Check for updates

Prioritization of risk genes for Alzheimer's disease: an analysis framework using spatial and temporal gene expression data in the human brain based on support vector machine

Shiyu Wang[1†], Xixian Fang[1†], Xiang Wen[2], Congying Yang[1], Ying Yang[1] and Tianxiao Zhang[1,3]*

[1]Department of Epidemiology and Biostatistics, School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an, China, [2]Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Beijing, China, [3]National Anti-Drug Laboratory Shaanxi Regional Center, Xi'an, China

Background: Alzheimer's disease (AD) is a complex disorder, and its risk is influenced by multiple genetic and environmental factors. In this study, an AD risk gene prediction framework based on spatial and temporal features of gene expression data (STGE) was proposed.

Methods: We proposed an AD risk gene prediction framework based on spatial and temporal features of gene expression data. The gene expression data of providers of different tissues and ages were used as model features. Human genes were classified as AD risk or non-risk sets based on information extracted from relevant databases. Support vector machine (SVM) models were constructed to capture the expression patterns of genes believed to contribute to the risk of AD.

Results: The recursive feature elimination (RFE) method was utilized for feature selection. Data for 64 tissue-age features were obtained before feature selection, and this number was reduced to 19 after RFE was performed. The SVM models were built and evaluated using 19 selected and full features. The area under curve (AUC) values for the SVM model based on 19 selected features (0.740 [0.690−0.790]) and full feature sets (0.730 [0.678−0.769]) were very similar. Fifteen genes predicted to be risk genes for AD with a probability greater than 90% were obtained.

Conclusion: The newly proposed framework performed comparably to previous prediction methods based on protein-protein interaction (PPI) network properties. A list of 15 candidate genes for AD risk was also generated to provide data support for further studies on the genetic etiology of AD.

KEYWORDS

Alzheimer's disease, risk gene prioritization, gene expression patterns, machine learning, genome-wide association analyses

1 Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disorder that is characterized by cognitive impairment and memory loss. It affected approximately 50 million people worldwide in 2020, which is expected to increase to 150 million by 2050 (Breijyeh and Karaman, 2020). Advanced age is the most important risk factor for AD (Knopman et al., 2021). A significant increase in the incidence rate of AD was observed in senior citizens after the age of 65 years (Knopman et al., 2021). Equal incidence rates of AD were identified for males and females after adjusting for age, indicating that sex might not be associated with the risk of AD (Knopman et al., 2021). The pathological features of AD include senile plaques formed by the accumulation of β-amyloid protein and neurofibrillary tangles composed of highly phosphorylated τ proteins. Several hypotheses have been proposed to explain the pathogenesis of AD, including oxidative stress (Yang et al., 2022), inflammation (Yang et al., 2022), and DNA damage (Tanaka et al., 2021). However, no consensus has yet been reached.

Previous studies have indicated that AD is a complex disorder, and its risk is attributed to multiple genetic and environmental factors (Carmona et al., 2018; Bertram and Tanzi, 2019). In the last decade, genome-wide association (GWA) analyses have significantly contributed to the genetic etiology of AD (Bertram and Tanzi, 2019). Jansen et al. confirmed 29 risk loci and several relevant pathways related to AD through a GWA meta-analysis (Jansen et al., 2019). In addition, Celeste et al. reviewed the relationship between several AD risk genes, including ABCA7, BIN1, CASS4, and CD33, and the cellular and neuropathological characteristics of AD (Karch et al., 2014). Nevertheless, a recent study indicated that approximately half of the heritability of AD remains unaccounted (Raybould and Sims, 2021). It is probable that a large number of susceptibility loci for AD have not yet been discovered. However, recent studies have indicated that larger-scale GWA studies in the future are less cost effective due to the intrinsic deficiency rooted in the study design of GWA studies; therefore, it might not be a preferable choice for unraveling these hidden genomic regions that contribute to the risk of AD (Escott-Price and Hardy, 2022). In this sense, prioritizing AD risk genes based on evidence gained from different perspectives and then validating these candidate risk genes in subsequent candidate gene-based association studies might be an effective strategy for discovering more relevant genes for AD risk. In a recent study, Cogill et al. applied machine-learning-based methods using brain developmental gene expression data to prioritize high-confidence candidate genes for autism spectrum disorder (Cogill and Wang, 2016). This study established a feasible analysis pipeline for prioritizing candidate risk genes for complex disorders, using spatial and temporal gene expression data.

Multiple lines of evidence have indicated that the expression of AD risk genes has specific spatial and temporal features (Moradifard et al., 2018; Grubman et al., 2019). Extracting and properly synthesizing information from these gene expression features might be an effective way to prioritize the risk genes for AD. In this study, we aimed to construct and evaluate a machine-learning-based model to identify high-confidence risk genes for AD using spatial and temporal gene expression data extracted from a publicly available database.

2 Materials and methods

The statistical analysis pipeline is shown in Figure 1. In this study, we propose an AD risk gene prediction framework based on spatial and temporal features of gene expression data (STGE). In this analysis framework, the gene expression data of providers of different tissues and ages were utilized as model features. Human genes were classified as AD risk or non-risk sets and randomly split into training and validation sets. Support vector machine (SVM) models were constructed to capture the expression patterns of genes that were believed to contribute to the risk of AD in the training set, which were then applied to the validation set to evaluate model performance. The STGE model was then applied to a gene set with an unknown status for AD risk, and a confidence score was assigned to each gene.

2.1 Data extraction

The data used in the present study were extracted from three publicly available databases: the GTEx database (https://gtexportal.org/home/) (GTEx Consortium, 2020), AlzData database (http://www.alzdata.org/) (Xu et al., 2018), and GWAS catalog (https://www.ebi.ac.uk/gwas/) (Buniello et al., 2019).

Spatial and temporal expression data for each gene were obtained from the GTEx database. Gene expression data related to tissues of the human brain (including the cerebellum, cortex, anterior cingulate cortex, hippocampus, substantia nigra, caudate, cerebellar hemisphere, frontal cortex, hypothalamus, nucleus accumbens, putamen, spinal cord, and amygdala) were extracted. Data from tissue sample providers under 20 or over 70 years of age were not included, and all these providers were healthy. In addition, we also removed tissue providers who scored 0 or 4 points on the death classification provided by GTEx database basing on the 4-point Hardy Scale (Hardy et al., 1985), because those scores represent the death of the provider is associated with chronic disease. Specifically, the score of 0 added by GTEx database stands for ventilator case (all cases on a ventilator immediately before death), and the score of 4 stands for slow death case (death after a long illness, with a terminal phase longer than 1 day; deaths that are not unexpected). Finally, gene expression data in 13 types of brain-related tissues for 14,697 genes were extracted from 317 tissue sample providers of various ages and genders (Supplementary Table S1 and Supplementary Figure S1).

AlzData is a database for scoring correlations between human genes and the risk of AD, based on evidence from high-throughput omics data. The CFG scores ranged from 0 to 5, with a higher score indicating a stronger correlation between the gene and AD. Genes with scores of 4–5 were extracted to form the AD risk gene set ("the right answer"). For genes with scores of 0–3, we supplemented the "DISEASE/TRAIT" (we always call it "trait" for short) from the GWAS catalog and excluded genes related to AD to obtain AD non-risk genes. Finally, 3,899 genes comprising 340 AD risk genes and 3,559 non-AD risk genes were identified, and these genes' CFG scores and GWAS traits are shown in Supplementary Table S2.

FIGURE 1
Analysis pipeline of the model construction and evaluations.
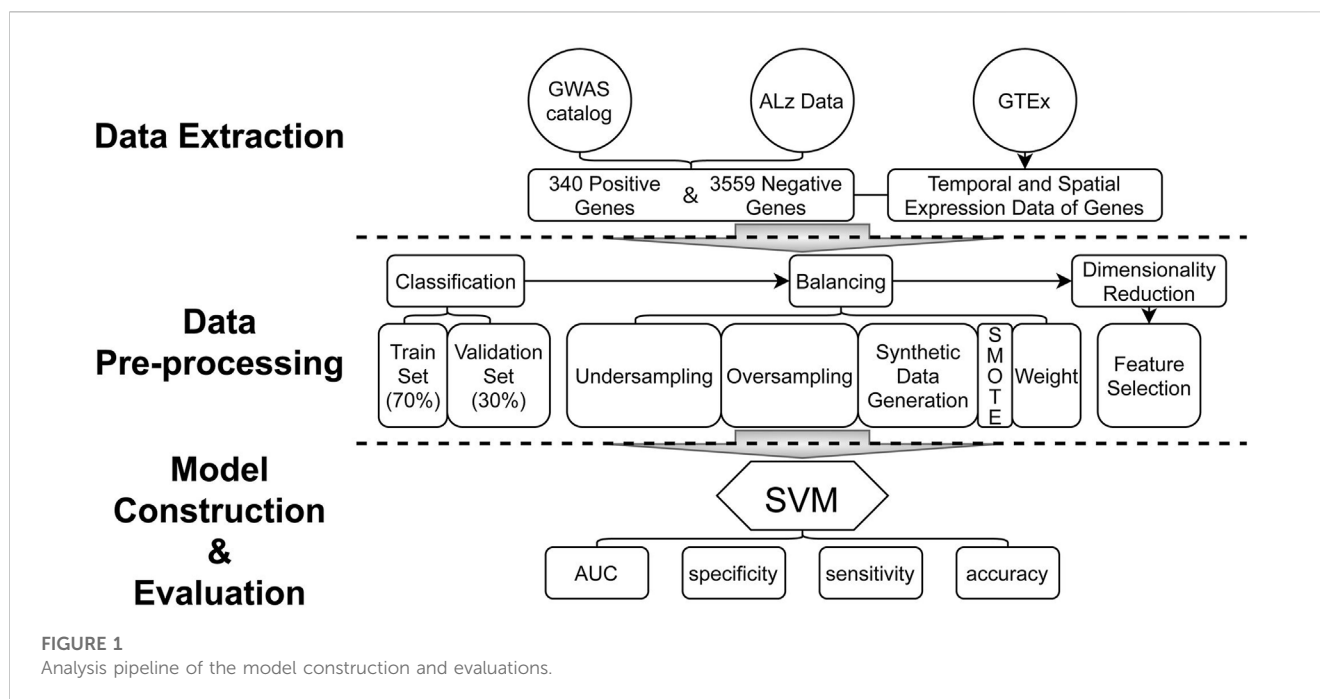
2.2 Model construction and evaluation

The SVM models were constructed based on spatial and temporal gene expression data extracted from relevant databases using the e1071 package of the R project, and both spatial and temporal aspects of the data are contained in the data features, which is going to be used for feature selection. Gene expression data were first grouped by the tissue type and age of the tissue providers. The median expression level of each gene in the tissue type age group was calculated and used as features in the SVM models. A total of 64 brain tissue-related features were obtained for model construction (Supplementary Table S3). The dataset was randomly divided into training and validation sets in a ratio of 7: 3. There were 238 AD risk genes and 2,491 AD non-risk genes in the training set. The SMOTE function in the DMwR package was used to balance gene numbers. Feature selection was conducted using the caret package, and 19 spatial and temporal features were selected based on recursive feature elimination (RFE). Accuracy and Kappa statistics were chosen as the evaluation indicators to estimate the performance of the selected features, and we chose the feature set with both the greatest value and least variance to build the SVM model. Parameter optimization was performed using a grid search strategy. Parameters including model accuracy, specificity, sensitivity, and area under the curve (AUC) were utilized to evaluate the performance of the SVM model. The R packages pROC and ROCR were used to draw the ROC curve and calculate the AUC, respectively. The R package ggplot2 was used for data visualization.

2.3 Results validation

After the genes with high confidence were predicted by SVM model, the normalized expression in AlzData database (http://www.alzdata.org/Normalized_differential1.php) will be used for providing these genes' differential expression data. Besides, KOBAS platform (http://kobas.cbi.pku.edu.cn/) will be used to do gene ontology (GO) and KEGG pathway enrichment analysis in all available databases (including OMIM, KEGG Disease and NHGRI GWAS Catalog).

3 Results

3.1 Feature selection based on recursive feature elimination

RFE was used for feature selection. Data for 64 tissue-age features were obtained before feature selection, and this number was reduced to 19 after RFE was performed. SVM models based on each of these 19 features (the gene expression levels were obtained by median values of samples) were built and evaluated for accuracy, specificity, sensitivity, and AUC (Table 1). The feature with the highest AUC was the human tissue of the brain cerebellum at the age of 40–49 (AUC = 0.688).

3.2 Comparison and validation of SVM models

SVM models were built and evaluated using 19 selected and full features (Table 2 and Figure 2). The AUC values for the SVM model based on 19 selected features (0.74 [0.690–0.790]) and full feature sets (0.730 [0.678–0.769]) were very similar. To evaluate model robustness, we also constructed these models based on the mean expression level of each gene in the tissue type age group. In addition, to examine the potential effects of sex, SVM models were constructed based on the expression data from male and female samples. The results are

TABLE 1 The mean accuracy, sensitivity, specificity and AUC of each model built by each selected feature from the RFE method.

Tissue	Age	Accuracy	Sensitivity	Specificity	AUC
Brain-Cerebellum	40–49	0.604	0.657	0.599	0.688
Brain-Amygdala	50–59	0.768	0.441	0.799	0.684
Brain-Frontal Cortex BA9	50–59	0.657	0.598	0.663	0.683
Brain-Anterior cingulate cortex BA24	30–39	0.67	0.559	0.681	0.683
Brain-Putamen basal ganglia	30–39	0.645	0.569	0.653	0.682
Brain-Anterior cingulate cortex BA24	40–49	0.701	0.529	0.717	0.678
Brain-Cerebellum	60–69	0.551	0.676	0.539	0.674
Brain-Cerebellar Hemisphere	60–69	0.689	0.549	0.702	0.667
Brain-Frontal Cortex BA9	60–69	0.644	0.520	0.656	0.665
Brain-Substantia nigra	40–49	0.715	0.500	0.736	0.664
Brain-Putamen basal ganglia	60–69	0.691	0.520	0.707	0.655
Brain-Caudate basal ganglia	30–39	0.695	0.559	0.708	0.640
Brain-Anterior cingulate cortex BA24	50–59	0.733	0.461	0.759	0.636
Brain-Cerebellum	30–39	0.770	0.461	0.800	0.633
Brain-Substantia nigra	60–69	0.774	0.461	0.803	0.628
Brain-Amygdala	60–69	0.736	0.480	0.760	0.626
Brain-Nucleus accumbens basal ganglia	40–49	0.564	0.598	0.561	0.621
Brain-Nucleus accumbens basal ganglia	60–69	0.561	0.539	0.563	0.606
Brain-Hypothalamus	30–39	0.555	0.52	0.558	0.604

TABLE 2 The average accuracy, sensitivity, specificity and AUC of the two models based on ten-fold cross validation.

	Selected feature set	Full feature set
Accuracy	0.756 ± 0.016	0.754 ± 0.023
Sensitivity	0.588 ± 0.069	0.500 ± 0.054
Specificity	0.772 ± 0.016	0.778 ± 0.021
AUC(95%CI)	0.740 (0.690–0.790)	0.730 (0.678–0.769)

summarized in Supplementary Table S4. There are no significant differences when mean values were utilized compared to median values. The model performance based on males or females was also very similar to that of models constructed using all samples. Finally, we chose the selected feature and median values to construct the SVM model because of its highest AUC. Besides, some known AD risk genes (such as APOE, PICALM and BIN1) were recovered with the final SVM model, and the probabilities of them being classified as AD risk genes are ranged from 0.723–0.783 and shown in Supplementary Table S5.

3.3 Risk genes of AD predicted by the SVM model

Based on the SVM models constructed using tissue-age-specific gene expression data, the risk contributions to AD onset and

development were evaluated for 10,798 genes that were not included in the model construction and evaluations (the external gene set). 15 genes predicted to be risk genes for AD with a probability greater than 90% were obtained (Table 3). Among these genes, GUCY1B3 had the highest confidence score as a risk gene for AD (0.93). To further investigate this gene set, we examined the gene expression patterns of these 15 genes in the human brain and made a heatmap showing in Supplementary Figure S2. In addition, 191 risk genes for AD with a probability greater than 80% are shown in Supplementary Table S6.

3.4 Differential gene expression analysis and pathway/ontology analysis

After the normalized differential gene expression analysis, there exist 8 genes among 15 candidate genes expressing differentially in AD. The differential expression data of these 8 genes are shown in Table 4. The GO and KEGG pathway enrichment analyses find out 15 pathways that are statistically correlated with candidate genes, which are shown in Supplementary Figures S3, S4.

4 Discussion

In the present study, we propose a novel machine-learning-based analysis pipeline using data extracted from the GTEx database
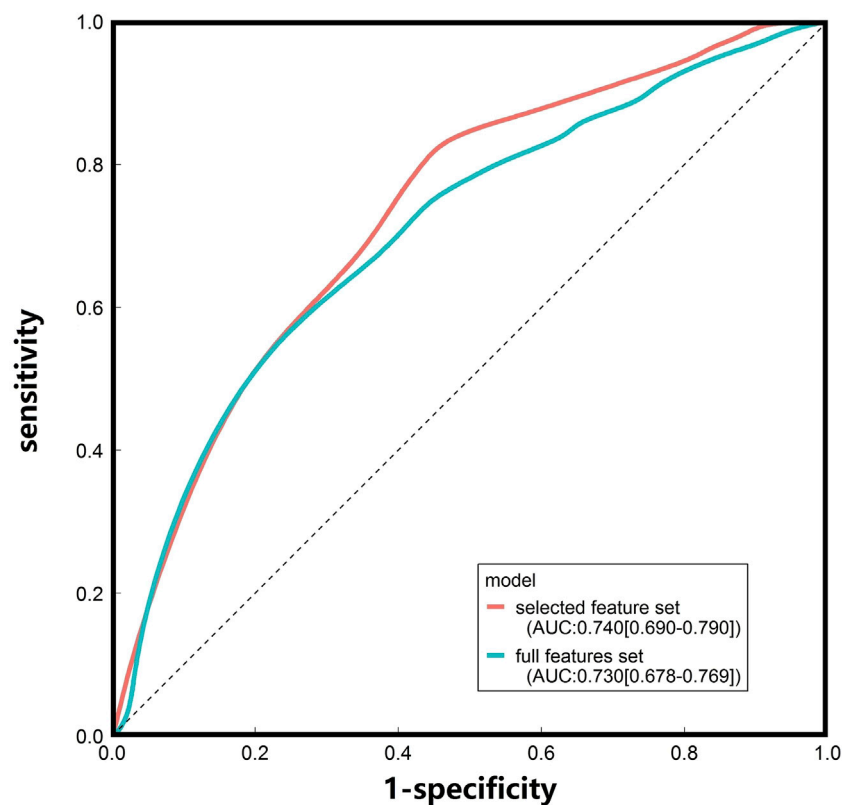
FIGURE 2
ROC curves of the SVM models constructed based on the median gene expression levels in different tissue-age groups.

to prioritize candidate AD risk genes. The performance measured by the AUC of the SVM models was promising, and a list of 15 candidate AD risk genes was presented according to the prediction model. In the last decade, several studies have been published to identify candidate AD risk genes, and most of these studies were based on protein–protein interaction (PPI) networks to identify hub genes using GWA data. The model performance measured by the AUC of these previous studies ranged from 0.63 to 0.84 depending on different settings (Luo et al., 2019; Lagisetty et al., 2022; Wang et al., 2022; Pei et al., 2023). The methods used in these comparative studies and their AUC are shown in the Supplementary Table S7. Unlike these previous studies, the STGE framework was used to predict AD candidate genes based on the spatial and temporal features of AD risk gene expression. The performance of our model (AUC = 0.74) was comparable to that of previous studies. In this sense, the present study proposed and validated an alternative framework for prioritizing risk genes for AD. In the future, an analysis framework integrating information from gene expression features and PPI network properties might be a promising method to further promote the accuracy and effectiveness of prediction models for prioritizing candidate AD risk genes.

Although most patients with AD experience the first symptom in their mid-60s, previous studies have indicated that changes in the molecular levels occur at a much earlier stage (Egan et al., 2019; Vermunt et al., 2019). A previously published family-based longitudinal study has shown that familial AD may have a long prodromal phase of several years (Chiotis et al., 2018). A recent cohort study also indicated that plasma phospho-tau181 levels were much higher from 16 years prior to the onset of AD symptoms in AD patients with specific DNA mutations (Wang et al., 2021; Karikari et al., 2022). The results of the current study offer new evidence at the gene expression level for prodromal changes in AD patients. Although AD is a late-onset disorder, more than half of the selected features were obtained from sample providers before the age of 60 years. Five of the 19 features, including tissues of the anterior cingulate cortex, putamen basal ganglia, caudate basal ganglia, cerebellum, and hypothalamus, were obtained from providers who are 30–39 years old. In accordance with multiple lines of previous evidence, these findings indicate that molecular-level changes might be identified several years before early symptoms appear in patients with AD. Nevertheless, since a couple of the AD risk genes used in this study were extracted from studies focusing on early-onset AD, we need to be cautious in interpreting these results. Future research using longitudinal data might provide more clues for identifying prodromal biomarkers for AD and, in turn, shed light on early screening and prevention of this complex neurodegenerative disorder.

Among the 15 candidate genes identified through STGE, a few are of particular interest. Sine oculis homeobox homolog 3 (SIX3) encodes a type of transcription factor belonging to the sine oculis homeobox transcription factor family (Steinmetz et al., 2010). Multiple lines of evidence based on animal models have linked this locus to brain development (Steinmetz et al., 2010; Schacht et al.,

TABLE 3 Genes predicted by the SVM model with their confidence score, location, length (bp) and biotype.

Genes	Confidence	Location	Length (bp)	Type
SIX3	0.911	2p21	4,370	protein coding
EFEMP1	0.904	2p16.1	58,197	protein coding
GUCY1B3	0.939	4p32.1	48,820	protein coding
MTPN	0.921	7q33	50,600	protein coding
ACTR3B	0.904	7q36.2	311,231	protein coding
BAG3	0.932	10q26.11	26,440	protein coding
INPP5A	0.924	10q26.3	245,697	protein coding
LRRC10B	0.917	11q12.2	2,270	protein coding
ELMOD1	0.914	11q22.3	75,771	protein coding
DRD2	0.900	11q23.2	66,087	protein coding
GABRA5	0.922	15q12	82,490	protein coding
PITPNM3	0.916	17p13.2-p13.1	105,293	protein coding
SEZ6	0.909	17q11.2	51,540	protein coding
ICAM5	0.922	19p13.2	7,428	protein coding
CSDC2	0.916	22q13.2	16,732	protein coding

TABLE 4 Differential expression results of candidate genes with FDRs < 0.05.

Genes	Brain region	log2 FoldChange	p-value	FDR
EFEMP1	Hippocampus	0.52	0.001	0.020
GUCY1B3	Hippocampus	−0.24	0.004	0.048
	Temporal Cortex	−0.73	0.000	0.000
ACTR3B	Temporal Cortex	−0.8	0.001	0.006
BAG3	Temporal Cortex	0.81	0.000	0.001
	Frontal Cortex	0.42	0.001	0.012
INPP5A	Temporal Cortex	−0.47	0.000	0.001
GABRA5	Hippocampus	−0.57	0.003	0.036
	Temporal Cortex	−1.37	0.000	0.000
SEZ6	Temporal Cortex	−1.37	0.000	0.000
ICAM5	Entorhinal Cortex	−0.61	0.001	0.015
	Frontal Cortex	−0.41	0.000	0.000

2020). A recent GWA study associated genetic polymorphisms of SIX3 with math ability, and its weakening was considered a sign of the progression of AD patients (Lee et al., 2018). Actin-related protein 3B (ACTR3B) encodes a member of the actin-related protein (ARP) family, which might regulate and induce cell shape changes and motility (Hu et al., 2018). Several previous studies have linked ACTR3B to brain aging progression, although no direct GWA study has validated the connection between genetic polymorphisms of these loci and AD (Hu et al., 2018; Seefelder and Kochanek, 2021). In addition, multiple animal models and population-based evidence have been published for dopamine receptor D2 (DRD2) and gamma-aminobutyric acid type A receptor subunit alpha 5 (GABRA5) being associated with brain-related disorders and traits, including schizophrenia, bipolar disorder, Parkinson's disorder, and neurotransmission (Prisciandaro et al., 2017; Escamilla et al., 2018; Mundorf et al., 2021; Zhang et al., 2021). In a recent study, Blum et al. concluded that the DRD2 Taq1A A1 allele might increase the risk of Alzheimer's aging in African Americans by integrating and reviewing previously published data (Blum et al., 2018). Additionally, the genes BAG Cochaperone 3 (BAG3), inositol polyphosphate-5-phosphatase A (INPP5A), seizure related 6 homolog (SEZ6), and intercellular adhesion molecule 5 (ICAM5) are involved in the progression of AD has been proposed in several functional studies using animal models (Hoarau et al., 2011; Paetau et al., 2017; Zhu et al., 2018; Zhou et al., 2020; Zhu et al., 2021). Within these genes, through proteomic study, BAG3 may affect AD by influencing the interpretation of Aβ and tau protein, and patients with AD have much lower levels of SEZ6 in their cerebrospinal fluid than those without dementia (Khoonsari et al., 2016; Gonzalez-Rodriguez et al., 2021). Further in vivo and in vitro studies are needed to validate the functional connections between the risk of AD and the genes on the predicted list.

Three of the 15 pathways identified by GO and KEGG pathway enrichment analyses are worthy of attention, including regulation of synapse structural plasticity, branching morphogenesis of a nerve and forced vital capacity. According to a review, synapse structural plasticity is related to the number of spines, and post-mortem reports of Alzheimer's brains showed reduced spine number in the hippocampus and cortex (Chidambaram et al., 2019). One research studying novel compounds' effect on neuronal branching morphogenesis of PC12 cells indicates that branching morphogenesis is one of the entry points for research to promote recovery of nerve regeneration following neurodegenerative diseases, like AD (Katebi et al., 2019). A prospective cohort study of 431,834 individuals shows that per unit decrease in lung function measure was each associated with increased risk for all-cause dementia (including AD). As for forced vital capacity, its hazard ratio (HR) is 1.16 and p-value is 2.04×10^{-5} (Ma et al., 2023).

The current study has several limitations. First, there is still much space for the promotion of STGE, although the performance of STGE is comparable to that of previous models based on PPI network properties. In addition, as bioinformatics data mining is based on publicly available databases, the completeness of the current work might be limited owing to data availability. The gene expression data in the brain substantia nigra in the age group of 30–39 years were unavailable from the database; therefore, this feature was not included in the model

construction and evaluation. Besides, although the data for training the model contains non-coding RNA, which have been shown to play an important role in the pathogenesis of complex disorders (Goyal et al., 2018), all candidate AD risk genes are protein-coding genes in the current study. Furthermore, the data we used in our research can only correlate to tissues, so we were unable to associate these genes with specific brain cell types.

In summary, in the present study, an efficient analysis framework based on spatial and temporal features of gene expression was proposed to prioritize AD risk genes. The newly proposed framework performed comparably to previous prediction methods based on PPI network properties. A list of 15 candidate genes for AD risk was also generated to provide data support for further studies on the genetic etiology of AD.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.gtexportal.org/home/datasets; https://www.ebi.ac.uk/gwas/docs/file-downloads; http://www.alzdata.org/CFG_rank1.php.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

TZ, SW, and XF designed the study. TZ, SW, and XF wrote the main manuscript text. SW, XF and XW conducted the statistical analysis. SW, XF, CY, and YY prepared all the tables, figures and

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1190863/full#supplementary-material

References

Bertram, L., and Tanzi, R. E. (2019). Alzheimer disease risk genes: 29 and counting. *Nat. Rev. Neurol.* 15 (4), 191–192. doi:10.1038/s41582-019-0158-4

Blum, K., Badgaiyan, R. D., Dunston, G. M., Baron, D., Modestino, E. J., McLaughlin, T., et al. (2018). The DRD2 Taq1A A1 allele may magnify the risk of Alzheimer's in aging african-Americans. *Mol. Neurobiol.* 55 (7), 5526–5536. doi:10.1007/s12035-017-0758-1

Breijyeh, Z., and Karaman, R. (2020). Comprehensive review on Alzheimer's disease: causes and treatment. *Molecules* 25 (24), 5789. doi:10.3390/molecules25245789

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47 (D1), D1005–D1012. doi:10.1093/nar/gky1120

Carmona, S., Hardy, J., and Guerreiro, R. (2018). The genetic landscape of Alzheimer disease. *Handb. Clin. Neurol.* 148, 395–408. doi:10.1016/B978-0-444-64076-5.00026-0

Chidambaram, S. B., Rathipriya, A. G., Bolla, S. R., Bhat, A., Ray, B., Mahalakshmi, A. M., et al. (2019). Dendritic spines: revisiting the physiological role. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 92, 161–193. doi:10.1016/j.pnpbp.2019.01.005

Chiotis, K., Saint-Aubert, L., Rodriguez-Vieitez, E., Leuzy, A., Almkvist, O., Savitcheva, I., et al. (2018). Longitudinal changes of tau PET imaging in relation to hypometabolism in prodromal and Alzheimer's disease dementia. *Mol. Psychiatry* 23 (7), 1666–1673. doi:10.1038/mp.2017.108

Cogill, S., and Wang, L. (2016). Support vector machine model of developmental brain gene expression data for prioritization of Autism risk gene candidates. *Bioinformatics* 32 (23), 3611–3618. doi:10.1093/bioinformatics/btw498

Egan, M. F., Kost, J., Voss, T., Mukai, Y., Aisen, P. S., Cummings, J. L., et al. (2019). Randomized trial of verubecestat for prodromal Alzheimer's disease. *N. Engl. J. Med.* 380 (15), 1408–1420. doi:10.1056/NEJMoa1812840

Escamilla, R., Camarena, B., Saracco-Alvarez, R., Fresán, A., Hernández, S., and Aguilar-García, A. (2018). Association study between COMT, DRD2, and DRD3 gene variants and antipsychotic treatment response in Mexican patients with schizophrenia. *Neuropsychiatr. Dis. Treat.* 14, 2981–2987. doi:10.2147/NDT.S176455

Escott-Price, V., and Hardy, J. (2022). Genome-wide association studies for Alzheimer's disease: bigger is not always better. *Brain Commun.* 4 (3), fcac125. doi:10.1093/braincomms/fcac125

Gonzalez-Rodriguez, M., Villar-Conde, S., Astillero-Lopez, V., Villanueva-Anguita, P., Ubeda-Banon, I., Flores-Cuadrado, A., et al. (2021). Neurodegeneration and

astrogliosis in the human CA1 hippocampal subfield are related to hsp90ab1 and bag3 in Alzheimer's disease. *Int. J. Mol. Sci.* 23 (1), 165. doi:10.3390/ijms23010165

Goyal, N., Kesharwani, D., and Datta, M. (2018). Lnc-ing non-coding RNAs with metabolism and diabetes: roles of lncRNAs. *Cell Mol. Life Sci.* 75 (10), 1827–1837. doi:10.1007/s00018-018-2760-9

Grubman, A., Chew, G., Ouyang, J. F., Sun, G., Choo, X. Y., McLean, C., et al. (2019). A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci.* 22 (12), 2087–2097. doi:10.1038/s41593-019-0539-4

GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369 (6509), 1318–1330. doi:10.1126/science.aaz1776

Hardy, J. A., Wester, P., Winblad, B., Gezelius, C., Bring, G., and Eriksson, A. (1985). The patients dying after long terminal phase have acidotic brains; implications for biochemical measurements on autopsy tissue. *J. Neural Transm.* 61 (3-4), 253–264. doi:10.1007/BF01251916

Hoarau, J. J., Krejbich-Trotot, P., Jaffar-Bandjee, M. C., Das, T., Thon-Hon, G. V., Kumar, S., et al. (2011). Activation and control of CNS innate immune responses in health and diseases: a balancing act finely tuned by neuroimmune regulators (NIReg). *CNS Neurol. Disord. Drug Targets* 10 (1), 25–43. doi:10.2174/187152711794488601

Hu, Y., Pan, J., Xin, Y., Mi, X., Wang, J., Gao, Q., et al. (2018). Gene expression analysis reveals novel gene signatures between Young and old adults in human prefrontal cortex. *Front. Aging Neurosci.* 10, 259. doi:10.3389/fnagi.2018.00259

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* 51 (3), 404–413. doi:10.1038/s41588-018-0311-9

Karch, C. M., Cruchaga, C., and Goate, A. M. (2014). Alzheimer's disease genetics: from the bench to the clinic. *Neuron* 83 (1), 11–26. doi:10.1016/j.neuron.2014.05.041

Karikari, T. K., Ashton, N. J., Brinkmalm, G., Brum, W. S., Benedet, A. L., Montoliu-Gaya, L., et al. (2022). Blood phospho-tau in alzheimer disease: analysis, interpretation, and clinical utility. *Nat. Rev. Neurol.* 18 (7), 400–418. Epub ahead of print. doi:10.1038/s41582-022-00665-2

Katebi, S., Esmaeili, A., Ghaedi, K., and Zarrabi, A. (2019). Superparamagnetic iron oxide nanoparticles combined with NGF and quercetin promote neuronal branching morphogenesis of PC12 cells. *Int. J. Nanomedicine* 14, 2157–2169. doi:10.2147/IJN.S191878

Khoonsari, P. E., Häggmark, A., Lönnberg, M., Mikus, M., Kilander, L., Lannfelt, L., et al. (2016). Analysis of the cerebrospinal fluid proteome in Alzheimer's disease. *PLoS One* 11 (3), e0150672. doi:10.1371/journal.pone.0150672

Knopman, D. S., Amieva, H., Petersen, R. C., Chételat, G., Holtzman, D. M., Hyman, B. T., et al. (2021). Alzheimer disease. *Nat. Rev. Dis. Prim.* 7 (1), 33. doi:10.1038/s41572-021-00269-y

Lagisetty, Y., Bourquard, T., Al-Ramahi, I., Mangleburg, C. G., Mota, S., Soleimani, S., et al. (2022). Identification of risk genes for Alzheimer's disease by gene embedding. *Cell Genom* 2 (9), 100162. doi:10.1016/j.xgen.2022.100162

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* 50 (8), 1112–1121. doi:10.1038/s41588-018-0147-3

Luo, P., Tian, L. P., Ruan, J., and Wu, F. X. (2019). Disease gene prediction by integrating PPI networks, clinical RNA-seq data and OMIM data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 16 (1), 222–232. doi:10.1109/TCBB.2017.2770120

Ma, Y. H., Shen, L. X., Li, Y. Z., Leng, Y., Yang, L., Chen, S. D., et al. (2023). Lung function and risk of incident dementia: A prospective cohort study of 431,834 individuals. *Brain Behav. Immun.* 109, 321–330. doi:10.1016/j.bbi.2023.02.009

Moradifard, S., Hoseinbeyki, M., Ganji, S. M., and Minuchehr, Z. (2018). Analysis of microRNA and gene expression profiles in Alzheimer's disease: A meta-analysis approach. *Sci. Rep.* 8 (1), 4767. doi:10.1038/s41598-018-20959-0

Mundorf, A., Kubitza, N., Hünten, K., Matsui, H., Juckel, G., Ocklenburg, S., et al. (2021). Maternal immune activation leads to atypical turning asymmetry and reduced

DRD2 mRNA expression in a rat model of schizophrenia. *Behav. Brain Res.* 414, 113504. doi:10.1016/j.bbr.2021.113504

Paetau, S., Rolova, T., Ning, L., and Gahmberg, C. G. (2017). Neuronal ICAM-5 inhibits microglia adhesion and phagocytosis and promotes an anti-inflammatory response in LPS stimulated microglia. *Front. Mol. Neurosci.* 10, 431. doi:10.3389/fnmol.2017.00431

Pei, Y., Chen, S., Zhou, F., Xie, T., and Cao, H. (2023). Construction and evaluation of Alzheimer's disease diagnostic prediction model based on genes involved in mitophagy. *Front. Aging Neurosci.* 15, 1146660. doi:10.3389/fnagi.2023.1146660

Prisciandaro, J. J., Tolliver, B. K., Prescot, A. P., Brenner, H. M., Renshaw, P. F., Brown, T. R., et al. (2017). Unique prefrontal GABA and glutamate disturbances in co-occurring bipolar disorder and alcohol dependence. *Transl. Psychiatry* 7 (7), e1163. doi:10.1038/tp.2017.141

Raybould, R., and Sims, R. (2021). Searching the dark genome for Alzheimer's disease risk variants. *Brain Sci.* 11 (3), 332. doi:10.3390/brainsci11030332

Schacht, M. I., Schomburg, C., and Bucher, G. (2020). six3 acts upstream of foxQ2 in labrum and neural development in the spider Parasteatoda tepidariorum. *Dev. Genes Evol.* 230 (2), 95–104. doi:10.1007/s00427-020-00654-9

Seefelder, M., and Kochanek, S. (2021). A meta-analysis of transcriptomic profiles of Huntington's disease patients. *PLoS One* 16 (6), e0253037. doi:10.1371/journal.pone.0253037

Steinmetz, P. R., Urbach, R., Posnien, N., Eriksson, J., Kostyuchenko, R. P., Brena, C., et al. (2010). Six3 demarcates the anterior-most developing brain region in bilaterian animals. *Evodevo* 1 (1), 14. doi:10.1186/2041-9139-1-14

Tanaka, H., Kondo, K., Fujita, K., Homma, H., Tagawa, K., Jin, X., et al. (2021). HMGB1 signaling phosphorylates Ku70 and impairs DNA damage repair in Alzheimer's disease pathology. *Commun. Biol.* 4 (1), 1175. doi:10.1038/s42003-021-02671-4

Vermunt, L., Sikkes, S. A. M., van den Hout, A., Handels, R., Bos, I., van der Flier, W. M., et al. (2019). Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype. *Alzheimers Dement.* 15 (7), 888–898. doi:10.1016/j.jalz.2019.04.001

Wang, Y., Chen, G., and Shao, W. (2022). Identification of ferroptosis-related genes in Alzheimer's disease based on bioinformatic analysis. *Front. Neurosci.* 16, 823741. doi:10.3389/fnins.2022.823741

Wang, Y. L., Chen, J., Du, Z. L., Weng, H., Zhang, Y., Li, R., et al. (2021). Plasma p-tau181 level predicts neurodegeneration and progression to Alzheimer's dementia: A longitudinal study. *Front. Neurol.* 12, 695696. doi:10.3389/fneur.2021.695696

Xu, M., Zhang, D. F., Luo, R., Wu, Y., Zhou, H., Kong, L. L., et al. (2018). A systematic integrated analysis of brain expression profiles reveals YAP1 and other prioritized hub genes as important upstream regulators in Alzheimer's disease. *Alzheimers Dement.* 14 (2), 215–229. doi:10.1016/j.jalz.2017.08.012

Yang, Y., Wang, L., Zhang, C., Guo, Y., Li, J., Wu, C., et al. (2022). Ginsenoside Rg1 improves Alzheimer's disease by regulating oxidative stress, apoptosis, and neuroinflammation through Wnt/GSK-3β/β-catenin signaling pathway. *Chem. Biol. Drug Des.* 99 (6), 884–896. doi:10.1111/cbdd.14041

Zhang, W., Xiong, B. R., Zhang, L. Q., Huang, X., Yuan, X., Tian, Y. K., et al. (2021). The role of the GABAergic system in diseases of the central nervous system. *Neuroscience* 470, 88–99. doi:10.1016/j.neuroscience.2021.06.037

Zhou, J., Chow, H. M., Liu, Y., Wu, D., Shi, M., Li, J., et al. (2020). Cyclin-dependent kinase 5-dependent BAG3 degradation modulates synaptic protein turnover. *Biol. Psychiatry* 87 (8), 756–769. doi:10.1016/j.biopsych.2019.11.013

Zhu, J. W., Jia, W. Q., Zhou, H., Li, Y. F., Zou, M. M., Wang, Z. T., et al. (2021). Deficiency of TRIM32 impairs motor function and purkinje cells in mid-aged mice. *Front. Aging Neurosci.* 13, 697494. doi:10.3389/fnagi.2021.697494

Zhu, K., Xiang, X., Filser, S., Marinković, P., Dorostkar, M. M., Crux, S., et al. (2018). Beta-site amyloid precursor protein cleaving enzyme 1 inhibition impairs synaptic plasticity via seizure protein 6. *Biol. Psychiatry* 83 (5), 428–437. doi:10.1016/j.biopsych.2016.12.023