Original Article

Exploration of mitochondrial autophagy related genes in the diagnosis model construction and molecular marker mining of Alzheimer's disease based on multi-omics integration

Shan Wang¹², Jiejie Zhang¹, Haitao Zhang¹, Yihan Yang¹, Ya Wen¹²*¹

¹ Department of Neurology, The Second Hospital of Hebei Medical University, Shijiazhuang 050000, Hebei, China
² Hebei Key Laboratory of Neurology, The Second Hospital of Hebei Medical University, Shijiazhuang 050000, Hebei, China

Article Info

Key features of Alzheimer's disease include neuronal loss, accumulation of beta-amyloid plaques, and formation of neurofibrillary tangles. These changes are due in part to abnormal protein metabolism, particularly the accumulation of amyloid beta. Mitochondria are the energy production centers within cells and are also the main source of oxidative stress. In AD, mitochondrial function is impaired, leading to increased oxidative stress and the production of more reactive oxidative substances, further damaging cells. Mitophagy is an important mechanism for maintaining mitochondrial health, helping to clear damaged mitochondria, prevent the spread of oxidative stress, and reduce abnormal protein aggregation. To this end, this article conducts an integrated analysis based on DNA methylation and transcriptome data of AD. After taking the intersection of the genes where the differential methylation sites are located and the differential genes, machine learning methods were used to build an AD diagnostic model. This article screened five diagnostic genes ATG12, CSNK2A2, CSNK2B, MFN1 and PGAM5 and conducted experimental verification. The diagnostic genes discovered and the diagnostic model constructed in this article can provide reference for the development of clinical diagnostic models for AD.

Keywords: DNA methylation, Alzheimer's disease, Diagnostic model, Mitochondrial autophagy, Molecular markers.

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disease with insidious and difficult-to-detect origins [1]. AD patients suffer from memory impairment, cognitive decline, and personality changes. On a global scale, the number of AD patients over 60 years of age is increasing year by year. Since AD patients require long-term home care and social security, they impose a large burden on society [2]. There is an urgent need to analyze the molecular mechanism of AD treatment from the perspective of genetic variation because the pathogenic mechanism of AD is not yet obvious [3].

Mitochondrial autophagy is a cellular process aimed at removing dysfunctional or redundant mitochondria, preserving energy metabolism, and regulating mitochondrial numbers in a clear process. Previous studies have demonstrated that mitochondrial dysfunction and early-onset bioenergetic deficits occur in neurons affected by AD. Jesse S Kerr et al. reviewed the cellular and molecular mechanisms of mitochondrial autophagy in AD [4]. They concluded that impaired mitochondrial autophagy triggers Aβ and Tau accumulation through increased oxidative damage and cellular energy deficits, which leads to exacerbation of AD. In addition, a large number of studies have shown that epigenetic factors perform an important role in the nervous system during aging. For example, numerous studies have indicated that 5-hydroxymethylcytosine (5hmC) plays a key function in the occurrence and development of AD. Liqi Shu et al. revealed significant differences in 5hmC levels in different brain regions of mice [5]. As mice aged, overall, 5hmC decreased significantly in the hippocampus but not in the cortex and cerebellum. Therefore, it is necessary to explore the biomarkers of AD from an epigenetic perspective.

In this paper, we have first downloaded DNA methylation and gene expression data from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) for AD and its control group and then performed the differential analysis of DNA methylation data to obtain methylation sites that were significantly different in the AD and control groups. Then we annotated these sites. On the other hand, the gene expression data were similarly subjected to differential expression analysis and then intersected with the mitochondrial autophagy-related genes (MRGs) collected previously in the literature [6], as well as genes annotated with differentially methylated loci and differentially expressed
genes (DEGs), and protein-protein interaction network (PPI) analysis and enrichment analysis were performed on the intersected genes. We also performed further screening of the intersected genes, including screening of diagnosis-related genes using the LASSO algorithm and constructing and validating diagnostic models. The diagnostic significance of the methylation sites corresponding to the diagnostic genes was also analyzed. Finally, we explored the immune landscape of AD and its controls based on the CIBERSORT algorithm and using diagnostic-related genes.

2. Materials and methods

2.1. Data acquisition

In this paper, DNA methylation data of AD and its controls (including 34 AD samples and 34 control samples) were downloaded from the GSE76105 dataset of the GEO database. Transcriptome data of AD and its controls were downloaded from the GSE5281 and GSE48350 datasets, where GSE5281 includes 87 AD samples and 74 control samples, and GSE48350 includes 87 AD samples and 173 control samples. We used GSE5281 as the training dataset and GSE48350 as the test dataset.

2.2. Differential expression analysis

In this paper, the differential expression analysis of DNA methylation and gene expression data was realized by using R package “ChAMP” and “limma”, respectively. Specifically, we took the methylation and gene expression matrix and sample labels as the input of the algorithm. Finally, we calculated the P value and logFC value of each methylation site/gene. For methylation data, we reserved the differential methylation sites with P<0.05. For gene expression data, we kept DEGs with P<0.05.

2.3. PPI network analysis

This study obtained files containing information on protein interactions by entering differentially expressed genes into the String database (https://cn.string-db.org/) with the default parameters of the website. Then, CIBERSORT software is used to visualize the PPI network.

2.4. Enrichment analysis

We implemented the KEGG and GO enrichment analysis of the intersecting genes with the R package "clusterProfiler". Then, we applied the R package "ggplot2" to plot the enrichment results with histograms and bubble plots.

2.5. LASSO regression analysis

LASSO regression is built on the machine tool of the generalized linear model and can screen discrete/continuous variables based on the input labels. Its complexity is controlled by the parameter α. The larger α is, the more significant the penalty. In this paper, the LASSO regression algorithm was utilized to screen genes associated with AD diagnosis. Specifically, the gene expression matrix and the corresponding sample labels were taken as the input of this algorithm, and finally, the genes with diagnostic significance were output.

2.6. Immunocorrelation analysis

In this study, the ssGSEA algorithm was adopted, and the expression matrix of diagnostic-related genes in AD and its control group determined by the LASSO algorithm was input into the algorithm, and then the infiltration abundance of multiple immune cells in the two groups was obtained. Immune cells with significantly different abundance in the two groups were analyzed based on the abundance information. In addition, the correlation between diagnostic genes, methylation sites corresponding to diagnostic genes and immune cells was also evaluated.

2.7. Construction of the nomogram model

The construction of column line graphs is of great value for diagnosing AD. Based on the diagnosis of relevant genes, the R package "rms" was chosen to construct column line graphs. The calibration curves were then used to evaluate the nomogram model. In particular, we evaluated the clinical utility of the lineage map and predicted high-risk probability stratification in a population of 1000 individuals by applying decision and clinical impact curves, respectively.

2.8. Experimental Validation of diagnosis-related genes

This article uses q-PCR to detect the expression of ATG12, CSNK2A2, CSNK2B, MFN1, and PGAM5 genes in BV2 microglia in normal culture and Alzheimer’s disease models. Data were analyzed and graphed using GraphPad Prism 9 (Version 9.4.0). Use AI to organize and combine pictures. All data are expressed as means ± SD. Statistical differences between groups were tested using t-test. P-values less than 0.05 were considered significant differences. The primer sequences are shown in Table 1.

2.9. Statistical analysis

A t-test or non-parametric test was used to compare
Mitochondrial genes in Alzheimer's diagnosis.

3. Results
3.1. Data acquisition and processing
The overall flowchart of the paper is given in Figure 1. We first performed quality control and differential expression analysis of DNA methylation data with the ChAMP package. In the QC process, CpG sites located on sex chromosomes and CpG sites that were null in more than 70% of the samples were removed, and CpG sites in the transcription start site from 2 kb upstream to 0.5 kb downstream were retained. Then 18052 differentially methylated sites with P<0.05 were retained during differential expression analysis. The volcano maps obtained from differential analysis and the expression heatmap of differentially methylated sites in AD and its controls are presented in Figure 2A-B, respectively.

3.2. Differential expression analysis of genes and analysis of intersecting genes
We conducted differential expression analysis of transcriptome data from the GSE5281 dataset with limma package and reserved the 9805 genes with significant expression differences in AD and control groups with P<0.05. The volcano maps obtained during differential expression analysis and the expression heatmap of the top 50 genes with significant expression differences in the two groups are given in Figure 3 A-B, respectively. The intersection Wayne plots of genes, DEGs, and MRGs where differentially methylated sites are located are shown in Figure 3C. In order to explore the interaction between the nine genes, the PPI network of the nine genes was constructed (Figure 3D). In order to explore the role of the pathways involved in the development and progression of AD, we performed multiple enrichment analyses for nine genes. The histograms and bubble plots obtained from GO enrichment analysis of the nine intersecting genes are given in Figure 4A-B. Figure 4C-D provides the pathway see interaction plots and histograms of the pathways obtained from me-
3.3. Diagnostic model construction and validation based on intersection genes

In this study, the LASSO algorithm was applied to screen genes associated with AD diagnosis from nine intersecting genes. The coefficient distribution and partial likelihood deviation plots of the LASSO algorithm are displayed in Figure 5A-B. Specifically, Figure 5A shows the partial likelihood deviation of the LASSO regression plotted in 10-fold cross-validation for the log (λ) variation. Figure 5B illustrates the selection of the lambda value with the lowest cross-validation error by performing 1000 cross-validations. Finally, we screened five diagnostic-related genes (ATG12, CSNK2A2, CSNK2B, MFN1, and PGAM5). The ROC curves of the diagnostic models constructed based on the logistic regression algorithm in the training and test sets are presented in Figure 5C-D, respectively. Among them, the AUC reached 0.902 for the training set and 0.746 for the test set. In addition, we validated the performance of each diagnostic-related gene individually. The ROC curves of ATG12 in the training and test sets are given in Figure 6A and Figure 6F, with AUCs of 0.785 and 0.537 in both datasets. The ROC curves of CSNK2A2 in the training and test sets are in Figure 6B and 6G, with AUCs of 0.699 and 0.592 in both datasets. The ROC curves of CSNK2B in the training and test sets are in Figure 6C and 6H, with AUCs of 0.823 and 0.611 in both datasets. Figure 6D and Figure 6I show the ROC curves of MFN1 in the training and test sets, with AUCs of 0.676 and 0.555 in the two data sets. The ROC curves of PGAM5 in the training and test sets are in Figure 6E and 6J, with AUCs of 0.651 and 0.683 in both datasets. Finally, we also verified the methylation sites corresponding to the diagnostic-related genes using ROC curves separately in Figure 7. The AUCs of cg26233209, cg20281729, cg11628282, cg10300057, cg07377098, and cg02515730 were 0.65, 0.622, 0.638, 0.639, 0.651 and 0.683, respectively.

3.4. Exploration of the immune landscape based on diagnosis-related genes

We used the transcriptome data of the training set of AD and its controls as input based on the ssGSEA algorithm to obtain the infiltration abundance of multiple immune cells in all samples. Box line plots and heatmap of immune infiltration abundance of these immune cells in AD and its controls are offered in Figure 8A-B, respectively. In the Discussion section, we will discuss the immune cells with significantly different infiltration abundance in the two groups. Figure 8C-V provide scatter plots of the correlation between the five diagnostic genes and immune cells. Figure 8C-E present the correlation scatter plots of immune cells with significant correlation with ATG12. Figure 8F-L provide scatter plots of the correlation of immune cells with important association with CSNK2A2. Figure 8M-T shows the scatter plot of the correlation of immune cells with substantial relevance to MFN1. Figure 8U-V give the correlation scatter plots of immune cells with essential correlation with PGAM5.

3.5. Construction of nomogram model

In this section, we constructed a nomogram model (Figure 9A) based on five intersecting genes (ATH12, CSNK2B, MFN1, PGAM5, and CSNK2A2) in the training set. The nomogram model was evaluated by the calibration curve in Figure 9B, and it can be seen from the figure that...
Mitochondrial genes in Alzheimer's diagnosis.

Cell. Mol. Biol. 2024, 70(6): 114-121

The nomogram model agrees well with the ideal model. In this regard, the DCA analysis in Figure 9C indicates that the nomogram model may have better clinical utility than the individual diagnosis-related diagnosis-related genes. The clinical impact curve analysis in Figure 9D reveals that the nomogram model has higher diagnostic effectiveness. In addition, as shown in Figure 10, compared with the BV2 group, the expression level of ATG12 mRNA in the BV2 AD cell model was significantly increased. However, the expression levels of CSNK2A2, CSNK2B, MFN1, and PGAM5 mRNA in the BV2 AD cell model were significantly lower than those in the BV2 group (Figure 11A-B).

4. Discussion

AD is a severe neurodegenerative disease. Mitochondrial autophagy is a biological process that results primarily in a decrease in the number of mitochondria. However, its role in AD is unclear. DNA methylation as an epigenetic plays a key role in the development and progression of AD. To this purpose, in this study, DNA methylation and gene expression data were obtained from the GEO database for AD and its controls separately, and their differential expression analysis was performed separately. Then the genes corresponding to differentially methylated sites differentially expressed genes and mitochondrial autophagy-related genes were intersected. The intersecting genes were then analyzed for enrichment, and some of these significant pathways are closely related to the occurrence and development of AD. In the pathways obtained by GO and KEGG enrichment in Figure 4, we can obviously observe that most of them are mitochondria-related biological processes. As a matter of fact, mitochondria are the main source of energy for neurons and are degraded through mitochondrial autophagy [7]. Amyloid plaques (Aβ) and tau tangles (tau) are typical features of AD. In
previous studies, it has been demonstrated that mitochondria are the direct sites of Aβ accumulation in AD neurons [8]. Aβ/tau pathology is manifested as mitochondrial dysfunction in AD [9,10]. Whereas mitochondrial autophagy is a crucial process for neuronal health and survival, impaired mitochondrial autophagy in AD may be caused by a dysfunctional fusion between autophagosomes and lysosomes [11]. The concept of microglia is crucial in the study of the early stages of AD and may even serve as an important potential therapeutic target [12]. Autophagy and mitochondrial autophagy also seem to be altered in microglia in AD [13]. Furthermore, the majority of AD also exhibit mitochondrial dysfunction. The translocase of the outer membrane (TOM) complex is an important translocase that maintains mitochondrial function under pathophysiological conditions, and TOM70 is one of the major receptor proteins of the TOM complex. In this recent study, Cao et al. determined that TOM70 expression is closely associated with the progression of AD, which opens a new pathway to investigate the mechanisms of mitochondrial abnormalities in AD [14]. The downregulation of mitochondrial ubiquitin ligases exacerbates Aβ pathology in a toxic Aβ oligomer-dependent manner [15]. The ubiquitin ligase COP1 can suppress neuroinflammation by degrading the transcription factor CCAAT/enhancer binding protein β (c/EBPβ) in microglia [16]. In addition, targeting the interaction between Aβ and amyloid-binding alcohol dehydrogenase (ABAD) is considered a potential future therapeutic direction for the treatment of AD [17].

It was further identified by LASSO regression of several genes with diagnostic associations (ATH12, CSNK2B, MFN1, PGAM5, and CSNK2A2). CSNK2B encodes a regulatory subunit of casein kinase II, which is highly expressed in the brain and is associated with development, neurogenesis, synaptic transmission, and plasticity [18,19]. Phosphorylation of this enzyme is involved in multiple signaling pathways (e.g., PI3K / Akt, WNT / β-catenin, JAK2 / STAT3), thus regulating cellular processes such as cell survival, proliferation, differentiation, migration, and cell cycle progression [20]. All these signaling pathways are inextricably linked to AD [21-23]. Ubiquitination of the mitochondrial fusion proteins MFN1 and MFN2 promotes mitochondrial degradability and prevents damaged mitochondria from fusing [24]. The accumulation of (Mutant APP) mAPP and Aβ affects mitochondria in hippocampal neurons and leads to AD neuronal dysfunction. Reddy PH showed by experimental results that mitochondrial split gene fusions in mAPP-HT22 cells compared to wild-type (WT)-(transfected primary mouse hippocampal neurons) HT22 cells (Mfn1, Mfn2, and Opa1) gene levels were reduced [9]. Kandimalla R et al. also found reduced levels of mitochondrial fusion proteins Mfn2, Mfn1, and Opa12 in tau mice relative to age-matched WT mice in a comparative experiment between WT and tau mice, suggesting aberrant mitochondrial dynamics in tau mice [25]. PGAM5-related pathways include selective autophagy and mitochondrial autophagy. CSNK2A2 is involved in the regulation of many cellular processes, such as cell cycle progression, apoptosis, and transcription. Cell cycle abnormalities often appear early in AD, and the cell cycle can be a therapeutic target for AD [26]. The relationship between ATH12 and AD is unclear.

The diagnostic model constructed in this paper was able to achieve AUCs of 0.902 and 0.746 for the internal and external test sets, respectively. Additionally, the DNA methylation sites corresponding to the diagnosis-related genes were diagnostically meaningful. We also explored the immune landscape of AD based on the ssGSEA algorithm. It was clear that the cells in Figure 8A have significantly different infiltration abundance in AD and its control group. Here we briefly discuss the role of immune cells (Activated CD8(+) T-cells, MDSC, and Macrophage) with significant differences between the two groups in Figure 8A in the development and progression of AD. The level of activated CD8(+) T-cells has been shown to correlate with clinical and structural markers of AD pathology [27]. Myeloid-derived suppressor cells (MDSCs) are the main immunosuppressive cells and have a good suppressive effect on inflammation [28]. Le Page et al. [29] detected a significant twofold increase in circulating monocytes and granulocyte MDSCs in amnestic mild cognitive impairment (aMCI), an early stage of AD. Salminen et al. provided extensive evidence to determine the role of MDSCs in a critical role in the pathogenesis of AD [30]. Microglia are resident macrophages in the central nervous system (CNS), and their critical role in AD is self-evident [12,31]. We also determined that diagnosis-related genes and their corresponding methylation sites are significantly correlated with multiple immune cells. Finally, we constructed a diagnostic model for AD by means of a columnar line graph model, which has better clinical utility.

5. Conclusion

In conclusion, our study identified DNA methylation loci and genes associated with AD diagnosis. These genes are involved in biological pathways related to AD. Among them, ATH12, CSNK2B, MFN1, PGAM5, and CSNK2A2 have high diagnostic accuracy and excellent AD biomarker potential and deserve further attention and study.

Conflict of interests

The author has no conflicts with any step of the article preparation.

Consent for publications

The author read and approved the final manuscript for publication.

Ethics approval and consent to participate

No human or animals were used in the present research.

Informed consent

The authors declare not used any patients in this research.

Availability of data and materials

In this paper, DNA methylation data of AD and its controls (including 34 AD samples and 34 control samples) were downloaded from the GSE76105 dataset of the GEO database. Transcriptome data of AD and its controls were downloaded from the GSE5281 and GSE48350 datasets.

Author Contributions

Shan Wang conceived the structure of the article and wrote the first draft. Jiejie Zhang and Haitao Zhang made detailed revisions to the article. Yihan Yang participated in part of the data analysis, and Ya Wen provided detailed reviews and guidance for the overall article. All authors read and approved the final manuscript.
Mitochondrial genes in Alzheimer's diagnosis.

Funding
This work was supported by the funding as follows: 1. National Natural Science Foundation of China (No. 82101378). 2. Key Project of Hebei Medical Science Research (No. 20211723). 3. Construction of provincial excellent characteristic disciplines (No. 2022LCTD-B18).

References
