



Screening for Core Genes Related to Pathogenesis of Alzheimer's Disease

Longxiu Yang¹, Yuan Qin¹ and Chongdong Jian^{2*}

¹ Department of Neurology, The First Affiliated Hospital of Guangxi Medical University, Nanning, China, ² Department of Neurology, The Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China

OPEN ACCESS

Edited by:

Tao Huang,
Shanghai Institute of Nutrition
and Health (CAS), China

Reviewed by:

Xiao Dan Shi,
The Fourth Military Medical University,
China

Wei Wei,
Jinan University, China

*Correspondence:

Chongdong Jian
ddchongjian@163.com

Specialty section:

This article was submitted to
Epigenomics and Epigenetics,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 17 February 2021

Accepted: 01 April 2021

Published: 22 April 2021

Citation:

Yang L, Qin Y and Jian C (2021)
Screening for Core Genes Related
to Pathogenesis of Alzheimer's
Disease.
Front. Cell Dev. Biol. 9:668738.
doi: 10.3389/fcell.2021.668738

Alzheimer's disease (AD), a nervous system disease, lacks effective therapies at present. RNA expression is the basic way to regulate life activities, and identifying related characteristics in AD patients may aid the exploration of AD pathogenesis and treatment. This study developed a classifier that could accurately classify AD patients and healthy people, and then obtained 3 core genes that may be related to the pathogenesis of AD. To this end, RNA expression data of the middle temporal gyrus of AD patients were firstly downloaded from GEO database, and the data were then normalized using *limma* package following a supplementation of missing data by k-Nearest Neighbor (KNN) algorithm. Afterwards, the top 500 genes of the most feature importance were obtained through Max-Relevance and Min-Redundancy (mRMR) analysis, and based on these genes, a series of AD classifiers were constructed through Support Vector Machine (SVM), Random Forest (RF), and KNN algorithms. Then, the KNN classifier with the highest Matthews correlation coefficient (MCC) value composed of 14 genes in incremental feature selection (IFS) analysis was identified as the best AD classifier. As analyzed, the 14 genes played a pivotal role in determination of AD and may be core genes associated with the pathogenesis of AD. Finally, protein-protein interaction (PPI) network and Random Walk with Restart (RWR) analysis were applied to obtain core gene-associated genes, and key pathways related to AD were further analyzed. Overall, this study contributed to a deeper understanding of AD pathogenesis and provided theoretical guidance for related research and experiments.

Keywords: Alzheimer's disease, IFS, feature selection, KNN model, protein-protein interaction network

INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that is almost incurable. According to the World Alzheimer Report 2018, there were approximately 50 million patients worldwide who suffered from AD, and AD became a major cause of death among old people (Patterson, 2018). Its main features are the deposit of β -amyloid (A β) plaques and neurofibrillary agglomerates (Dos Santos Picanco et al., 2018). A recent genetic study unearthed that A β deposition frequently occurs in people with ApoE4 (Genin et al., 2011). People with ApoE4 gene have high plasma cholesterol, which in turn stimulates the deposition of A β and tau proteins in the brain, thereby leading to AD (Greenberg et al., 2020). It is reported that the pathogenesis of AD is associated with heredity and gene expression like TREM2, PLCG2, ABI3 (Sims et al., 2017;

Ulland and Colonna, 2018). Besides, circRNA and miRNA are also found to be related to the pathogenesis of AD (Dube et al., 2019; Iranifar et al., 2019).

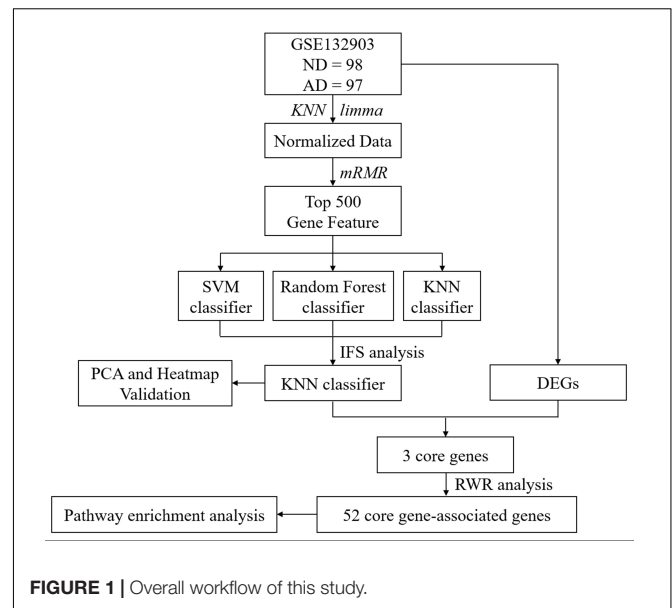
Gene expression regulation is the most critical way of life regulating. Aberrant gene expression in brain tissue accounts for diverse diseases. For instance, ROCK1 gene expression is relevant to AD progression (Li X. et al., 2020). While aberrant expression of genes such as LRRK induces the occurrence of Parkinson's disease (Wang et al., 2017). Researchers disclosed that gene expression in the MTG is probably closely related to the pathogenesis of AD, and the blockage of GABA signaling pathway in the MTG may result in cognitive decline (Govindpani et al., 2020). Given the above studies, this study surmised that the gene expression of the MTG is closely related to the pathogenesis of AD. This study attempted to probe into the critical factors affecting the pathogenesis of AD by analyzing the gene expression related to the MTG of AD sufferers.

Machine learning is a pivotal means of modern medical research, by which researchers always explore core genes that affect the occurrence of diseases. In the field of bioinformatics, machine learning is mainly applied in construction of diagnostic or prognostic models for disease, screening for biomarkers indicating disease outcome, etc., while linear-regression analysis is the common one in prognostic model establishment. For example, Feng and Jin (2018) constructed a risk model for prognostic prediction of patients with breast cancer through bioinformatics methods. Additionally, in algorithms that help for model construction, Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN), and K-Nearest Neighbor (KNN) are frequently used (Zhang, 2016; Huang et al., 2018; Chowdhury et al., 2019; Kulkarni et al., 2021). A previous study combined the KNN with genetic algorithm to greatly improve the accuracy of heart disease diagnosis (Jabbar et al., 2013). Max-Relevance and Min-Redundancy (mRMR) is an effective analytical method used to identify core genes in diseases. For example, Xu et al. (2014) identified core genes with mRMR to establish a model for determining malignant thyroid epithelioma. Regarding the identification of feature genes, the genes that are screened out to establish prognostic models or classifiers for disease are recognized potential biomarkers for outcome prediction. Besides, methods like mRMR, Boruta and ReliefF are also practicable (Zhang et al., 2008; Degenhardt et al., 2019). The mRMR method is instrumental for discovering core genes that affect Guillain-Barré syndrome (GBS) (Xu et al., 2016). This study selected the best AD classifier among SVM, RF, and KNN classifiers following the mRMR analysis and incremental feature selection (IFS) algorithm. Afterwards, functions of related genes in the optimal classifier were further explored. These findings may provide a deeper insight into the research and treatment of AD.

ANALYTICAL METHODS

Dataset Preparing

As presented in **Figure 1**, the overall workflow of this study was drawn to clarify our research design. RNA expression



data (GSE132903) of the AD MTG were downloaded from the Gene Expression Omnibus (GEO) database¹. The samples in this dataset were collected from Brain and Body Donation Program (BBDP) volunteers, including 98 healthy subjects (ND) and 97 AD patients. The corresponding platform annotation file was downloaded to annotate the RNA expression dataset, and an expression matrix with gene ids was created with the probe annotation categories. Afterwards, the missing data were supplemented with KNN ($K = 10$) (Trojanskaya et al., 2001; de Brevern et al., 2004), and the final data were standardized using the R package *limma* for further analysis.

Feature Selection by mRMR

The mRMR algorithm was implemented to rank feature importance in standardized cohort as previously reported (Xu et al., 2014, 2016; He et al., 2019). The mRMR feature selection method can compute relevance between features and sample phenotypes, and can comprehensively rank features according to the redundancy between features. Features at the top have a better trade-off between the relevance and redundancy than features at the bottom. Here, feature genes with the maximum relevance with AD and the minimum redundancy with other features were found through the mRMR algorithm (Peng et al., 2005). Mutual information (MI) represents the relevance between a random variable and another random variable. MI function was applied to assess the relevance between features and quantify the relevance and redundancy. The MI function was defined as follows (1),

$$I(x, y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (1)$$

where x and y represent two vectors, $p(x, y)$ represents joint probabilistic density, $p(x)$ and $p(y)$ represent marginal probabilistic densities. Thereafter, the relevance between genes

¹<https://www.ncbi.nlm.nih.gov/geo/>

and sample phenotypes were computed following the MI function (2),

$$D = I(f, c) \quad (2)$$

where D represents the relevance between genes and phenotypes, f represents gene, and c represents phenotype. The redundancy between genes was identified as R and was computed by the following formula (3),

$$R = \frac{1}{m} \sum_{f_i \in T} I(f, f_i) \quad (3)$$

where m represents the total number of genes in the dataset, and T represents the gene set containing all genes. Then, the trade-off between the relevance and redundancy was computed by the following formula (4),

$$\max_f(D - R) \quad (4)$$

After repeated computation per the above formula, the trade-off of each feature gene in the dataset was sorted. A new gene list was obtained (5),

$$S = \{f'_1, f'_2, \dots, f'_h, \dots, f'_N\} \quad (5)$$

where the subscript index of each feature gene in S was selected. A feature that was selected earlier had a smaller index and could have high feature importance. Finally, the top 500 genes in the ranked feature list were selected for subsequent research.

Classifier Selection by IFS Method

Following mRMR analysis, IFS was sequentially applied to identify genes for the optimal AD classifier (Chen et al., 2017; Li M. et al., 2020). Firstly, based on the ranked feature list, a series of feature subsets were set as $F_1, F_2, F_3, \dots, F_n$, where $F_i = \{f_1, f_2, \dots, f_i\}$ and f_i refers to the top 500 genes in the ranked feature list. Secondly, the Python package *sklearn* was applied to establish a series of AD classifiers using the above feature subsets with SVM, RF, and KNN algorithms. SVM, RF, and KNN classifiers all can compute the expression of feature genes to identify AD patients (Zhang, 2016; Sarica et al., 2017; Huang et al., 2018). IFS curves were then plotted under 10-fold cross-validation to obtain the Matthews correlation coefficient (MCC), a parameter able to reflect classifier effectiveness (Chicco and Jurman, 2020), of each candidate classifier. Eventually, the classifier with the greatest MCC value was identified as the optimal AD classifier, and the genes involved in were taken as the optimal feature genes.

Principal Component Analysis (PCA) and Heatmap Construction

The R package *FactoMineR* was applied for dimensionality reduction of the two downloaded cohort based on the features of the optimal classifier following PCA. In brief, PCA can reduce the dimensionality of the data in two datasets and map the data into 2 representative dimensions. A scatter plot was drawn based

on the distribution of samples in the two dimensions to present the variance between samples and between groups. Furthermore, the expression of feature genes of the classifier in ND and AD populations was compared through clustering analysis using the R package *heatmap*.

Random Walk With Restart (RWR) and Enrichment Analyses

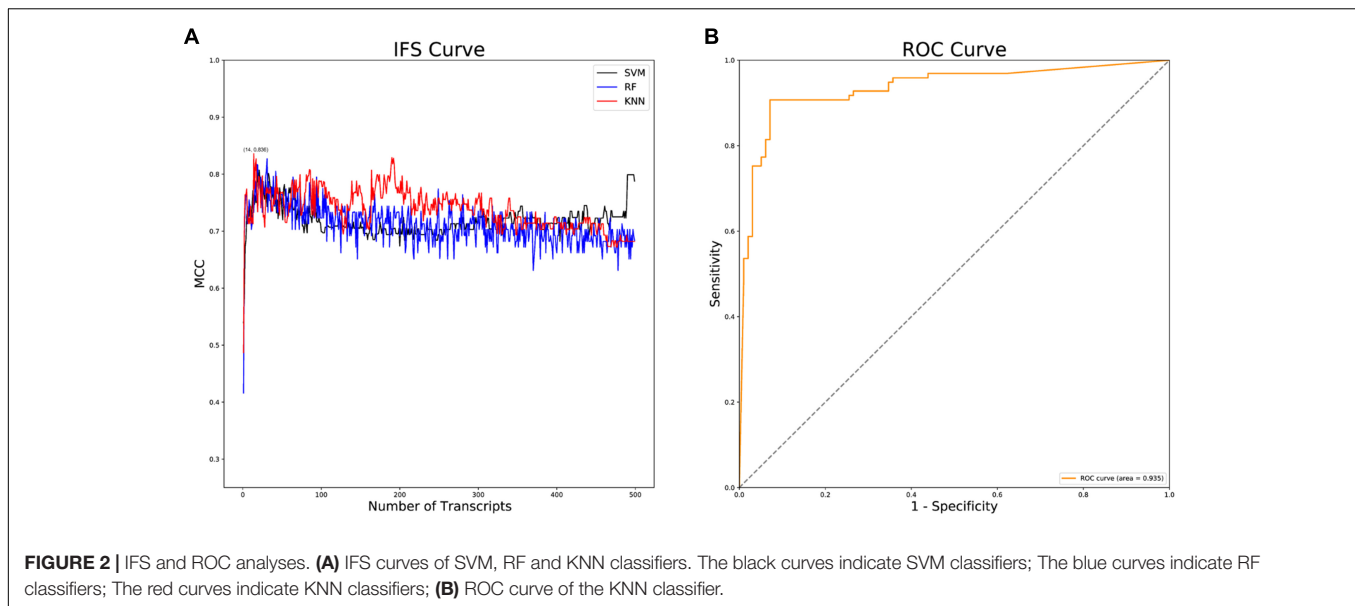
To explore core genes from the optimal classifier and their potential functions, R package *limma* was firstly employed to analyze the difference in gene expression between the ND and AD groups in GEO, and differentially expressed genes (DEGs) were screened ($|\text{Log}_2\text{FC}| > 0.585$, $\text{FDR} < 0.05$). The DEGs were then intersected with the feature genes in the identified classifier to obtain core genes. Wilcox test was implemented to test the differential expression of the core genes in ND and AD samples. Additionally, the DEGs were projected onto the STRING website to construct a protein-protein interaction (PPI) network (Interaction Score > 0.4). RWR algorithm stimulates a random walking starting from a seed node or several seed nodes to a randomly selected neighbor node or to return to the origin in a constructed network. This walking is iterative and terminates when all nodes in the network are walked, and finally a relevance score between each node and the seed node is obtained (Kohler et al., 2008; Zhang et al., 2018). Here, RWR algorithm was run to calculate the relevance score between each node gene and the seed node gene, and the node gene with a score $> 10^{-5}$ was taken as the core gene-associated gene.

Gene Ontology (GO) function annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were sequentially carried out for the core gene-associated genes using the R package *ClusterProfiler*, thereby to explore the critical functions that may affect the pathogenesis of AD. The results were finally visualized using the R package *enrichplot*.

RESULTS

Results of the mRMR and IFS Analyses

Following data downloading and normalization, 28,844 genes were obtained for feature importance analysis through mRMR analysis. The top 500 feature genes in the mRMR analysis were selected (**Supplementary Table 1**), by which a series of AD classifiers were constructed. Then, IFS analysis was implemented to select the optimal classifier. As illustrated in **Figure 2A**, the KNN classifier composed of 14 feature genes had the highest Matthews correlation coefficient (MCC) value. Then, the diagnostic efficacy of the KNN classifier was validated using receiver operation characteristic (ROC) curves. The results presented the sensitivity was 0.907, the specificity was 0.929, the accuracy was 0.918, the MCC value was 0.836, and the area under the curve (AUC) value was 0.935, indicating the high diagnostic efficacy of the KNN classifier in classifying AD patients accurately (**Figure 2B**).



Results of PCA and Heatmap Analysis

Principal Component analysis was conducted for two groups of patients (ND/AD) according to the expression of 14 feature genes in the optimal KNN classifier. The results revealed that PCA analysis could markedly classify AD patients and normal subjects (**Figure 3A**). Besides, a clustering heat map was drawn to analyze the expression of the 14 feature genes in different populations. The results denoted that the 14 feature genes in the KNN model could distinguish AD patients from healthy subjects (**Figure 3B**). These findings manifested that the 14 feature genes in the KNN model exhibited a favorable performance in classifying AD patients from normal individuals, indicating an outstanding diagnostic efficacy.

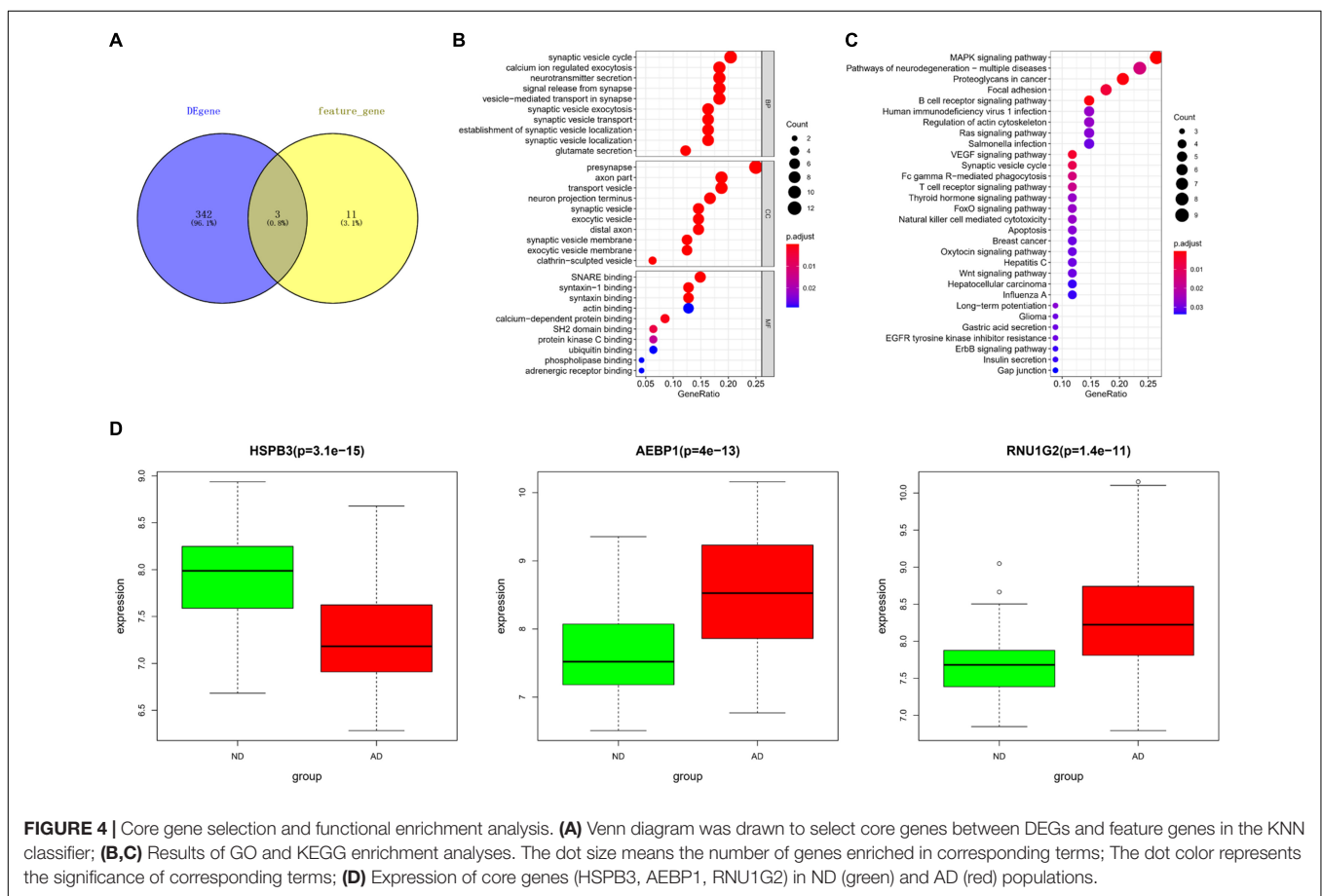
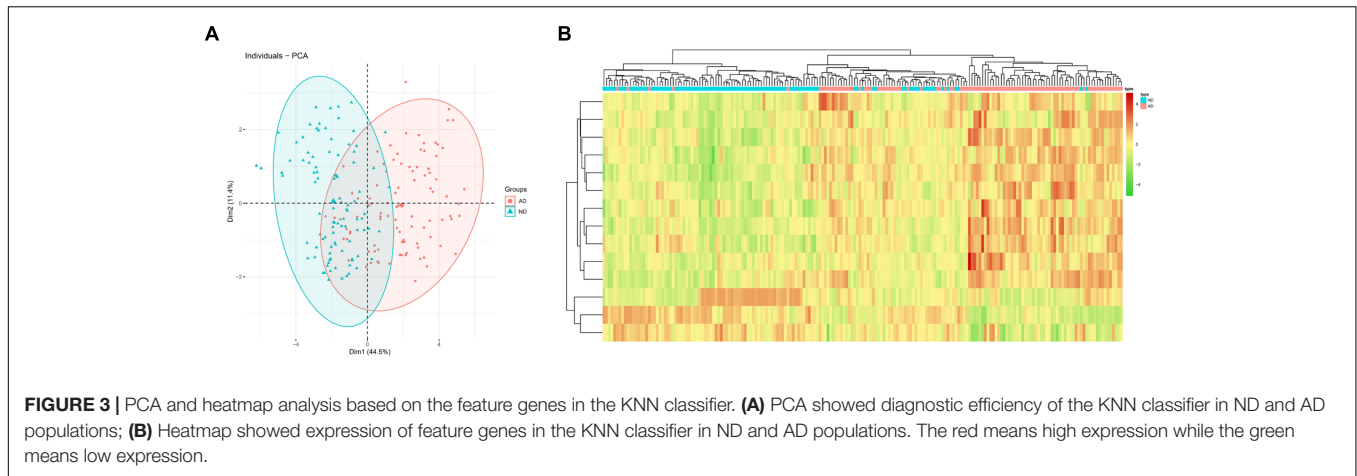
Results of RWR and Enrichment Analysis

Differentially expressed genes screened from the downloaded gene expression data in GEO and feature genes in the KNN classifier were intersected to obtain 3 core genes, including heat shock protein family B (small) member 3 (HSPB3), adipocyte enhancer binding protein 1 (AEBP1), RNA U1 Small Nuclear 4 (RNU1G2) (**Figure 4A**). Based on the DEGs, a PPI network was constructed. Since the Interaction Score of RNU1G2 in the PPI network was less than 0.4, AEBP1 and HSPB3 were picked up as seed nodes to perform RWR algorithm. Eventually, 52 core gene-associated genes were obtained (**Supplementary Table 2**). Then, the 52 core gene-associated genes were subjected to enrichment analysis. As analyzed, the genes were related to biological processes, such as synaptic vesicle cycle, transport vesicle, protein kinase C binding (**Figure 4B**), and activated in pathways such as MAPK signaling pathway, B cell receptor signaling pathway, and T cell receptor signaling pathway (**Figure 4C**). Finally, expression of the 3 core genes in AD was detected. As shown in **Figure 4D**, HSPB3 was conspicuously down-regulated, while AEBP1 and RNU1G2 were notably up-regulated in the AD group. Taken together, these results demonstrated that the 3 core

genes were closely associated with the pathogenesis of AD, and were mainly related to cell functions involved in immunity and cell transportation.

DISCUSSION

The cause of AD is thought to be correlated with A β deposition and the hyperphosphorylation of tau proteins, whereas the cause of the above processes remains a mystery, which puzzles researchers for treatment and prevention of AD (Park et al., 2019; Busche and Hyman, 2020). There are many hypotheses about AD, including A β cascade hypothesis, tau hypothesis, inflammation hypothesis, cholinergic and oxidative stress hypothesis and glucose hypometabolism (Du et al., 2018). A β hypothesis considered as a major cause of AD believes that A β deposition is the major cause of AD, and A β deposition in the nervous system makes nerve cells lack necessary nutrients and cell apoptosis (Du et al., 2018). Neurofibrillary tangles are another pathological feature of AD patients, besides, phosphorylated tau aggregates proteins to cause neuron damage, and drugs targeting tau are promising for AD therapy (Brier et al., 2016; Gauthier et al., 2016; Li and Gotz, 2017; Novak et al., 2017). In addition, several investigations considered that inflammation and oxidative stress are central to AD pathogenesis. The sustained activation of the brain's resident macrophages (microglia) exacerbates both A β and tau pathology and hastens AD pathogenesis (Kinney et al., 2018). Besides, the A β clearance disorder caused by oxidative stress also hastens AD pathogenesis (Cheignon et al., 2018). The deepening of study on the pathogenesis of AD has revealed numerous key genes affecting the pathogenesis of AD. For instance, MS4A and somatic APP are involved in the above pathways to influence AD pathogenesis (Lee et al., 2018; Deming et al., 2019). Moreover, lncRNA and miRNA also participate in AD pathogenesis (Dube et al., 2019; Iranifar et al., 2019). In this study, publicly available gene expression data of AD patients



were analyzed, and the top 500 feature genes that may affect AD pathogenesis were screened out from 28,844 genes with the mRMR algorithm. This study speculated that expression of these genes may be closely related to AD pathogenesis.

Based on mRMR analysis, this study constructed an optimal AD classifier that could accurately classify AD patients and healthy individuals among SVM, RF, and KNN classifiers via IFS analysis, and then a 14-gene signature was obtained. These

feature genes were then intersected with DEGs to obtain 3 core genes (HSPB3, AEBP1, RNU1G2) which may function in AD pathogenesis. Heat shock proteins (HSPs) are important molecular chaperones that prevent protein misfolding and promote the degradation of improperly folded proteins (van Noort et al., 2017). HSPs play a role in protecting multiple sclerosis, protein folding diseases, and genetic white matter diseases (van Noort et al., 2017). Although there is no direct

evidence that HSPB3, a member of the HSP family, participates in AD pathogenesis, this study observed that HSPB3 was conspicuously lowly expressed and may be critical in AD (Boelens et al., 1998). Consulting to other members of the HSPs in the nervous system, HSPB3 may protect the nervous system from A β by degrading A β , whereas HSPB3 deletion may lead to AD (Calderwood and Murshid, 2017). AEBP1 plays an important role in lipid metabolism, which activates inflammatory responses through the NF- κ B pathway and regulates adipogenesis in preadipocytes (Majdalawieh and Ro, 2010; Shijo et al., 2018; Gerhard et al., 2019). AEBP1 is up-regulated in AD patients, which promotes the inflammatory response around the nucleus in hippocampal pyramidal neurons, the formation of neurofibrillary tangles, and the progression of AD (Shijo et al., 2018). This study revealed that AEBP1 was up-regulated in the MTG of AD patients. This result denoted that the inflammatory stress and adipogenesis in the MTG may result in the pathogenesis of AD. RNU1G2 is a kind of small nuclear RNA molecule (snRNA) that cannot translate itself into protein, but it participates in pre-mRNA processing. So far, there has been few discussions about the mechanism of RNU1G2 and its biological functions. However, a study manifested the changes of RNU1G2 expression in the brain of AD patients (Piras et al., 2019). This study disclosed that RNU1G2 was highly expressed in the brain of AD patients and may be critical in AD pathogenesis, indicating that alternative RNA splicing is promising to disclose the pathogenesis of AD. In conclusion, whilst some research on HSPB3 and AEBP1 has presented their roles in the pathogenesis of AD, these investigations are still insufficient. It is worth exploring the role of the above three genes in the occurrence and progression of AD.

Furthermore, RWR analysis was conducted here on a DEGs-based PPI network with the above core genes as seed genes, and the core gene-associated genes screened out were then subjected to GO and KEGG enrichment analyses. The results illustrated that these core gene-associated genes were mainly related to biological processes such as synaptic vesicle cycle, transport vesicle, and protein kinase C binding, and activated in functional pathways such as MAPK signaling pathway, B cell receptor signaling pathway, and T cell receptor signaling pathway. The above results indicated that AD may be associated with neurotransmitter transmission. The decrease of neurotransmitter and activity is an essential phenotype of AD, whereas neurotransmitter supplementation is pivotal to the treatment of AD (Kandimalla and Reddy, 2017). In this study, the three core genes we identified were related to neurotransmitter transmission, suggesting that these genes may affect AD pathogenesis by modulating neurotransmitter secretion. Moreover, the results of enrichment analysis clarified that AD was associated with immune cell stress response. Inflammation is considered to be a key factor influencing the progression of AD, and microglia activation exacerbates both A β and tau deposition (Kinney et al., 2018). T cells function in AD and multiple sclerosis. An investigation displayed that hippocampal T cell infiltration leads to neuroinflammation and cognitive impairments (Laurent et al., 2017). Another study suggested that A β serves as an antigenic factor of T cells to

potentiate encephalitis (Leoutsakos et al., 2018). As such, this study exhibited that the core genes we identified were related to B cell and T cell receptor signaling pathways, indicating that the three core genes may regulate immune cell activity, thereby affecting the pathogenesis of AD.

In summary, gene expression data of AD were firstly downloaded here from GEO database. Next, an AD classifier with favorable diagnostic efficacy was screened out through mRMR and IFS. The feature genes in the classifier were intersected with DEGs to obtain three core genes (HSPB3, AEBP1, RNU1G2) closely related to AD pathogenesis. Among the three core genes, HSPB3 may regulate protein folding processes and degrade misfolded proteins. AEBP1 stimulates adipogenesis in preadipocytes to induce inflammation and activates inflammatory responses to facilitate the pathogenesis of AD. Although our studies are reliable, these results are still in need of verification by molecular biological experiments. Further research could explore the biological functions of the three genes and the pathogenesis of AD to produce findings that account more for clinical therapy, thereby benefiting AD patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

All authors contributed to data analysis, drafting and revising the article, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work.

FUNDING

This study was supported by the funds from Guangxi Natural Science Foundation for Young Scientists (No. 2017JJB140090z), the National Natural Science Foundation of China (No. 81860244), the Natural Science Foundation of Guangxi (No. 2018GXNSFAA281051 and 2020GXNSFAA259057), and the First Batch of High-level Talent Scientific Research Projects of the Affiliated Hospital of Youjiang Medical University for Nationalities in 2019 (No. R20196308).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2021.668738/full#supplementary-material>

Supplementary Table 1 | Top 500 genes selected by mRMR analysis.

Supplementary Table 2 | Fifty-two core gene-associated genes obtained by RWR analysis.

REFERENCES

- Boelens, W. C., Van Boekel, M. A., and De Jong, W. W. (1998). HspB3, the most deviating of the six known human small heat shock proteins. *Biochim. Biophys. Acta* 1388, 513–516. doi: 10.1016/s0167-4838(98)00215-5
- Brier, M. R., Gordon, B., Friedrichsen, K., McCarthy, J., Stern, A., and Christensen, J. (2016). Tau and Abeta imaging, CSF measures, and cognition in Alzheimer's disease. *Sci. Transl. Med.* 8:338ra366. doi: 10.1126/scitranslmed.aaf2362
- Busche, M. A., and Hyman, B. T. (2020). Synergy between amyloid-beta and tau in Alzheimer's disease. *Nat. Neurosci.* 23, 1183–1193. doi: 10.1038/s41593-020-0687-6
- Calderwood, S. K., and Murshid, A. (2017). Molecular chaperone accumulation in cancer and decrease in Alzheimer's disease: the potential roles of HSF1. *Front. Neurosci.* 11:192. doi: 10.3389/fnins.2017.00192
- Cheignon, C., Tomas, M., Bonnefont-Rousselot, D., Faller, P., Hureau, C., and Collin, F. (2018). Oxidative stress and the amyloid beta peptide in Alzheimer's disease. *Redox Biol.* 14, 450–464. doi: 10.1016/j.redox.2017.10.014
- Chen, L., Zhang, Y. H., Wang, S., Zhang, Y., Huang, T., and Cai, Y. D. (2017). Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways. *PLoS One* 12:e0184129. doi: 10.1371/journal.pone.0184129
- Chico, D., and Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21:6. doi: 10.1186/s12864-019-6413-7
- Chowdhury, A. R., Chatterjee, T., and Banerjee, S. A. (2019). Random Forest classifier-based approach in the detection of abnormalities in the retina. *Med. Biol. Eng. Comput.* 57, 193–203. doi: 10.1007/s11517-018-1878-0
- de Brevern, A. G., Hazout, S., and Malpertuy, A. (2004). Influence of microarrays experiments missing values on the stability of gene groups by hierarchical clustering. *BMC Bioinformatics* 5:114. doi: 10.1186/1471-2105-5-114
- Degenhardt, F., Seifert, S., and Szymczak, S. (2019). Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform.* 20, 492–503. doi: 10.1093/bib/bbx124
- Deming, Y., Filipello, F., Cignarella, F., Cantoni, C., Hsu, S., Mikesell, R., et al. (2019). The MS4A gene cluster is a key modulator of soluble TREM2 and Alzheimer's disease risk. *Sci. Transl. Med.* 11:eaau2291. doi: 10.1126/scitranslmed.aau2291
- Dos Santos Picanco, L. C., Ozela, P. F., de Fatima de Brito Brito, M., Pinheiro, A. A., Padilha, E. C., Braga, F. S., et al. (2018). Alzheimer's disease: a review from the pathophysiology to diagnosis, new perspectives for pharmacological treatment. *Curr. Med. Chem.* 25, 3141–3159. doi: 10.2174/0929867323666161213101126
- Du, X., Wang, X., and Geng, M. (2018). Alzheimer's disease hypothesis and related therapies. *Transl. Neurodegener.* 7:2. doi: 10.1186/s40035-018-0107-y
- Dube, U., Del-Aguila, J. L., Li, Z., Budde, J. P., Jiang, S., Hsu, S., et al. (2019). An atlas of cortical circular RNA expression in Alzheimer disease brains demonstrates clinical and pathological associations. *Nat. Neurosci.* 22, 1903–1912. doi: 10.1038/s41593-019-0501-5
- Feng, L., and Jin, F. (2018). Screening of differentially methylated genes in breast cancer and risk model construction based on TCGA database. *Oncol. Lett.* 16, 6407–6416. doi: 10.3892/ol.2018.9457
- Gauthier, S., Feldman, H. H., Schneider, L. S., Wilcock, G. K., Frisoni, G. B., Hardlund, J. H., et al. (2016). Efficacy and safety of tau-aggregation inhibitor therapy in patients with mild or moderate Alzheimer's disease: a randomised, controlled, double-blind, parallel-arm, phase 3 trial. *Lancet* 388, 2873–2884. doi: 10.1016/S0140-6736(16)31275-2
- Genin, E., Hannequin, D., Wallon, D., Slegers, K., Hiltunen, M., Combarros, O., et al. (2011). APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol. Psychiatry* 16, 903–907. doi: 10.1038/mp.2011.52
- Gerhard, G. S., Hanson, A., Wilhelmsen, D., Piras, I. S., Still, C. D., Chu, X., et al. (2019). AEBP1 expression increases with severity of fibrosis in NASH and is regulated by glucose, palmitate, and miR-372-3p. *PLoS One* 14:e0219764. doi: 10.1371/journal.pone.0219764
- Govindpani, K., Turner, C., Waldvogel, H. J., Faull, R. L. M., and Kwakowsky, A. (2020). Impaired expression of GABA signaling components in the Alzheimer's disease middle temporal Gyrus. *Int. J. Mol. Sci.* 21:8704. doi: 10.3390/ijms21228704
- Greenberg, S. M., Bacskai, B. J., Hernandez-Guillamon, M., Pruzin, J., Sperling, R., and van Veluw, S. J. (2020). Cerebral amyloid angiopathy and Alzheimer disease—one peptide, two pathways. *Nat. Rev. Neurol.* 16, 30–42. doi: 10.1038/s41582-019-0281-2
- He, Z., Zhang, J., Yuan, X., Xi, J., Liu, Z., and Zhang, Y. (2019). Stratification of breast cancer by integrating gene expression data and clinical variables. *Molecules* 24:631. doi: 10.3390/molecules24030631
- Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., and Xu, W. (2018). Applications of Support Vector Machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics* 15, 41–51. doi: 10.21873/cgp.20063
- Iranifar, E., Seresht, B. M., Momeni, F., Fadaei, E., Mehr, M. H., Ebrahimi, Z., et al. (2019). Exosomes and microRNAs: new potential therapeutic candidates in Alzheimer disease therapy. *J. Cell. Physiol.* 234, 2296–2305. doi: 10.1002/jcp.27214
- Jabbar, M. A., Deekshatulu, B. L., and Chandra, P. J. P. T. (2013). Classification of heart disease using K- nearest neighbor and genetic algorithm. 10, 85–94.
- Kandimalla, R., and Reddy, P. H. (2017). Therapeutics of neurotransmitters in Alzheimer's disease. *J. Alzheimers Dis.* 57, 1049–1069. doi: 10.3233/JAD-161118
- Kinney, J. W., Bemiller, S. M., Murtishaw, A. S., Leisgang, A. M., Salazar, A. M., and Lamb, B. T. (2018). Inflammation as a central mechanism in Alzheimer's disease. *Alzheimers Dement. (N Y)* 4, 575–590. doi: 10.1016/j.trci.2018.06.014
- Kohler, S., Bauer, S., Horn, D., and Robinson, P. N. (2008). Walking the interactome for prioritization of candidate disease genes. *Am. J. Hum. Genet.* 82, 949–958. doi: 10.1016/j.ajhg.2008.02.013
- Kulkarni, H., Thangam, M., and Amin, A. P. (2021). Artificial neural network-based prediction of prolonged length of stay and need for post-acute care in acute coronary syndrome patients undergoing percutaneous coronary intervention. *Eur. J. Clin. Invest.* 51:e13406. doi: 10.1111/eci.13406
- Laurent, C., Dorothée, G., Hunot, S., Martin, E., Monnet, Y., Duchamp, M., et al. (2017). Hippocampal T cell infiltration promotes neuroinflammation and cognitive decline in a mouse model of tauopathy. *Brain* 140, 184–200. doi: 10.1093/brain/aww270
- Lee, M. H., Siddoway, B., Kaeser, G. E., Segota, I., Rivera, R., Romanow, W. J., et al. (2018). Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* 563, 639–645. doi: 10.1038/s41586-018-0718-6
- Leoutsakos, J. S., Yan, H., Anderson, W. S., Asaad, W. F., Baltuch, G., Burke, A., et al. (2018). Deep brain stimulation targeting the fornix for mild Alzheimer Dementia (the ADvance Trial): a two year follow-up including results of delayed activation. *J. Alzheimers Dis.* 64, 597–606. doi: 10.3233/JAD-180121
- Li, C., and Gotz, J. (2017). Tau-based therapies in neurodegeneration: opportunities and challenges. *Nat. Rev. Drug Discov.* 16, 863–883. doi: 10.1038/nrd.2017.155
- Li, M., Chen, F., Zhang, Y., Xiong, Y., Li, Q., and Huang, H. (2020). Identification of post-myocardial infarction blood expression signatures using multiple feature selection strategies. *Front. Physiol.* 11:483. doi: 10.3389/fphys.2020.00483
- Li, X., Wang, S. W., Li, X. L., Yu, F. Y., and Cong, H. M. (2020). Knockdown of long non-coding RNA TUG1 depresses apoptosis of hippocampal neurons in Alzheimer's disease by elevating microRNA-15a and repressing ROCK1 expression. *Inflamm. Res.* 69, 897–910. doi: 10.1007/s00011-020-01364-8
- Majdalawieh, A., and Ro, H. S. (2010). PPARgamma and LXRalpha face a new regulator of macrophage cholesterol homeostasis and inflammatory responsiveness, AEBP1. *Nucl. Recept. Signal.* 8:e004. doi: 10.1621/nrs.08004
- Novak, P., Schmidt, R., Kontsekova, E., Zilka, N., Kovacech, B., Skrabana, R., et al. (2017). Safety and immunogenicity of the tau vaccine AADvac1 in patients with Alzheimer's disease: a randomised, double-blind, placebo-controlled, phase 1 trial. *Lancet Neurol.* 16, 123–134. doi: 10.1016/S1474-4422(16)30331-3
- Park, J. S., Lee, J., Jung, E. S., Kim, M. H., Kim, I. B., Son, H., et al. (2019). Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. *Nat. Commun.* 10:3090. doi: 10.1038/s41467-019-11000-7
- Patterson, C. (2018).
- Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern. Anal. Mach. Intell.* 27, 1226–1238. doi: 10.1109/TPAMI.2005.159
- Piras, I. S., Krate, J., Delvaux, E., Nolz, J., Mastroeni, D. F., Persico, A. M., et al. (2019). Transcriptome changes in the Alzheimer's disease middle temporal Gyrus: importance of RNA metabolism and mitochondria-associated membrane genes. *J. Alzheimers Dis.* 70, 691–713. doi: 10.3233/JAD-181113

- Sarica, A., Cerasa, A., and Quattrone, A. (2017). Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. *Front. Aging Neurosci.* 9:329. doi: 10.3389/fnagi.2017.00329
- Shijo, M., Honda, H., Suzuki, S. O., Hamasaki, H., Hokama, M., Abolhassani, N., et al. (2018). Association of adipocyte enhancer-binding protein 1 with Alzheimer's disease pathology in human hippocampi. *Brain Pathol.* 28, 58–71. doi: 10.1111/bpa.12475
- Sims, R., van der Lee, S. J., Naj, A. C., Bellenguez, C., Badarinarayan, N., Jakobsdottir, J., et al. (2017). Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat. Genet.* 49, 1373–1384. doi: 10.1038/ng.3916
- Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., et al. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* 17, 520–525. doi: 10.1093/bioinformatics/17.6.520
- Ulland, T. K., and Colonna, M. (2018). TREM2 - a key player in microglial biology and Alzheimer disease. *Nat. Rev. Neurol.* 14, 667–675. doi: 10.1038/s41582-018-0072-1
- van Noort, J. M., Bugiani, M., and Amor, S. (2017). Heat shock proteins: old and novel roles in neurodegenerative diseases in the central nervous system. *CNS Neurol. Disord. Drug Targets* 16, 244–256. doi: 10.2174/1871527315666161031125317
- Wang, S., Zhang, X., Guo, Y., Rong, H., and Liu, T. (2017). The long noncoding RNA HOTAIR promotes Parkinson's disease by upregulating LRRK2 expression. *Oncotarget* 8, 24449–24456. doi: 10.18632/oncotarget.15511
- Xu, J., Gao, C., Zhang, F., Ma, X., Peng, X., Zhang, R., et al. (2016). Differentially expressed lncRNAs and mRNAs identified by microarray analysis in GBS patients vs healthy controls. *Sci. Rep.* 6:21819. doi: 10.1038/srep21819
- Xu, Y., Deng, Y., Ji, Z., Liu, H., Liu, Y., Peng, H., et al. (2014). Identification of thyroid carcinoma related genes with mRMR and shortest path approaches. *PLoS One* 9:e94022. doi: 10.1371/journal.pone.0094022
- Zhang, J., Suo, Y., Liu, M., and Xu, X. (2018). Identification of genes related to proliferative diabetic retinopathy through RWR algorithm based on protein-protein interaction network. *Biochim. Biophys. Acta Mol. Basis Dis.* 1864, 2369–2375. doi: 10.1016/j.bbdis.2017.11.017
- Zhang, Y., Ding, C., and Li, T. (2008). Gene selection algorithm by combining reliefF and mRMR. *BMC Genomics* 9(Suppl. 2):S27. doi: 10.1186/1471-2164-9-S2-S27
- Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. *Ann. Transl. Med.* 4:218. doi: 10.21037/atm.2016.03.37

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.

Copyright © 2021 Yang, Qin and Jian. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.