Original Article
Classifying mild cognitive impairment and Alzheimer’s disease by constructing a 14-gene diagnostic model

Jing Han1*, Gang-Hua Feng2*, Hua-Wu Liu1, Ji-Ping Yi2, Ji-Bao Wu2, Xiao-Xi Yao2

1School of Basic Medical Sciences, Xiangnan University, Chenzhou 423000, Hunan, China; 2Department of Neurology, Chenzhou First People’s Hospital, Chenzhou 423000, Hunan, China. *Equal contributors.

Received January 14, 2022; Accepted May 16, 2022; Epub July 15, 2022; Published July 30, 2022

Abstract: Background: Alzheimer’s disease (AD) and mild cognitive impairment (MCI) are two neurodegenerative diseases. Most patients with MCI will develop AD. Early detection of AD and MCI is a crucial issue in terms of secondary prevention. Therefore, more diagnostic models need to be developed to distinguish AD patients from MCI patients. Methods: In our research, the expression matrix and were screened from Gene Expression Omnibus (GEO) databases. A 14-gene diagnostic model was constructed with lasso logistic analysis. The efficiency and accuracy of diagnostic model have also been validated. In order to clarify the expression differences of 14 genes in health donor, AD and MCI, the blood samples of patients and healthy individuals were collected. The mRNA expression of the 14 genes in blood sample were detected. The SH-SY5Y cell injury model was constructed and biological function of POU2AF1 and ANKRD22 in SH-SY5Y have been proved. Results: We obtained 16 genes which have an area under curve (AUC) ≥0.6. After that, a diagnostic model based on 14 genes was constructed. Validation in independent cohort showed that the diagnostic model has a good diagnostic efficiency. The expressions of 6 genes in AD patients were significantly lower than those in healthy individuals and MCI patients, while the expressions of 8 genes in AD patients were significantly higher than those in healthy individuals and MCI patients. In in vitro experiments, we found that two key genes POU2AF1 and ANKRD22 could regulate neuronal development by regulating cell viability and IL-6 expression. Conclusion: The diagnostic model established in this study has a good diagnose efficiency. Most of these genes in diagnostic model also showed diagnostic value in AD patients. This research also can help doctors make better diagnosis for the treatment and prevention of AD.

Keywords: Alzheimer’s disease, mild cognitive impairment, unsupervised clustering, SH-SY5Y cells, risk factors

Introduction

Alzheimer’s Disease (AD) is a common degenerative disease of the central nervous system. AD is clinically characterized by decline in memory and cognitive function, and may be accompanied by psychiatric symptoms and behavioral disorders [1]. Currently, there are about 50 million people with AD worldwide, and the number is expected to reach 152 million by 2050 [2]. AD not only seriously endangers the health of patients, but also affects their quality of life. At the same time, it brings heavy economic burden to families and society. Therefore, AD has always been a research hotspot in the field of international neurological diseases. It is unlikely that a single biomarker can be used to make an adequate distinction between cases and non-cases of AD [3, 4]. In addition, no specific biomarker can reliably predict the occurrence of clinical symptoms in asymptomatic individuals [5]. Therefore, it is important to search for specific biomarkers and construct diagnostic model for the early diagnosis of AD.

Mild cognitive impairment (MCI) is an intermediate state that occurs during the transition from normal to AD. Early detection of MCI is a crucial issue in terms of secondary prevention. It can act as a transitional state of evolving dementia with a range of conversion of 10-15% per year [6]. More than 80% MCI patients convert to dementia in 6 years [7]. At the same time, some MCI patients can become cognitively normal or stable without further deterioration with intervention and effective treatment [8]. Therefore, early screening, intervening and treatment in MCL patients can not only reduce the incidence
Constructing a diagnostic model of Alzheimer’s disease

Table 1. Clinical characteristics of patients (GSE63063)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>States (years)</td>
<td></td>
</tr>
<tr>
<td>CTL</td>
<td>136 (35.05%)</td>
</tr>
<tr>
<td>MCI</td>
<td>113 (29.12%)</td>
</tr>
<tr>
<td>AD</td>
<td>139 (35.82%)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>154 (39.69%)</td>
</tr>
<tr>
<td>Female</td>
<td>234 (60.31%)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>≤80</td>
<td>242 (62.37%)</td>
</tr>
<tr>
<td>&gt;60</td>
<td>146 (37.62%)</td>
</tr>
</tbody>
</table>

of AD, but also delay the occurrence of AD. In the study of early diagnosis of AD and MCI, many specific biomarkers have been reported. Some researchers found that three plasma metabolites could differentiate MCI and AD in elderly patients [9]. Cerebrospinal fluid (CSF) biomarkers have also been proved that they were better at classifying controls from both dementia and MCI patients [10]. Several cell-free microRNAs (miRNAs) might also have key roles in MCI and dementia progression. They have also been regarded as biomarkers to detect MCI and dementia [11]. However, owing to the characteristic of significant heterogeneity across AD individuals, multimodal approach and multi-gene diagnostic model might be more effective in predicting disease progression of AD.

With the development of next-generation sequence, more differentially expressed genes (DEGs) related to AD and other neurodegenerative diseases have been proved by whole RNA sequencing (RNA-seq) analysis. At the same time, multimodal and multi-gene diagnostic model has also been constructed for the early diagnosis of AD. Some researchers constructed risk prediction model by using the proportion of neutrophils and the most important hub genes. The model was demonstrated to be effective in prospective AD risk prediction [12]. In addition, some miRNAs in plasma have also been used to construct machine learning model for AD status [13]. These reports suggest that multi-gene diagnostic models are accurate classifiers to differentiate MCI and AD in patients. In the present study, a diagnostic model based on 14 genes was first constructed to differentiate AD patients from MCI patients and healthy individuals. The risk prediction model using those potential biomarkers achieved a high area under curve (AUC) in a validation cohort and effectively determined AD risk in a prospective cohort.

Materials and methods

Patients and datasets

For bioinformatics analyze, expression matrix and patients’ clinical characteristics (Table 1) were obtained from Gene Expression Omnibus (GEO) database (GSE63063), and the mRNA expression level of patients’ blood mononuclear cells were sequenced using GPL10558 platform. Patients were separated to three groups: health donor (CTL), mild cognitive impairment (MCI) and Alzheimer’s disease (AD) according to their clinical characteristics. For clinical experiments, patients diagnosed with MCI and AD in our hospital were enrolled [14]. Blood samples were collected with patients’ consent and stored at -70°C in refrigerator. CTL blood samples were collected from 30 health donors during physical examinations. The Ethics Committee of Chenzhou First People’s Hospital approved this study (2021 [064]). Written informed consents were obtained from all patients.

Data process

The expression profile was downloaded from GEO database. Expression matrix was first annotated using GPL10558 annotation file. Standard deviation of each gene expression was first calculated using R software, and genes with top 25% standard deviation were obtained for further analysis. Data were separated randomly to training cohort and validation cohort at a ratio of 7:3 (270 for training cohort, 118 for validation cohort).

Principal component analysis (PCA) and unsupervised clustering

PCA was undertaken using R package “psych” and visualized by R package “ggplot2”. Further, unsupervised clustering was undertaken using R package “pheatmap” by K-Means method.

Lasso logistic analysis

Lasso logistic analysis was used to filter biosignatures and construct diagnostic model. Lasso
Constructing a diagnostic model of Alzheimer’s disease

Analysis was undertaken using “glmnet” package. Receiver operating characteristic curve (ROC) was used to identify diagnose efficiency of gene signatures and diagnostic model.

Pathway enrichment analysis

Metascape utilizes the well-adopted hypergeometric test and Benjamini-Hochberg p-value correction algorithm to identify ontology terms [15]. By default, gene list filtered by Lasso logistic analysis were listed in Metascape webpage, the website automatically enriched the pathways in Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, MSigDB, etc. [16, 17]. Results were then downloaded from the webpage of Metascape website.

Real-time polymerase chain reaction

Total RNA was extracted from 90 blood samples using QIAamp RNA Blood Mini Kit (Qiagen, German), and total RNA was first reverse transcribed to cDNA using Advantage RT-for-PCR Kit (Takara, Japan). Relative expression of each gene was further detected by RT2 SYBR® Green qPCR Mastermixes (Qiagen, German). Glyceraldehyde phosphate dehydrogenase (GAPDH) was used as an endogenous control. The relative expression levels were calculated using the \( 2^{-\Delta\Delta Ct} \) method. Primer sets for qPCR were listed in Table S1.

Cell culture and treatment

Human SH-SY5Y hippocampal neuronal cell line was obtained from American Type Culture Collection (ATCC, Manassas, VA, USA) and was cultured in DMEM medium with 10% fetal bovine serum (FBS), 100 U/mL penicillin and 100 mg/mL streptomycin. Cells were cultured under 5% CO\(_2\) at 37°C. To establish AD model in vitro, SH-SY5Y cells were incubated with A\(_\beta\)1-42 for 24 h.

Cell transfection

The si-RNA targeting POU2AF1 and overexpression plasmid oe-ANKRD22 with corresponding si-NC and oe-NC plasmid were purchased from GenePharma (Shanghai, China). Transfection was performed using Lipofectamine 3000 (Invitrogen, USA) with manufacturer’s instruction. qPCR was further used to validate transfection efficiency.

Lactate dehydrogenase (LDH) cytotoxicity assay

LDH assay was processed using LDH-Cytotoxicity Assay Kit (Abcam, USA) following manufacture instructions. Cells transfected were incubated with A\(_\beta\)1-42 for 2 h, followed by detection of LDH release rate. Results were shown as (%).

CCK-8 assay

Cell proliferation was detected using Cell Counting kit-8 (CCK-8) according to the previous study [18]. Briefly, cells were added in each well of a 96-well plate for 24 h. After treatment, CCK-8 solution (Apexbio, USA, 10 µL) was added to each well of the 96-well plate, and the cells were incubated at 37°C for 2 h. The absorbance was measured at 450 nm.

Statistical analysis

All data were expressed as mean ± standard deviation (SD). Multiplex group comparison was undertaken for continuous variables using one-way ANOVA, while comparison between two groups was performed using Student’s t test. Above analyses were undertaken using R software 4.04 and GraphPad Prism 8. All statistical tests were 2-sided. P<0.05 indicated statistically significant difference.

Results

PCA and unsupervised clustering

Sequencing data from AD, MCI patients and health donors were first downloaded from GEO database and then annotated. After annotating the expression matrix, genes with top 25% standard error of mean (SEM) were filtered. To investigate expression difference among the three groups, unsupervised clustering was undertaken (Figure 1A). Results showed that unsupervised clustering could not cluster participates into different groups (Figure 1B). Further, we used PCA to investigate difference among the three groups. Results indicated that mRNA expression data could not distinguish AD patients from those of MCI and health controls (Figure 1C, 1D).

4479
Constructing a diagnostic model of Alzheimer’s disease

Filtering gene signatures by lasso logistic method

After analyzing the gene expression difference of blood mononuclear cells between AD, MCI and health control, we combined MCI and health control data together as non-AD patients, and then used the expression data of training cohort to construct a lasso Logistic diagnostic model. This model could help filter AD gene signatures (Figure 2A). Lasso logistic model showed 30 gene signatures which could diagnose the occurrence of AD (Figure 2B). Further we used Metascape to analyze the pathways that were related to gene signatures. The results indicated that the gene signatures related to AD could regulate interleukin-6 (IL-6) production, inflammatory reaction and humoral immune response, which also indicated immune states of blood mononuclear cells different between AD and other participates (Figure 2C).

ROC curve filter key-gene signatures for AD diagnostic model

To reduce the complexity of diagnostic model, we used ROC curve to evaluate whether filtered gene signatures could diagnose AD patients in all participates after filtering gene signatures using lasso logistic model (Figures 3 and 4). According to previous studies, AUC >0.6 as a screening standard has a good predictive value, so we also set the screening criteria as log-rank with P<0.05 and AUC >0.6 following previews studies [19, 20]. Finally, 16 genes with ROC’s AUC ≥0.6 were retained for further diagnostic model construction. The retained genes were the triggering receptors expressed on myeloid cells (TREM1), transmembrane protein 205 (TMEM205), hepatitis a virus cellular receptor 2 (HAVCR2), ankyrin repeat domain 22 (ANKRD22), ghrelin (GHRL), ectonucleoside triphosphate diphosphohydrolase 1 (ENTPD1), lysophosphatidylcholine acyltransferase 2 (LP- CAT2), CD5, ubiquitin specific peptidase 36
Constructing a diagnostic model of Alzheimer’s disease

Figure 2. Filtering of gene signatures by lasso logistic method. A. Tuning parameter (Lambda) selection in the lasso model used 10-fold cross-validation. B. Lasso coefficient profiles plot. C. Metascape enriched pathways.

GO:0032675: regulation of interleukin-6 production
GO:0006954: inflammatory response
ko05169: Epstein-Barr virus infection
GO:0006959: humoral immune response
Constructing a diagnostic model of Alzheimer’s disease

Figure 3. ROC curve of the first 15 gene signatures. (A-O) ROC curve of (A) PRPF40A, (B) TMED3, (C) TREM1, (D) FAM13A, (E) MGC16703, (F) TMEM205, (G) HAVCR2, (H) ANKRD22, (I) KCTD12, (J) CD163, (K) GTSCR1, (L) DNAJB12, (M) GHRL, (N) C10ORF73 and (O) VAMP5.
Constructing a diagnostic model of Alzheimer's disease

Figure 4. ROC curve of the other 15 gene signatures. (A-O) ROC curve of (A) ZRANB2, (B) ENTPD1, (C) LPCAT2, (D) CD5, (E) USP36, (F) LOC728411, (G) RDH11, (H) HS.555252, (I) POU2AF1, (J) CCR6, (K) NCOR2, (L) VPREB3, (M) LYPD2, (N) TFCP2 and (O) LRRN3.
Constructing a diagnostic model of Alzheimer’s disease

(USP36), retinol dehydrogenase 11 (RDH11), HS.5555252, POU domain, class 2, associating factor 1 (POU2AF1), C-C chemokine receptor 6 (CCR6), V-Set pre-B cell surrogate light chain 3 (VPREB3), transcription factor CP2 (TFCP2), and leucine-rich repeat neuronal protein 3 (LRRN3) (Figures 3 and 4).

Construction and validation of AD diagnostic model

After filtering gene signatures, lasso logistic analysis was used to construct a diagnostic model for distinguishing AD patients from those of MCI and health participates. A diagnostic model was constructed based on training cohort, and 14 genes were maintained as key-gene signatures (Figure 5A, 5B). The formula of diagnostic model constructed was as follow:

RISKSCORE = GHRL * -1.21067788 + HAVCR2 * -0.93385388 + TMEM205 * -0.80577310 + ANKRD22 * -0.56002346 + LPCAT2 * -0.24643750 + TREM1 * -0.15400542 + HS.5555252 * 0.08323193 + VPREB3 * 0.33281970 + CCR6 * 0.60404361 + RDH11 * 0.81924596 + POU2AF1 * 0.99940487 + LRRN3 * 1.45216397 + USP36 and TFCP2 in AD patients were significantly higher than those of health donors (Figure 6G). Besides, there was no significant difference in expression level of key-gene signatures between MCI patients and health donors (Figure 6G). This indicated that diagnostic model constructed can diagnose AD patients from other groups but can’t diagnose MCI patients from health donors.

Validate AD diagnostic model in clinical sample

After construction of AD diagnostic model, this study also validated the diagnostic model by using clinical samples. Patients’ and health donors’ blood samples were collected, and expressions of 14 key-gene signatures were identified. Results indicated similar conclusion in bioinformatics analysis. The expressions of GHRL, HAVCR2, TMEM205, ANKRD22, LPCAT2 and TREM1 in AD patients were significantly lower than those of health donors and MCI patients, while the expressions of HS.5555252, VPREB3, CCR6, RDH11, POU2AF1, LRRN3, USP36 and TFCP2 in AD patients were significantly higher than those of health donors and MCI patients (Figure 7A). This indicates that the AD diagnostic model we constructed also has a good diagnose efficiency at clinical level. Finally, we processed pathway enrichment analysis using these 14 biomarkers, and the results showed these genes enriched at immune effector process, production of molecular mediator of immune response and regulation of IL-6 production pathway (Figure 7B). These results indicated that IL-6 production pathway was the key pathway in AD development and these biomarkers might regulate IL-6 production and affect AD development indirectly.

POU2AF1 and ANKRD22 could regulate proliferation and IL-6 expression of SH-SY5Y cell

To further research into the molecular mechanism of these genes affecting the development of AD, in vitro experiments were performed. Considering a previous study which found that IL-6 production and inflammation response were enriched by AD biomarkers, we first calculated the correlation of expression between IL-6 and 14 biomarkers. Four genes were chosen first: ANKRD22, LPCAT, POU2AF1, and TFCP2. Considering that expressions...
Constructing a diagnostic model of Alzheimer's disease

Figure 5. Construction of AD diagnostic model by lasso logistic method. A. Tuning parameter (Lambda) selection in the lasso model used 10-fold cross-validation. B. Lasso coefficient profiles plot.
Constructing a diagnostic model of Alzheimer’s disease

Figure 6. Validation of AD diagnostic model. A, B. ROC analyze validated diagnostic model in both training cohort and validation cohort. C, D. ROC analyze validated diagnostic model between health control and MCI patients. E. Unsupervised clustering indicated that participates can be divided into 2 major groups. F. PCA result shows distinction for three groups. G. Expression of 14 gene signatures between 3 groups. n=136 in group of CTL, n=112 in group of MCI, n=140 in group of AD, and *P<0.05 in comparison with the CTL group.

of ANKRD22 and POU2AF1 in AD patients have maximum fold change compared to normal control, ANKRD22 and POU2AF1 were chosen for further research (Figure 8A). Considering that these genes may regulate IL-6 expression, we measured IL-6 expression in AD cell model. The result showed that IL-6 production in AD model cell was significantly higher than that in normal cell (Figure 8B). Moreover, we silenced POU2AF1 expression and upregulated ANKRD22 expression in AD model cell (Figure 8C, 8G). Then LDH cell cytotoxicity assay was processed, and results showed that POU2AF1 could reduce cytotoxic effect of Aβ, while ANKRD22 could stimulate cytotoxic effect of Aβ (Figure 8D, 8H). The results of CCK-8 assay showed that POU2AF1 could upregulate cell viability of hippocampal neuronal cell while ANKRD22 could reduce hippocampal neuronal cell viability (Figure 8E, 8I). Finally, we mea-
Constructing a diagnostic model of Alzheimer’s disease

Figure 7. Biomarkers identification. A. Identification of expression levels of 14 key genes in CTL, MCI and AD groups. B. Pathway enrichment result of 14 biomarkers. n=30 per group, *P<0.05 in comparison with the CTL group. & P<0.05 in comparison with the MCI group.

sured IL-6 expression in hippocampal neuronal cells, and results showed that POU2AF1 could suppress IL-6 expression, while ANKRD22 could upregulate IL-6 expression (Figure 8F, 8J). These results revealed that POU2AF1 and ANKRD22 were important regulators of in vitro model of AD.

Discussion

Early prediction of the risk of MCI-to-AD transformation is important to provide appropriate early intervention and better control of the disease. Potential MCI/AD related biomarkers have been used for diagnosis and prognosis [21-24]. Over the past few decades, several core CSF biomarkers for AD have been identified and tested. For early AD, 10 CSF proteins associated with AD pathology in Aβ+ individuals were identified [25]. Other CSF biomarkers in AD including tau and p-tau at threonine 181 (ptau181) are also associated with future decline in some noncognitive AD symptoms studied [26]. However, individual studies of biomarker validity vary greatly. Furthermore, CSF samples are not suitable for all patients compared with blood samples. Therefore, additional work is needed to determine new biomarkers in other samples including blood samples. In this study, we explored diagnostic signatures of AD and established a diagnostic model in blood to differentiate AD patients from MCI patients and healthy individuals.

We investigated the expression difference of genes between AD patients, MCI patients and health donors and construct a 30 genes lasso logistic model. The genes in this model were enrichment in pathways of interleukin 6 (IL-6) production, inflammatory response and EBV (Epstein-Barr virus) infection. Increasing evidence suggests that Alzheimer’s disease pathogenesis is not restricted to the neuronal compartment, but includes strong interactions with immunological mechanisms in the brain. Some researchers have found that the increase of plasma IL-6 level plays an important role in promoting the development of AD [27]. The relationship between viral infection and AD has also been concerned for a long time. The level of EBV and other viruses demonstrated specificity to AD brains [28]. These reports suggested that most of the genes in the model are closely related to the development and diagnosis of AD.

After filtering gene signatures, we used ROC analysis and lasso logistic analysis to construct a diagnostic model for distinguishing AD patients from MCI patients and health partici-
POU2AF1 and ANKRD22 could regulate proliferation and IL-6 expression of SH-SY5Y cell. A. Correlation analysis between 14 biomarkers and IL-6 expression. B. IL-6 expression level in AD model cell. C. POU2AF1 expression in transfected cell. D. LDH assay shows difference in Aβ induced cytotoxicity after transfection. E. CCK-8 assay shows the effect of POU2AF1 on cell viability. F. qPCR shows the effect of POU2AF1 on IL-6 expression. G. ANKRD22 expression in transfected cell. H. LDH assay shows difference in Aβ induced cytotoxicity after transfection. I. CCK-8 assay shows the effect of ANKRD22 on cell viability. J. qPCR shows the effect of ANKRD22 on IL-6 expression. n=3 for each group, *P<0.05.

Figure 8. POU2AF1 and ANKRD22 could regulate proliferation and IL-6 expression of SH-SY5Y cell. A. Correlation analysis between 14 biomarkers and IL-6 expression. B. IL-6 expression level in AD model cell. C. POU2AF1 expression in transfected cell. D. LDH assay shows difference in Aβ induced cytotoxicity after transfection. E. CCK-8 assay shows the effect of POU2AF1 on cell viability. F. qPCR shows the effect of POU2AF1 on IL-6 expression. G. ANKRD22 expression in transfected cell. H. LDH assay shows difference in Aβ induced cytotoxicity after transfection. I. CCK-8 assay shows the effect of ANKRD22 on cell viability. J. qPCR shows the effect of ANKRD22 on IL-6 expression. n=3 for each group, *P<0.05.

The AD diagnostic model includes 14 genes. Among them, TREM1, CCR6 and TFCP2 have been found to be closely associated with the development of AD. TREM1 is a potential therapeutic target for AD. Some researchers revealed that the concentrations of soluble TREM1 showed higher expression in AD patients [29]. For CCR6, it has been regarded as a biomarker for AD in a triple transgenic mouse model [30]. As a host transcription factor, TFCP2 has been found that it could bind to some viruses such as Herpes simplex (HSV-1) and regulate some genes associated with AD [31]. Beside these three genes, some of other
11 genes have also been regarded as diagnostic or prognostic factors [32-35]. Therefore, these studies further confirmed the robustness and utility of the 14-gene signature.

Highly specific and sensitive blood biomarkers have been used in the diagnosis of many diseases. Genes and proteins in the blood of some patients with AD are also thought to be molecular markers that distinguish AD patients from healthy individuals. Plasma P-TAU181 (P-tau-181) has been identified to be increased in preclinical AD and further increased at the MCI and dementia stages. It has been considered as a non-invasive diagnostic and prognostic biomarker for AD [36]. Several independent studies suggested that the plasma Aβ42/40 ratio could be used to distinguish AD patients from healthy individuals [37-39]. These reports confirm the feasibility of using blood molecules as diagnostic markers for AD or MCI. Compared with single-factor biomarkers, establishing blood multi-gene biomarkers is of great significance for improving the diagnostic accuracy of AD patients. In the present study, 6 genes showed lower expression in blood of AD patients compared with that in CTL and MCI patients. Other 8 genes showed higher expression in blood sample of AD patients. Most of these genes in diagnostic model also showed diagnostic value in AD patients. These results further prove that it is necessary to develop blood multigene diagnostic markers.

In order to explore mechanism of the key genes in 14 biomarkers to AD development. We calculated the correlation of expression between IL-6 and 14 biomarkers and found 2 key genes. ANKRD22 has been reported to be involved in the progression of many types of cancers [40, 41]. It has also been regarded as a biomarker of Parkinson's disease [42]. POU2AF1 is a potential immune-related biomarker involved in some diseases [43, 44]. However, the effect of these two genes on the development of Alzheimer’s disease has not been reported. In the present study, we found that ANKRD22 and POU2AF1 could regulate cell activity of AD cell model. Targeting these genes could be an avenue for AD treatment in the future.

There are also some limitations in this research. The model established in this study could accurately distinguish AD patients from MCI/CTL. However, this polygenic model could not be used to distinguish AD patients from MCI patients or MCL patients from CTL. In addition, the diagnosis of MCI requires further analysis in the future. In addition, more clinical samples need to be collected to verify expression difference of the 14 genes between different types of AD patients or MCI patients. The regulatory mechanism of the 14 genes in the development of AD will also be further studied.

Conclusions

In total, a 14 optimal feature genes classifier was acquired via a series of bioinformatics analysis to distinguish AD patients from MCI patients/CTL. ROC curve analysis suggested that the diagnostic performance of the classifier was good and could accurately distinguish MCI patients/CTL from AD patients. The expression of these genes was then testified via clinical samples. Our future work will focus on more advanced feature selection methods to improve classification accuracy under the clinical diagnosis.

Acknowledgements

This research was supported by the Research Foundation of Bureau of Public Health of Hunan Province, China (2020SK50304). The study was designed and performed in agreement with the Declaration of Helsinki and was approved by Ethics Committee of Chenzhou First People's Hospital. Written informed consent was obtained from individual participants or their guardians.

Disclosure of conflict of interest

None.

Abbreviations

AD, Alzheimer’s disease; MCI, mild cognitive impairment; GEO, Gene Expression Omnibus; AUC, area under curve; CSF, cerebrospinal fluid; miRNAs, microRNAs; DEGs, differentially expressed genes; RNA-seq, RNA sequencing; CTL, health donor; PCA, principal component analysis; ROC, receiver operating characteristic curve; GO, Gene Ontology; KEGG, Kyoto encyclopedia of genes and genomes; GAPDH, glyceraldehyde phosphate dehydrogenase; SD, standard deviation; SEM, standard error of mean; TREM1, triggering receptors expressed
on myeloid cells; TMEM205, transmembrane protein 205; HAVCR2, hepatitis A virus cellular receptor 2; ANKRD22, ankyrin repeat domain 22; GHRL, ghrelin; ENTPD1, ectonucleoside triphosphate diphosphohydrolase 1; LPCAT2, lysophosphatidylcholine acyltransferase 2; USP36, ubiquitin specific peptidase 36; RDH11, retinol dehydrogenase 11; POU2AF1, POU domain, class 2, associating factor 1; CCR6, C-C chemokine receptor 6; VPREB3, V-Set pre-B cell surrogate light chain 3; TFCP2, transcription factor CP2; LRRN3, leucine-rich repeat neuronal protein 3; IL-6, interleukin 6; EBV, Epstein-Barr virus; HSV-1, herpes simplex; TFCP2, transcrip
tion factor CP2; LRRN3, leucine-rich repeat neuronal protein 3; IL-6, interleukin 6; EBV, Epstein-Barr virus; HSV-1, herpes simplex; P-tau181, P-TAU181.

Address correspondence to: Xiao-Xi Yao, Department of Neurology, Chenzhou First People’s Hospital, Chenzhou 423000, Hunan, China. E-mail: yaoxxtt@126.com

References

Constructing a diagnostic model of Alzheimer’s disease

Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011; 27: 1739-1740.


[37] Janelidze S, Pannee J, Mikulskis A, Chiao P, Zetterberg H, Blennow K and Hansson O. Concordance between different amyloid immunoassays and visual amyloid positron emission...
Constructing a diagnostic model of Alzheimer’s disease


Constructing a diagnostic model of Alzheimer’s disease

Table S1. The primer sequences of genes applied in this study

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>GHRL</td>
<td>GCAGAGGATGAAGATGAAATGGAAGTCC</td>
<td>CTCTTCCAGAGGATGTCTGAG</td>
</tr>
<tr>
<td>HAVCR2</td>
<td>GACTCTAGCAGACGATGGAAGTC</td>
<td>GGTGGTAAAGCATCCTTGAAAGG</td>
</tr>
<tr>
<td>TMEM205</td>
<td>GCAAATGTGGTGGACCTTCGGTCTC</td>
<td>GCAGAGTTGGAAGGACAGGACAG</td>
</tr>
<tr>
<td>ANKRD22</td>
<td>GACAAGCAGAATGAGGCTTCTTG</td>
<td>AGCAGAGGATAAAGAGACTGTT</td>
</tr>
<tr>
<td>LPCAT2</td>
<td>GATGGCAAGCATTGACTTCCGAG</td>
<td>CCTCCGTATGAGCATCCTC</td>
</tr>
<tr>
<td>TREM1</td>
<td>CGATGTCTCCACTCTCGACTCT</td>
<td>CAGCAACAGGACAGAGAGACC</td>
</tr>
<tr>
<td>HS.555252</td>
<td>ACAAGGAATTGTAAGGCCAGA</td>
<td>GGGCTTTTTCAATTAGTGGA</td>
</tr>
<tr>
<td>VPREB3</td>
<td>ACCATCAGGGACTACCTGGTGTTG</td>
<td>CTCATCCTGTACCTGCAAGTT</td>
</tr>
<tr>
<td>CCR6</td>
<td>CTGAAACCTGTGCCTACGCTT</td>
<td>CACAGGAGAACCCTGAGAAGT</td>
</tr>
<tr>
<td>RDH11</td>
<td>AGCAGGTGTGTTGCGGAAACT</td>
<td>CCGACACATCATCCTGCAAG</td>
</tr>
<tr>
<td>POU2AF1</td>
<td>CCAGTGAGGAAGACATGCGGAGA</td>
<td>CAGAACCCTCAGCTGGAGAG</td>
</tr>
<tr>
<td>LRRN3</td>
<td>GTGACTGTGTCATCGTCCTGGATG</td>
<td>CTTGCGGAAACATCGCCTCCG</td>
</tr>
<tr>
<td>USP36</td>
<td>AGCAGATGTCCGATGAGGAGG</td>
<td>GATGCCTGTGAGGATGCGAG</td>
</tr>
<tr>
<td>TFCP2</td>
<td>GGAACCGACCTCGAGAAGAAT</td>
<td>CAGACATCCCTTCATCGCAG</td>
</tr>
<tr>
<td>IL-6</td>
<td>GTACATCCTGACGGCATCT</td>
<td>GTGCCTGCTTGTGCTTTGAC</td>
</tr>
<tr>
<td>GAPDH</td>
<td>AGGTCGGTGTGAACTGGATAG</td>
<td>TGTAGACCATGTAGTTGGCA</td>
</tr>
</tbody>
</table>

GHRL, ghrelin; HAVCR2, hepatitis A virus cellular receptor 2; TMEM205, transmembrane protein 205; ANKRD22, ankyrin repeat domain 22; LPCAT2, lysophosphatidylcholine acyltransferase 2; TREM1, the triggering receptors expressed on myeloid cells 1; VPREB3, V-Set Pre-B cell surrogate light chain 3; CCR6, C-C chemokine receptor 6; RDH11, retinol dehydrogenase 11; POU2AF1, POU domain, class 2, associating factor 1; LRRN3, leucine-rich repeat neuronal protein 3; USP36, ubiquitin specific peptidase 36; TFCP2, transcription factor CP2; IL-6, interleukin 6; GAPDH, glyceraldehyde phosphate dehydrogenase.