants that overlap with microglia-specific enhancers, which interact with the active promoters of these genes[19]. Understanding whether these genetic risk variants overlap with specific regulatory elements provides deeper insights into the cell-type-specific mechanisms underlying gene expression regulation.

In this study, we systematically integrated AD GWAS summary statistics with cell type-specific eQTL data to enhance our understanding of the genetic mechanisms underlying AD. We employed Summary Data-Based Mendelian Randomization (SMR) and Bayesian colocalization (COLOC) methods to identify and prioritize potential disease-causing genes. Our analysis included five recent AD GWAS datasets and three cell type-specific eQTL datasets derived from single-cell sequencing of AD brain samples, as well as a tissue-level metabrain eQTL dataset from previous studies. We focused on prioritizing candidate causal genes for follow-up functional studies in the future. We examined their associated variants and the possible effects on enhancers in a cell type-specific manner. By comparing our results with existing studies, we identified novel cell type-specific candidate genes and used tools such as eQTpLot to visualize their colocalization. Additionally, we used differential gene expression analysis data to investigate the associations between these novel candidate causal genes and AD pathology and

cognitive function. This comprehensive approach aims to improve our understanding of AD's genetic basis at the molecular and cellular level and identify potential therapeutic targets.

Methods

Datasets

We utilized summary statistics from 5 latest GWAS studies on AD involving European ancestry, downloaded from the NHGRI-EBI GWAS Catalog. As shown in Additional file 1: **Table S1**, Kunkle et al. 2019 included 21,982 AD cases, and 41,944 controls from the U.S., Canada, France, Germany, Netherlands, Iceland, U.K., Greece, and other regions, totaling 63,926 samples[6]. Jansen et al. 2019 involved 24,087 AD cases, 47,793 proxy cases, and 383,378 controls, with a total of 455,258 samples from the U.S., Norway, Sweden, U.K., and other regions (Additional file 1: **Table S1)[7]**. Wightman et al. 2021 analyzed 39,918 AD cases, 46,613 proxy cases, and 676,386 controls, with a total sample size of 762,917 from Finland, Iceland, Norway, Spain, Sweden, U.K., U.S., and other regions (Additional file 1: **Table S1)[9]**. Schwartzentruber et al. 2021 included 21,982 AD cases, 53,000 proxy cases, and 419,944 controls, totaling 472,868 samples from Greece, Canada, U.S., U.K., France, and Germany (Additional file 1: **Table S1)[8]**. Bellenguez et al. 2022 provided data on 39,106 clinically diagnosed AD cases, 46,828 proxy cases, and 401,577 controls, amounting to 487,511 samples from Portugal, Switzerland, Spain, Greece, Czech Republic, Netherlands, Sweden, U.S., Belgium, Norway, Finland, Denmark, Italy, U.K., Bulgaria, France, and Germany (Additional file 1: **Table S1)[4]**.

We utilized multiple cis-eQTL datasets predominantly from individuals of European ancestry, including both tissue and cell type levels datasets in brain cortex (Additional file 1: **Table S2**). The Metabrain eQTL dataset offered a tissue-level cis-eQTL dataset derived from a meta-analysis of 14 bulk RNA-seq datasets focused on the brain cortex[10] (see Additional file 1: **Table S2**). Additionally, three cell type-specific cis-eQTL datasets were obtained from single-cell sequencing data. The research conducted by Fujita et al. 2024 provided a cell type-specific eQTL dataset sourced from the dorsolateral prefrontal cortex (DLPFC) (Additional file 1: **Table S2**)[14]. Furthermore, Bryois et al. 2021 provided a cell type-specific eQTL dataset encompassing the temporal cortex, cortex, white matter, DLPFC, and prefrontal cortex (PFC)[11] (see Additional file 1: **Table S2**). Moreover, we performed eQTL analysis and generated a cell type-specific eQTL dataset utilizing the snRNA dataset from the DLPFC region as reported by the previous study from Mathys et al., 2023[20] (Additional file 1: **Table S2**).

eQTL analysis

To conduct eQTL analysis for the Mathys et al., 2023 snRNA dataset from the ROSMAP cohort, we generated pseudobulk expression profiles. We focused on seven main cell types (Excitatory neurons, Inhibitory neurons, Oligodendrocytes, Oligodendrocyte Progenitor Cells (OPCs), Astrocytes, Immune cells, Vasculature cells). Pseudobulk UMI count matrices for each cell type were generated by summing UMI counts per gene across all cells within each individual using Seurat (Version 5.0.1). Low-expression genes were filtered out using the `filterByExpr` function from edgeR (version 3.40.2) with default parameters. The remaining pseudobulk counts were normalized using the trimmed mean of M-values (TMM) method from edgeR, and log2 counts

per million (CPM) were computed and then quantile normalized with the `voom` function from limma (version 3.54.2) as a previous study[14].

To identify cis-eQTLs within 1 Mb of the transcription start site of each gene, we used Matrix EQTL (version 2.3) for analysis. Bi-allelic SNPs were retained if they had a minor allele frequency >0.05, a call rate >95%, and Hardy-Weinberg equilibrium p > 10^-6 using PLINK2 as a previous study[14]. Gene expression was modeled using a linear regression with SNP allele counts and several covariates, and significance was determined by t-statistics. To account for population structure, the top 3 genotype principal components (PCs) were included as covariates as a previous study[14]. Additionally, the top 40 expression PCs, calculated within each cell type, were used to control for non-genetic structure as . Covariates including age, sex, post-mortem interval, study cohort (ROS or MAP), and total number of genes detected were also included as a previous study[14].

Summary data-based Mendelian Randomization

We performed SMR and Heterogeneity in Dependent Instruments (HEIDI) tests to investigate pleiotropic associations between gene expression and AD within cis-regions, using the SMR software tool (version 1.3.1). The SMR method, as detailed in the previous study[21], enables the testing of whether the effect size of a SNP on a phenotype is mediated through gene expression. This tool facilitates the prioritization of candidate causal genes underlying GWAS hits for further functional studies by leveraging summary-level data from both GWAS and eQTL datasets (as mentioned above). For our analysis, we used default parameters in the SMR software with a p-value threshold of 5.0e-8 to select the top associated eQTLs for the SMR test, focusing

exclusively on cis-regions. The HEIDI test, which assesses heterogeneity among dependent instruments, was conducted using a default eQTL p-value threshold of 1.57e-3 to filter SNPs for each probe, corresponding to a chi-squared value (df =1) of 10. The association between gene expression and AD was determined as P-value of SMR < 0.05/number of probes tested[21]. For the HEIDI test, significance was determined as P-value of HEIDI > 0.05 as previous studies[21].

Bayesian colocalization analysis

We conducted colocalization analysis using the Coloc package (version 5.2.3)[22] to investigate whether AD phenotype and gene expression share common causal variants in a given region. The input data comprised SNP effect sizes and associated p-values from both the AD GWAS and eQTL datasets (as mentioned above), formatted according to the package's requirements. Using the coloc.abf function in the package, we tested the hypothesis of a shared causal variant under the assumption of at most one causal variant per trait. Colocalization analysis calculates posterior probabilities (PPs) of the five hypotheses: 1) PPH0; no association with either gene expression or phenotype; 2) PPH1; association with gene expression, not with the phenotype; 3) PPH2; association with the phenotype, not with gene expression; 4) PPH3; association with gene expression and phenotype by independent SNVs; and 5) PPH4; association with gene expression and phenotype by shared causal SNVs. As a large PP for H4 strongly supports shared causal variants affecting both gene expression and phenotype, we considered PPH4 > 0.75 and PPH4/PPH3 >3 as strong evidence for colocalization as previous studies[23].

Network analysis of cell type specific candidate causal genes

For each cell type, we utilized the identified candidate causal genes as input to construct a cell type-specific protein–protein interaction (PPI) network. This network was generated using STRING (version 12.0) with a confidence score threshold of 0.4 as the minimum required interaction score and default settings for all other parameters. The resulting network was then visualized using Cytoscape (version 3.10.2). In the network, nodes represent genes, proteins, or other molecular entities, while edges illustrate the interactions between these molecules.

Pathway enrichment of all candidate causal genes

To perform pathway enrichment analysis, we utilized the all-candidate causal genes in Metascape v3.5.20240101[24]. We conducted pathway and process enrichment analyses using various ontology sources, including KEGG Pathway, GO Biological Processes, Reactome Gene Sets, Canonical Pathways, CORUM, WikiPathways, and PANTHER Pathway. The entire genome was used as the background for enrichment calculations. Terms with a p-value < 0.01, a minimum count of 3, and an enrichment factor > 1.5 (where the enrichment factor is the ratio of observed to expected counts) were selected for further analysis. To group similar terms, we calculated kappa similarity between enriched term pairs and performed hierarchical clustering based on kappa scores. Clusters were defined with a similarity threshold > 0.3. The most statistically significant term within each cluster was identified to represent that cluster. P-values were determined using the cumulative hypergeometric distribution, and q-values were adjusted for multiple comparisons using the Benjamini-Hochberg procedure in Metascape. We showed the top 10 clusters with their representative enriched terms (one per cluster) in the results.

eQTplot analysis for visualizing colocalization

We utilized the eQTpLot (version 0.0.0.9000) R package to visualize the colocalization between AD GWAS data and eQTL data[25]. This tool enables comprehensive visualization of gene-trait interactions by generating a series of customizable plots. Using eQTpLot, we produced visualizations that highlight the overlap between AD GWAS and eQTL signals, the correlation between their p-values, and the enrichment of eQTLs among trait-significant variants. Additionally, the tool provided insights into the linkage disequilibrium (LD) landscape of the locus and the relationship between the directions of effect for eQTL signals and colocalizing GWAS peaks, which help us to better understand the genetic relationships between gene expression and AD.

Cell-type-specific enhancer activity analysis

GWAS risk variants located in noncoding regions can influence phenotypic outcomes by affecting transcriptional gene promoters and enhancers[19]. Clusters of enhancers, known as super-enhancers, play a vital role in regulating cell-identity genes and are key to establishing cell-type-specific gene expression patterns[19]. In this study, we evaluated the impact of disease variants on cis-gene expression in specific cell types by evaluating whether disease variants are located within or next to regulatory elements, including enhancers and promoters. A previous study highlights that although active promoters are typically conserved across different brain cell types, active enhancers show marked cell-type specificity[19]. Thus, we focused on variant-enhancer analysis. We used a publicly available dataset, including ATAC-seq, which identifies open chromatin regions, and ChIP-seq, which marks active enhancers (H3K27ac) and promoters (H3K4me3) for each brain cell type, accessed through the UCSC genome browser session

(hg19). This dataset was generated from nuclei isolated from brain tissue resected during epilepsy treatment in 10 individuals[19]. This approach allowed us to identify which enhancers are active in specific cell types, thereby elucidating the cell-type-specific effects of disease variants on gene expression.

Druggability analysis

To identify druggable genes, we classified the identified candidate causal genes into three tiers based on druggability confidence according to a previous study[26]. Tier 1 included genes whose protein products are targets of approved small molecule, and biotherapeutic drugs were identified using manually curated efficacy target information from release 17 of the ChEMBL database. Tier 2 comprised proteins closely related to Tier 1 targets, identified through a BLASTP search of Ensembl peptide sequences against approved drug efficacy targets. Tier 3 encompassed proteins with more distant relationships to drug targets, identified by BLASTP with ≥25% identity over ≥75% of the sequence and E-value ≤0.001. Additionally, to prioritize alternative targets for non-druggable candidate causal genes, we utilized data from EpiGraphDB to identify directly AD related interacting genes that are indicated to be druggable with Tier1 druggability[27] based on protein-protein interaction (PPI) networks (IntAct[28], STRING[29]) and with literature or xQTL evidence for AD relevance[27].

Potential drug/compound prediction

To identify potential pharmacological drug/compound that could modulate the expression of candidate causal genes for AD, we utilized the Drug Signatures Database (DSigDB)[30]. This resource includes 22,527 gene sets and 17,389 unique compounds linked to 19,531 genes. We

accessed and downloaded the annotated drug/compound gene sets from DSigDB's official website[31]. Using the enricher function from the R package clusterProfiler (version 4.10.1), we performed enrichment analysis to explore connections between our target genes—either druggable causal genes or tier 1 interacting genes—and potential drugs, aiming for AD drug repurposing. We set an Benjamini-Hochberg adjusted p-value threshold of <0.01 to identify drugs significantly associated with these target genes. The top 10 enriched drugs/compounds were visualized using a dot plot, and an interaction network was generated with Cytoscape (version 3.10.2) to illustrate the relationships between the target genes and the enriched drugs/compounds.

Results

Workflow

To identify and prioritize genes associated with AD, we integrated summary-level data from GWAS with eQTL data. As shown in **Figure 1**, we incorporated data from five recent AD GWAS datasets and three cell type-specific eQTL datasets from single-cell sequencing of AD brain samples, along with a tissue-level Metabrain eQTL dataset from previous research, as described in the **Methods**. As outlined in **Figure 1**, we first employed SMR to evaluate how SNPs associated with AD risk influence gene expression. Subsequently, we used Coloc to validate the colocalization of genetic variants within specific genomic regions. We identified 33 candidate causal genes that met our rigorous criteria (**Figure 2)**. These genes were then examined across multiple cell type-specific datasets to assess their replicability. We explored how associated variants might regulate gene expression in a cell type-specific manner, utilizing previous data on cell type-specific enhancers or promoters in brain tissue. Additionally, we compared our findings

with prior studies to highlight novel candidate genes with less previous support as shown in **Figure 1**. For these novel genes, we visualized colocalization results and derived differential gene expression data from earlier studies to confirm their association with AD. Finally, we assessed the druggability of the prioritized candidate causal genes to explore potential therapeutic targets.

Summary results of detected candidate causal genes

We integrated data from five recent AD GWAS datasets and three cell type-specific eQTL datasets obtained from single-cell sequencing of AD brain samples, along with a metabrain tissue-level eQTL dataset from prior research. Utilizing SMR and HEIDI as well as Coloc analyses, we identified 33 candidate causal genes across these datasets that met the filtering criteria: SMR FDR < 0.05, HEIDI p-value > 0.05, Coloc PPH4 < 0.75, and Coloc PPH4/PPH3 > 3, as shown in **Figure 2** and **Additional file 1: Table S3-S6**. Out of the 33 candidate causal genes, two (AL355353.1 and AL137789.1) are lncRNA genes, while the remaining 31 genes are mRNA genes. 27 candidate causal genes were observed in cell type-specific eQTL datasets, combining results from all GWAS datasets, as shown in **Figure 2**. As shown in **Additional file 2: Figure S1-S5**, the Bellenguez AD GWAS summary statistics revealed the highest number of candidate causal genes compared to the other AD GWAS datasets. With the combined results from all GWAS datasets, of the 27 cell type level candidate causal genes, 21 were found to be causal in only one cell type (**Figure 2)**. While genes including *ACE, CD2AP, JAZF1, APH1B*, *ARL17B* and *SCIMP* were shared across multiple cell types, as shown in **Figure 2**. The majority consistently show the same sign in their MR beta values across different cell types. For example, *CD2AP* was detected with a positive MR beta value in both excitatory neurons and microglia in

the Fujita eQTL dataset (**Figure 2)**. Interestingly, there is one gene, *JAZF1*, that exhibits an inconsistent MR beta value sign across different cell types. Specifically, *JAZF1* shows a negative MR beta value in microglia in all the Fujita, Mathys and Bryois eQTL datasets (**Figure 2)**. However, it displays a positive MR beta value in OPCs in the Fujita eQTL dataset (**Figure 2)**. Furthermore, we noted concordant MR beta signs across single-nucleus eQTL and bulk eQTL datasets. *CD2AP*, *EGFR*, *SNX31*, *PABPC1*, *ACE*, *ARL17B*, *APH1B*, *PRSS36*, *GRN*, and *LRRC37A* are genes that are shared between the metabrain and cell type level candidate causal genes (**Figure 2)**. The MR values of these genes consistently displayed the same sign in both the metabrain dataset and the cell type level dataset (**Figure 2)**. Additionally, *TSPAN14*, *SLC39A13*, *FCER1G*, *CR1*, *NDUFAF6*, *TP53INP1* were identified exclusively as candidate causal genes in the bulk metabrain eQTL dataset (**Figure 2)**. 17 genes were identified exclusively as candidate causal genes in the cell type eQTL datasets (**Figure 2)**.

Table 1. Novel discoveries, and functional analysis of candidate causal genes

-	**Causal genes (combined results with 5 GWAS summary data)**	
Celltypes	**Identified in 1 snRNA dataset**	**Identified in at least 2 snRNA datasets**
Astrocytes	*SCIMP, HS3ST5, KANSL1*	*EGFR, SNX31, PABPC1*
Excitatory Neurons	*APH1B, GRN, PRSS36, AL355353.1, ACE, LRRC37A, CD2AP*	*SCIMP*
Immune Cells or Microglia	*ZYX, CCDC6, RIN3, ARL17B, FERMT2, CD2AP*	*USP6NL, CASS4, PICALM, JAZF1, RABEP1, BIN1*
Inhibitory Neurons	*CELF1, ACE*	*SCIMP*
Oligodendrocytes	*MINDY2, AL137789.1, APH1B*	
OPCs	*ARL17B, JAZF1*	

As mentioned earlier, a total of 27 candidate causal genes were observed in cell type-specific eQTL datasets (**Table 1**). The highest number of candidate causal genes was detected in microglia, followed by excitatory neurons, astrocytes, inhibitory neurons, oligodendrocytes, and OPCs (**Figure 2 and Table 1)**. We identified 10 cell type-specific candidate causal genes (*EGFR*, *SNX31*, *PABPC1*, *SCIMP*, *USP6NL*, *CASS4*, *PICALM*, *JAZF1*, *RABEP1* and *BIN1*), which were detected in at least two snRNA datasets (**Table 1**). Among these, genes *CASS4*, *PICALM*, *USP6NL*, *BIN1* and *RABEP1* were previously nominated by Agora[32, 33], while genes *JAZF1* and *SCIMP* were identified as colocalized genes in previous studies[14, 15]. *EGFR* is a recently prioritized causal gene with genetic regulation[4]. *SNX31* was identified as a colocalized gene in an earlier study with limited supporting evidence[34-36]. Additionally, *PABPC1*, located nearby *SNX31* emerged as a novel candidate causal gene with limited supporting evidence.

To analyze interactions among candidate causal genes for each cell type, we first constructed cell type-specific PPI networks as described in the Methods. Our PPI analysis revealed there are interactions among the corresponding proteins of the candidate causal genes in astrocytes and microglia, as illustrated in **Figure 3**. In astrocytes, we identified interactions among protein *PABPC1*, *EGFR*, and *KANSL1*, with *EGFR* serving as a central node that connects *PABPC1* and *KANSL1* (**Figure 3A)**. As shown in **Figure 3B,** the PPI network for microglia showed a more intricate interaction landscape, with 12 nodes and 16 edges. Specifically, *BIN1*, *FERMT2*, *PICALM*, *RIN3*, *CD2AP*, and *CASS4* were interconnected, indicating a complex network of interactions that could play a significant role in microglial functions related to AD.

To identify the enriched pathways and processes, we used Metascape to perform a comprehensive enrichment analysis of the 31 candidate causal mRNA genes. **Figure 3C** displays the Top 10 clusters with their representative enriched terms, one per cluster. As displayed in **Figure 3C and Additional file 1: Table S7**, the representative enriched term of the top 1 cluster was "regulation of aspartic-type peptidase activity" (GO term) with q-value < 0.05 (q-value: 0.00178). The other clusters are not significantly enriched with q-value > 0.05.

Visualization of colocalization for the novel astrocyte-specific candidate causal gene

We used eQTpLot to visualize the colocalization between eQTL (Astrocyte specific eQTL from Fujita et al 2024) and AD GWAS (Bellenguez et al., 2022) signals for the novel candidate causal gene, PABPC1. As shown in **Figure 4A-C**, *PABPC1* is indicated to be affected by the lead GWAS significant loci rs1693551 (GWAS P-value: 1.785e-08; Beta: 0.0459 from Bellenguez et al., 2022 AD GWAS summary statistics data). Our analysis indicates that rs1693551 may also affect the other nearby gene *SNX31* **(Figure 4B)**. We observed a tendency for eQTL to be overrepresented in the lists of significant variants from the AD GWAS (p-value = 4e-5 for *PABPC1* in astrocyte) (**Figure 4D).** Congruous SNPs effect on the gene expression in astrocyte and AD risk were also observed for *PABPC1* (**Figure 4A, 4E, 4F)**. eQTpLot P-value correlation analysis further confirms the colocalization between the *PABPC1* gene expression in astrocyte and AD risk as shown in **Figure 4E (**r = 0.81, p = 1.36e-49**).** The variant rs1693551 with reference allele of T and alternative allele of C is not identified as a new risk locus in the latest GWAS study[4]. However, our analysis reveals that it surpasses the genome-wide significance threshold, as illustrated by the Manhattan plot for chromosome 8 shown in **Additional file 2: Figure S6**. Additionally, we also observed colocalization of shared causal variant for *PABPC1*

gene expression and AD risk with eQTL datasets from Mathys et al 2023 and Metabrain **(Additional file 2: Figure S7 and S8).** We also visualized the colocalization for the causal gene *EGFR* in astrocyte (**Additional file 2: Figure S9**).

The MR and colocalization analyses identified a causal link between *PABPC1* gene expression in astrocytes and AD risk. To further explore this relationship, we examined *PABPC1* expression in both astrocytes and astrocyte subtypes, and its association with AD pathology and cognitive function. Specifically, we utilized differential gene expression (DEG) results from a previous study[20] focused on the DLPFC region and applied multiple testing corrections. The findings, presented in **Additional file 2: Figure S10**, indicate that *PABPC1* expression in astrocytes is significantly associated with perceptual orientation. Additionally, expression in the astrocyte subtype GRM3 shows a suggestive association with tangle density.

Enhancers harboring AD risk variants regulate cell-type-specific gene expression

Our results reveal that certain genes, such as *PABPC1*, was identified as candidate causal gene exclusively in astrocytes, but not in other brain cell types. This highlights that many candidate causal genes may be specific to a single cell type. To further understand this cell-type-specific effect, it is crucial to investigate how these variants influence gene expression and the underlying regulatory mechanisms. Enhancers are genomic regions that regulate gene expression, often in a cell-specific manner. A previous study [19] analyzed enhancer and promoter activity in human brain cell nuclei, revealing that genetic variants associated with brain traits and diseases exhibit cell-specific enhancer enrichment patterns. To determine if the cell-type-specific causal genes identified in our study are regulated by cell-type-specific enhancer activity, we analyzed a

publicly available dataset, including ATAC-seq for open chromatin regions and ChIP-seq for active enhancers (H3K27ac) and promoters (H3K4me3) for each brain cell type, as detailed in the Methods section.

As illustrated in **Figure 5**, for the candidate causal gene *PABPC1* in astrocytes, the associated disease variant is rs1693551 (chr8, hg19_position: 10675584 bp), which is located just 59 bp from the boundary of an astrocyte-specific enhancer (chr8, hg19_position: 101675643-101676301 bp) identified in the previous study. Given its proximity to the enhancer boundary, it is possible that the enhancer region extends beyond what was detected, especially considering the dynamic nature of enhancers and technical limitations of current detection methods. **Figure 5** shows that this enhancer, located downstream of the *PABPC1* gene, is active only in astrocytes—evidenced by prominent H3K27ac and ATAC-seq peaks—while not active in other cell types. This suggests that the variant likely influences gene expression through a cell-type-specific enhancer, which may explain why *PABPC1* was detected as a causal gene exclusively in astrocytes.

Druggability analysis and drug/compound prediction

To identify druggable genes from our candidate causal genes, we categorized them based on a prior drug tier classification[26]. Tier 1 includes targets of approved drugs and clinical candidates; Tier 2 includes targets with known drug-like interactions or high similarity to approved drug targets; and Tier 3 includes proteins with distant similarities to drug targets or those in key druggable families, as mentioned in the Methods. As detailed in **Additional file 1: Table S8**, we identified three candidate causal genes—*EGFR*, *ACE*, and *APH1B*—as Tier 1

449 druggable, and three genes—*GRN*, *PRSS36*, and *CR1*—as Tier 3 druggable. The remaining
450 candidate causal genes were not classified as druggable based on the previous study[26]. For
451 these non-druggable genes, we used EpiGraphDB to prioritize potential alternative drug targets
452 within the same PPI network. We identified directly AD related interacting genes with Tier 1
453 druggability using PPI networks from IntAct and STRING databases, shown in **Additional file**
454 **1: Table S8**.
455
456 To identify drugs targeting the causal genes identified in this study and to broaden the scope of
457 potential drug targets, we conducted a drug/compound enrichment analysis using DSigDB. This
458 analysis aimed to find potential drugs for 74 target genes, which include both the druggable
459 causal genes identified in this study and directly interacting genes with Tier 1 druggability, as
460 detailed in **Additional file 1: Table S8**. The results of the enrichment analysis are presented in
461 **Additional file 1: Table S9**. We focused on drugs with an adjusted p-value of less than 0.01 and
462 selected the top 10 most significant potential drugs/compound based on their adjusted p-value
463 (**Additional file 1: Table S9 and Figure 6A**). **Figure 6A** presents the drugs grouped by gene
464 ratio (the percentage of target genes overlapping with the drug gene set). Within each group, the
465 drugs are ranked by their adjusted p-value significance. The results highlight that 3-(1-
466 methylpyrrolidin-2-yl)pyridine targets the highest number of genes, with 17 target genes
467 including *EEF2*, *ADRB2*, *CD4*, *EGFR*, *APP*, *TFRC*, *ITGAL*, *PLD1*, *FYN*, *PIK3CA*, *RAF1*, *SRC*,
468 *TP53*, *VEGFA*, *MAPK1*, *TNFRSF1A*, and *ACE* (**Additional file 1: Table S9**). In the second
469 group, Imatinib mesylate is the most significant drug, targeting 14 genes, followed by
470 Dinoprostone and Capsaicin. In the third group, histamine is the most significant drug, targeting
471 13 genes, followed by Gefitinib. Imatinib mesylate is detected as the most significant drug across

groups. These top 10 enriched drugs (**Figure 6A**) show promise for therapeutic applications in AD and need further investigation.

To illustrate the interactions between drugs and target genes—both causal genes identified in this study and directly interacting genes (AD related) with Tier 1 druggability—we constructed an interaction network using Cytoscape, as shown in **Figure 6B**. This network highlights that Tier 1 druggable genes, such as *EGFR* (targeted by all top 10 drugs) and *ACE* (targeted by 5 of the top 10 drugs) (**Additional file 1: Table S9** and **Figure 6B**), are directly targeted by multiple drugs. Additionally, the Tier 3 druggable gene *CR1* is directly targeted by Imatinib mesylate. In the network, druggable and non-druggable causal genes are represented by blue circles; interacting genes are shown in green circles, and drugs/compounds are depicted in pink (**Figure 6B**). The central area of the network features drugs and Tier 1 druggable genes, indicating direct targeting, while the surrounding groups represent interacting genes and non-druggable causal genes, which are indirectly targeted through these interactions. This visualization demonstrates the role and significance of the top 10 drugs in targeting multiple causal genes, both directly and indirectly (**Figure 6B**).

Discussion

Many disease-associated loci exert effects that are specific to cell types[11, 14, 37, 38]. Brain diseases are influenced by genetic effects that are specific to both cell types and brain regions[11, 14, 39]. Previous GWAS studies often identify risk variants that impact disease phenotypes by regulating genes in specific tissues, yet the precise cell types involved are often not well characterized[10, 40]. Our study addresses this knowledge gap by using brain single-cell eQTL

data to reveal how genetic variants impact AD at the cellular level, offering crucial insights into cell-type-specific regulation driving the disease. In this study, we combined data from five recent AD GWAS with three cell-type-specific eQTL datasets from single-cell RNA sequencing and one bulk tissue eQTL dataset from a prior meta-analysis. Through SMR and colocalization analyses, we identified candidate causal genes at both bulk and cell-type levels, uncovering novel genes and confirming known ones. We investigated gene regulation in specific cell types by analyzing enhancer activity using previous H3K27ac and ATAC-seq data. Network and pathway enrichment analyses provided additional insights into the biological relevance of these genes. To facilitate drug repurposing for AD, we performed a drug/compound enrichment analysis using the DSigDB, mapping drug interactions with both causal and interacting druggable genes. This integrated approach highlights the importance of cell-type specific functional evidence in genetic research, revealing how AD GWAS variants contribute to disease through cell-specific gene expression. By examining genetic effects at the cellular level, we gain clearer insights into AD molecular mechanism and identify promising targets for drug discovery.

In recent years, there has been growing recognition of the context-specific nature of eQTLs, extending from tissue types to functional, environmental, and cellular contexts[11, 14, 41-43]. Our study underscores the critical value of cell-type-specific eQTL datasets in identifying candidate causal genes for AD. Specifically, we identified 17 genes exclusively as candidate causal genes within the cell-type eQTL datasets (**Figure 2**). This finding highlights the limitations of bulk tissue analyses, which often aggregate signals across various cell types and may miss gene-regulatory effects that are specific to cellular contexts. By focusing on cell-type-specific eQTL data, we can uncover gene associations that are masked when only bulk tissue

data is used. Furthermore, of the 27 candidate causal genes identified through cell-type-specific eQTL datasets, 21 were found to be causal in only one cell type (**Figure 2**). This cell-type specificity highlights the importance of considering cellular heterogeneity in genetic studies of complex diseases like AD.

Our study reveals that the gene JAZF1 exhibits discordant MR beta value signs across different cell types. Specifically, JAZF1 shows a negative MR beta value in microglia and a positive MR beta value in OPCs (**Figure 2**). The negative MR beta value in microglia aligns with the known downregulation of JAZF1 in multiple brain regions[44]. This discrepancy could be attributed to technical limitations, as OPCs are less prevalent in brain single-cell datasets, leading to less reliable expression measurements. However, it is also possible that the discordant MR values reflect distinct functional roles of JAZF1 in these cell types. Microglia plays a key role in immune responses and neuroinflammation, while OPCs are critical for oligodendrocyte maturation and myelination[45, 46]. The differential impact of JAZF1 on these processes could explain its varied effects across cell types. Future research should focus on validating findings in independent datasets to resolve this discordancy.

In our analysis, *PABPC1* emerged as a novel candidate causal gene for AD, highlighting its potential role in disease mechanisms. Specifically, the MR and colocalization analyses identified a causal link between *PABPC1* gene expression in astrocytes and AD risk. We found that *PABPC1* expression in astrocytes is significantly linked to perceptual orientation and shows a suggestive association with tangle density in the GRM3 astrocyte subtype. *PABPC1* is known to bind tau proteins[47]. It also regulates translation and mRNA stability[48]. Additionally,

PABPC1 is involved in stress granules and RNA splicing, critical for managing cellular stress and maintaining protein synthesis[49]. Its associations with neurofilament light chain (NF-L)[50], along with its co-localization with small tau inclusions in tauopathy[51], underscore its relevance in AD pathology. These findings warrant further investigation into *PABPC1* as a potential therapeutic target. The AD risk loci rs1693551, which achieved GWAS significance only in the latest AD GWAS summary statistics[4], has been less studied. It is the leading GWAS locus associated with the expression of the causal genes SNX31 and PABPC1 in astrocytes, underscoring its potential significance in AD. This highlights the need for further investigation into its role and relevance in the disease.

In our results, 21 of the 27-cell type level candidate causal genes were found to be causal in only one specific cell type (**Figure 2)**. Previous research indicates that cell-type-specific enhancers harboring AD-risk variants can drive such cell type-specific gene regulation[19]. For example, while *PICALM* and *BIN1* are expressed in multiple cell types, they contain microglia-specific enhancers with AD-risk variants[19]. Consistent with the previous findings, our study reveals astrocyte-specific enhancers harboring AD-risk variants associated with *PABPC1* gene expression, although interactions between these enhancers and gene promoters remain unconfirmed due to the lack of PLAC-seq data in astrocytes[19]. In addition to microglia, which are well-known for their roles in AD, our study highlights the importance of astrocytes. We provide more molecular evidence showing that astrocytes are critically involved in AD through specific gene expression and enhancer activity associated with AD-risk variants.

Our DSigDB enrichment analysis identified several drugs/compounds with potential therapeutic relevance for AD, including Imatinib mesylate, histamine, Dinoprostone, 3-(1-methylpyrrolidin-2-yl)pyridine, Gefitinib, Crystal violet, cerivastatin, and hexachlorophene. Imatinib mesylate was highlighted as the most significant drug (**Additional file 1: Table S9**). Imatinib mesylate is notable for its role as a tyrosine kinase inhibitor and has been shown to reduce Aβ production in various experimental models[52]. Research suggests it may be effective in treating neurodegenerative disorders, including AD[53]. However, further studies are needed to fully understand its effects on the brain, particularly its ability to cross the blood-brain barrier. Some research has explored how imatinib interacts with brain transporters such as breast cancer resistance protein and P-glycoprotein[54], which is important for optimizing its use in neurodegenerative diseases. 3-(1-methylpyrrolidin-2-yl)pyridine (Nicotine) stands out for targeting the highest number of analyzed genes. Nicotine, an alkaloid in tobacco, functions by activating nicotinic acetylcholine receptors (nAChRs), which are widely expressed throughout the nervous system[55]. It has dual effects on oxidative stress and neuroprotection[56], suppresses neuroinflammation[57], and prevents Aβ aggregation[58]. Despite these benefits, its use in AD is limited by cardiovascular risks[59], addiction and negative associations with smoking[60]. However, Nicotine's gene targeting profile found in this study suggests it could impact multiple pathways involved in AD, potentially offering a therapeutic approach through nicotinic derivatives that mitigate these adverse effects.

There are several limitations in this study. The study incorporated multiple datasets, including the three cell-type-specific eQTL datasets with partial overlap of participants from the ROSMAP cohort (see **Additional file 1: Table S2**). This partial overlap may introduce biases, potentially

affecting the robustness of our findings. Furthermore, the study analyzed data from various brain regions across multiple datasets, including the cortex from the bulk metabrain eQTL dataset, the DLPFC region from the Fujita 2024 and Mathys_2023 snRNA eQTL datasets, and a range of regions such as the temporal cortex, white matter, and PFC from the Bryois 2021 snRNA eQTL dataset. The variability in brain regions might limit the generalizability of our findings, as genetic effects can be region-specific. Also, the GWAS and eQTL datasets primarily included individuals of European ancestry, which limits the generalizability of the findings to other ethnic groups. Additionally, our analysis was limited to cis-eQTLs, which reflect direct effects on genes. Cis-eQTLs do not capture the full spectrum of genetic influences, as trans-eQTLs could reveal downstream gene sets and pathways affected by disease variants. Future studies should explore available cell-specific trans-eQTL data to better understand the causal effects of genetic variants acting in trans. Furthermore, future research should use independent snRNA eQTL datasets for validation. Lastly, while our study identified potential drug targets through enrichment analysis, their clinical efficacy remains unconfirmed. Experimental validation and clinical trials are necessary to establish their therapeutic potential. Moreover, since the candidate causal genes were identified from brain tissue data and in drugs that face challenges in crossing the blood-brain barrier, further investigation is needed to evaluate the viability of these targets for drug development. In addition, despite the common challenge that smaller gene sets pose in pathway enrichment analysis due to reduced statistical power, our results with 31 input genes demonstrate that meaningful enrichments can still be detected. As shown in **Additional file 1: Table S7**, the p-value of 8.13×10^{-8} of the one significantly enriched pathway (regulation of aspartic-type peptidase activity) indicates a highly significant enrichment, suggesting that the observed pathway association is unlikely to have occurred by chance. Furthermore, the q-value

of 1.78 × 10^(-3) of this enriched pathway confirms that the result is robust, with a very low false discovery rate, even after correcting for multiple testing. These findings indicate that, while larger gene sets generally provide more power, our analysis can still yield reliable, statistically significant results when the genes are biologically relevant.

Our analysis identified both novel and established candidate causal genes, elucidating their roles in AD molecular mechanisms and highlighting the significance of cell-type specificity in gene expression regulation and enhancer activity.

Declarations

Availability of data and materials

GWAS summary statistics for AD were downloaded from https://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST007001-GCST008000/GCST007511/, https://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST013001-GCST014000/GCST013197/, http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90027001-GCST90028000/GCST90027158/, http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90012001-GCST90013000/GCST90012877/, http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST007001-GCST008000/GCST007320/.
Publicly available summary statistics of metabrain eQTLs was obtained from MetaBrain website (https://www.metabrain.nl/). Fujita and Bryois Single cell eQTL datasets were obtained from Synapse: syn52335807 and https://doi.org/10.5281/zenodo.5543734, respectively. Mathys et al., 2023 snRNA dataset from ROSMAP cohort (downloaded from Synapse: syn52293442). The publicly available dataset, including ATAC-seq, which identifies open chromatin regions, and

636 ChIP-seq, which marks active enhancers (H3K27ac) and promoters (H3K4me3) for each brain
637 cell type, accessed through the UCSC genome browser session (hg19) at:
638 https://genome.ucsc.edu/s/nottalexi/glassLab_BrainCellTypes_hg19
639

640 **Competing interests**

641 A.S. has received support from Avid Radiopharmaceuticals, a subsidiary of Eli Lilly (in kind
642 contribution of PET tracer precursor) and participated in Scientific Advisory Boards (Bayer
643 Oncology, Eisai, Novo Nordisk, and Siemens Medical Solutions USA, Inc) and an Observational
644 Study Monitoring Board (MESA, NIH NHLBI), as well as several other NIA External Advisory
645 Committees. He also serves as Editor-in-Chief of Brain Imaging and Behavior, a Springer-Nature
646 Journal.
647 S. L., T. R., P. B., D. C., D. B., N. T., K. N., S. C., M. C., Y. H., and T. P. have no interest to
648 declare.
649 The funders had no role in the study's design, the collection, analyses, or interpretation of data,
650 the writing of the manuscript, or the decision to publish the results.
651

652 **Funding**

653 A.S. receives support from multiple NIH grants (P30 AG010133, P30 AG072976, R01
654 AG019771, R01 AG057739, U19 AG024904, R01 LM013463, R01 AG068193, T32 AG071444,
655 U01 AG068057, U01 AG072177, and U19 AG074879).
656 K.N receives support from NIH grants (R01LM012535, U01AG072177, and U19AG0748790).
657 U19AG074879).
658 S.C was supported by ADNI Health Equity Scholarship (ADNI HESP) a sub-award of NIA grant
659 (U19AG024904).

S.L was supported by CLEAR-AD Diversity Scholarship of U19AG074879.

Acknowledgements

We thank the participants of the ROS/MAP study for their valuable contributions and generous donation of brain samples. We also appreciate Dr. Masashi Fujita and Dr. Hansruedi Mathys for their assistance in accessing the single-cell datasets.

Figure legend.

Figure 1. Study workflow.

Figure 2. SMR beta value signs for candidate causal genes from SMR and colocalization analysis. Note: all five GWAS datasets results are combined. The candidate causal genes are filtered by SMR FDR < 0.05, HEIDI > 0.05, Coloc PPH4 < 0.75, Coloc PPH4/PPH3 > 3.

Figure 3. Candidate causal genes network analysis and pathway enrichment. A. STRING PPI network of Astrocyte candidate causal genes. B. STRING PPI network of Microglia candidate causal genes. C. Pathway enrichment of all 31 detected candidate causal (mRNA) genes

Figure 4. eQTpLot for colocalization between eQTLs for the gene PABPC1 and a GWAS signal for AD. The GWAS dataset is from Bellenguez et al., 2022 and the cell type eQTL dataset of astrocyte is from Fujita et al., 2024. **A** shows the locus of interest, containing the PABPC1 gene, with chromosomal space indicated along the horizontal axis. The position of each point on the vertical axis corresponds to the p-value of association for that variant with AD, while the color scale for each point corresponds to the magnitude of that variant's p-value for association with PABPC1 expression. Variants with congruous effects are plotted using a blue color scale,

683 while variants with incongruous effects are plotted using a red color scale. The directionality of
684 each triangle corresponds to the GWAS direction of effect, while the size of each triangle
685 corresponds to the effect size for the eQTL data. The default genome-wide p-value significance
686 threshold for the GWAS analysis, 5e-8, is depicted with a horizontal red line. **B** displays the
687 genomic positions of all genes within AD. **C** depicts a heatmap of LD information of all
688 PABPC1 eQTL variants, displayed in the same chromosomal space as panels A and B for ease of
689 reference (R2min=0.1, LDmin = 10). **D** depicts the enrichment of PABPC1 eQTLs among
690 GWAS-significant variants, while **E** and **F** depicts the correlation between P_{GWAS} and P_{eQTL} for
691 PABPC1 and AD, with the computed Pearson correlation coefficient (r) and p-value (p)
692 displayed on the plot. For **E**, the analysis is confined only to variants with congruous directions
693 of effect, while for **F** the analysis includes only variants with incongruous directions of effect. A
694 lead variant is indicated in both **E** and **F**, and both are also labeled in panel **A**.

695 **Figure 5. Brain cell-type-specific chromatin profiles by UCSC Genome Browser (hg19).** A.
696 H3K27ac and ATAC-seq data for *PABPC1*, showing active enhancer regions and open chromatin
697 specific to astrocytes, with a yellow vertical line marking the location of the associated disease
698 variant and a dashed square showing the region of active enhancer.

699 **Figure 6. Potential drugs enrichment analysis and gene-drug interaction network.** A. Top
700 10 enriched drug/compounds based on DSigDB predictions. B. Interaction network illustrating
701 connections between enriched drugs/compounds and target genes. Blue circles indicate
702 druggable/non-druggable causal genes identified in this study, green circles represent druggable
703 interacting genes linked to non-druggable causal genes, and pink nodes denote the top 10
704 enriched drugs/compounds.

705

Additional files

Additional file 1: Supplementary Tables

Table S1. Alzheimer's disease GWAS studies. Table S2. Brain cortex region cis-eQTL datasets. Table S3. SMR and Coloc analysis results for metabrain eQTL and AD GWAS summary statistics. Table S4. SMR and Coloc results for Bryois cell type specific eQTL and AD GWAS summary statistics. Table S5. SMR and Coloc results for Fujita cell type specific eQTL and AD GWAS summary statistics. Table S6. SMR and Coloc results for Mathys cell type specific eQTL and AD GWAS summary statistics. Table S7. Pathway enrichment of candidate causal genes. Table S8. Druggability of candidate causal genes. Table S9. Drug/compound enrichment analysis results.

Additional file 2: Supplementary Figures

Figure S1. SMR beta value and significance for candidate causal genes from SMR and colocalization analysis. Figure S2. SMR beta value and significance for candidate causal genes from SMR and colocalization analysis. Figure S3. SMR beta value and significance for candidate causal genes from SMR and colocalization analysis. Figure S4. SMR beta value and significance for candidate causal genes from SMR and colocalization analysis. Figure S5. SMR beta value and significance for candidate causal genes from SMR and colocalization analysis. Figure S6. Manhattan plot of AD GWAS (Bellenguez et al., 2022) on chromosome 8. Figure S7. eQTpLot for colocalization between eQTLs for the gene PABPC1 and a GWAS signal for AD. Figure S8. eQTpLot for colocalization between eQTLs for the gene PABPC1 and a GWAS signal for AD. Figure S9. eQTpLot for colocalization between eQTLs for the gene EGFR and a GWAS signal for AD. Figure S10: DEGs detection of PABPC1 with pathology and cognitive function.

729

730 ## References

731 1. Iqbal K, Grundke-Iqbal I. Alzheimer's disease, a multifactorial disorder seeking
732 multitherapies. Alzheimer's & Dementia. 2010;6(5):420-4.
733 2. Tanzi RE, Bertram L. New Frontiers in Alzheimer's Disease Genetics. Neuron.
734 2001;32(2):181-4.
735 3. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al.
736 Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer's Disease in Late Onset
737 Families. Science. 1993;261(5123):921-3.
738 4. Bellenguez C, Küçükali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, et al. New
739 insights into the genetic etiology of Alzheimer’s disease and related dementias. Nature Genetics.
740 2022;54(4):412-36.
741 5. Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al.
742 Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.
743 Nature Genetics. 2013;45(12):1452-8.
744 6. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-
745 analysis of diagnosed Alzheimer’s disease identifies new risk loci and implicates Aβ, tau,
746 immunity and lipid processing. Nature Genetics. 2019;51(3):414-30.
747 7. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-
748 wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease
749 risk. Nature Genetics. 2019;51(3):404-13.

8. Schwartzentruber J, Cooper S, Liu JZ, Barrio-Hernandez I, Bello E, Kumasaka N, et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. Nature Genetics. 2021;53(3):392-402.

9. Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. Nature Genetics. 2021;53(9):1276-82.

10. de Klein N, Tsai EA, Vochteloo M, Baird D, Huang Y, Chen C-Y, et al. Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. Nature Genetics. 2023;55(3):377-88.

11. Bryois J, Calini D, Macnair W, Foo L, Urich E, Ortmann W, et al. Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. Nature Neuroscience. 2022;25(8):1104-12.

12. Patel D, Zhang X, Farrell JJ, Chung J, Stein TD, Lunetta KL, et al. Cell-type-specific expression quantitative trait loci associated with Alzheimer disease in blood and brain tissue. Translational Psychiatry. 2021;11(1):250.

13. Lee B, Yao X, Shen L, for the Alzheimer's Disease Neuroimaging I. Integrative analysis of summary data from GWAS and eQTL studies implicates genes differentially expressed in Alzheimer's disease. BMC Genomics. 2022;23(4):414.

14. Fujita M, Gao Z, Zeng L, McCabe C, White CC, Ng B, et al. Cell subtype-specific effects of genetic variation in the Alzheimer's disease brain. Nature Genetics. 2024;56(4):605-14.

15. Pan S, Kang H, Liu X, Lin S, Yuan N, Zhang Z, et al. Brain Catalog: a comprehensive resource for the genetic landscape of brain-related traits. Nucleic Acids Research. 2023;51(D1):D835-D44.

16. Majewski J, Pastinen T. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. Trends in Genetics. 2011;27(2):72-9.

17. Trabzuni D, Ryten M, Walker R, Smith C, Imran S, Ramasamy A, et al. Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. Journal of Neurochemistry. 2011;119(2):275-82.

18. Consortium TG, Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318-30.

19. Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, et al. Brain cell type–specific enhancer–promoter interactome maps and disease-risk association. Science. 2019;366(6469):1134-9.

20. Mathys H, Peng Z, Boix CA, Victor MB, Leary N, Babu S, et al. Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer’s disease pathology. Cell. 2023;186(20):4365-85.e27.

21. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nature Genetics. 2016;48(5):481-7.

22. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLOS Genetics. 2014;10(5):e1004383.

23. Kia DA, Zhang D, Guelfi S, Manzoni C, Hubbard L, Reynolds RH, et al. Identification of Candidate Parkinson Disease Genes by Integrating Genome-Wide Association Study, Expression, and Epigenetic Data Sets. JAMA Neurology. 2021;78(4):464-72.

24. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523.

25. Drivas TG, Lucas A, Ritchie MD. eQTpLot: a user-friendly R package for the visualization of colocalization between eQTL and GWAS signals. BioData Mining. 2021;14(1):32.

26. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. Science Translational Medicine. 2017;9(383):eaag1166.

27. Liu Y, Elsworth B, Erola P, Haberland V, Hemani G, Lyon M, et al. EpiGraphDB: a database and data mining platform for health data science. Bioinformatics. 2021;37(9):1304-11.

28. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Research. 2014;42(D1):D358-D63.

29. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51(D1):D638-D46.

30. Yoo M, Shin J, Kim J, Ryall KA, Lee K, Lee S, et al. DSigDB: drug signatures database for gene set analysis. Bioinformatics. 2015;31(18):3069-71.

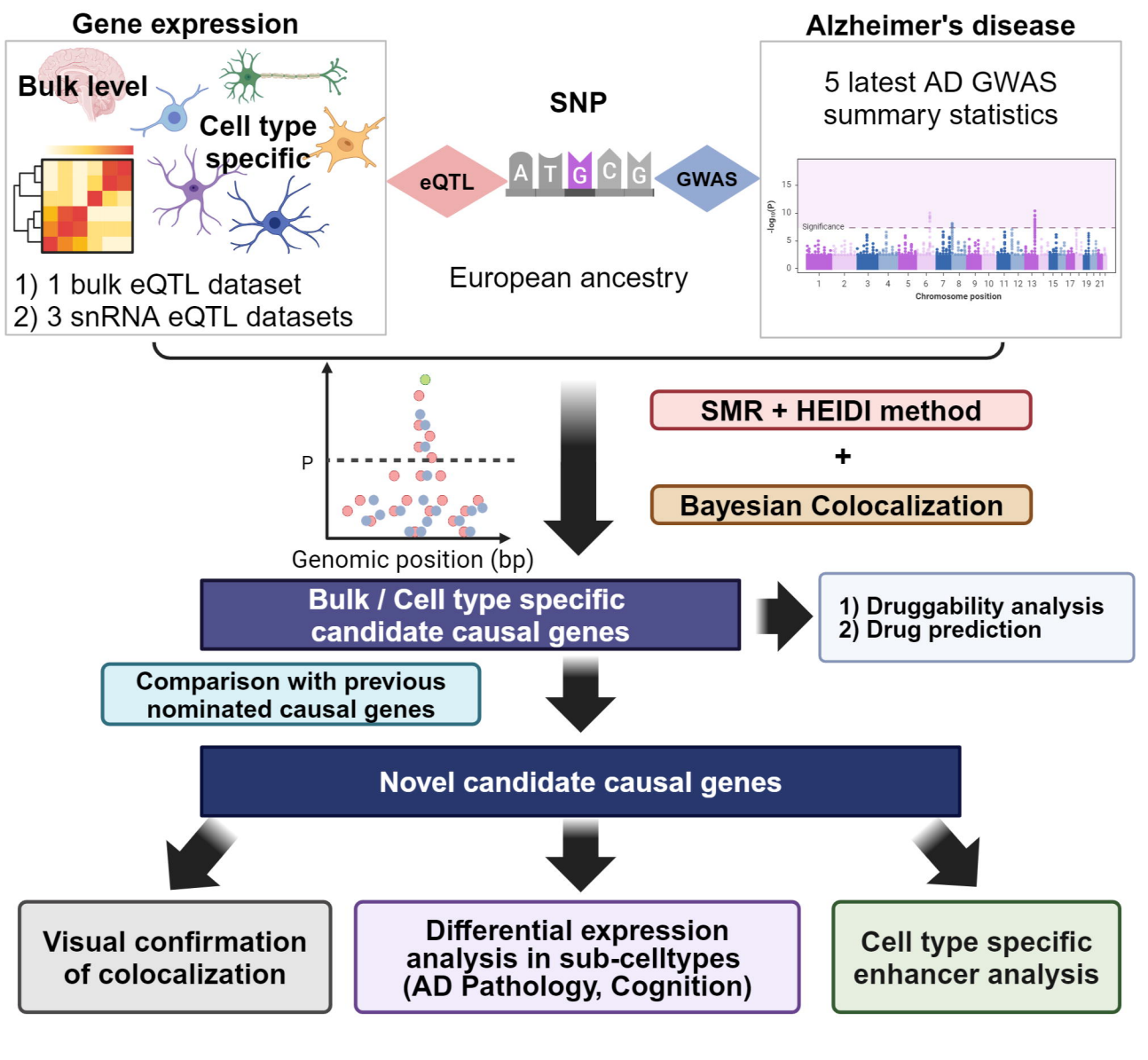
31. Drug SIGnatures DataBase. http://dsigdb.tanlab.org/DSigDBv1.0/. Accessed 20 Aug 2024.

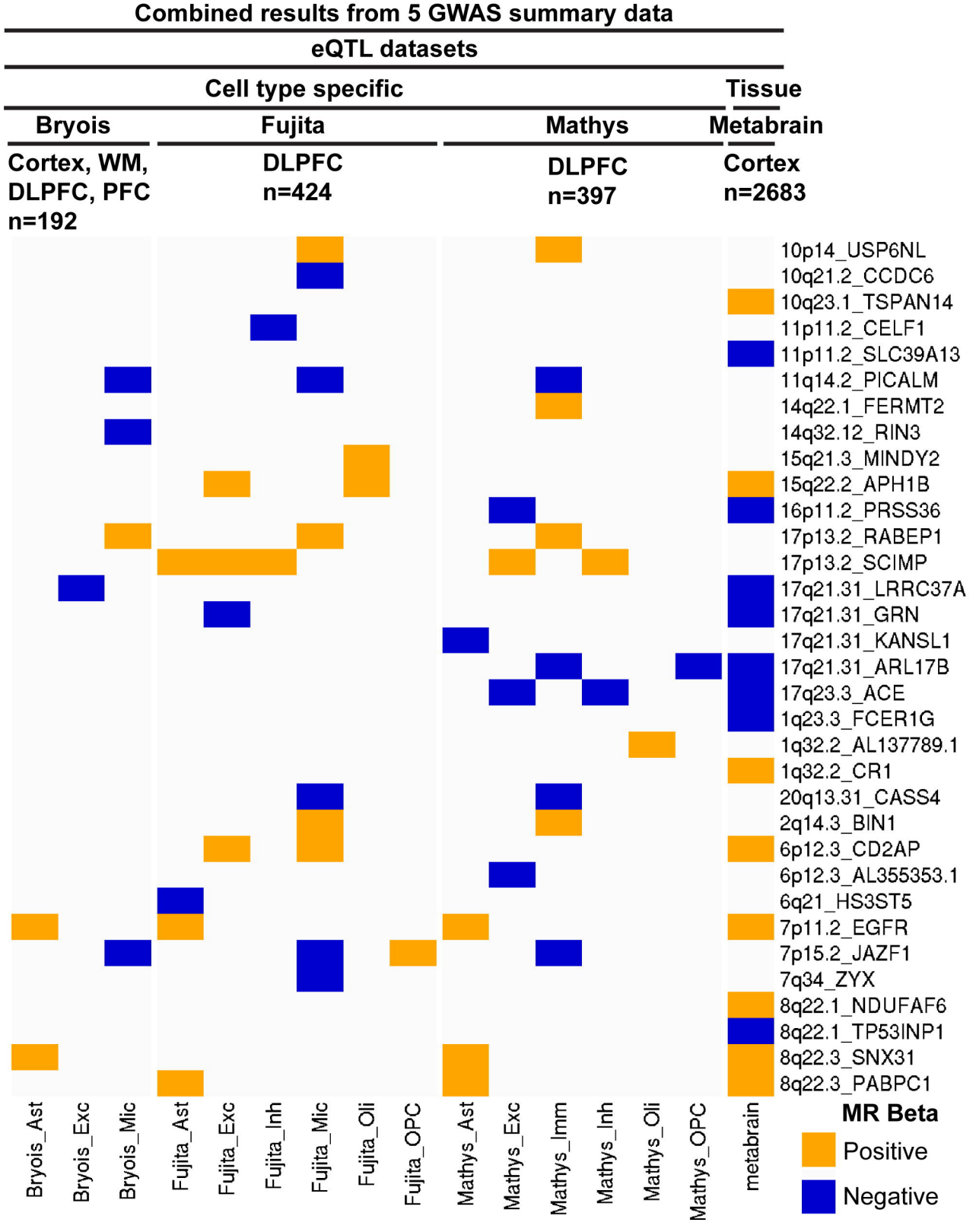
817 32. Hodes RJ, Buckholtz N. Accelerating Medicines Partnership: Alzheimer's Disease
818 (AMP-AD) Knowledge Portal Aids Alzheimer's Drug Discovery through Open Data Sharing.
819 Expert Opinion on Therapeutic Targets. 2016;20(4):389-91.
820 33. Agora AD, Knowledge Portal n.d. (accessed May 26, 2024) doi: 10.57718/agora-
821 adknowledgeportal.
822 34. Haglund A, Zuber V, Yang Y, Abouzeid M, Feleke R, Ko JH, et al. Single-cell Mendelian
823 randomisation identifies cell-type specific genetic effects on human brain disease and behaviour.
824 bioRxiv. 2022:2022.11.28.517913.
825 35. Küçükali F, Sleegers K, Eadb. Functional Interpretation of Alzheimer's Disease Genetic
826 Risk and Systematic Gene Prioritization. Alzheimer's & Dementia. 2022;18(S4):e069298.
827 36. Alvarado CX, Makarious MB, Weller CA, Vitale D, Koretsky MJ, Bandres-Ciga S, et al.
828 omicSynth: An open multi-omic community resource for identifying druggable targets across
829 neurodegenerative diseases. The American Journal of Human Genetics. 2024;111(1):150-64.
830 37. Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, et al.
831 Single-cell eQTL mapping identifies cell type–specific genetic control of autoimmune disease.
832 Science.376(6589):eabf3041.
833 38. Siletti K, Hodge R, Mossi Albiach A, Lee KW, Ding S-L, Hu L, et al. Transcriptomic
834 diversity of cell types across the adult human brain. Science.382(6667):eadd7046.
835 39. Pyun J-M, Park YH, Wang J, Bennett DA, Bice PJ, Kim JP, et al. Transcriptional risk
836 scores in Alzheimer's disease: From pathology to cognition. Alzheimer's & Dementia.
837 2024;20(1):243-52.
838 40. Qi T, Wu Y, Fang H, Zhang F, Liu S, Zeng J, et al. Genetic control of RNA splicing and
839 its distinct role in complex trait variation. Nature Genetics. 2022;54(9):1355-63.

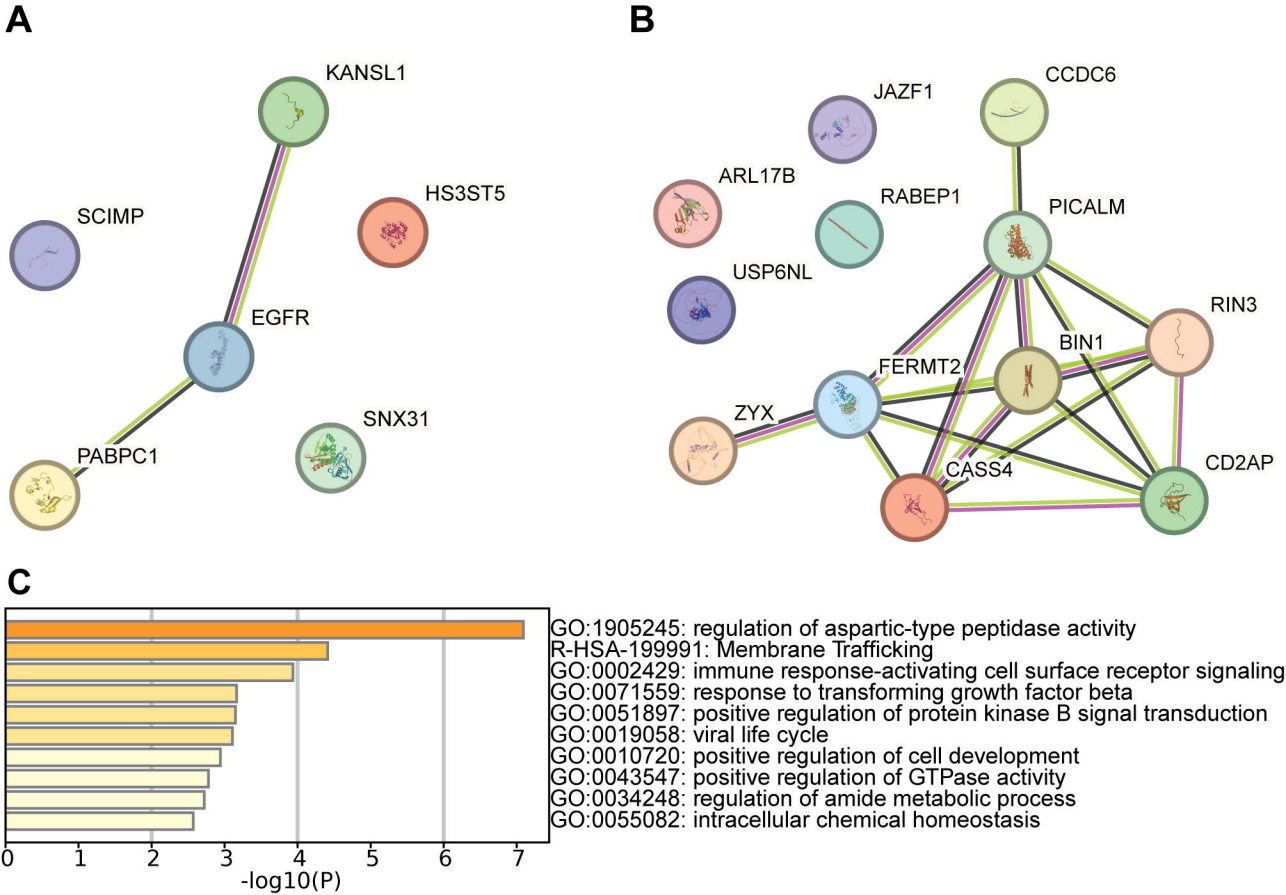
840 41. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue
841 Expression (GTEx) project. Nature Genetics. 2013;45(6):580-5.
842 42. Strober BJ, Elorbany R, Rhodes K, Krishnan N, Tayeb K, Battle A, et al. Dynamic
843 genetic regulation of gene expression during cellular differentiation. Science.
844 2019;364(6447):1287-90.
845 43. Nathan A, Asgari S, Ishigaki K, Valencia C, Amariuta T, Luo Y, et al. Single-cell eQTL
846 models reveal dynamic T cell state dependence of disease loci. Nature. 2022;606(7912):120-8.
847 44. Britton JS, Wiley JC, Beck J, Yi L, Bradic L, Do K, et al. Agora: An open-access
848 platform for the exploration of nascent targets for Alzheimer's disease therapeutics. Alzheimer's
849 & Dementia. 2023;19(S12):e079328.
850 45. Perry VH, Nicoll JAR, Holmes C. Microglia in neurodegenerative disease. Nature
851 Reviews Neurology. 2010;6(4):193-201.
852 46. Levine JM, Reynolds R, Fawcett JW. The oligodendrocyte precursor cell in health and
853 disease. Trends in Neurosciences. 2001;24(1):39-47.
854 47. Maziuk B, Ballance HI, Wolozin B. Dysregulation of RNA Binding Protein Aggregation
855 in Neurodegenerative Disorders. Frontiers in Molecular Neuroscience. 2017;10.
856 48. Qi Y, Wang M, Jiang Q. PABPC1——mRNA stability, protein translation and
857 tumorigenesis. Frontiers in Oncology. 2022;12.
858 49. Kavanagh T, Halder A, Drummond E. Tau interactome and RNA binding proteins in
859 neurodegenerative diseases. Molecular Neurodegeneration. 2022;17(1):66.
860 50. Stefanizzi I, Cañete-Soler R. Coregulation of light neurofilament mRNA by poly(A)-
861 binding protein and aldolase C: Implications for neurodegeneration. Brain Research.
862 2007;1139:15-28.

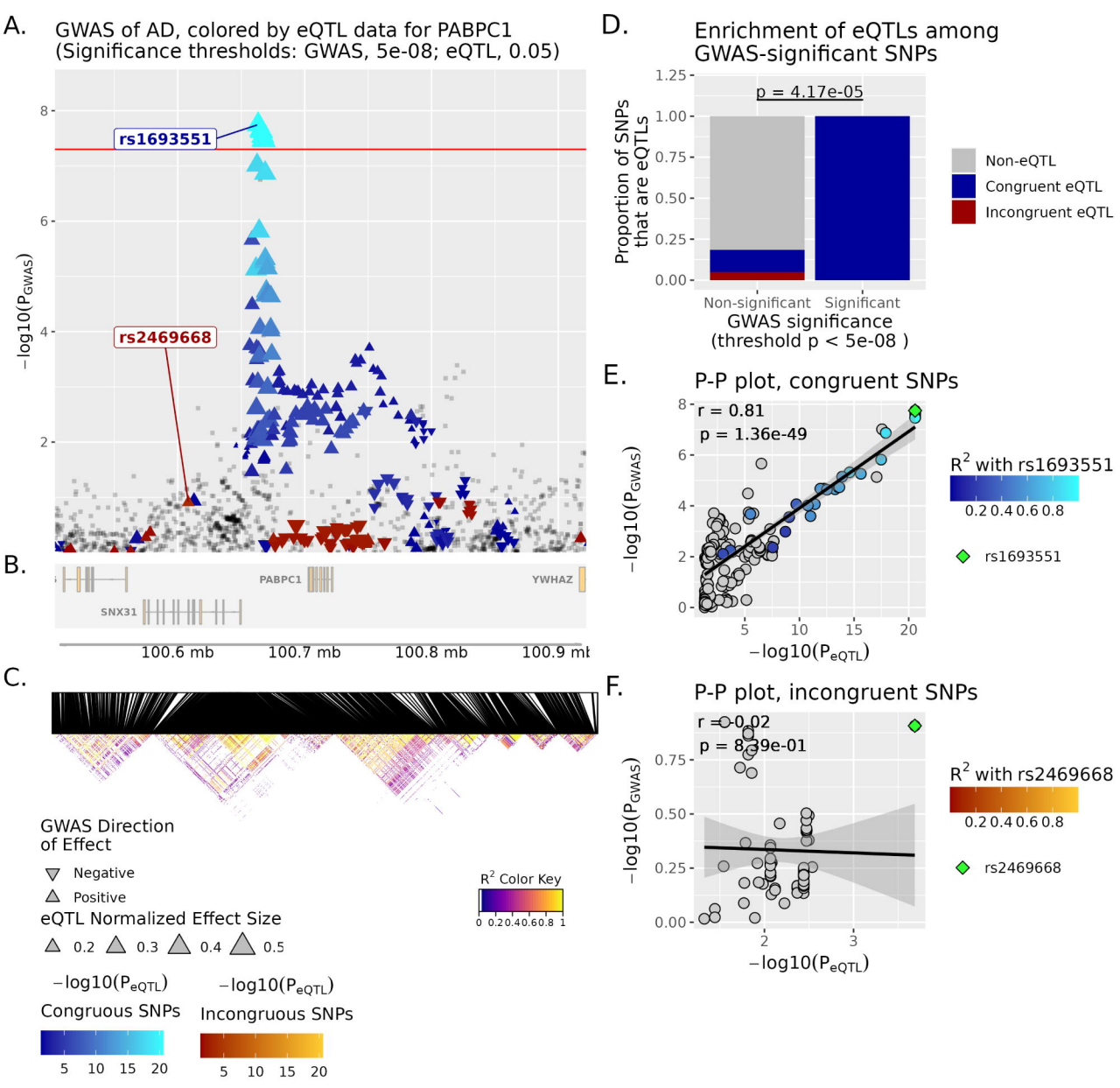
863 51. Maziuk BF, Apicco DJ, Cruz AL, Jiang L, Ash PEA, da Rocha EL, et al. RNA binding
864 proteins co-localize with small tau inclusions in tauopathy. Acta Neuropathologica
865 Communications. 2018;6(1):71.
866 52. Netzer WJ, Dou F, Cai D, Veach D, Jean S, Li Y, et al. Gleevec inhibits β-amyloid
867 production but not Notch cleavage. Proceedings of the National Academy of Sciences.
868 2003;100(21):12444-9.
869 53. Kumar M, Kulshrestha R, Singh N, Jaggi AS. Expanding spectrum of anticancer drug,
870 imatinib, in the disorders affecting brain and spinal cord. Pharmacological Research.
871 2019;143:86-96.
872 54. Bihorel S, Camenisch G, Lemaire M, Scherrmann J-M. Influence of breast cancer
873 resistance protein (Abcg2) and p-glycoprotein (Abcb1a) on the transport of imatinib mesylate
874 (Gleevec®) across the mouse blood–brain barrier. Journal of Neurochemistry. 2007;102(6):1749-
875 57.
876 55. Zarrindast M-R, Khakpai F. The modulatory role of nicotine on cognitive and non-
877 cognitive functions. Brain Research. 2019;1710:92-101.
878 56. Guan Z-Z, Yu W-F, Nordberg A. Dual effects of nicotine on oxidative stress and
879 neuroprotection in PC12 cells. Neurochemistry International. 2003;43(3):243-9.
880 57. Nizri E, Irony-Tur-Sinai M, Lory O, Orr-Urtreger A, Lavi E, Brenner T. Activation of the
881 Cholinergic Anti-Inflammatory System by Nicotine Attenuates Neuroinflammation via
882 Suppression of Th1 and Th17 Responses. The Journal of Immunology. 2009;183(10):6681-8.
883 58. Nordberg A, Hellström-Lindahl E, Lee M, Johnson M, Mousavi M, Hall R, et al. Chronic
884 nicotine treatment reduces β-amyloidosis in the brain of a mouse model of Alzheimer's disease
885 (APPsw). Journal of Neurochemistry. 2002;81(3):655-8.

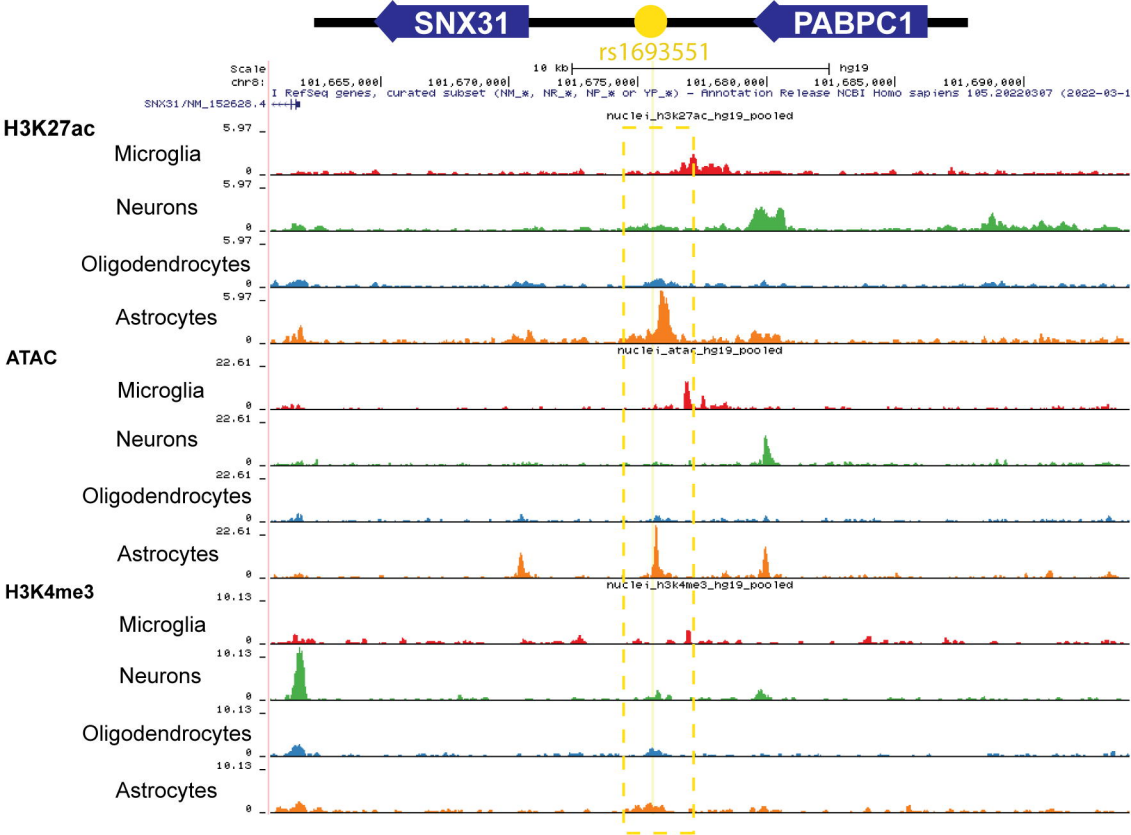
886 59. Benowitz NL, Fraiman JB. Cardiovascular effects of electronic cigarettes. Nature
887 Reviews Cardiology. 2017;14(8):447-56.
888 60. Swan GE, Lessov-Schlaggar CN. The Effects of Tobacco Smoke and Nicotine on
889 Cognition and the Brain. Neuropsychology Review. 2007;17(3):259-73.
890

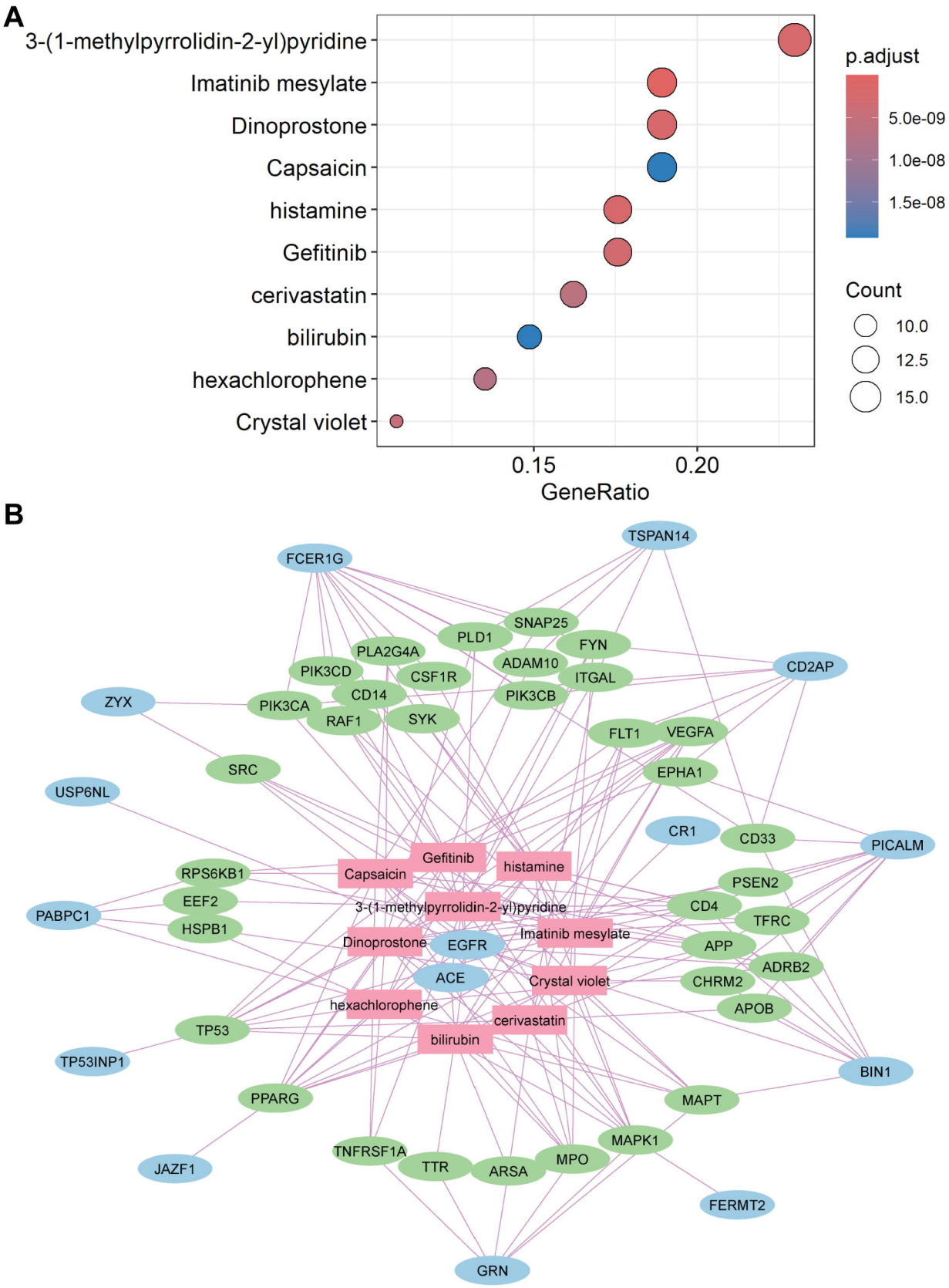

Gene expression
Bulk level
Cell type specific
1) 1 bulk eQTL dataset
2) 3 snRNA eQTL datasets
SNP
eQTL
GWAS
European ancestry
Alzheimer's disease
5 latest AD GWAS summary statistics
Significance
Chromosome position
P
Genomic position (bp)
SMR + HEIDI method
+
Bayesian Colocalization
Bulk / Cell type specific candidate causal genes
1) Druggability analysis
2) Drug prediction
Comparison with previous nominated causal genes
Novel candidate causal genes
Visual confirmation of colocalization
Differential expression analysis in sub-celltypes (AD Pathology, Cognition)
Cell type specific enhancer analysis


Combined results from 5 GWAS summary data
eQTL datasets
Cell type specific
Tissue
Bryois
Fujita
Mathys
Metabrain
Cortex, WM, DLPFC, PFC n=192
DLPFC n=424
DLPFC n=397
Cortex n=2683
10p14_USP6NL
10q21.2_CCDC6
10q23.1_TSPAN14
11p11.2_CELF1
11p11.2_SLC39A13
11q14.2_PICALM
14q22.1_FERMT2
14q32.12_RIN3
15q21.3_MINDY2
15q22.2_APH1B
16p11.2_PRSS36
17p13.2_RABEP1
17p13.2_SCIMP
17q21.31_LRRC37A
17q21.31_GRN
17q21.31_KANSL1
17q21.31_ARL17B
17q23.3_ACE
1q23.3_FCER1G
1q32.2_AL137789.1
1q32.2_CR1
20q13.31_CASS4
2q14.3_BIN1
6p12.3_CD2AP
6p12.3_AL355353.1
6q21_HS3ST5
7p11.2_EGFR
7p15.2_JAZF1
7q34_ZYX
8q22.1_NDUFAF6
8q22.1_TP53INP1
8q22.3_SNX31
8q22.3_PABPC1
Bryois_Ast
Bryois_Exc
Bryois_Mic
Fujita_Ast
Fujita_Exc
Fujita_Inh
Fujita_Mic
Fujita_Oli
Fujita_OPC
Mathys_Ast
Mathys_Exc
Mathys_Imm
Mathys_Inh
Mathys_Oli
Mathys_OPC
metabrain
MR Beta
Positive
Negative


A
KANSL1
SCIMP
HS3ST5
EGFR
SNX31
PABPC1
B
JAZF1
CCDC6
ARL17B
RABEP1
PICALM
USP6NL
RIN3
BIN1
FERMT2
ZYX
CASS4
CD2AP
C
GO:1905245: regulation of aspartic-type peptidase activity
R-HSA-199991: Membrane Trafficking
GO:0002429: immune response-activating cell surface receptor signaling
GO:0071559: response to transforming growth factor beta
GO:0051897: positive regulation of protein kinase B signal transduction
GO:0019058: viral life cycle
GO:0010720: positive regulation of cell development
GO:0043547: positive regulation of GTPase activity
GO:0034248: regulation of amide metabolic process
GO:0055082: intracellular chemical homeostasis
0
1
2
3
4
5
6
7
-log10(P)


A.
GWAS of AD, colored by eQTL data for PABPC1
(Significance thresholds: GWAS, 5e-08; eQTL, 0.05)
rs1693551
rs2469668
−log10(PGWAS)
B.
PABPC1
YWHAZ
SNX31
100.6 mb
100.7 mb
100.8 mb
100.9 mb
C.
GWAS Direction of Effect
Negative
Positive
eQTL Normalized Effect Size
0.2 0.3 0.4 0.5
R2 Color Key
0 0.2 0.4 0.6 0.8 1
−log10(PeQTL)
Congruous SNPs
5 10 15 20
−log10(PeQTL)
Incongruous SNPs
5 10 15 20
D.
Enrichment of eQTLs among GWAS-significant SNPs
p = 4.17e-05
Proportion of SNPs that are eQTLs
Non-eQTL
Congruent eQTL
Incongruent eQTL
Non-significant
Significant
GWAS significance
(threshold p < 5e-08)
E.
P-P plot, congruent SNPs
r = 0.81
p = 1.36e-49
−log10(PGWAS)
−log10(PeQTL)
R2 with rs1693551
0.2 0.4 0.6 0.8
rs1693551
F.
P-P plot, incongruent SNPs
r = -0.02
p = 8.39e-01
−log10(PGWAS)
−log10(PeQTL)
R2 with rs2469668
0.2 0.4 0.6 0.8
rs2469668


SNX31
PABPC1
rs1693551
Scale
chr8:
10 kb
hg19
101,665,000
101,670,000
101,675,000
101,680,000
101,685,000
101,690,000
SNX31/NM_152628.4
RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - Annotation Release NCBI Homo sapiens 105.20220307 (2022-03-1
H3K27ac
nuclei_h3k27ac_hg19_pooled
Microglia
Neurons
Oligodendrocytes
Astrocytes
5.97
ATAC
nuclei_atac_hg19_pooled
22.61
H3K4me3
nuclei_h3k4me3_hg19_pooled
10.13


A
3-(1-methylpyrrolidin-2-yl)pyridine
Imatinib mesylate
Dinoprostone
Capsaicin
histamine
Gefitinib
cerivastatin
bilirubin
hexachlorophene
Crystal violet
0.15
0.20
GeneRatio
p.adjust
5.0e-09
1.0e-08
1.5e-08
Count
10.0
12.5
15.0
B
TSPAN14
FCER1G
SNAP25
PLD1
FYN
PLA2G4A
ADAM10
CD2AP
PIK3CD
CSF1R
ITGAL
ZYX
CD14
PIK3CB
PIK3CA
RAF1
SYK
FLT1
VEGFA
SRC
EPHA1
USP6NL
CR1
CD33
PICALM
Gefitinib
RPS6KB1
Capsaicin
histamine
PSEN2
EEF2
3-(1-methylpyrrolidin-2-yl)pyridine
CD4
PABPC1
TFRC
HSPB1
Imatinib mesylate
Dinoprostone
EGFR
APP
ADRB2
ACE
Crystal violet
CHRM2
hexachlorophene
APOB
cerivastatin
TP53
bilirubin
TP53INP1
BIN1
PPARG
MAPT
MAPK1
TNFRSF1A
JAZF1
TTR
ARSA
MPO
FERMT2
GRN