



OPEN Identification of novel risk genes for Alzheimer's disease by integrating genetics from hippocampus

Jie Li^{1,5}, Lingfang Li^{2,5}, Shanshan Cai¹, Kun Song³✉ & Shenghui Hu⁴✉

Alzheimer's disease (AD) stands as the most prevalent neurodegenerative ailment, presently lacking a definitive cure. Given that primary medications for AD patients in the early or middle stages demonstrate optimal efficacy, it becomes crucial to delve into the identification of risk genes associated with early onset. In our study, we compiled and integrated three transcriptomics datasets (GSE48350, GSE36980, GSE5281) originating from the hippocampus of 37 AD patients and 66 healthy controls (CTR) for comprehensive bioinformatics analysis. Comparative analysis with CTR revealed 25 up-regulated genes and 291 down-regulated genes in AD. Those down-regulated genes were notably enriched in processes related to the transmission and transport of synaptic signals. Intriguingly, 27 differentially expressed genes implicated in AD were also correlated with the Braak stage, establishing a connection with various immune cell types that exhibit differences in AD, including cytotoxic T cells, neutrophils, CD4 T cells, Th1, Th2, and Tfh. Significantly, a Cox model, constructed using nine feature genes, effectively stratified AD samples (HR = 2.72, 95% CI 1.94 ~ 3.81, P = 3.6e-10), highlighting their promising potential for risk assessment. In conclusion, our investigation sheds light on novel genes intricately linked to the onset and progression of AD, offering potential biomarkers for the early detection of this debilitating condition. This study contributes valuable insights toward enhancing the strategies for preventing and treating AD.

Keywords Alzheimer's disease, Hippocampus, Transcriptomics, Gene, Biomarker

Alzheimer's disease (AD) stands as a progressive neurodegenerative disorder influenced by an interplay of genetic and epigenetic factors, alongside gene-environment interactions that potentially contribute to its onset¹. AD holds the prominent position as the world's foremost prevalent neurodegenerative disorder, commanding significant attention and concern globally. Clinically, AD manifests as a profound impairment of executive and cognitive functions². Pathologically, the disease progression is characterized by escalating hippocampal and cortical atrophy, observable through neuroimaging and visual examination. This is accompanied by the presence of intracellular neurofibrillary tangles (NFTs) and the extracellular deposition of hyperphosphorylated amyloid-beta (A β) 1-42 peptides, leading to neuronal and synaptic loss, as well as reactive glial hyperplasia^{1,3-5}. As the aging population burgeons, AD has emerged as the predominant cause of dementia, accounting for 50-75% of cases, with its incidence doubling approximately every 5 years after the age of 65⁶. The escalating prevalence of AD, coupled with its growing social and economic burden, has positioned it as a significant societal challenge, imperiling the health of both urban and rural residents in China⁷. A recent national cross-sectional study in China reported 15.07 million dementia patients aged 60 and above, among whom 9.83 million are afflicted with AD⁷. Given that most FDA-approved drugs exhibit optimal efficacy in the early or middle stages of AD, there is an imperative need to explore the risks associated with the early onset of this debilitating condition.

Modern high-throughput sequencing technologies have empowered the generation of vast datasets, providing a robust methodology for delving into the etiology of AD. Molecular genetic investigations have revealed key genes implicated in AD, including amyloid precursor protein (APP)^{8,9}, presenilin 1 (PSEN1)^{9,10}, and

¹School of Basic Medical Sciences, Hunan University of Medicine, Huaihua, Hunan, China. ²Department of Cardiovascular Medicine, Xiangya Hospital, Central South University, Changsha, Hunan, China. ³Department of Gastrointestinal Surgery & National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, Hunan, China. ⁴Department of Orthopaedics, Xiangya Second Hospital, Central South University, Changsha, Hunan, China. ⁵These authors contributed equally: Jie Li and Lingfang Li. ✉email: sking1986@csu.edu.cn; 190847484@qq.com

presenilin 2 (PSEN2)^{9,11}, identified as causative factors. Several risk factors influencing AD development have also been identified, encompassing elements such as smoking¹², stress¹³, depression¹⁴, and insufficient sleep¹⁵. Furthermore, extensive family studies have pinpointed a robust risk gene—the E4 allele of Apolipoprotein E (ApoE), which significantly heightens the risk of AD across diverse populations^{16–18}. Additionally, the triggering receptor expressed on myeloid cells 2 (TREM2) gene has emerged as another noteworthy contributor, elevating the risk ratio by 2.9% for the development of AD^{19,20}.

The upregulation or downregulation of genes can instigate changes in metabolic, immune, and other physiological processes, contributing to the onset of diseases^{21,22}. Therefore, the identification of differentially expressed genes (DEGs) is a crucial avenue for unraveling altered biological pathways in various diseases, including neurological disorders and cancers²³. Given that the selective vulnerability of specific brain regions plays a pivotal role in neurodegenerative disorders, and AD is characterized by profound neuronal damage in the hippocampus situated in the medial temporal lobe²⁴, exploring DEGs becomes particularly pertinent. While the occurrence of AD is known to increase with age, we contend that abnormal transcriptional alterations may also contribute to disease-related mechanisms^{25,26}. In essence, it is imperative to investigate whether these DEGs in the hippocampus may exert an influence on the onset of AD.

Consistent with our previous research, the integration of uniform data from multiple studies has been shown to increase the statistical power of genetic analysis²⁷. In this study, we performed a thorough search and retrieved three transcriptomic datasets of the hippocampus from the Gene Expression Omnibus. By leveraging integrated data from matched baselines between AD patients and controls, we conducted a meticulous analysis of DEGs to precisely investigate the pivotal role of gene transcription in the pathogenesis of AD. The objective of this investigation is to uncover novel risk genes associated with the pathogenesis of AD.

Materials and methods

Data source and processing

We conducted a thorough search on the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo>) using the keywords “Alzheimer” and “hippocampus”. Subsequently, we identified three microarray datasets utilizing the GPL570 platform, namely GSE48350, GSE36980, and GSE5281 (Table S1). The raw “CEL” files for each dataset were procured from the GEO database and subjected to processing and normalization using the R package *affy* (version 1.68.0). Following this, the datasets were integrated to mitigate batch effects, employing the R package *sva* (version 3.38.0). The merged dataset encompassed 37 AD samples and 66 CTR samples. We have retrieved all the RNA sequencing data related to the hippocampus of AD patients from the GEO database. Although the sample size is limited, the sample we utilized is invaluable as it was derived from the hippocampus of actual patients. The hippocampus, a crucial region of the brain, is responsible for memory processing and spatial orientation. AD patients often exhibit hippocampal damage and functional decline. During the progression of AD, the hippocampus undergoes a range of pathological changes, including neuronal death, neurofibrillary tangles, and amyloid protein deposition. These alterations lead to hippocampal atrophy and loss of function, which subsequently impacts the patient’s memory and cognitive abilities. Hence, there is a close association between AD and the hippocampus.

The procedural details are visually represented in Figure S1, elucidating the workflow undertaken in our analysis.

Identification of DEGs and enrichment analysis

In our analysis, the representation of a gene’s expression value involves calculating the average value across corresponding probes when multiple probes are associated with the same gene. Conversely, when a probe corresponds to multiple genes, the expression values of those respective genes are uniformly represented by the expression value of the probe. DEGs between AD and CTR samples were identified using the R package *limma* (version 3.54.2). We applied stringent criteria, setting thresholds at an adjusted *P* value less than 0.05 and an absolute log₂ fold change (log₂FC) greater than 0.585. Subsequently, down-regulated and up-regulated DEGs were separately subjected to Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses^{28–30} using the R package *clusterProfiler* (version 4.6.2).

Similarly, DEGs between Braak stages (i.e., III_IV vs V_VI) were computed using the R package *limma* (version 3.54.2) and screened based on thresholds of a *P* value less than 0.05 and an absolute log₂FC greater than 0.585.

Correlation analysis of DEGs

The Pearson’s correlations among pivotal DEGs were computed and visually presented using the R package *Corrplot* (version 0.92).

Protein–protein interaction (PPI) network analysis

The DEGs between AD and CTR samples were subjected to imputation using the STRING database (<https://string-db.org/>) to elucidate PPI with a combined score exceeding 0.7, indicative of high confidence. Subsequently, the PPI network was retrieved and visualized through Cytoscape software (version 3.6.1). In the resulting network, nodes represented genes, and their distinct colors delineated up-regulated or down-regulated genes. Edges symbolized interactions between genes, with their sizes reflecting the strength of these relationships.

Evaluation of immune cells infiltration

The transcriptome expression matrix, post-removal of batch effects, was utilized to assess immune cell infiltration using the web-based tool ImmCellAI (<http://bioinfo.life.hust.edu.cn/ImmCellAI/#!/analysis>). This analysis involved estimating the abundance of twenty-four immune cell types within the microenvironment of subjects’

hippocampus. These twenty-four cell types were categorized into two layers: Layer 1, encompassing DC, B cell, Monocyte, Macrophage, NK, Neutrophil, CD4 T, CD8 T, NKT, Tgd and Layer 2, comprising CD4 naive, CD8 naive, Tc, Tex, Tr1, nTreg, iTreg, Th1, Th2, Th17, Tfh, Tcm, Tem, MAIT.

Construction of classification model

The DEGs were individually utilized as features to construct classification models using the R package caret (version 6.0–94). In this context, three distinct machine learning methods were employed to develop predictive models, each exploring different tuning parameters. These methods comprised Random Forest (rf), Neural Network (nnet), and Support Vector Machines with Radial Basis Function Kernel (svmRadial). Details regarding these methods, including their associated R packages and tuning parameters, can be found in Table S2. For model training, 70% of the samples were randomly selected as the training set, with the remaining 30% constituting the test set. Receiver Operating Characteristic (ROC) curves were generated using the R package pROC (version 1.18.0), and the Area Under Curves (AUC) was employed to assess the predictive performance of the classifiers.

Construction of prognostic model

The identified pivotal DEGs were employed as features in a multivariable Cox proportional hazards regression model, executed using the R package survival (version 3.5–3). Kaplan–Meier survival curves were then constructed with the R package survminer (version 0.4.9), and the significance of survival differences was assessed through log-rank tests. The risk score was computed as follows:

$$Riskscore = \sum_{i=1}^8 Coef_i \times Feature_i = \sum_{i=1}^8 \ln HR_i \times Feature_i$$

where ‘i’ denotes the *i*-th feature, ‘Feature’ represents its expression value, and ‘Coef’ stands for its coefficient in the fitted Cox model. The ‘Coef’ corresponds to the natural logarithm of the hazard ratio (HR).

In our dataset, the median of the risk score was utilized to categorize samples into high- and low-risk groups. Subsequently, univariable Cox proportional hazards regression analysis was performed to ascertain the HR, along with its 95% confidence interval (CI) and associated *P*-value, comparing the high- and low-risk groups.

Results

Identification of differentially expressed genes in AD compared with CTR

As detailed in Table S1, three datasets involving the sequencing of hippocampal tissue were acquired from GEO. All samples were amalgamated after mitigating batch effects. To discern the differentially expressed genes in AD relative to CTR, we specifically opted for age- and sex-matched samples. This subset comprised 37 AD samples and 46 CTR samples. Notably, only those samples with an age exceeding 60 were retained for subsequent analyses (Table 1).

In this analysis, a total of 316 DEGs were discerned in AD compared with CTR, comprising 25 up-regulated and 291 down-regulated genes (Fig. 1A, Table S3). The prevalence of down-regulated DEGs suggests that the biological pathways associated with these genes may be suppressed, potentially linking them with the pathogenesis of AD. Subsequently, the up-regulated and down-regulated genes underwent KEGG enrichment analysis and GO enrichment analysis, respectively. Due to the limited number of up-regulated genes, no significant pathways were enriched. Delving into the biology pathways influenced by the down-regulated DEGs, KEGG enrichment analysis and GO enrichment analysis were conducted. As anticipated, the results revealed significant enrichment in pathways such as Alzheimer’s disease and Pathways of neurodegeneration—multiple diseases, aligning with prior studies^{31,32}. GO enrichment analysis further highlighted biological processes (BP) related to AD, including vesicle-mediated transport in synapses, neurotransmitter transport, learning or memory, exocytosis, and axonogenesis (Fig. 1B, Table S4B). Concurrently, significantly enriched cellular components (CC) encompassed distal axon, neuronal cell body, and synaptic membrane, while molecular functions (MF) were associated with transmembrane transporter binding, channel regulator activity, and metal ion transmembrane transporter activity (Fig. 1B). These findings underscore the potential association of these down-regulated DEGs with the transmission and transport of synaptic signals, potentially contributing to the onset of disease in AD patients.

	AD (n = 37)	CTR (n = 66)	<i>P</i> value
Number (age > 60)	37	46	
Age ± sd (years)	83.51 ± 8.75	80.74 ± 9.06	0.1615#
Sex (Male/Female)	18/19	26/20	0.5132*
Braak stage	AD (III_IV)	AD (V_VI)	
Number	7	11	/
Age ± sd (years)	86.57 ± 5.71	81.09 ± 9.79	0.1535#
Sex (Male/Female)	4/3	4/7	0.6305*

Table 1. The characteristics of all subjects used in this study. # represents two-tailed Welch’s *t*-test. *stands for Fisher’s test.

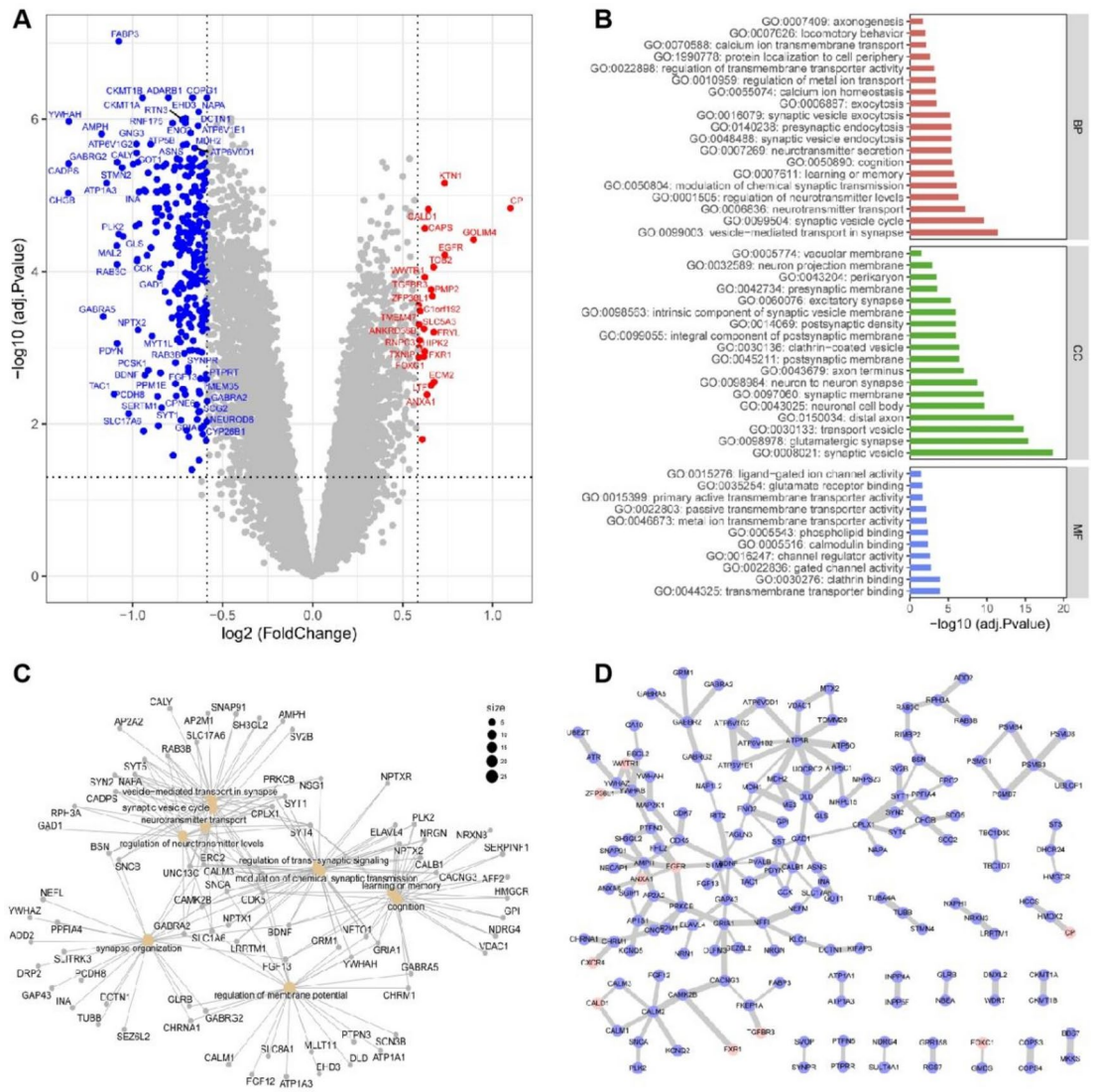


Fig. 1. Identification and analysis of DEGs between age-matched AD patients and controls. **(A)** Volcano plot shows the DEGs between AD samples and controls. **(B)** The significant GO BP, CC, and MF enriched by down-regulated genes in AD. **(C)** Those down-regulated DEGs involved in the top 10 BP. **(D)** PPI network with high confidence constructed by all DEGs. Nodes in red and in blue represented by up-regulated and down-regulated genes, respectively. Edges stand for the interaction score between proteins.

Furthermore, we highlighted the down-regulated DEGs involved in the top 10 BP (Fig. 1C), revealing numerous genes participating in multiple pathways. This observation may stem from the shared functionality of these signaling pathways, indirectly bolstering the credibility of the association between these DEGs and AD. For instance, *SNCA* has been identified as a key player in a multitude of BPs, encompassing vesicle-mediated transport in synapses, synaptic vesicle cycle, neurotransmitter transport, synapse organization, and regulation of membrane potential. Similarly, *YWHAZ*, *TUBB*, and *GLRB* were all implicated in synapse organization. Furthermore, we constructed a PPI network using all DEGs with a combined score exceeding 0.7 (Fig. 1D). This underlines the high-confidence interactions among these DEGs at the protein level.

Screening of Braak stage-related DEGs in AD

Furthermore, building upon the analysis of AD samples categorized by Braak stage (i.e., III_IV, V_VI), we delved deeper into the exploration of DEGs associated with the Braak stage (Table 1). In this context, we identified 46 up-regulated genes and 98 down-regulated genes (Fig. 2A, Table S5). Subsequently, these genes underwent intersection analysis with the DEGs derived from AD and CTR samples. Intriguingly, we observed that 4 DEGs consistently exhibited up-regulation, while 23 DEGs consistently displayed down-regulation (Fig. 2B). This suggests that these genes are not only implicated in the onset of AD but also play a role in the disease's progressive development.

As depicted in Fig. 2C, the 27 consistent DEGs were visually presented. Notably, intriguing patterns of positive correlations among the down-regulated DEGs and negative correlations between the down-regulated DEGs

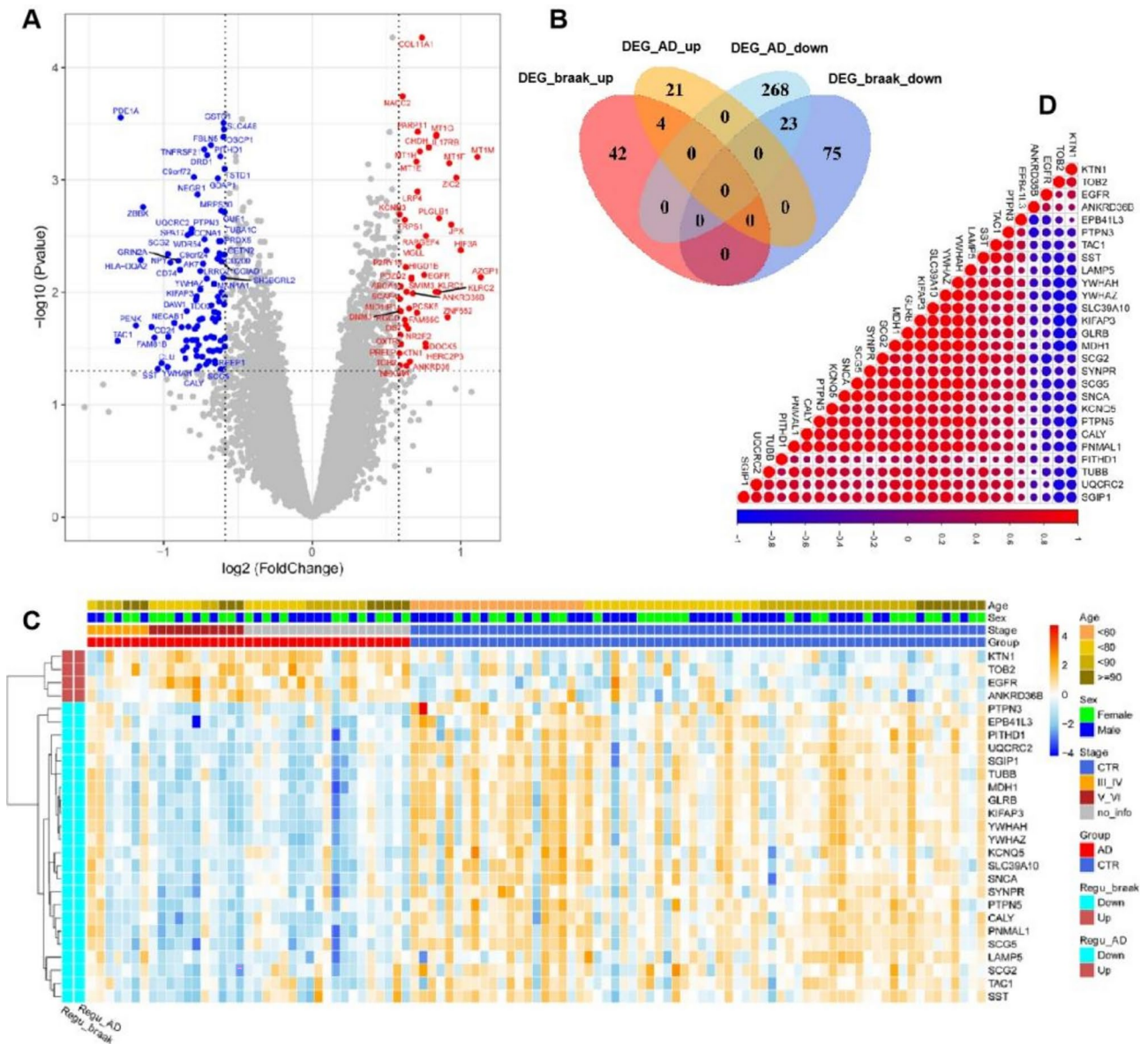


Fig. 2. Identification and analysis of DEGs between Braak stage of AD patients. **(A)** Volcano plot shows the DEGs between Braak stage of AD patients. **(B)** Venn diagram presents those overlapped DEGs. **(C)** The abundance of 27 overlapped DEGs (*KTN1*, *TOB2*, *EGFR*, *ANKRD36B*, *EPB41L3*, *PTPN3*, *TAC1*, *SST*, *LAMP5*, *YWHAH*, *YWHAZ*, *SLC39A10*, *KIFAP3*, *GLRB*, *MDH1*, *SCG2*, *SYNPR*, *SCG5*, *SNCA*, *KCNQ5*, *PTPN5*, *CALY*, *PNMAL1*, *PITHD1*, *TUBB*, *UQCRC2*, *SGIP1*). **(D)** Pearson's correlation at the transcriptomic level among the 27 overlapped DEGs.

and the up-regulated DEGs were observed (Fig. 2D). This observation hints at a potential regulatory interplay wherein the four up-regulated DEGs may inhibit the expression of the down-regulated genes. This, in turn, could suppress the activity of certain pathways, contributing to the onset and progression of AD. Simultaneously, this intriguing correlation pattern suggests the potential of these 27 genes as biomarkers, indicating their candidacy for further exploration in the context of AD.

Evaluation and comparison of immune cells infiltration in *hippocampus* from AD subjects

The transcriptomic profiling derived from age- and sex-matched samples, comprising 37 AD samples and 46 CTR samples, was employed to assess immune cell infiltration in the microenvironment of the subjects' hippocampus tissue. Utilizing ImmuCellAI, twenty-four cell types were estimated and categorized into two layers: layer 1 (DC, B cell, monocyte, macrophage, NK, neutrophil, CD4 T, CD8 T, NKT, Tgd) and layer 2 (CD4 naive, CD8 naive, Tc, Tex, Tr1, nTreg, iTreg, Th1, Th2, Th17, Tfh, Tcm, Tem, MAIT) (Fig. S2). Initially, the percentages of these twenty-four immune cells were individually compared between AD and CTR samples, revealing significant differences in six cell types (Fig. 3A). Specifically, neutrophil cells and cytotoxic T cells were

observed to be enriched in AD, while CD4 T cells, Th1, Th2, and follicular helper T cells (Tfh) were found to be depleted in AD. Moreover, when comparing these immune cells based on the Braak stage of AD patients, no significant associations were identified. This suggests that the variations in immune cells appear to be linked to the AD disease itself rather than being correlated with the Braak stage of AD patients.

Furthermore, we conducted an exploration of Pearson's correlations between the twenty-four immune cells and the aforementioned twenty-seven DEGs. Notably, as illustrated in Fig. 3B, we observed positive correlations between the down-regulated genes and the depleted immune cells (i.e. Th1, Th2, central memory cell), while negative correlations were evident with the enriched immune cells (i.e. neutrophil). This intriguing finding suggests that the differential infiltration of immune cells in the microenvironment may be a consequence of the altered expression patterns of these specific genes.

Construction of classifiers and Survival analysis based on the key DEGs in AD

To assess their potential as discriminative features for AD, the 27 identified key DEGs were employed to construct classification models using three machine learning algorithms—random forest (rf), neural network (nnet), and support vector machines with radial basis function kernel (svmRadial). The AUC for each model was then individually evaluated. Simultaneously, we discovered that fourteen classifiers, each characterized by one of the genes (*KTN1*, *YWHAZ*, *KIFAP3*, *PNMAL1*, *MDH1*, *PITHD1*, *SCG5*, *UQCRC2*, *YWHAH*, *SLC39A10*, *TUBB*, *SNCA*, *GLRB*, and *PTPN3*), consistently achieved an AUC exceeding 0.7 across all three machine

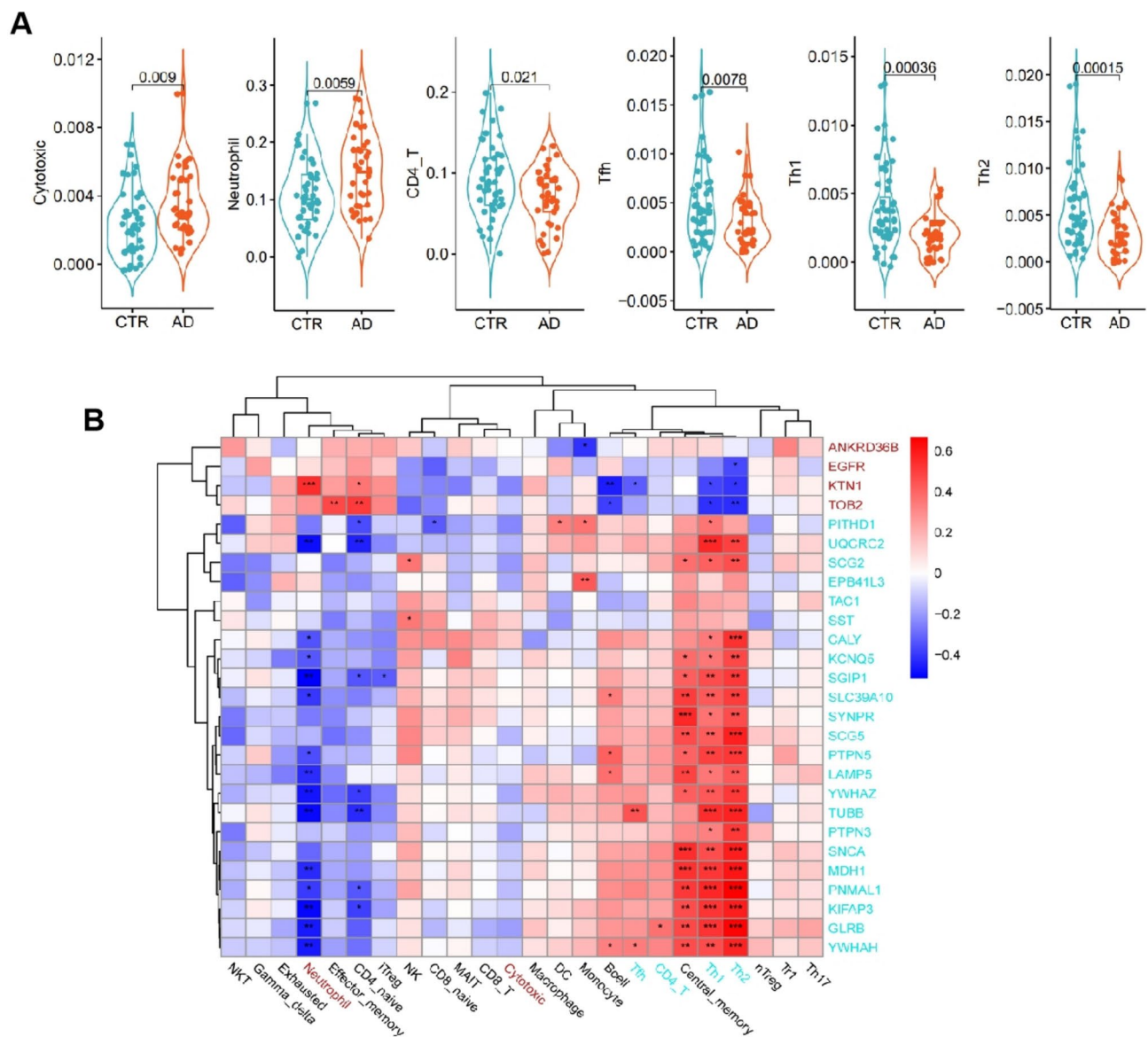


Fig. 3. The proportion of immune cells in the subjects' hippocampus evaluated by immuCellAI. **(A)** Comparison of several immune cells between AD patients and CTR. **(B)** Pearson's correlation between the 24 immune cells and the 27 overlapped DEGs.

learning algorithms. This observation underscores the promising potential of these feature genes in effectively characterizing AD samples from controls (Fig. 4A).

Given that the hippocampal tissues used in this study were sampled within a few hours post-mortem, we proceeded to explore the potential of the identified key DEGs for survival analysis in AD. Specifically, we selected AD samples, utilizing the patients' age as the survival time. Leveraging the fourteen feature genes mentioned earlier, we conducted a multivariable Cox proportional hazards regression analysis and employed a backward stepwise algorithm for model selection based on the Akaike Information Criterion (AIC). Subsequently, an exceptional prognostic model featuring nine genes, with a concordance index of 0.76 and a significance level less than 0.001, was identified (Fig. 4B). Notably, genes *YWHAZ*, *PITHD1*, *SCG5*, *YWHAH*, *TUBB* exhibited a Hazard Ratio (HR) less than 1, indicating that higher expression levels of these genes are associated with a lower risk of AD and a slower disease progression. Conversely, genes *PNMAL1*, *SLC39A10*, *GLRB*, *PTPN3* presented a significant HR greater than 1, signifying that higher expression of these genes is linked to a greater risk of AD and a faster disease progression.

After calculating the risk score for each AD sample using the nine characteristic genes, we observed that the median risk score effectively stratified the AD samples into distinct high- and low-risk groups (Fig. 4C, HR = 2.72, 95% CI 1.94 ~ 3.81, $P = 3.6 \times 10^{-10}$). Notably, the high-risk group exhibited a significantly poorer overall survival age.

Similarly, leveraging the six immune cells (i.e. Cytotoxic, Neutrophil, CD4⁺ T, Tfh, Th1, Th2) that demonstrated significant differences in AD, we constructed a Cox prognostic model. Consistently, based on the median risk score, these AD samples were notably categorized into high- and low-risk groups with statistical significance (Fig. 4D, HR = 2.72, 95% CI 1.37 ~ 5.38, $P = 0.0035$).

Discussion

In this study, we conducted an integrative analysis of three datasets, employing bioinformatics methods to identify DEGs in the hippocampus between AD patients and controls (Figs. 1, 2). A predominant proportion of those DEGs exhibited down-regulation, and these down-regulated genes were notably enriched in processes related to the transmission and transport of synaptic signals, including neurotransmitter secretion and transport, synaptic vesicle cycle, and vesicle-mediated transport in synapses. Significantly, a multitude of studies has underscored the pivotal role of declined neurotransmission function and synaptic degeneration in the initiation and progression of cognitive decline^{33–35}. Moreover, a recent scientific report in 2022 highlighted the interplay between extracellular endocytosis and autophagosome biogenesis at presynaptic sites, influencing activity-dependent synaptic vesicular cycling³⁶. Additionally, research has demonstrated that the loss of neural network connections contributes to synaptic loss³⁴.

Here, a comprehensive analysis revealed a total of 27 DEGs (Fig. 2C,D), comprising 23 down-regulated and 4 up-regulated genes (Fig. 2B), identified not only in AD but also correlated with the Braak stage of AD patients (Fig. 2C). This implies a potential association of these genes with both the onset and progression of AD. Specifically, *KTNI*, *YWHAZ*, *KIFAP3*, *PNMAL1*, *MDH1*, *PITHD1*, *SCG5*, *UQCRC2*, *YWHAH*, *SLC39A10*, *TUBB*, *SNCA*, *GLRB*, and *PTPN3* were individually employed to construct classifiers, achieving an Area Under the AUC exceeding 0.7 (Fig. 4A). This underscores their potential for characterizing AD samples effectively. Furthermore, a nine-gene-feature prognostic model was established, significantly stratifying AD patients (Fig. 4B). The risk score derived from these nine genes demonstrated a capacity to indicate the progression of AD. Notably, *SNCA*, recognized as a causative gene for Parkinson's disease, has also been reported to be associated with AD^{37–39}. *SNCA* is intricately involved in the binding process with A β peptides to facilitate their aggregation⁴⁰, and modulating the activity of BACE1 to regulate APP processing^{37,41}. Dysregulated level of *SNCA* has been linked to cognitive performance⁴², and the effect of *SNCA* protein on memory has been documented⁴³. Moreover, *SST* and *TAC1*, were identified as hub genes through PPI analysis of DEGs in the hippocampus of AD patients⁴⁴, adding another layer to the complexity of AD. Additionally, we observed that *TUBB*, *GLRB*, and *YWHAZ* were enriched in the synaptic organization pathway.

Several prior studies have highlighted the pivotal roles of *YWHAZ*^{45,46} and *TUBB*^{47,48} in the development of AD, with reduced protein expression of *YWHAZ* reported in the hippocampus of AD patients⁴⁶. Previous studies also showed that EGFR is a potential dual molecular target for AD⁴⁹. This strategy can treat AD through EGFR protein degradation⁵⁰. *TAC1* is identified as the hub gene and may be related to synaptic function and inflammation which was also identified as a key gene in the frontal cortex of AD⁵¹. And more abundant *TAC1* was showed in AD-resilient than AD-dementia brain⁵². Excessive SST-14 release accumulates near SST-positive interneurons (SST-INs) in the form of amyloids, which bind to A β to form toxic mixed oligomers. Conversely, chronic stimulation of postsynaptic *SST2/4* on glutamatergic neurons by hyperactive SST-INs promotes intense MAPK p38 activity, leading to somatodendritic p-tau staining and apoptosis/neurodegeneration⁵³. Loss of *LAMP5* interneurons drives neuronal network dysfunction in Alzheimer's disease⁵⁴. Copper metabolism biomarkers *SCG5* altered AD progression⁵⁵. *KCNQ5* was identified as attractive drug targets in neuropsychiatric diseases including AD⁵⁶. *PITHD1* exhibits a strong association with AD. *PITHD1*^{-/-} mice exhibit olfactory bulb (OB) proteome changes related to synaptic transmission, cognition, and memory. OB *PITHD1* expression increases with age in wild-type (WT) mice and decreases in Tg2576 AD mice at late stages⁵⁷.

The infiltration of immune cells and the subsequent neuroinflammatory response within the brain parenchyma and adjacent structures are believed to play a pivotal role in the onset and progression of AD⁵⁸. Recent discoveries, notably the association of immune receptor genes such as *TREM2* and *CD33* with AD, further underscore the significance of immune-related mechanisms^{59,60}. Clinical analyses of pre-AD conditions, including Mild Cognitive Impairment (MCI), provide additional evidence of early and substantial involvement of inflammation in the disease pathogenesis^{61,62}. Consequently, our analysis revealed distinct infiltration patterns of several immune cell types in the hippocampus of AD patients compared to controls, encompassing cytotoxic T

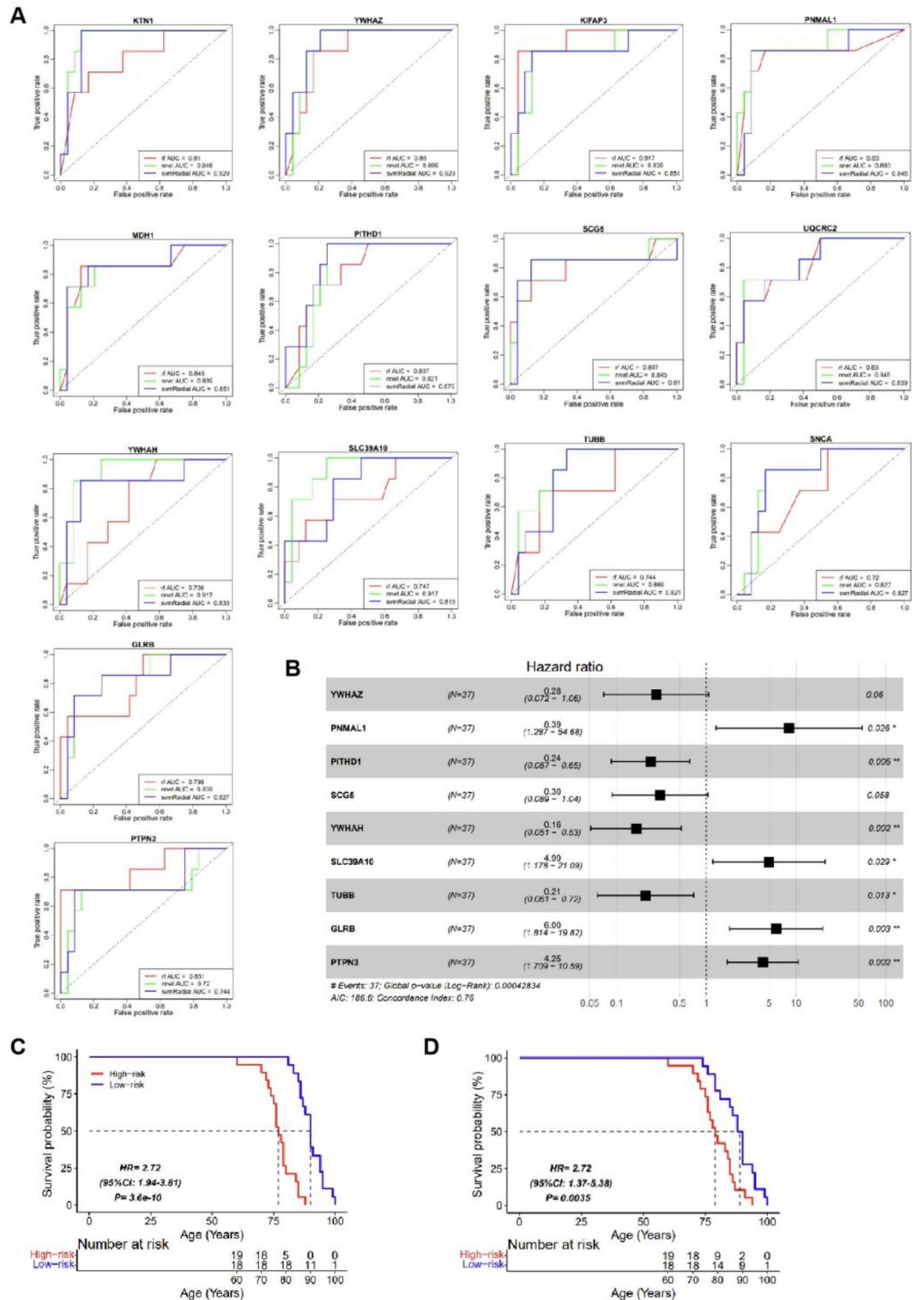


Fig. 4. Classification model and Cox regression model constructed by the characteristic genes. **(A)** Fourteen classifiers featured by genes *KTN1*, *YWHAZ*, *KIFAP3*, *PNMAL1*, *MDH1*, *PITHD1*, *SCG5*, *UQCRC2*, *YWHAH*, *SLC39A10*, *TUBB*, *SNCA*, *GLRB*, and *PTPN3*, respectively. Three machine learning algorithms were used, including random forest (rf), neural network (nnet), and support vector machines with radial basis function kernel (svmRadial). **(B)** Forest plot of the Cox regression model constructed by nine feature genes. **(C)** Kaplan–Meier survival curve fitted by nine feature genes. **(D)** Kaplan–Meier survival curve fitted by six immune cells, including cytotoxic, neutrophil, CD4 T, Tfh, Th1, Th2.

cells, neutrophils, CD4 T cells, Th1, Th2, and Tfh (Fig. 3A). Intriguingly, these immune cells exhibited significant correlations with the aforementioned DEGs, implying a potential essential role of these DEGs in immune infiltration (Fig. 3B). Notably, no differences were observed in these immune cells based on the Braak stage of AD patients. However, this finding warrants further validation in larger cohorts to solidify its significance.

In addition, there are two limitations in this study. The sample size used in this study was relatively small and all of them were downloaded from GEO. It is necessary to obtain samples from real world in a larger sample sizes and test our findings. At the same time, Lab-based experiments is also needed to design and verify the above results in the future.

Conclusion

We explored novel genes linked to the onset and progression of AD, anticipating that they may prove to be effective biomarkers for AD onset. The novel genes identified are linked to the Braak stage in AD patients and hold the potential to effectively characterize AD. They can also significantly stratify AD patients and indicate the progression of AD. Considering that the major drugs for AD exhibit optimal efficacy in the early or intermediate stages, these risk genes associated with early onset have significant implications for guiding clinical medication. This study offers a new perspective that could contribute to enhancing strategies for the prevention and treatment of AD.

Data availability

The datasets used and analysed during the current study available from the corresponding author on reasonable request. All data generated or analysed during this study are included in this published article and its supplementary information files.

Received: 25 April 2024; Accepted: 29 October 2024

Published online: 11 November 2024

References

- Braak, H. & Braak, E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* **82** (4), 239–259. <https://doi.org/10.1007/BF00308809> (1991).
- Masters, C. L. et al. Alzheimer's disease. *Nat. Rev. Dis. Primers* **1**, 15056. <https://doi.org/10.1038/nrdp.2015.56> (2015).
- De Strooper, B. & Karran, E. The cellular phase of Alzheimer's disease. *Cell* **164** (4), 603–615. <https://doi.org/10.1016/j.cell.2015.12.056> (2016).
- Cosacak, M. I., Bhattarai, P. & Kizil, C. Alzheimer's disease, neural stem cells and neurogenesis: cellular phase at single-cell level. *Neural Regen. Res.* **15** (5), 824–827. <https://doi.org/10.4103/1673-5374.268896> (2020).
- Canter, R. G., Penney, J. & Tsai, L. H. The road to restoring neural circuits for the treatment of Alzheimer's disease. *Nature* **539** (7628), 187–196. <https://doi.org/10.1038/nature20412> (2016).
- Alzheimer's disease facts and figures. *Alzheimers Dement.* **17** (3) 327–406. <https://doi.org/10.1002/alz.12328> (2021).
- Zhou, M. et al. Mortality, morbidity, and risk factors in China and its provinces, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet* [https://doi.org/10.1016/S0140-6736\(19\)30427-1](https://doi.org/10.1016/S0140-6736(19)30427-1) (2019).
- Prasher, V. P. et al. Molecular mapping of Alzheimer-type dementia in Down's syndrome. *Ann. Neurol.* **43** (3), 380–383. <https://doi.org/10.1002/ana.410430316> (1998).
- Bagyinszky, E., Youn, Y. C., An, S. S. & Kim, S. The genetics of Alzheimer's disease. *Clin. Interv. Aging* **9**, 535–551. <https://doi.org/10.2147/CLIA.S51571> (2014).
- Cruts, M., Hendriks, L. & Van Broeckhoven, C. The presenilin genes: a new gene family involved in Alzheimer disease pathology. *Hum. Mol. Genet.* https://doi.org/10.1093/hmg/5.supplement_1.1449 (1996).
- Wolfe, M. S. et al. Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and gamma-secretase activity. *Nature* **398** (6727), 513–517. <https://doi.org/10.1038/19077> (1999).
- Durazzo, T. C., Mattsson, N. & Weiner, M. W. Alzheimer's disease neuroimaging initiative. Smoking and increased Alzheimer's disease risk: a review of potential mechanisms. *Alzheimers Dement.* <https://doi.org/10.1016/j.jalz.2014.04.009> (2024).
- Justice, N. J. et al. Posttraumatic stress disorder-like induction elevates β -amyloid levels, which directly activates corticotropin-releasing factor neurons to exacerbate stress responses. *J. Neurosci.* **35** (6), 2612–2623. <https://doi.org/10.1523/JNEUROSCI.3333-14.2015> (2015).
- Vilalta-Franch, J. et al. Depression subtypes and 5-year risk of dementia and Alzheimer disease in patients aged 70 years. *Int. J. Geriatr. Psychiatry* **28** (4), 341–350. <https://doi.org/10.1002/gps.3826> (2013).
- Shi, L. et al. Sleep disturbances increase the risk of dementia: A systematic review and meta-analysis. *Sleep Med. Rev.* **40**, 4–16. <https://doi.org/10.1016/j.smrv.2017.06.010> (2018).
- Martens, Y. A. et al. ApoE cascade hypothesis in the pathogenesis of Alzheimer's disease and related dementias. *Neuron* **110** (8), 1304–1317. <https://doi.org/10.1016/j.neuron.2022.03.004> (2022).
- Raulin, A. C. et al. ApoE in Alzheimer's disease: pathophysiology and therapeutic strategies. *Mol. Neurodegener.* **17** (1), 72. <https://doi.org/10.1186/s13024-022-00574-4> (2022).
- Koutsodendris, N., Nelson, M. R., Rao, A. & Huang, Y. Apolipoprotein E and Alzheimer's disease: Findings, hypotheses, and potential mechanisms. *Annu. Rev. Pathol.* **17**, 73–99. <https://doi.org/10.1146/annurev-pathmechdis-030421-112756> (2022).
- Jin, S. C. et al. TREM2 is associated with increased risk for Alzheimer's disease in African Americans. *Mol. Neurodegener.* **10**, 19. <https://doi.org/10.1186/s13024-015-0016-9> (2015).
- Kim, B. et al. TREM2 risk variants are associated with atypical Alzheimer's disease. *Acta Neuropathol.* **144** (6), 1085–1102. <https://doi.org/10.1007/s00401-022-02495-4> (2022).
- Rodríguez-Esteban, R. & Jiang, X. Differential gene expression in disease: a comparison between high-throughput studies and the literature. *BMC Med. Genom.* **10** (1), 59. <https://doi.org/10.1186/s12920-017-0293-y> (2017).
- Jiang, H. et al. Identification of a selective and direct NLRP3 inhibitor to treat inflammatory disorders. *J. Exp. Med.* **214** (11), 3219–3238. <https://doi.org/10.1084/jem.20171419> (2017).
- Borragero, G., Haylett, W., Seedat, S., Kuivaniemi, H. & Bardien, S. A review of genome-wide transcriptomics studies in Parkinson's disease. *Eur. J. Neurosci.* **47** (1), 1–16. <https://doi.org/10.1111/ejn.13760> (2018).
- Akila Parvathy Dharshini, S., Taguchi, Y. H. & Michael, Gromiha M. Exploring the selective vulnerability in Alzheimer disease using tissue specific variant analysis. *Genomics* **111** (4), 936–949. <https://doi.org/10.1016/j.ygeno.2018.05.024> (2019).
- Ciryam, P., Tartaglia, G. G., Morimoto, R. I., Dobson, C. M. & Vendruscolo, M. Widespread aggregation and neurodegenerative diseases are associated with supersaturated proteins. *Cell Rep.* **5** (3), 781–790. <https://doi.org/10.1016/j.celrep.2013.09.043> (2013).

26. Liang, W. S. et al. Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. *Physiol. Genom.* **33** (2), 240–256. <https://doi.org/10.1152/physiolgenomics.00242.2007> (2008).
27. Li, B. et al. De novo mutation of cancer-related genes associates with particular neurodevelopmental disorders. *J. Mol. Med. (Berlin)* **98** (12), 1701–1712. <https://doi.org/10.1007/s00109-020-01991-y> (2020).
28. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28** (1), 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
29. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28** (11), 1947–1951. <https://doi.org/10.1002/pro.3715> (2019).
30. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51** (D1), D587–D592. <https://doi.org/10.1093/nar/gkac963> (2023).
31. Wu, C. I. et al. APP and DYRK1A regulate axonal and synaptic vesicle protein networks and mediate Alzheimer's pathology in trisomy 21 neurons. *Mol. Psychiatry* **27** (4), 1970–1989. <https://doi.org/10.1038/s41380-022-01454-5> (2022).
32. Zhou, Z. et al. Downregulation of ATP6V1A involved in Alzheimer's disease via synaptic vesicle cycle, phagosome, and oxidative phosphorylation. *Oxid. Med. Cell Longev.* **2021**, 5555634. <https://doi.org/10.1155/2021/5555634> (2021).
33. Tzioras, M., McGeachan, R. I., Durrant, C. S. & Spires-Jones, T. L. Synaptic degeneration in Alzheimer disease. *Nat. Rev. Neurol.* **19** (1), 19–38. <https://doi.org/10.1038/s41582-022-00749-z> (2023).
34. Kashyap, G. et al. Synapse loss and progress of Alzheimer's disease -A network model. *Sci. Rep.* **9** (1), 6555. <https://doi.org/10.1038/s41598-019-43076-y> (2019).
35. Bastrikova, N., Gardner, G. A., Reece, J. M., Jeromin, A. & Dudek, S. M. Synapse elimination accompanies functional plasticity in hippocampal neurons. *Proc. Natl. Acad. Sci. U. S. A.* **105** (8), 3123–3127. <https://doi.org/10.1073/pnas.0800027105> (2008).
36. Yang, S. et al. Presynaptic autophagy is coupled to the synaptic vesicle cycle via ATG-9. *Neuron* **110** (5), 824–840.e10. <https://doi.org/10.1016/j.neuron.2021.12.031> (2022).
37. Uddin, M. S. et al. Autophagy and Alzheimer's disease: From molecular mechanisms to therapeutic implications. *Front. Aging Neurosci.* **10** 04. <https://doi.org/10.3389/fnagi.2018.00004> (2018).
38. Matsubara, M. et al. Genetic association between Alzheimer disease and the alpha-synuclein gene. *Dement. Geriatr. Cogn. Disord.* **12** (2), 106–109. <https://doi.org/10.1159/000051243> (2001).
39. Wang, Q., Tian, Q., Song, X., Liu, Y. & Li, W. SNCA gene polymorphism may contribute to an increased risk of Alzheimer's disease. *J. Clin. Lab. Anal.* **30** (6), 1092–1099. <https://doi.org/10.1002/jcla.21986> (2016).
40. Yoshimoto, M. et al. NACP, the precursor protein of the non-amyloid beta/A4 protein (A beta) component of Alzheimer disease amyloid, binds A beta and stimulates A beta aggregation. *Proc. Natl. Acad. Sci. U. S. A.* **92** (20), 9141–9145. <https://doi.org/10.1073/pnas.92.20.9141> (1995).
41. Roberts, H. L., Schneider, B. L. & Brown, D. R. α -Synuclein increases β -amyloid secretion by promoting β - γ -secretase processing of APP. *PLoS ONE* **12** (2), e0171925. <https://doi.org/10.1371/journal.pone.0171925> (2017).
42. Korff, A., Liu, C., Ghingina, C., Shi, M. & Zhang, J. Alzheimer's Disease neuroimaging Initiative α -Synuclein in cerebrospinal fluid of Alzheimer's disease and mild cognitive impairment. *J. Alzheimers Dis.* <https://doi.org/10.3233/JAD-130458> (2013).
43. Larson, M. E. et al. Soluble α -synuclein is a novel modulator of Alzheimer's disease pathophysiology. *J. Neurosci.* **32** (30), 10253–10266. <https://doi.org/10.1523/JNEUROSCI.0581-12.2012> (2012).
44. Liu, Y. J. et al. Identification of hub genes associated with cognition in the hippocampus of Alzheimer's disease. *Bioengineered* **12** (2), 9598–9609. <https://doi.org/10.1080/21655979.2021.1999549> (2021).
45. Yang, F. et al. Identification of key regulatory genes and pathways in prefrontal cortex of Alzheimer's disease. *Interdiscip. Sci.* **12** (1), 90–98. <https://doi.org/10.1007/s12539-019-00353-8> (2020).
46. Ho Kim, J. et al. Proteome-wide characterization of signalling interactions in the hippocampal CA4/DG subfield of patients with Alzheimer's disease. *Sci. Rep.* **5**, 11138. <https://doi.org/10.1038/srep11138> (2015).
47. Zahid, S., Khan, R., Oellerich, M., Ahmed, N. & Asif, A. R. Differential S-nitrosylation of proteins in Alzheimer's disease. *Neuroscience* **256**, 126–136. <https://doi.org/10.1016/j.neuroscience.2013.10.026> (2014).
48. Yaghoobi, A. & Malekpour, S. A. Unraveling the genetic architecture of blood unfolded p-53 among non-demented elderlies: novel candidate genes for early Alzheimer's disease. *BMC Genom.* **25** (1), 440. <https://doi.org/10.1186/s12864-024-10363-6> (2024).
49. Choi, H. J., Jeong, Y. J., Kim, J. & Hoe, H. S. EGFR is a potential dual molecular target for cancer and Alzheimer's disease. *Front. Pharmacol.* **14**, 1238639. <https://doi.org/10.3389/fphar.2023.1238639> (2023).
50. Kargbo, R. B. Treatment of cancer and Alzheimer's disease by PROTAC degradation of EGFR. *ACS Med. Chem. Lett.* **10** (8), 1098–1099. <https://doi.org/10.1021/acsmchemlett.9b00283> (2019).
51. Zhu, M., Tang, M. & Du, Y. Identification of TAC1 associated with Alzheimer's disease using a robust rank aggregation approach. *J. Alzheimers Dis.* **91** (4), 1339–1349. <https://doi.org/10.3233/JAD-220950> (2023).
52. Morgan, G. R. & Carlyle, B. C. Interrogation of the human cortical peptidome uncovers cell-type specific signatures of cognitive resilience against Alzheimer's disease. *Sci. Rep.* **14** (1), 7161. <https://doi.org/10.1038/s41598-024-57104-z> (2024).
53. Almeida, V. N. Somatostatin and the pathophysiology of Alzheimer's disease. *Ageing Res. Rev.* **96**, 102270. <https://doi.org/10.1016/j.arr.2024.102270> (2024).
54. Deng, Y. et al. Loss of LAMP5 interneurons drives neuronal network dysfunction in Alzheimer's disease. *Acta Neuropathol.* **144** (4), 637–650. <https://doi.org/10.1007/s00401-022-02457-w> (2022).
55. Zhuang, X. et al. SCG5 and MITF may be novel markers of copper metabolism immunorelevance in Alzheimer's disease. *Sci. Rep.* **14** (1), 13619. <https://doi.org/10.1038/s41598-024-64599-z> (2024).
56. Baird, D. A. et al. Identifying drug targets for neurological and psychiatric disease via genetics and the brain transcriptome. *PLoS Genet.* **17** (1), e1009224. <https://doi.org/10.1371/journal.pgen.1009224> (2021).
57. Lachén-Montes, M. et al. Smelling the dark proteome: Functional characterization of PITH domain-containing protein 1 (Clorf128) in olfactory metabolism. *J. Proteome Res.* **19** (12), 4826–4843. <https://doi.org/10.1021/acs.jproteome.0c00452> (2020).
58. Chen, X. et al. Microglia-mediated T cell infiltration drives neurodegeneration in tauopathy. *Nature* **615** (7953), 668–677. <https://doi.org/10.1038/s41586-023-05788-0> (2023).
59. Zhao, L. CD33 in Alzheimer's disease—biology, pathogenesis, and therapeutics: A mini-review. *Gerontology* **65** (4), 323–331. <https://doi.org/10.1159/000492596> (2019).
60. Qin, Q. et al. TREM2, microglia, and Alzheimer's disease. *Mech. Ageing Dev.* **195**, 111438. <https://doi.org/10.1016/j.mad.2021.111438> (2021).
61. Newcombe, E. A. et al. Inflammation: the link between comorbidities, genetics, and Alzheimer's disease. *J. Neuroinflamm.* **15** (1), 276. <https://doi.org/10.1186/s12974-018-1313-3> (2018).
62. Heneka, M. T. et al. Neuroinflammation in Alzheimer's disease. *Lancet Neurol.* **14** (4), 388–405. [https://doi.org/10.1016/S1474-4422\(15\)70016-5](https://doi.org/10.1016/S1474-4422(15)70016-5) (2015).

Acknowledgements

The funder provided support in the form of salaries of all authors.

Author contributions

Kun Song and Shenghui Hu designed the study and revised the manuscript. Jie Li, Lingfang Li, and Shanshan Cai collected and organized the data, Jie Li and Lingfang Li conducted the statistical analysis, and created graphical representation of data. Jie Li and Lingfang Li wrote the initial draft and followed by subsequent revision and refinement.

Funding

This research was supported by the Key Scientific Research Project of Hunan Provincial Department of Education (No.23A0722) and the Doctoral New Investigator Grants of Hunan University of Medicine (No.2020122003).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-78181-0>.

Correspondence and requests for materials should be addressed to K.S. or S.H.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024