

Deep Post-GWAS Analysis Identifies Potential Risk Genes and Causal Variants for Alzheimer's Disease, Providing New Insights Into Its Disease Mechanisms

ZHEN WANG

Yeshiva University Albert Einstein College of Medicine <https://orcid.org/0000-0002-1896-3716>

Quanwei Zhang

Albert Einstein College of Medicine

Jih-Rong Lin

Albert Einstein College of Medicine

M.Reza Jabalamel

Albert Einstein College of Medicine

Joydeep Mitra

Albert Einstein College of Medicine

Nha Nguyen

Albert Einstein College of Medicine

Zhengdong Zhang (✉ zhengdong.zhang@einstein.yu.edu)

Yeshiva University Albert Einstein College of Medicine

Research

Keywords: Alzheimer's disease, neurodegeneration, post-GWAS analysis, causal variants

Posted Date: October 28th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-96273/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Alzheimer's disease (AD) is a genetically complex, multifactorial neurodegenerative disease. It affects more than 45 million people worldwide and currently remains untreatable. Although genome-wide association studies (GWAS) have identified many AD-associated common variants, only about 25 genes are currently known to affect the risk of developing AD, despite its highly polygenic nature. Moreover, the causal variants underlying GWAS AD-association signals remain unknown.

Methods: We developed a computational pipeline that integrates 936 AD-associated SNPs, linkage disequilibrium and genomic data from multiple sources – e.g., disease genes databases, functional annotation of genetic variants, GTEx, and the 1000 Genomes Project – to predict both AD risk genes and their causal variants.

Results: We identified 342 putative AD risk genes in 203 risk regions spanning 502 AD-associated common variants. 246 AD risk genes have not been identified as AD risk genes by previous GWAS, and 115 of them are outside the risk regions, likely under the regulation of transcriptional regulatory elements contained therein. Even more significantly, for 109 AD risk genes, we predicted 150 causal variants, of both coding and regulatory (in promoters or enhancers) types, and 85 (57%) of them are supported by functional annotation. In-depth functional analyses showed that AD risk genes were overrepresented in AD-related pathways or GO terms – e.g., the complement and coagulation cascade and phosphorylation and activation of immune response – and their expression was relatively enriched in microglia, endothelia, and pericytes of the human brain. We found nine AD risk genes – e.g., *IL1RAP*, *PMAIP1*, *LAMTOR4* – as predictors for the prognosis of AD survival and genes such as *ARL6IP5* with altered network connectivity between AD patients and normal individuals involved in AD progression.

Conclusions: Our findings provide novel biological insights into the genetic architecture, expression profiles, functional pathways involved in the AD etiology, and open new strategies for developing therapeutics targeting AD risk genes or causal variants to influence AD pathogenesis.

Background

Alzheimer's disease (AD) is a progressive, chronic neurodegenerative disorder with a long prodromal phase [1]. Thus far, the amyloid cascade hypothesis has provided the most compelling explanation for the onset and development of AD. Although the mechanism by which A β exerts its toxic activity remains unclear, the hypothesis posits that the A β aggregation first impacts synaptic function, then leads progressively to tau hyperphosphorylation, aggregation, and intracellular deposition, triggering neuroinflammation, and finally causes neurodegeneration [2, 3]. Regardless of the biological evidence supporting the amyloid cascade hypothesis, major clinical trials that tested anti-A β drugs so far failed to improve patients' cognition as the primary outcome [4]. Another hallmark of AD is the aggregation of the microtubule-associated protein tau. It is, however, still unclear how tau aggregation is triggered in neurons and which modification is critical for the initiation of its aggregation [5]. Currently, multiple strategies

targeting tau pathology are actively being pursued to study the causal connection between tau and neurodegeneration [2, 5].

With a complex genetic etiology and a high heritability, estimated ranging from 60–80% [6], AD is usually divided into two subgroups of diseases: the familial early-onset AD, caused by mutations in single genes including *APP*, *PSEN1*, and *PSEN2*, and the late-onset AD (LOAD), influenced by multiple common variants with low effect sizes [7]. Over the past decade, GWAS revealed a large number of AD-associated genetic loci (**Supplementary Fig. S1** and **Table S1**), including *SORL1*, *ABCA7*, *CLU*, *CR1*, *INPP5D*, *CD33*, *BIN1*, *PICALM*, *PTK2B*, and *APOE*, a locus that has been repeatedly validated across different studies [8]. Two recent meta-analyses of large cohorts of LOAD ($n = 455,258$ and $94,437$) identified 29 [9] and 25 [10] risk loci, respectively. Interpretation of these results, however, remains elusive, because GWAS only detect statistical associations among a subset of all variants and 86% of AD associated SNPs are non-coding (either intronic or intergenic, **Supplementary Fig. S1**). To better understand the biological mechanisms underlying AD etiology, the functional impact of genetic association signals needs to be extensively investigated to identify disease genes and causal variants underlying the genetic signals detected by GWAS.

To this end, we sought to integrate genomic data from multiple sources – e.g., GWAS signals from the GWAS Catalog, disease genes databases, functional annotation of genetic variants, and the 1000 Genomes Project – to predict AD risk genes and causal variants. In this study, we aimed to first compile a list of high-confidence AD risk genes derived from association signals, then systematically uncover the characteristics of the identified AD risk genes, including the level and variation of their expression in different types of cells, and finally use a computational framework that we developed to identify putative causal variants connected to AD risk genes. Our results provide novel biological insights into the genetic architecture, expression profiles, functional pathways involved in the AD etiology, and ultimately a basis for future therapeutic development for the disease.

Material And Methods

In this study, we developed a computational pipeline (**Fig. 1**) that integrates multiple data sets to predict AD risk genes as well as their causal variants, and systematically investigated their characteristics and functions. We first used the PGA software tool [11] to identify AD risk genes on the whole-genome level, and then examined their expression patterns across different human cells and tissues and during brain development, their co-expression network, and survival curves at the transcriptomic level. Next, using an analytical framework that we developed, we predicted potential causal variants for identified AD risk genes.

Data sets and their sources

AD GWAS SNPs. From 58 AD GWAS collected in the GWAS Catalog [12] (as of December 2018) and the latest study [9], we collected 936 AD-associated SNPs ($P < 9E-06$, **Fig. 2** and **Supplementary Table S1**).

AD genes for PGA training. 338 AD genes (**Supplementary Table S2**) were collected from three disease-gene databases: MalaCards [13], DISEASES [14], and DisGeNET v5.0 [15]. They were separated into two groups: a stringent set of 98 AD genes supported by at least two of these three disease-gene databases and a lenient set of 240 AD genes supported by only one of the three databases for AD.

Differentially expressed genes (DEGs). We assembled a list of 10,314 genes that have been shown in three studies [16-18] to be differentially expressed between AD patients and normal controls (**Supplementary Table S3**).

Variant and genome annotation. We used multiple annotation metrics for various data analyses in this study, including scores from LINSIGHT [19], ExPecto [20], and PrimateAI [21], enhancer-gene connection from HEDD [22], eQTL from GTEx v.6.p [23], and Synapse (syn17015233).

Identification of genomic risk regions and putative candidate genes

Using PGA [11], we first systematically identified genomic AD-risk regions and AD risk gene candidates and then integrated both gene network and annotation data with GWAS signals to score all genes for AD risk. Gene regulatory information about enhancers and eQTLs in 44 human tissues and their target genes from HEDD [22] and GTEx (v6p) [23] was used to identify distal AD risk gene candidates. Risk gene candidates were scored with aforementioned training gene sets, and high-scoring candidates with scores greater than the threshold estimated with either the stringent (= 21.4) or the lenient (= 13.1) gene set were taken as (putative) AD risk genes for downstream analysis.

Characterization of AD risk genes

Functional enrichment analysis. We used DAVID [24] to identify KEGG pathways and GO terms enriched with AD risk genes. *P*-values were adjusted for multiple tests using the Bonferroni or FDR methods.

Expression Weighted Cell-type Enrichment (EWCE) analysis. We used EWCE (v1.2.0) [25] to identify brain cell types, which are more likely to be affected by AD risk genes. Single-cell RNA-seq data from 33 types of cells (**Supplementary Table S4**) from human adult brains (20-51 years old; a total of 35,289 cells; from the visual cortex, frontal cortex, and cerebellum) [26] were used in this analysis. The average expression across all samples for each cell type was used to determine enriched expression by EWCE.

Tissue gene expression analysis. To investigate gene-tissue expression specificity, we examined the expression profiles of AD risk genes across different tissues from the Gene Enrichment Profiler (<http://xavierlab2.mgh.harvard.edu/EnrichmentProfiler/help.html>) [27], which catalogs normalized expression values of ~12,000 genes across 126 primary human tissues. We grouped the putative risk

genes into different clusters according to their different expression profile across tissues using the Euclidean distance and the Ward's clustering method [28].

Analysis of expression of AD risk genes during human brain development. Temporal gene expression patterns in different brain tissues and their regulation across the lifespan reveal molecular mechanisms involved in the formation, maturation, and degeneration of the human brain. Here, we examined the spatiotemporal expression patterns of AD risk genes in adult human brain. We first compiled gene expression profiles across ten regions of adult human brains from individuals of 19 to 40 years old, using RNA-seq data from BrainSpan [29] (<http://www.brainspan.org/>, as of April 2018). We then explored the temporal dynamics of gene expression in a more specific human brain region (i.e., prefrontal cortex) during brain development from fetuses to older adults, using spotted oligonucleotide microarrays yielding data from Brain Cloud [30] (<http://braincloud.jhmi.edu/>, as of May 2018).

Gene co-expression networks. We built gene co-expression networks using transcriptomic and proteomic data. From Synapse (syn7391833), we collected expression profiles of ~23,200 genes (included 315 AD risk genes) in four human brain regions from normal controls and AD patients with the definitive disease status: frontal pole (Brodmann area 10, BM10) with 111 cases and 76 controls, superior temporal gyrus (BM22) with 102 cases and 65 controls, parahippocampal gyrus (BM36) with 90 cases and 68 controls, and inferior frontal gyrus (BM44) with 90 cases and 64 controls. From Synapse (syn10239444), we also collected proteomic profiles of ~10,000 gene products in two human brain regions: frontal cortex and anterior cingulate gyrus (ACG), both with 10 cases and 10 controls. We first removed proteins with less than three samples in AD patients or normal controls, and then imputed missing values (0 or '#N/A') by replacing them with averaged values from AD and control samples, separately. We constructed gene co-expression networks using the Pearson's correlation ($r > 0.7$) among AD patients and normal controls for each brain region. The co-expression networks were drawn using the 'igraph' (v1.2.4.1) R package.

Survival analysis. Among 315 AD risk genes, we carried out survival analysis (using the 'survival' R package) to identify ones that may predict by their expression levels the disease prognosis among AD patients. For each AD risk gene, using its expression profile (described above), we first separated AD patients into two groups, with either high (\geq the median) or low ($<$ the median) expression of the gene. We then used the Kaplan-Meier estimator to compute estimated survivor functions, plotted them to directly compare survival between these two groups, and used the log-rank test to formally test whether their survival curves are identical.

Prediction of AD causal variants

Causal variants contribute to the disease etiology and pathology mainly through two major mechanisms: coding variants may alter the function of gene products and non-coding variants may alter the transcription of genes by changing regulatory elements. Recently developed methods for variant annotation, such as LINSIGHT [19], ExPecto [20], and PrimateAI [21], can quantify or provide useful information about the functional impact of genetic variants in the human genome, regardless of their

potential connection to a specific disease or trait. By integrating such functional annotation of variants as a part of the computational framework that we developed for the post-GWAS analysis (**Fig. 1**), we are able to not only identify disease risk genes but also further predict disease causal variants. In this framework, we first use the variant type as the guide. For coding variants, we assume causality through the functional alteration of gene products and use their PrimateAI scores to prioritize them. For non-coding variants, we assume causality is through gene regulation and prioritize variants in promoters by their ExPecto scores and ones in enhancers by their LINSIGHT scores.

We systematically fine-mapped the genomic neighborhoods of predicted causal variants by analyzing all genotyped and imputed variants from the latest AD GWAS [9]. Only variants mapped to risk region and coding sequences and regulatory elements (i.e., promoters and enhancers) of AD risk genes were selected for downstream analysis. We used the following criteria to predict causal variants: PrimateAI scores, which measure the pathogenicity of coding variants from 0 (less pathogenic) to 1 (more pathogenic), are greater than 0.7; ExPecto scores, which predict functional impact of non-coding variants in promoters, are greater than $\ln(\text{fold change} = 1.2)$; and LINSIGHT scores, which measure the probability of negative selection on non-coding sites and can be used to prioritize SNVs associated with genetic diseases, are greater than 0.9.

Transcription factor binding site motif analysis

The reference and the alternative alleles of each variant, along with ± 25 bp of flanking sequences, were analyzed using HOMER (v4.9.1) [31]. We used the findMotifs.pl program with the default parameters to find transcription factor binding site (TFBS) motifs in the reference and the alternative allele sequences, which were used as the background for each other to control the nucleotide context. A library of 364 vertebrate TFBS motifs in the format of position weight matrices was searched for matches. The matched motifs with scores greater than 7.5 were regarded as gain or loss of TFBS motifs.

Results

AD risk regions and risk genes

Using 936 GWAS AD SNPs and linkage disequilibrium, we identified 598 genomic risk regions, spanning ~ 55.0 Mb of the human genome (**Fig. 1**). Based on the genomic annotation of genes and regulatory elements, we could connect 1,445 genes to 432 risk regions (**Supplementary Fig. S2** and **Table S5**). Among AD risk gene candidates, 967 are proximal genes, overlapping AD risk regions, and 506 are distal genes, linked to AD risk regions through long-range gene regulatory elements (e.g., enhancers) (**Supplementary Table S5**). 28 genes are both proximal for some risk regions and distal for other risk regions. 35 candidates were not scored due to their lack of GO annotation and/or exclusion from the gene functional linkage network. Among 1,410 scored candidates, 342 loci, distributed in 203 risk regions (**Fig. 2** and **Supplementary Table S6**), surpassed the “high” threshold (see Method) and thus were considered

as (putative) AD risk genes. They included 233 (68.1%) candidates proximal to AD risk regions, and additional 115 (33.6%) distal genes, which are likely to be regulated by regulatory elements in the risk regions (**Fig. 1** and **Supplementary Table S6**).

KEGG pathways and GO biological processes enriched with AD risk genes

AD risk genes were overrepresented in 10 KEGG pathways (FDR < 0.05) and with 151 GO terms ($P < 0.01$ after the Bonferroni correction) (**Fig. 3A-B** and **Supplementary Table S7**), all highly relevant to the disease pathology. The most significantly enriched KEGG pathway is the complement and coagulation cascades. Increasing evidence suggests that deregulation of the complement cascade is a contributing factor leading to chronic inflammation and neurodegeneration observed in AD. The complement system plays an important role in the innate and adaptive immune responses, restricts amyloid plaque formation, and helps clearance of plaque components associated with AD [32]. AD risk genes are also enriched in MAPK signaling pathway, which contributes to the AD pathogenesis through multiple mechanisms, including the regulation of neuronal apoptosis and phosphorylation of APP and tau [33]. Additional enriched pathways, such as GnRH signaling, PI3K-Akt signaling, neurotrophin signaling, and calcium signaling, all have been shown to likely play a role in AD [34-37], whereas Ras signaling may play an important role in aging [38]. Many enriched GO terms – e.g., “nervous system development”, “neurogenesis”, “neuron differentiation”, “neuron development” – are related to the development and differentiation of neurons. Some are directly related to AD: e.g., “phosphorylation ($p < 0.01$)”, “activation of immune response ($p < 0.01$)”, and “learning memory” are part of the currently predominant hypothesis of the pathogenesis of AD. In addition, several blood-related GO terms are enriched: “blood vessel development”, “vasculature development”, “negative regulation of blood coagulation”, and “hemopoiesis”. Increasing evidence has shown that the hematopoietic system may contribute to the initiation and/or progression of AD [39].

AD risk genes and human aging

Eight AD risk genes – *ABCA1*, *TREM2*, *PLD2*, *MAPK12*, *NR1H3*, *CDK5RAP3*, *IRAK2*, and *ABCG1* – were also implicated by rare variants for connection with human lifespan in a whole exome sequencing study of a longevity cohort (data not shown). *IRAK2* is informative for AD prognosis. AD patients with high expression of *IRAK2* have significantly better survival outcome than patients with low expression of it ($P < 0.05$) in superior temporal gyrus and inferior frontal gyrus. Another gene, *ABCA1*, was one of the top genes with the highest score in our longevity study. *ABCA1* gene contributes to lipid processing and the formation of A β . Studies of transgenic mouse models revealed that deficiency or overexpression of *ABCA1* is associated with increased or decreased A β deposition [40, 41], respectively, implying that ABCA1-mediated lipidation influences amyloid degradation. Moreover, ABCA1 also plays a crucial role in ApoE – a major genetic risk factor for LOAD – lipoprotein metabolism in the brain [42]. These data suggest that *ABCA1* may be a therapeutic target for AD and aging.

Clustered expression of AD risk genes in different human tissues

Based on their expression profiles in different human tissues, AD risk genes can be clustered into three groups (**Fig. 3C** and **Supplementary Fig. S3**). The first group of 32 genes was expressed almost exclusively in the central nervous system (CNS), especially the frontal and the prefrontal cortices. Many genes in this group, such as *BIN1*, *MAPT*, and *CNTNAP2*, have been implicated in the pathogenesis of AD [43-45]. The second group included 131 genes actively expressed in the immune cells such as B and T lymphocytes. Recent studies showed that inflammation contributes to the pathogenesis of AD [46]. It is important to note that many immune related genes, such as *TREM2*, *INPP5D*, *CD34*, and *CD55*, were included in this group. *TREM2* is a cell surface receptor of the immunoglobulin superfamily protein expressed in microglia in the CNS. As a potential key molecule in AD pathogenesis, it might protect against neurodegeneration by promoting phagocytosis to clear apoptotic neurons [47] and a broad array of microglial functions in response to A β deposition [48]. *INPP5D* plays a significant role in inflammatory responses and has been implicated in the pathogenesis of late-onset AD through the regulation of microglial and myeloid cell function [7]. This suggests that immune processes may directly contribute to the pathology and progression of AD, rather than being the consequence of the neurodegeneration. Cholesterol metabolism-associated genes in this group, such as *SORL1* and *ABCA7*, have been linked to AD in previous studies. Studies showed that the suppression of *SORL1* expression contribute to the overexpression of A β and an increased risk of AD [49]. *ABCA7* is a genetic risk factor for late-onset AD and may participate in the regulation of A β homeostasis in the brain [50]. Moreover, this group also includes tauopathy-associated AD risk genes, such as, *PTK2B* and *PICALM*. Previous studies in *Drosophila* indicated that *PTK2B* acts as an early marker and *in vivo* modulator of tau toxicity [51]. Cell-based and *in vivo* data showed that perturbations of *PICALM* levels might be a key for the regulation of autophagy and tau levels and therefore essential for modulating tau toxicity [52]. 115 genes in the third group were expressed across a wide range of different tissues, including the CNS. Many genes in this group, such as *ApoE*, *CR1*, and *EPHA1*, are known to be associated with AD. Human studies clearly indicate that ApoE isoforms differentially affect A β aggregation and clearance [8], and *CR1* may play a role in the clearance of A β [53].

High expression of AD risk genes in microglia, endothelia, and pericytes of human brain

The expression of AD risk genes was significantly enriched in microglia, endothelia and pericytes in the frontal and the visual cortices and cerebellum from human adults (**Fig. 3D** and **Supplementary Table S4**). The over-expression profile was evident for many genes (**Supplementary Fig. S4**), which were enriched for microglial markers (e.g., *HLA-DRA* and *TREM2*) and endothelial markers (e.g., *CD34*). Previous studies showed that microglia are the primary cells contributing to the initiation of the immune response to AD pathology, and the aberrant microglial activation is a causal factor for the development of AD [54]. Recent

studies also suggests endothelial dysfunction may be involved in the pathogenesis of AD [55]. Pericytes, cells in the blood-brain barrier, degenerate in AD and are reported to control multiple steps of AD-alike neurodegeneration cascade in mice overexpressing A β -precursor protein [56]. Moreover, oligodendrocytes are only significantly enriched with overexpressed AD risk genes in cerebellum. The major function of oligodendrocytes is the formation of myelin, whose breakdown is associated with AD [57].

Connectivity of AD risk genes in co-expression network

AD is a progressive neurodegenerative disease that involves alteration of gene expression at the whole transcriptome level. The perturbation in the sub-networks of co-expression involving AD risk genes can partially reflect AD progression. Finding the altered network hub genes involved in AD progression may help identify AD biomarkers. We carried out the gene co-expression network analysis across four brain regions to examine the gene regulation patterns among AD risk genes. Overall, we observed that connections among AD risk genes are less in AD patients compared to normal controls (**Supplementary Fig. S5 and S6**). For each brain region, we considered top 20 genes with most interactions with other genes as network hubs. We found that hub genes with high connectivity in AD patients – e.g., *ARL6IP5* (BM10), *RNF6* (BM22), *TP53INP* (BM36), and *GGH* (BM44) – tended to have low connectivity in healthy individuals (**Fig. 3E and Supplementary Fig. S5**). *Rnf6*, a ring-finger-dependent ubiquitin ligase, functions for proteasomal degradation in axonal growth cones of primary hippocampal neurons in mice by regulating the levels of *Limk1*, which play a crucial role in neurodevelopment and synaptic plasticity [58]. *TP53INP1*, a major regulator of p53 in response to oxidative stress [59], is a tumor suppressor associated with malignant tumor metastasis in breast, liver, pancreas, and stomach and plays a critical role in cancer progression. Interestingly, previous studies showed inverse correlation between cancer and AD [60]. It has been reported that tripeptide GGH might be used for Cu chelation therapy for AD treatment as Cu ion level was reported to be elevated in AD brains and accumulation of amyloid plaques leading to metal homeostasis dysregulation [61]. On the other hand, many hub genes in normal people – e.g., *LMTK2* (BM10), *SPPL2A* (BM22), *MAPT* (BM44), and *USP8* (BM36) – usually had low connectivity in AD patients. *LMTK2* may contribute to the neurodegenerative process by disrupting axonal transport, tau hyperphosphorylation and enhancing apoptosis [62]. Its expression is decreased in a tau mouse model of AD [63]. As one of the deubiquitinases, which play a critical role in regulating synaptic function and whose dysfunction results in several neurological disorders, *USP8* has been shown to be associated with AD [64], Parkinson's disease, and Lewy body disease. *MAPT* encodes tau protein, whose hyperphosphorylation and subsequent intracellular neurofibrillary entanglement is one of definitive neuropathological hallmarks of AD. *SPPL2a* is an intramembrane protease of lysosomes/late endosomes and plays a critical role in regulation of intramembrane proteolysis in B cells and the regulation of innate and adaptive immunity [65].

We also analyzed co-expression of AD risk genes at the proteomic level and observed similar patterns that the AD risk genes were less connected among AD patients than normal controls in the ACG region, while opposite pattern in the FC region (**Supplementary Fig. S7**). For example, *ARL6IP5* and *GGH* were

network hubs in AD patients but less connected in controls in the FC region (**Supplementary Fig. S8**). In the ACG region, we observed network hubs such as *PTK2B*, *SPARC*, and *RAD50* showing large alteration between AD patients and controls (**Supplementary Fig. S8**). *SPARC* is a matricellular protein which can facilitate the migration of immune cells (e.g., blood-derived dendritic cells). Although its role in AD-related neuroinflammation is still not clear, a study has shown that there are significant alterations in its expression and it collocates to A β protein deposits in AD brain tissues [66].

Expression of AD risk genes in human brain and its connection to disease survival

Using data from three studies of differential gene expression between AD cases and controls in different brain regions [16-18], we found 171 (50%) AD risk genes were differentially expressed in at least one brain region, including 102 up-regulated genes, 64 down-regulated genes, and 5 genes showing both up- and down-regulation in different brain regions (**Supplementary Table S8**). Differential expression of AD risk genes was either widespread, occurring in multiple brain regions, or limited to a specific brain region. *TGFB2*, the highest ranked risk gene, was up-regulated in frontal cortex (FC), central nervous system (CNS), temporal cortex (TCX), superior temporal gyrus (STG) and parahippocampal gyrus (PHG), while *PTK2B* was down-regulated in brain cerebellum (CBE), TCX, and PHG. *COL25A1*, the second highest ranked risk gene, and *PMAIP1* were separately down- and up-regulated only in the TCX region. Differential expression of some AD risk genes was discordant in different brain regions. For example, *ApoE* and *CST3* in AD patients were up-regulated in TCX region but down-regulated in cerebellum.

Since AD is mainly a late-onset neurodegenerative disorder, we specifically examined how AD risk genes are expressed among adults. Using a binarization procedure [67], we analyzed their spatiotemporal expression patterns using RNA-seq data from BrianSpan. Although no strong pattern was found (**Supplementary Fig. S9**), the proportion of AD risk genes with dramatically suppressed expression was increased at the age of 40 compared to early ages. The proportion of AD risk genes that tend to be transcriptionally actively was relatively higher at the early ages (of 23 and 30) in comparison to the old ages (of 36 and 40). We next examined the spatiotemporal expression pattern of AD risk genes during the development of the frontal cortex across an extended range of ages (from 18.05 to 78.23). We did not observe any distinct expression pattern across this range of ages.

As a chronic neurodegenerative disease, AD starts slowly and gradually worsens overtime. We hypothesized that genes whose expression correlates with AD progression may mark AD severity and thus can be used to predict AD prognosis. To test this hypothesis, we assessed the impact of AD genes on survival using the Kaplan-Meier analysis. Based on expression levels, nine genes – *NRG3*, *IL1RAP*, *PMAIP1*, *STRADA*, *SGK3*, *LAMTOR4*, *MAPK12*, *PHB*, and *GRB2* – separated AD patients into low- and high-risk groups with different disease survival ($P < 0.05$). Their expression also trends differently with age between healthy individuals and AD patients in at least one brain region ($P < 0.05$) (**Fig. 4** and **Supplementary Fig. S10**).

Predicted AD causal variants

Using the computational framework that we developed for this project, we predicted 150 unique potential causal variants for 109 AD risk genes (**Supplementary Table S9** and **Fig. S11A**). To evaluate this prediction, we analyzed their effect on sequence motifs of transcription factors binding sites (TFBS) and compared these to the eQTL data from GTEx. Motif analysis revealed that 69 predicted causal variants (46%, **Supplementary Table S9**) cause either gain or loss of TFBS motifs, likely affecting TF binding. Among them, 32 (21%, **Supplementary Table S9**) have also been identified as eQTLs. Together, 85 (57%) of the predicted causal variants can be functionally annotated (**Supplementary Fig. S11B**). Three modules were developed in our computational framework to predict causal variants in different functional genomic regions:

Coding variants. We predicted 54 causal coding variants. For example, rs7412 and rs4147934 are two missense coding SNPs, each in high LD ($r^2 > 0.5$) with one of the AD GWAS lead SNPs in its corresponding risk region, were predicted as causal variants (**Fig. 5 A and B, Supplementary Table S9**). rs7412, in *APOE* with a PrimateAI score = 0.80, is a well-known variant reported to be associated with AD. rs4147934, in *ABCA7* with a PrimateAI score = 0.78, has been proposed as a functional candidate variant accounting for the GWAS signal at *ABCA7* locus in Caucasians [68]. Although AD risk from rs4147934 is probably population-specific since its association signal was not replicated in the African American cohort [69], our analysis provides additional evidence in support of its causal role in AD and thus its impact in non-European ancestry populations merits further investigation.

Non-coding variants in promoters. We also predicted 33 causal promotor variants in brain tissues or cells (**Fig. 5C-E**), including rs76516995 (ExPecto score = 0.194) for *BIN1* in astrocytes, rs4292 (ExPecto score = 0.390) for *ACE* in neural cells, and rs12691088 (ExPecto score = 0.224) for *APOC1* in astrocytes. rs76516995 and rs4292 have also been identified as eQTLs by GTEx, while rs12691088 has been shown to be associated with AD-related phenotypes in multiple brain regions [70].

Non-coding variants in enhancers. We predicted 64 causal enhancer variants. Two of them, rs2271920 (LINSIGHT score = 0.951) and rs117423666 (LINSIGHT score = 0.966) **Fig. 5F** and **Supplementary Table S9**), are causal variants for risk gene *PTK2B*. rs2271920 is an AD GWAS lead SNP itself [71]. As a GTEx eQTL, it changes the expression of *PTK2B*, likely by altering the binding sites motif of BCL6 and ZNF467. For the *SNX1* locus, we also predicted two causal enhancer variants: rs146600064 (LINSIGHT score = 0.914) and rs60226406 (LINSIGHT score = 0.977) (**Fig. 5G** and **Supplementary Table S9**). rs146600064 is in total LD ($r^2 = 1$) with the lead AD SNP rs74615166 [7]. In another risk region indexed by the AD lead SNP rs7207400 [71], several predicted causal variants – rs2049515, rs4341787, rs549929529, and rs242557 – are connected to multiple AD-risk genes – *NMT1*, *CRHR1*, *SPPL2C*, *MAPT*, and *NSF* (**Fig. 5H** and **Supplementary Table S9**). Although a disease risk gene could be influenced by multiple causal variants, we examined additional data – e.g., TFBS motifs, eQTL, and ClinVar – to further prioritize them. rs549929529 is most likely a causal variant for *MAPT* as it is reported in ClinVar to be associated with 'MAPT-Related_Spectrum_Disorders'. Compared with other SNPs, rs4341787 is in the highest LD ($r^2 =$

0.779) with the AD lead SNP rs7207400 and is an eQTL identified by GTEx for *NMT1*. Also, in high LD ($r^2 = 0.775$) with rs7207400, rs2049515 could be a causal variant for multiple genes as it is an eQTL for *CRHR1*, *SPPL2C*, and *NSF* (**Fig. 5H** and **Supplementary Table S9**). Moreover, rs1522388, a predicted causal variant (LINSIGHT score = 0.979) for *FLNB*, was also identified as a reporter assay QTL in HepG2 cell line [72], which experimentally demonstrated its functional impact as a regulatory variant.

To date, the functional impact of the aforementioned variants is still poorly understood in AD etiology, but our findings provide promising causal variant candidates for further experimental validation, which will in turn identify potential drug targets for the development of AD treatment. To explore the molecular function of the predicted causal variants, we examined eQTLs (FDR < 0.05) identified in three brain regions (temporal cortex, dorsolateral prefrontal cortex, and cerebellum) in an AD cohort (Synapse: syn17015233). We found 16 (11%, **Supplementary Table S10**) causal variants were AD-related eQTLs. Several causal variants – e.g., rs2236393 for *CDH3* and rs12752439 for *HSPG2* – were not identified as eQTLs by GTEx but instead are AD-related eQTLs. This analysis provides direct evidence for their involvement in AD pathogenesis.

Discussions

GWAS have uncovered thousands of genetic variants that influence risk for complex human diseases. However, there is still a large gap between the statistical associations linking locus and trait and the functional impact of causal variants underlying disease risk. Multiple factors have made it difficult to bridge this gap. First, the association of a locus with disease does not reveal the underlying causal variant as many co-inherited variants in strong LD with one another at the locus often have statistically indistinguishable disease associations. Second, genes affected by the causal variants are usually unknown. Consequently, functional studies aimed at determining the causal genetic variants and the biological mechanisms underlying the observed disease association have lagged. Thus, it is important to not only identify disease risk genes but also explore their causal variants. Responding to this need, we proposed a post-GWAS computational framework that could predict the causal variants for a specific disease.

Using this framework for AD, we predicted 342 AD risk genes and, for 109 of them, 150 causal variants. Many of the predicted AD risk genes with the highest risk scores are well-known AD risk genes. For examples, genes such as *ApoE* and *SORL1* have already been shown to functionally underlie the disease pathology [73] (Fig. 2). The role of the *TGFB2* gene in AD is still poorly understood, although it was found to autocrinally induce the apoptosis of primary cortical neurons [74] and up-regulated in multiple brain regions among AD patients. TGF β signaling and *COL25A1* play important roles in the pathogenesis of AD [75, 76], while *CST3* may offer neuroprotection against AD [77]. *USP8* gene depletion leads to decreased levels and activity of *BACE1*, the rate-limiting enzyme in the production of amyloid- β [64].

Several recent studies explored the connection between diseases and genetic variants in coding sequences [21], non-coding regions [19], and promoters [20] in human genome. Our framework for post-

GWAS analysis integrates these data to predict disease causal variants. Over 90% of disease-associated variants found in GWAS are located in non-coding regions, and aggregate analysis of them has shown that they are significantly enriched in transcription-factor-occupied regions and DNase I hypersensitive sites, most of which overlap enhancer regions [78], suggesting that many disease loci identified by GWASs influence transcriptional output of one or more target genes through enhancer variants. However, it has been difficult, computationally or experimentally, to identify causal variants in enhancers for any disease. For AD, we predicted 62 causal enhancer variants. For over 66% of them, there is additional support for their functional impact on enhancer: 16 are known eQTLs and 29 can alter transcriptional factor binding through gain or loss of TFBS motifs.

AD risk genes are expressed in many different types of brain cells. We found that they are over-expressed in microglia, endothelia, and pericytes from three different regions of adult brains, consistent with previous reports showing that these cell types are likely to be associated with AD pathology. We also observed over-expression of AD risk genes in specific cell types or brain regions or both. For example, only oligodendrocytes from cerebellum showed active expression of AD risk genes. When frontal cortex with visual cortex were compared, only In4a and In7 cells from the former and In1a cells from the later showed enriched expression for AD risk genes. These findings were replicated using a single-cell transcriptome data set [79] (Fig. 3D). A previous expression weighted cell-type enrichment analysis of a set of 178 AD risk genes using the same data set failed to identify significantly enriched cell types [80], likely due to the incomplete list AD risk genes used in that analysis.

As expected, many predicted AD risk genes were also identified by survival analysis as predictors for AD prognosis. They include *IL1RAP*, *PMAIP1*, *LAMTOR4*, and *GRB2*. *IL1RAP*, a key immune signaling factor, impacts amyloid accumulation by modulating the activity of microglia and is crucial in clearing brain amyloid and limiting plaque growth [81]. *PMAIP1* is an essential mediator of p53-dependent apoptosis, an important biological process in neurodegenerative disorders [82]. *LAMTOR4* is a component of the Rag-Regulator complex and an essential regulator of lysosomes in microglia. Its absence has been shown to result in diminished number of microglia in Zebrafish [83]. *GRB2*, a cytoplasmic protein, are involved in protecting the cytoskeletal architecture in AD-like conditions [84] and interacts with the C-terminal fragment or tyrosine-phosphorylated APP. This interaction intensifies significantly in neuronal cells and AD brains [85]. Finally, it is worth noting that the predictor genes for AD prognosis are region-specific, as none was identified in BM44 and only one (*GRB2*) in BM33.

Conclusions

Recent GWAS have identified many AD-associated common variants. For most of the reported associations, however, the underlying biological mechanisms are not clear. The critical first step for their elucidation is to identify the most likely disease genes and causal variants as the source of the association signals. Toward this goal, we developed an integrated computational framework to predict both disease genes and corresponding causal variants. Using this approach, we analyzed AD GWAS signals and examined different functional aspects of AD risk genes and causal variants that we

identified. Our study provides novel biological insights into the genetic architecture, expression profiles, and functional pathways involved in the AD etiology, and ultimately new strategies for developing therapeutics for the disease.

Abbreviations

ACG

anterior cingulate gyrus; AD = Alzheimer's disease; BM = Brodmann area; EWCE = expression weighted cell-type enrichment; FC = frontal cortex; GWAS = genome-wide association studies; LD = linkage disequilibrium; LOAD = late-onset Alzheimer's disease; PHG = parahippocampal gyrus; STG = superior temporal gyrus; TCX = temporal cortex; TFBS = Transcription factor binding site

Declarations

Acknowledgements

We thank Yousin Suh of the Department of Genetics at the Albert Einstein College of Medicine and Michael G. Rosenfeld of the School of Medicine at the University of California, San Diego, for their expert advice on the biology of enhancers.

Authors' contributions

Z.D.Z. conceived and designed this study. Z.W. carried out the analyses and also participated in the study design. Z.W. and Z.D.Z. wrote the manuscript. Q.Z., J.R.L., M.R.J., J.M., N.N. provided useful input for the analyses and helped edit the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Institutes of Health grants R01 HG008153 from the National Human Genome Research Institute and R01 AG057909 from the National Institute on Aging to Z.D.Z.

Availability of data and materials

Only publicly available data were used in this study. See the 'Data sets' subsection above for their availability. Computer code for data processing is available from the authors upon request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ College of Animal Sciences, Zhejiang University, Hangzhou, Zhejiang, China.

² Department of Genetics, Albert Einstein College of Medicine, Bronx, New York, USA.

References

1. Masters CL, Bateman R, Blennow K, Rowe CC, Sperling RA, Cummings JL: **Alzheimer's disease.***Nat Rev Dis Primers* 2015, **1**:15056.
2. Frigerio CS, De Strooper B: **Alzheimer's Disease Mechanisms and Emerging Roads to Novel Therapeutics.***Annu Rev Neurosci* 2016, **39**:57-79.
3. Selkoe DJ, Hardy J: **The amyloid hypothesis of Alzheimer's disease at 25years.***Embo Mol Med* 2016, **8**:595-608.
4. Karran E, Hardy J: **A Critique of the Drug Discovery and Phase 3 Clinical Programs Targeting the Amyloid Hypothesis for Alzheimer Disease.***Ann Neurol* 2014, **76**:185-205.
5. Wang YP, Mandelkow E: **Tau in physiology and pathology.***Nature Reviews Neuroscience* 2016, **17**:5-21.
6. Cacace R, Sleegers K, Van Broeckhoven C: **Molecular genetics of early-onset Alzheimer's disease revisited.***Alzheimers Dement* 2016, **12**:733-748.
7. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, Jun G, DeStefano AL, Bis JC, Beecham GW, et al: **Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.***Nature Genetics* 2013, **45**:1452-U1206.
8. Liu CC, Liu CC, Kanekiyo T, Xu H, Bu G: **Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy.***Nat Rev Neurol* 2013, **9**:106-118.
9. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, Sealock J, Karlsson IK, Hagg S, Athanasiu L, et al: **Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk.***Nat Genet* 2019.
10. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, Boland A, Vronskaya M, van der Lee SJ, Amlie-Wolf A, et al: **Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk**

- loci and implicates Abeta, tau, immunity and lipid processing.*Nat Genet* 2019, **51**:414-430.**
11. Lin JR, Jaroslawicz D, Cai Y, Zhang Q, Wang Z, Zhang ZD: **PGA: post-GWAS analysis for disease gene identification.***Bioinformatics* 2017.
 12. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, et al: **The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog).***Nucleic Acids Res* 2017, **45**:D896-D901.
 13. Rappaport N, Twik M, Plaschkes I, Nudel R, Iny Stein T, Levitt J, Gershoni M, Morrey CP, Safran M, Lancet D: **MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search.***Nucleic Acids Res* 2017, **45**:D877-D887.
 14. Pletscher-Frankild S, Palleja A, Tsafou K, Binder JX, Jensen LJ: **DISEASES: text mining and data integration of disease-gene associations.***Methods* 2015, **74**:83-89.
 15. Pinero J, Bravo A, Queralt-Rosinach N, Gutierrez-Sacristan A, Deu-Pons J, Centeno E, Garcia-Garcia J, Sanz F, Furlong LI: **DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants.***Nucleic Acids Res* 2017, **45**:D833-D839.
 16. Li X, Long J, He T, Belshaw R, Scott J: **Integrated genomic approaches identify major pathways and upstream regulators in late onset Alzheimer's disease.***Sci Rep* 2015, **5**:12393.
 17. Ciryam P, Kundra R, Freer R, Morimoto RI, Dobson CM, Vendruscolo M: **A transcriptional signature of Alzheimer's disease is associated with a metastable subproteome at risk for aggregation.***P Natl Acad Sci USA* 2016, **113**:4753-4758.
 18. syn14237651: <https://www.synapse.org/#!/Synapse:syn14237651>.
 19. Huang YF, Gulko B, Siepel A: **Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data.***Nat Genet* 2017, **49**:618-624.
 20. Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, Troyanskaya OG: **Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk.***Nature Genetics* 2018, **50**:1171+.
 21. Sundaram L, Gao H, Padigepati SR, McRae JF, Li YJ, Kosmicki JA, Fritzilas N, Hakenberg J, Dutta A, Shon J, et al: **Predicting the clinical impact of human mutation with deep neural networks.***Nature Genetics* 2018, **50**:1161+.
 22. Wang Z, Zhang Q, Zhang W, Lin JR, Cai Y, Mitra J, Zhang ZD: **HEDD: Human Enhancer Disease Database.***Nucleic Acids Res* 2018, **46**:D113-D120.
 23. Consortium GT: **The Genotype-Tissue Expression (GTEx) project.***Nat Genet* 2013, **45**:580-585.
 24. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.***Nat Protoc* 2009, **4**:44-57.
 25. Skene NG, Grant SG: **Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment.***Front Neurosci* 2016, **10**:16.
 26. Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, Duong TE, Gao D, Chun J, Kharchenko PV, Zhang K: **Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain.***Nat Biotechnol* 2018, **36**:70-80.

27. Benita Y, Cao Z, Giallourakis C, Li C, Gardet A, Xavier RJ: **Gene enrichment profiles reveal T-cell development, differentiation, and lineage-specific transcription factors including ZBTB25 as a novel NF-AT repressor.***Blood* 2010, **115**:5376-5384.
28. Murtagh F, Legendre P: **Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?***J Classif* 2014, **31**:274-295.
29. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, van de Lagemaat LN, Smith KA, Ebbert A, Riley ZL, et al: **An anatomically comprehensive atlas of the adult human brain transcriptome.***Nature* 2012, **489**:391-399.
30. Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, Colantuoni EA, Elkahloun AG, Herman MM, Weinberger DR, Kleinman JE: **Temporal dynamics and genetic control of transcription in the human prefrontal cortex.***Nature* 2011, **478**:519-523.
31. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.***Mol Cell* 2010, **38**:576-589.
32. Morgan BP: **Complement in the pathogenesis of Alzheimer's disease.***Semin Immunopathol* 2018, **40**:113-124.
33. Zhu X, Lee HG, Raina AK, Perry G, Smith MA: **The role of mitogen-activated protein kinase pathways in Alzheimer's disease.***Neurosignals* 2002, **11**:270-281.
34. Wang L, Chadwick W, Park SS, Zhou Y, Silver N, Martin B, Maudsley S: **Gonadotropin-releasing hormone receptor system: modulatory role in aging and neurodegeneration.***CNS Neurol Disord Drug Targets* 2010, **9**:651-660.
35. Zhang B, Wang Y, Li H, Xiong R, Zhao ZB, Chu XK, Li QQ, Sun SY, Chen SD: **Neuroprotective effects of salidroside through PI3K/Akt pathway activation in Alzheimer's disease models.***Drug Des Dev Ther* 2016, **10**:1335-1343.
36. Mitre M, Mariga A, Chao MV: **Neurotrophin signalling: novel insights into mechanisms and pathophysiology.***Clin Sci* 2017, **131**:13-23.
37. Berridge MJ: **Calcium signalling and Alzheimer's disease.***Neurochem Res* 2011, **36**:1149-1156.
38. Slack C: **Ras signaling in aging and metabolic regulation.***Nutr Healthy Aging* 2017, **4**:195-205.
39. Lampron A, Gosselin D, Rivest S: **Targeting the hematopoietic system for the treatment of Alzheimer's disease.***Brain Behav Immun* 2011, **25**:S71-S79.
40. Koldamova R, Staufenbiel M, Lefterov I: **Lack of ABCA1 considerably decreases brain ApoE level and increases amyloid deposition in APP23 mice.***J Biol Chem* 2005, **280**:43224-43235.
41. Wahrle SE, Jiang H, Parsadanian M, Kim J, Li A, Knoten A, Jain S, Hirsch-Reinshagen V, Wellington CL, Bales KR, et al: **Overexpression of ABCA1 reduces amyloid deposition in the PDAPP mouse model of Alzheimer disease.***J Clin Invest* 2008, **118**:671-682.
42. Krimbou L, Denis M, Haidar B, Carrier M, Marcil M, Genest J: **Molecular interactions between apoE and ABCA1: impact on apoE lipidation.***Journal of Lipid Research* 2004, **45**:839-848.

43. Chapuis J, Hansmannel F, Gistelincq M, Mounier A, Van Cauwenberghe C, Kolen KV, Geller F, Sottejeau Y, Harold D, Dourlen P, et al: **Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology.***Mol Psychiatr* 2013, **18**:1225-1234.
44. Allen M, Kachadorian M, Quicksall Z, Zou FG, Chai HS, Younkin C, Crook JE, Pankratz VS, Carrasquillo MM, Krishnan S, et al: **Association of MAPT haplotypes with Alzheimer's disease risk and MAPT brain gene expression levels.***Alzheimers Res Ther* 2014, **6**.
45. Hirano A, Ohara T, Takahashi A, Aoki M, Fuyuno Y, Ashikawa K, Morihara T, Takeda M, Kamino K, Oshima E, et al: **A genome-wide association study of late-onset Alzheimer's disease in a Japanese population.***Psychiat Genet* 2015, **25**:139-146.
46. Heppner FL, Ransohoff RM, Becher B: **Immune attack: the role of inflammation in Alzheimer disease.***Nat Rev Neurosci* 2015, **16**:358-372.
47. Jiang T, Tan L, Zhu XC, Zhang QQ, Cao L, Tan MS, Gu LZ, Wang HF, Ding ZZ, Zhang YD, Yu JT: **Upregulation of TREM2 ameliorates neuropathology and rescues spatial cognitive impairment in a transgenic mouse model of Alzheimer's disease.***Neuropsychopharmacology* 2014, **39**:2949-2962.
48. Wang Y, Cella M, Mallinson K, Ulrich JD, Young KL, Robinette ML, Gilfillan S, Krishnan GM, Sudhakar S, Zinselmeyer BH, et al: **TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model.***Cell* 2015, **160**:1061-1071.
49. Andersen OM, Reiche J, Schmidt V, Gotthardt M, Spoelgen R, Behlke J, von Arnim CAF, Breiderhoff T, Jansen P, Wu X, et al: **Neuronal sorting protein-related receptor sorLA/LR11 regulates processing of the amyloid precursor protein.***P Natl Acad Sci USA* 2005, **102**:13461-13466.
50. Li H, Karl T, Garner B: **Understanding the function of ABCA7 in Alzheimer's disease.***Biochem Soc Trans* 2015, **43**:920-923.
51. Dourlen P, Fernandez-Gomez FJ, Dupont C, Grenier-Boley B, Bellenguez C, Obriot H, Caillierez R, Sottejeau Y, Chapuis J, Bretteville A, et al: **Functional screening of Alzheimer risk loci identifies PTK2B as an in vivo modulator and early marker of Tau pathology.***Mol Psychiatr* 2017, **22**:874-883.
52. Moreau K, Fleming A, Imarisio S, Lopez Ramirez A, Mercer JL, Jimenez-Sanchez M, Bento CF, Puri C, Zavodszky E, Siddiqi F, et al: **PICALM modulates autophagy activity and tau accumulation.***Nat Commun* 2014, **5**:4998.
53. Crehan H, Holton P, Wray S, Pocock J, Guerreiro R, Hardy J: **Complement receptor 1 (CR1) and Alzheimer's disease.***Immunobiology* 2012, **217**:244-250.
54. Hansen DV, Hanson JE, Sheng M: **Microglia in Alzheimer's disease.***J Cell Biol* 2018, **217**:459-472.
55. Kelleher RJ, Soiza RL: **Evidence of endothelial dysfunction in the development of Alzheimer's disease: Is Alzheimer's a vascular disorder?***Am J Cardiovasc Dis* 2013, **3**:197-226.
56. Sagare AP, Bell RD, Zhao Z, Ma Q, Winkler EA, Ramanathan A, Zlokovic BV: **Pericyte loss influences Alzheimer-like neurodegeneration in mice.***Nat Commun* 2013, **4**:2932.
57. Cai Z, Xiao M: **Oligodendrocytes and Alzheimer's disease.***Int J Neurosci* 2016, **126**:97-104.

58. Cuberos H, Vallee B, Vourc'h P, Tastet J, Andres CR, Benedetti H: **Roles of LIM kinases in central nervous system function and dysfunction.***FEBS Lett* 2015, **589**:3795-3806.
59. Cano CE, Gommeaux J, Pietri S, Culcasi M, Garcia S, Seux M, Barelier S, Vasseur S, Spoto RP, Pebusque MJ, et al: **Tumor Protein 53-Induced Nuclear Protein 1 Is a Major Mediator of p53 Antioxidant Function.***Cancer Res* 2009, **69**:219-226.
60. Driver JA, Beiser A, Au R, Kreger BE, Splansky GL, Kurth T, Kiel DP, Lu KP, Seshadri S, Wolf PA: **Inverse association between cancer and Alzheimer's disease: results from the Framingham Heart Study.***Brit Med J* 2012, **344**.
61. Miller LM, Wang Q, Telivala TP, Smith RJ, Lanzirotti A, Miklossy J: **Synchrotron-based infrared and X-ray imaging shows focalized accumulation of Cu and Zn co-localized with beta-amyloid deposits in Alzheimer's disease.***J Struct Biol* 2006, **155**:30-37.
62. Bencze J, Morotz GM, Seo W, Bencs V, Kalman J, Miller CCJ, Hortobagyi T: **Biological function of Lemur tyrosine kinase 2 (LMTK2): implications in neurodegeneration.***Mol Brain* 2018, **11**:20.
63. Matarin M, Salih DA, Yasvoina M, Cummings DM, Guelfi S, Liu W, Nahaboo Solim MA, Moens TG, Paublete RM, Ali SS, et al: **A genome-wide gene-expression analysis and database in transgenic mice during development of amyloid or tau pathology.***Cell Rep* 2015, **10**:633-644.
64. Yeates EF, Tesco G: **The Endosome-associated Deubiquitinating Enzyme USP8 Regulates BACE1 Enzyme Ubiquitination and Degradation.***J Biol Chem* 2016, **291**:15753-15766.
65. Friedmann E, Hauben E, Maylandt K, Schleegeer S, Vreugde S, Lichtenthaler SF, Kuhn PH, Stauffer D, Rovelli G, Martoglio B: **SPPL2a and SPPL2b promote intramembrane proteolysis of TNFalpha in activated dendritic cells to trigger IL-12 production.***Nat Cell Biol* 2006, **8**:843-848.
66. Strunz M, Jarrell JT, Cohen DS, Rosin ER, Vanderburg CR, Huang X: **Modulation of SPARC/Hevin Proteins in Alzheimer's Disease Brain Injury.***J Alzheimers Dis* 2019, **68**:695-710.
67. Lin JR, Cai Y, Zhang Q, Zhang W, Nogales-Cadenas R, Zhang ZD: **Integrated Post-GWAS Analysis Sheds New Light on the Disease Mechanisms of Schizophrenia.***Genetics* 2016, **204**:1587-1600.
68. Holton P, Ryten M, Nalls M, Trabzuni D, Weale ME, Hernandez D, Crehan H, Gibbs JR, Mayeux R, Haines JL, et al: **Initial assessment of the pathogenic mechanisms of the recently identified Alzheimer risk Loci.***Ann Hum Genet* 2013, **77**:85-105.
69. N'Songo A, Carrasquillo MM, Wang X, Burgess JD, Nguyen T, Asmann YW, Serie DJ, Younkin SG, Allen M, Pedraza O, et al: **African American exome sequencing identifies potential risk variants at Alzheimer disease loci.***Neurol-Genet* 2017, **3**.
70. Tao Yang PT, Sihai Zhao, Jieping Ye: **Identifying Genetic Risk Factors via Sparse Group Lasso with Group Graph Structure.***arXiv preprint arXiv:170903645* 2017.
71. Jun G, Ibrahim-Verbaas CA, Vronskaya M, Lambert JC, Chung J, Naj AC, Kunkle BW, Wang LS, Bis JC, Bellenguez C, et al: **A novel Alzheimer disease locus located near the gene encoding tau protein.***Mol Psychiatry* 2016, **21**:108-117.
72. van Arensbergen J, Pagie L, FitzPatrick VD, de Haas M, Baltissen MP, Comoglio F, van der Weide RH, Teunissen H, Vosa U, Franke L, et al: **High-throughput identification of human SNPs affecting**

- regulatory element activity.*Nat Genet* 2019, **51**:1160-1169.**
73. Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, Katayama T, Baldwin CT, Cheng R, Hasegawa H, et al: **The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease.***Nature Genetics* 2007, **39**:168-177.
74. Hashimoto Y, Nawa M, Chiba T, Aiso S, Nishimoto I, Matsuoka M: **Transforming growth factor beta2 autocrinally mediates neuronal cell death induced by amyloid-beta.***J Neurosci Res* 2006, **83**:1039-1047.
75. von Bernhardi R, Cornejo F, Parada GE, Eugenini J: **Role of TGFbeta signaling in the pathogenesis of Alzheimer's disease.***Front Cell Neurosci* 2015, **9**:426.
76. Tong Y, Xu Y, Scarce-Levie K, Ptacek LJ, Fu YH: **COL25A1 triggers and promotes Alzheimer's disease-like pathology in vivo.***Neurogenetics* 2010, **11**:41-52.
77. Kaur G, Levy E: **Cystatin C in Alzheimer's disease.***Front Mol Neurosci* 2012, **5**:79.
78. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shores N, Whitton H, Ryan RJ, Shishkin AA, et al: **Genetic and epigenetic fine mapping of causal autoimmune disease variants.***Nature* 2015, **518**:337-343.
79. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, Gephart MGH, Barres BA, Quake SR: **A survey of human brain transcriptome diversity at the single cell level.***P Natl Acad Sci USA* 2015, **112**:7285-7290.
80. Wang P, Zhao D, Lachman HM, Zheng D: **Enriched expression of genes associated with autism spectrum disorders in human inhibitory neurons.***Transl Psychiatry* 2018, **8**:13.
81. Condello C, Yuan P, Schain A, Grutzendler J: **Microglia constitute a barrier that prevents neurotoxic protofibrillar A beta 42 hotspots around plaques.***Nature Communications* 2015, **6**.
82. Mattson MP: **Apoptosis in neurodegenerative disorders.***Nat Rev Mol Cell Biol* 2000, **1**:120-129.
83. Shen K, Sidik H, Talbot WS: **The Rag-Ragulator Complex Regulates Lysosome Function and Phagocytic Flux in Microglia.***Cell Rep* 2016, **14**:547-559.
84. Majumder P, Roy K, Singh BK, Jana NR, Mukhopadhyay D: **Cellular levels of Grb2 and cytoskeleton stability are correlated in a neurodegenerative scenario.***Dis Model Mech* 2017, **10**:655-669.
85. Russo C, Dolcini V, Salis S, Venezia V, Zambrano N, Russo T, Schettini G: **Signal transduction through tyrosine-phosphorylated C-terminal fragments of amyloid precursor protein via an enhanced interaction with Shc/Grb2 adaptor proteins in reactive astrocytes of Alzheimer's disease brain.***J Biol Chem* 2002, **277**:35282-35288.

Figures

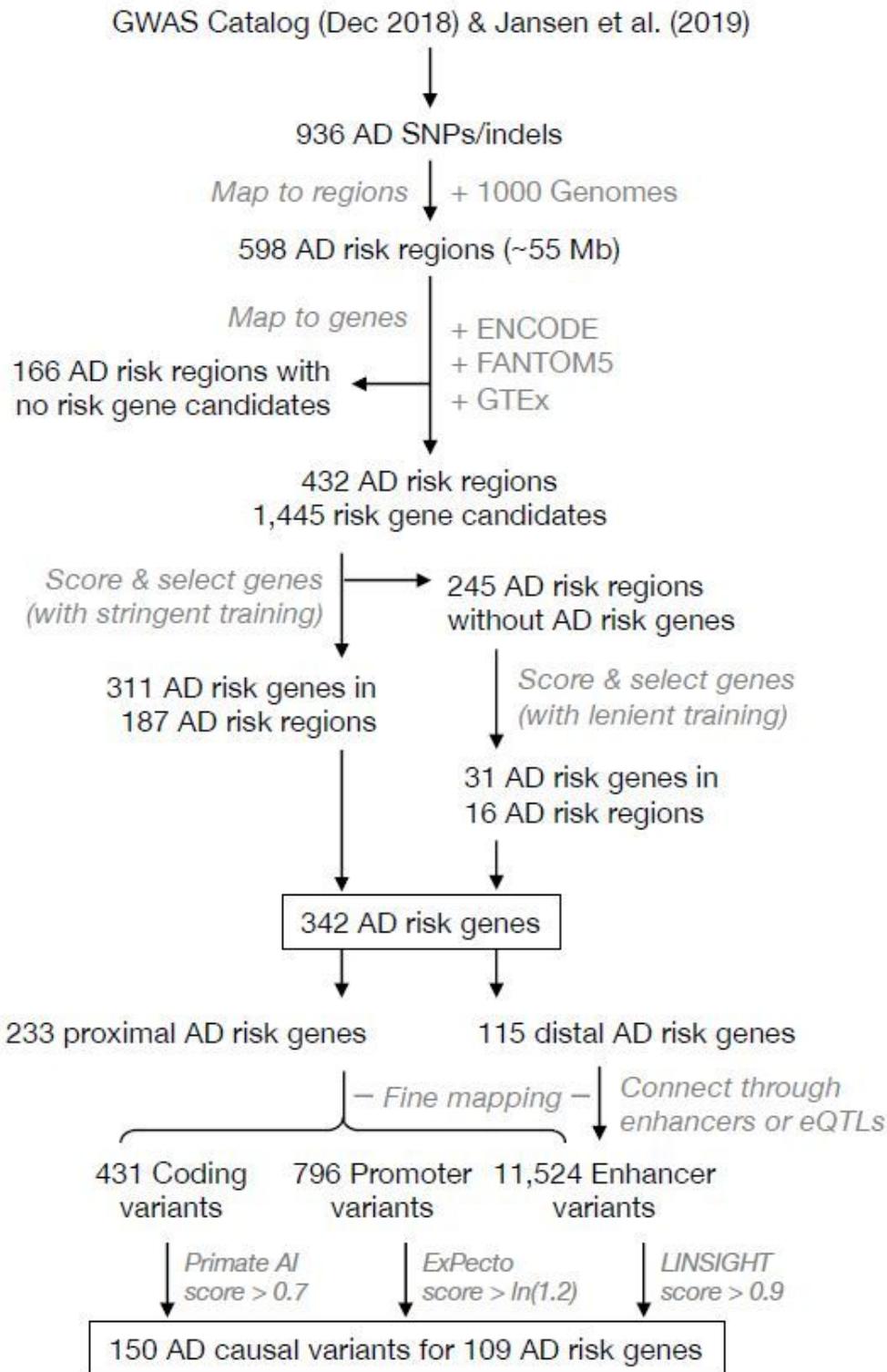


Figure 1

Flowchart of AD risk genes and causal variants prediction.

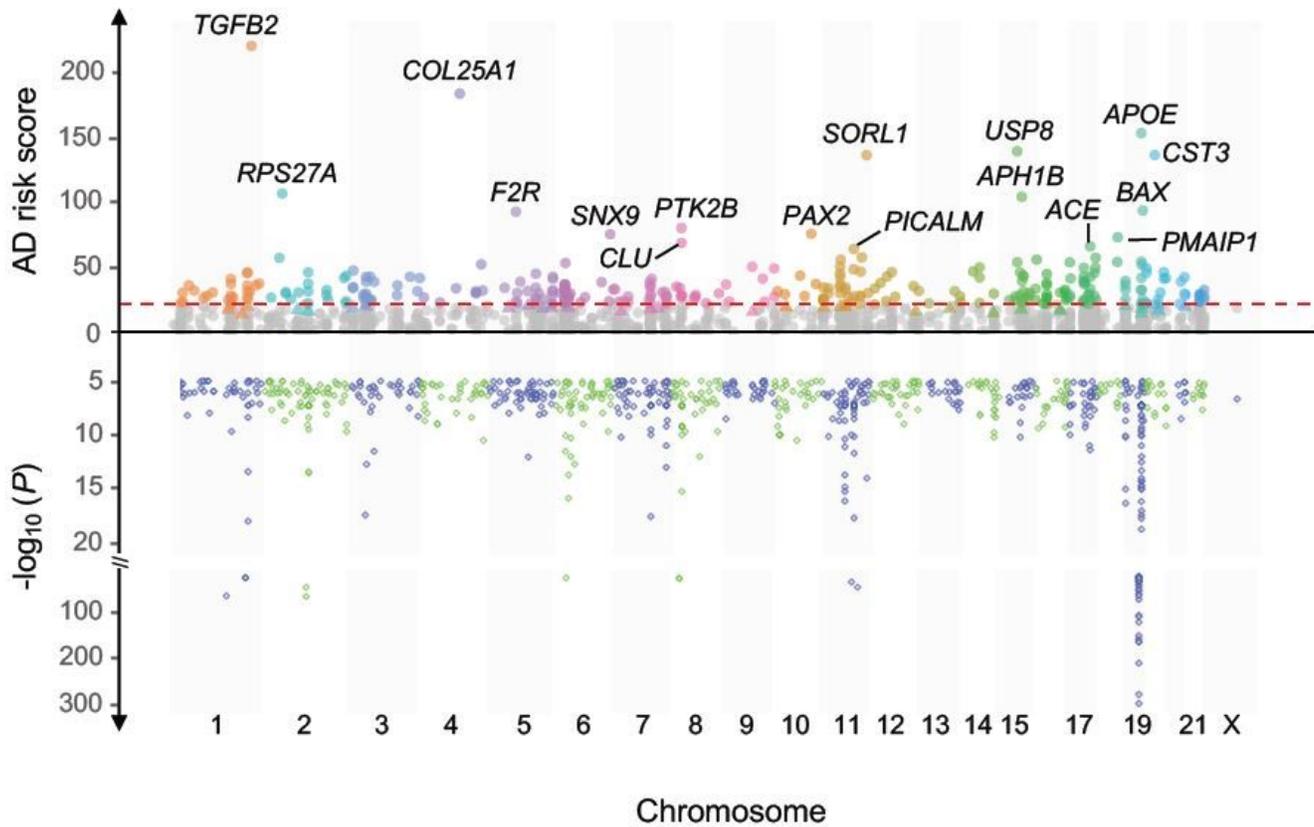


Figure 2

AD risk genes and AD-associated variants across chromosomes. The upper part of the figure shows the scores of AD risk gene candidates, which were calculated based on the stringent training gene set. Colored and gray dots represent AD risk gene candidates above and below the threshold (= 21.4, the red dashed line), respectively. The colored small triangles below the threshold represent additional AD risk genes predicted with the lenient gene training set. Several top AD risk genes are labelled with their gene symbols. The upper part of the figure is the Manhattan plot of the 936 AD-associated SNPs that we collected from the GWAS Catalog and Jansen et al [9].

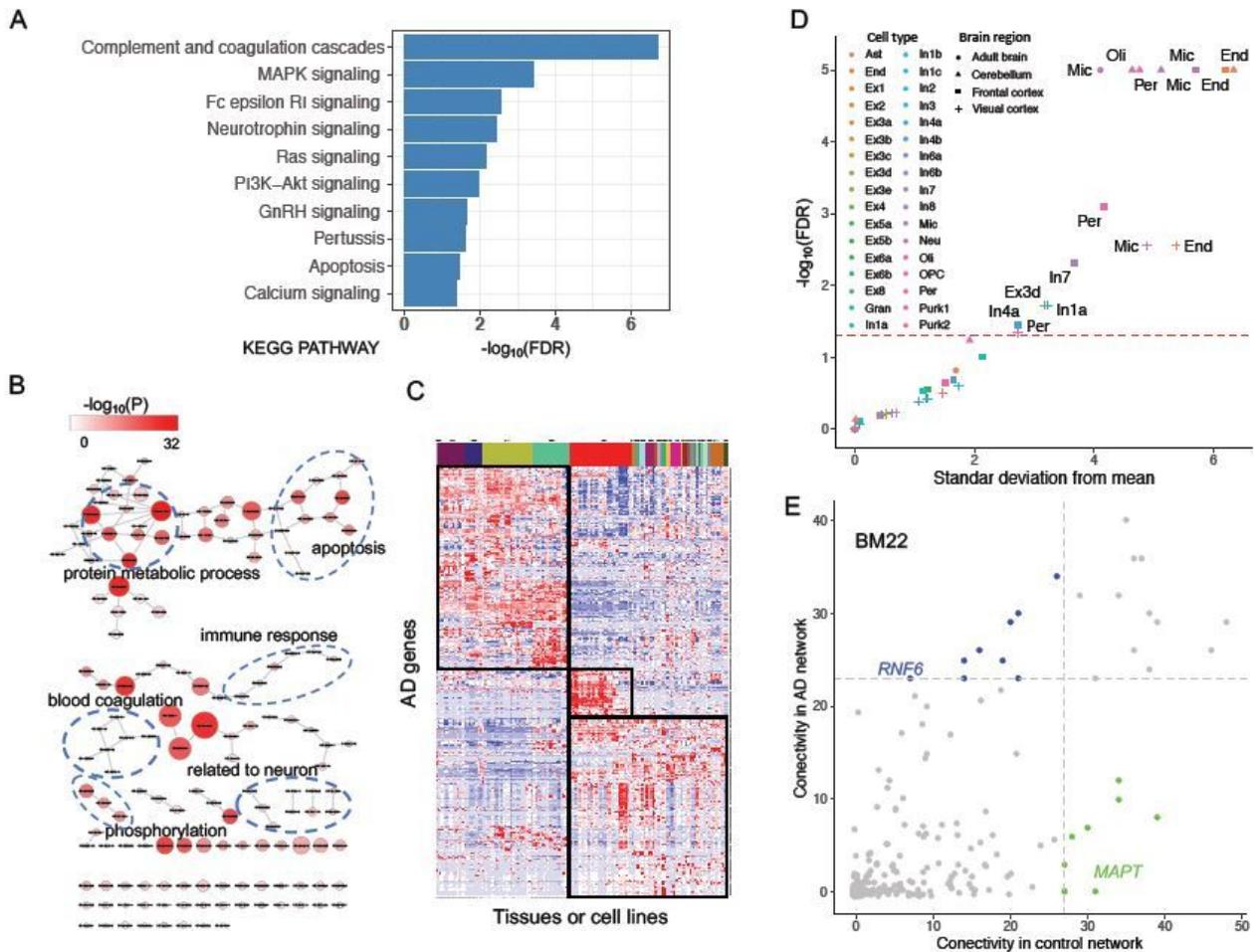


Figure 3

Functional annotation of AD risk genes. (A) KEGG pathways and (B) GO terms of biological processes enrichment of AD risk genes. In (B), each dot represents a significantly enriched GO term, whose $-\log_{10}(\text{P-value})$ and AD risk gene count are indicated by the color and size of the dot, respectively. (C) Expression of AD risk genes in different tissues. Three expression clusters are indicated by the black boxes. See Supplementary Figure S3 for details. (D) Cell type enrichment analysis. We analyzed the expression of AD risk genes in different types of single cells from frontal and visual cortices, cerebellum, and adult brain. The red dashed line represents $\text{FDR} = 0.05$, and cell types whose transcriptomes were significantly enriched with the expression of AD risk genes are shown, with their brain regions indicated. Used cell type abbreviations: Ast, astrocytes; End, endothelial; Ex, excitatory; Gran, granule; In, inhibitory; Mic, microglia; Neu, neuron; Oli, oligodendrocytes; OPC, oligodendrocytes precursor; Per, pericytes; Purk, purkinje. (E) Network connectivity of AD risk genes in the co-expression network of both AD patients and normal controls in the brain region BM22. Each dot represents an AD risk gene. The gray dashed line marks the threshold for network hub genes either in the co-expression network of AD patients or normal controls. Blue and green dots represent network hub genes only in the co-expression network of AD

patients and only in the co-expression network of normal controls, respectively. See Supplementary Figure S6 for details and other brain regions.

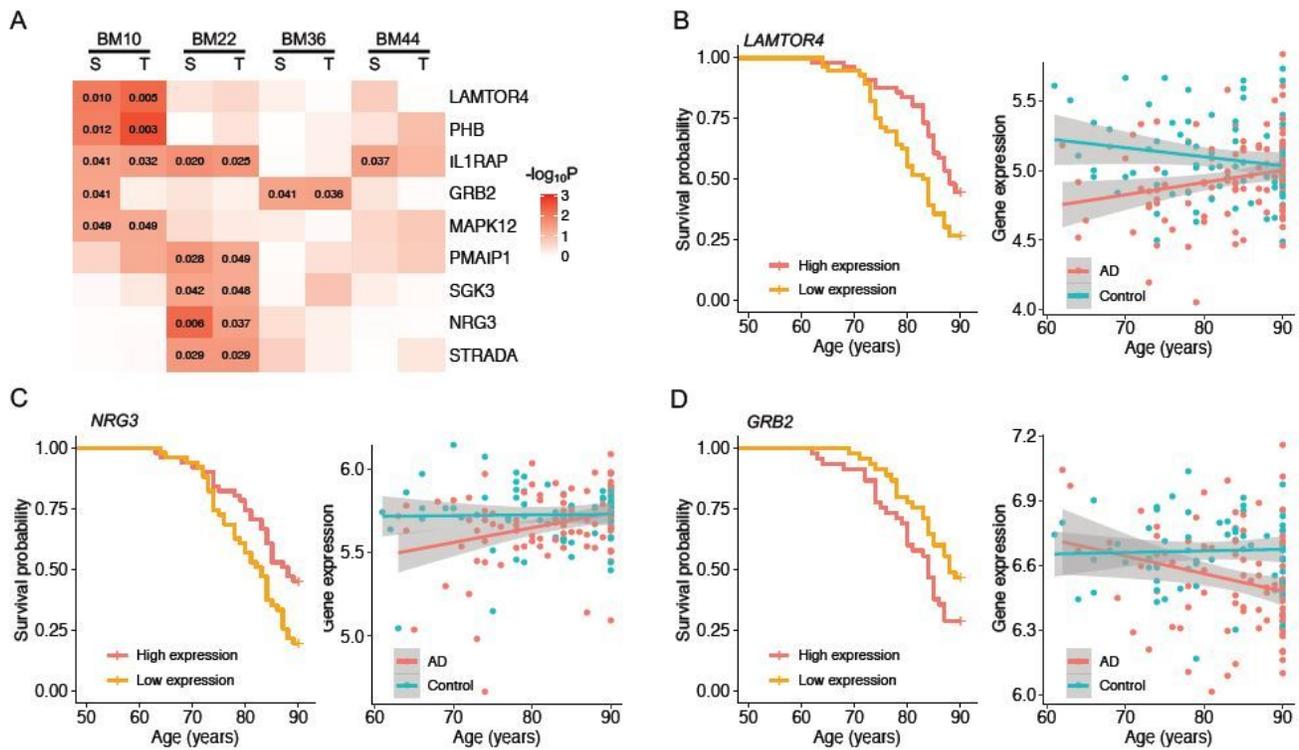


Figure 4

AD survival curves and gene expression trajectories with age. We analyzed the expression of each AD risk gene to see how it may affect the survival of AD patients with low and high expression levels and how it trends with age in AD patients and normal controls. (A) The heatmap of P-values of Kaplan-Meier survival analyses and gene expression trend tests across four brain regions (see Methods). Nine AD risk genes with at least one P-value less than 0.05 (labeled) are included in the heatmap. This figure includes AD survival curves and gene expression trajectories with age for three genes with $P < 0.05$: (B) *LAMTOR4* in the brain region BM10, (C) *NRG3* in the brain region BM22, and (D) *GRB2* in the brain region BM36. See Supplementary Figure S10 for plots of other significant results.

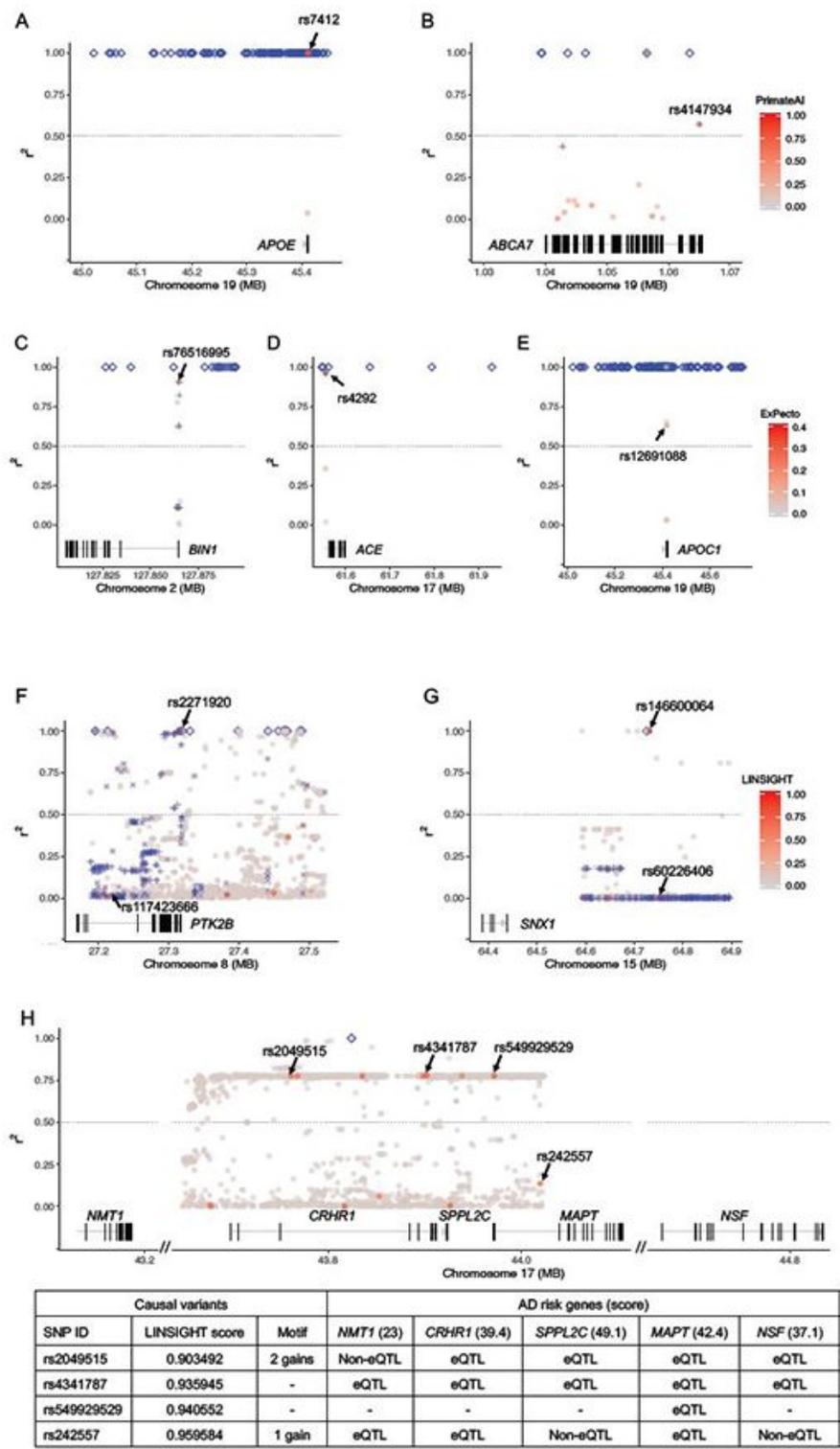


Figure 5

Predicted causal variants. Plotted are examples of coding, promoter, and enhancer causal variants identified using PrimateAI (A-B), ExPecto (C-E), and LINSIGHT (F-H), respectively, with their affected AD risk genes. Each dot represents a variant, whose annotation score is indicated by the color of the dot. Predicted causal variants are marked by arrows with their SNP IDs. Blue diamonds represent AD-

associated variants identified by GWAS. Plus and cross signs represent eQTLs and variants located in enhancers. The gray dashed line represents LD $r^2 = 0.5$.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTables.xlsx](#)
- [SupplementaryFigures.pdf](#)
- [SupplementaryFiguresandTablesLegends.docx](#)