RESEARCH

Open Access

Identification of metabolism-related subtypes and feature genes in Alzheimer's disease

Piaopiao Lian^{1†}, Xing Cai^{2†}, Cailin Wang¹, Ke Liu¹, Xiaoman Yang¹, Yi Wu¹, Zhaoyuan Zhang¹, Zhuoran Ma¹, Xuebing Cao^{1*} and Yan Xu^{1*}

Abstract

Background Owing to the heterogeneity of Alzheimer's disease (AD), its pathogenic mechanisms are yet to be fully elucidated. Evidence suggests an important role of metabolism in the pathophysiology of AD. Herein, we identified the metabolism-related AD subtypes and feature genes.

Methods The AD datasets were obtained from the Gene Expression Omnibus database and the metabolism-relevant genes were downloaded from a previously published compilation. Consensus clustering was performed to identify the AD subclasses. The clinical characteristics, correlations with metabolic signatures, and immune infiltration of the AD subclasses were evaluated. Feature genes were screened using weighted correlation network analysis (WGCNA) and processed via Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway analyses. Furthermore, three machine-learning algorithms were used to narrow down the selection of the feature genes. Finally, we identified the diagnostic value and expression of the feature genes using the AD dataset and quantitative reverse-transcription polymerase chain reaction (qRT-PCR) analysis.

Results Three AD subclasses were identified, namely Metabolism Correlated (MC) A (MCA), MCB, and MCC subclasses. MCA contained signatures associated with high AD progression and may represent a high-risk subclass compared with the other two subclasses. MCA exhibited a high expression of genes related to glycolysis, fructose, and galactose metabolism, whereas genes associated with the citrate cycle and pyruvate metabolism were downregulated and associated with high immune infiltration. Conversely, MCB was associated with citrate cycle genes and exhibited elevated expression of immune checkpoint genes. Using WGCNA, 101 metabolic genes were identified to exhibit the strongest association with poor AD progression. Finally, the application of machine-learning algorithms enabled us to successfully identify eight feature genes, which were employed to develop a nomogram model that could bring distinct clinical benefits for patients with AD. As indicated by the AD datasets and qRT-PCR analysis, these genes were intimately associated with AD progression.

[†]Piaopiao Lian and Xing Cai have equally contributed to this work.

*Correspondence: Xuebing Cao caoxuebing@126.com Yan Xu xuyanwxf@126.com Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



Conclusion Metabolic dysfunction is associated with AD. Hypothetical molecular subclasses of AD based on metabolic genes may provide new insights for developing individualized therapy for AD. The feature genes highly correlated with AD progression included *GFAP*, *CYB5R3*, *DARS*, *KIAA0513*, *EZR*, *KCNC1*, *COLEC12*, and *TST*.

Keywords Alzheimer's disease, Metabolic subclass, Immune infiltration, Characteristic genes, Machine learning

Background

Alzheimer's disease (AD) is the most prevalent type of dementia and affect>50 million individuals worldwide [1]. The primary pathological characteristics of AD are the buildup of amyloid- β (A β) plaque and intraneuronal neurofibrillary tangle (NFT) [2]. Aß plaques occur owing to the successive enzymatic breakdown of amyloid precursor protein by β -secretase and γ -secretase [3]. Despite decades of research, the pathogenic mechanism of AD remains unclear and the current treatments are unsatisfactory let alone curative [4]. Therefore, early diagnosis and intervention are necessary for patients with AD. However, AD diagnosis has long been a challenge, and current biomarkers are inadequate to provide personalized genetic-level treatments. Thus, molecular subtypes may help identify the heterogeneity among patients with AD and facilitate the discovery of targeted therapies for AD.

Mounting evidence suggests that AD is a wide-ranging metabolic disorder characterized by disrupted glycolipid and energy metabolism. These metabolic abnormalities may contribute to the severity of AD neuropathology and the eventual manifestation of AD symptoms [5–9], thus emphasizing the crucial role of metabolism in AD and elevating the prominence metabolism dysfunction in AD research. Therefore, it is necessary to explore the metabolism-related subtypes and feature genes of AD.

In this study, we integrated eight AD datasets, including 737 patients with AD, into a single dataset for clustering analysis based on metabolic genes. Through consensus clustering, we identified three distinct subclasses of AD, which were designated as Metabolism Correlated (MC) A (MCA), MCB, and MCC subclasses. Subsequently, we evaluated the clinical characteristics, correlations with metabolic signatures, immune infiltration patterns, and prognostic implications of these AD subclasses. Weighted correlation network analysis (WGCNA) R package was employed to identify the module most associated with poor AD progression, and we performed a functional enrichment analysis of the genes associated with this module. To further narrow down the selection of the feature genes, three machinelearning algorithms were employed, including Support Vector Machines (SVM), least absolute shrinkage and selection operator (LASSO) regression, and Random Forest (RF). Thus, we successfully identified eight core genes exhibiting outstanding diagnostic potential and serving as promising therapeutic targets for AD.

Methods

Data collection and processing

The gene expression data of patients with AD were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) [10]. The following eight datasets were selected: GSE48350, GSE5281, GSE28146, GSE122063, GSE118553, GSE8442201 (GSE84422 includes three subsets and GSE8442201 was annotated by GPL570), GSE132903, and GSE106241. A detailed description of these datasets is provided in Additional file 3: Table S1. We performed data filtering, background correction, log2 transformation, and normalization of these datasets. In addition, we merged the datasets and applied a batch correction using the Combat method from the "sva" package.

Identification of AD subclasses

For consensus clustering [11], we utilized a previously published compilation of 2,752 metabolism-relevant genes [12], which encode all known human metabolic enzymes and transporters. Our aim was to classify the AD samples into distinct subclasses using consensus clustering. The maximum number of clusters was 5 and a filter was applied based on a cluster consensus score threshold of > 0.8.

Gene set variation analysis

Gene set variation analysis (GSVA) represents an unsupervised and nonparametric approach to gene set enrichment analysis that estimates the score attributed to a particular pathway or signature based on transcriptomic data [13]. We acquired 84 metabolism-relevant gene signatures from previously published study [12]. By utilizing the GSVA R package, we calculated 120 scores for each sample corresponding to these 84 metabolism signatures.

Evaluation of immune infiltration

Various algorithms are employed to assess the status of immune infiltration. The XCELL package was used to quantify the relative abundance of immune and stromal cells between the AD subclasses based on their gene expression profiles. The EPIC [14], ssGSEA [15], quan-TIseq [16], TIMER [17], CIBERSORT [18], MCPCounter

[19], XCELL [19], and ESTIMATE [20] algorithms were employed to calculate the ESTIMATE score and relative abundance of immune cells.

Weighted correlation network analysis

The WGCNA package was used to establish a WGCNA network to identify gene modules associated with the three AD subclasses and the clinical characteristics of patients with AD [21]. To determine the optimal soft-threshold power, we employed a scale-free topology standard. Subsequently, we generated a weighted adjacency matrix and transformation of a topological overlap matrix. Hierarchical clustering and tree analysis was performed to screen modules containing >50 genes. Each module was visually represented using an arbitrary color. The module eigengene represented each of the distinct modules. The traits examined in this study included the AD subclasses and several clinical features, such as NFTs and Braak.

Functional enrichment analysis

R package "clusterProfiler" [22] was used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses to identify the functions and pathways of hub genes in the cyan module.

Machine learning

Stable and robust features play a crucial role in forecasting the onset and advancement of AD. We developed three machine-learning models: RF, LASSO regression, and SVM. The RF algorithm, known for its effectiveness and popularity, utilizes a majority voting approach to combine decision trees, resulting in high precision and rapid autonomous learning across diverse datasets. The LASSO regression algorithm, a well-established linear prediction method, makes predictions based on regression coefficients and has been extensively applied in various fields [23]. The SVM algorithm, a widely used machine-learning technique, projects input data into a higher-dimensional feature space by mapping a kernel function, thus facilitating classification compared with the original feature space [24]. Through an iterative learning process, SVM converges to the optimal hyperplane that maximizes interclass span. These machine-learning models were built based on an earlier study [25].

Establishment and assessment of a nomogram

The combined dataset comprised 1262 samples, including 525 normal samples and 737 AD samples. These samples were randomly partitioned into testing (20%, N=252) and training (80%, N=1010) datasets. The feature genes were used to develop a nomogram using the "rms" package with the training set. The effectiveness of the nomogram was assessed separately for the test and training datasets. Calibration curves were employed to assess the predictive performance of the nomogram model. Finally, the clinical value of the model was assessed via decision curve analysis (DCA) and by examining the area under the curve (AUC) values.

Assessment of the diagnostic significance of feature genes in AD

To assess the discriminative capacity of the feature genes for non-AD controls and patients with AD, we used eight datasets: GSE5281, GSE48350, GSE118553, GSE28146, GSE122063, GSE132903, GSE8442201, and GSE1297. The diagnostic performance of these feature genes was visualized by plotting the AUC using the R package of "pROC".

Animals

The P301S mouse, which carries the human tau gene with the P301S mutation, is a well-characterized mouse model used to study AD. P301S transgenic mice were a gift from Professor Gang Li at the Department of Neurology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology [26]. This transgenic mouse has a C57Bl/6 J background. All transgenic and nontransgenic mice were littermates of P301S mice. The 8-month-old P301S mice (male, n=3) were used as an in vivo AD model and age-matched male C57BL/6 J mice (n=3) were used as controls. The mice were housed under standard laboratory conditions and maintained in an artificial 12/12 h light/dark cycle. Food and water were provided ad libitum. All animal experiments were reviewed and approved by the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology.

Quantitative reverse-transcription polymerase chain reaction

The cortices of the mice were surgically removed and stored at – 80 °C for subsequent biochemical analysis. The total RNA was extracted using TRIzol reagent. The mRNA was reverse transcribed to cDNA using a reverse transcription kit (Takara, Japan) according to the manufacturer's instructions. The cDNA, primers, and ChamQ SYBR qPCR Master Mix (Vazyme, China) were combined into a polymerase chain reaction (PCR) reaction plate and the mRNA levels of GFAP, CYB5R3, DARS, KIAA0513, EZR, KCNC1, COLEC12, and TST were measured using StepOnePlus real-time PCR System. All experiments were repeated thrice and the primer sequences are listed in Additional file 3: Table S2.

Statistical analysis

Statistical analyses were conducted using R language (version 4.2.0). Between-group comparisons were conducted via Wilcoxon test. A P-value of < 0.05 was considered statistically significant.

Results

Consensus clustering identifies three AD subclasses

The flowchart systematically describes our study (Fig. 1). Based on the previously reported 2,752 metabolismrelated genes, consensus clustering classified the gene expression profiles for 743 AD samples after removing the batch effect (Fig. 2A, B) into distinct subclasses. They were categorized into two to five subclasses (Additional file 1: Fig. S1). After comprehensive consideration, k=3was determined as the optimal number of clusters. When k=3, the CDF plot displayed the minimum fluctuation and the consensus matrix heatmap exhibited clear and distinct boundaries (Fig. 2C, D). Both principal component analysis (PCA) and a metabolism-associated genes expression heatmap unveiled significant discrepancies in the expression profiles between the three subclasses (Fig. 2E, F).

Clinical characteristics of the AD subclasses

The gamma-secretase activity in MCA was notably higher compared with that in MCB and MCC (P < 0.001or 0.05; Fig. 3A). Compared with MCB, beta-secretase activity, NFTs, and Braak were elevated in MCA and MCC (P < 0.01; Fig. 3B, D, E). The PH in MCC was higher compared with that in MCA and MCB (P < 0.05; Fig. 3F). The three AD subclasses contained a greater proportion of women than men (Fig. 3H). Furthermore, the proportion of one and two APOE 4 alleles was significantly higher in MCA compared with that in MCB and MCC (Fig. 3I). With respect to age and alpha-secretase activity, there was no difference between the three AD subclasses (Fig. 3C, G). The tissue origin of MCA, MCB, and MCC is shown in Additional file 2: Fig. S2.

Association between the AD subclasses and metabolism-associated signatures

Given that the AD subclasses were established based on metabolism genes, we investigated whether the different subclasses exhibited varying metabolic signatures. Initially, 84 metabolism processes were measured utilizing the "GSVA" R package. Next, we performed a differential analysis to identify the subclass-specific metabolic signatures, which were identified as signatures with a greater GSVA score in the relevant subclasses. The results indicated that only MCA and MCB exhibited distinct metabolism signatures of 40 and 30, respectively, whereas MCC exhibited negligible distinct metabolism signatures. Notably, 7 of the 40 distinct metabolism signatures in MCA were associated with carbohydrate metabolism (Fig. 4).

MCA was primarily associated with gene signatures for carbohydrate and lipid metabolism, with carbohydrate metabolism primarily comprising genes related to glycolysis, fructose, mannose, and galactose metabolism, whereas the genes related to citrate cycle and pyruvate metabolism were downregulated compared with the other two subclasses. Lipid metabolism in MCA mainly included fatty acid degradation. MCB was primarily associated with amino acid biosynthesis, nucleotides biosynthesis, and the citric acid cycle.

Association between the AD subclasses and immune infiltration

To determine the characteristics of the AD subclasses, the ESTIMATE algorithm was applied to calculate the immune and stromal scores. The immune scores displayed a marked difference across the three groups, whereby MCA demonstrated a higher immune score compared with MCB and MCC (P<0.0001; Fig. 5B). Furthermore, MCA exhibited a higher stromal score compared with those of MCB and MCC (P < 0.0001; Fig. 5B). Owing to the observed difference in the immune scores among the AD subclasses, immune infiltration was further examined to characterize the immunological landscape. We quantified the abundance of 24 microenvironment cell and analyzed the samples for the expression of immune checkpoints (Fig. 5A). Compared with other subclasses, we observed higher expression of several immune checkpoint genes in MCB, which may serve as targets for immunotherapy, including CD274 (PDL1) and PDCD1 (PDL2; Fig. 5C). In addition, MCA exhibited higher abundance of 18 immune cell populations (regulatory T cells, CD4+T cells, nature B cells, memory B cells, activated dendritic cells, M1 macrophages, activated natural killer cells, memory CD4+T cells, activated mast cells, resting natural killer cells, M0 macrophages, M2 macrophages, eosinophils, resting dendritic cells, resting mast cells, neutrophils, endothelial cells, and fibroblasts) compared with MCB or MCC (Fig. 5D). Notably, MCA demonstrated a higher infiltration of endothelial cells and fibroblasts (Fig. 5D). Therefore, we quantified the various types of cancer-associated fibroblasts (CAFs) and observed that MCA exhibited an enrichment of all the distinct subtypes of fibroblasts. Furthermore, MCA exhibited a depletion of normal fibroblasts (Fig. 5E).

WGCNA to identify poor AD progression-associated module and hub genes

We conducted WGCNA using the merged dataset to identify the module associated with poor AD progression.



Fig. 1 Flowchart of the research



Fig. 2 Identification of the AD subclasses. A, B Principal component analysis (PCA) of eight datasets before A and after B batch effect removal. C Consensus clustering utilizing metabolism-associated genes. CDF curve for k=2-5 is shown. D Consensus clustering matrix when k=3 E PCA supported stratification into three AD subclasses. F Metabolism-associated genes expression heatmap for the three AD subclasses. Red indicates high expression. Blue indicates low expression. White indicates no difference



Fig. 3 Clinical characteristics of the AD subclasses. **A** Comparison of gamma-secretase activity between each subclass. **B** Comparison of beta-secretase activity between each subclass. **C** Comparison of alpha-secretase activity between each subclass. **D** Comparison of Braak between each subclass. **E** Comparison of NFTs between each subclass. **F** Comparison of pH between each subclass. **G** Comparison of age between each subclass. **H** Proportion of sex between each subclass. **I** Proportion of APOE4 alleles in each subclass. (0 indicates no APOE4 allele, 1 represents one APOE4 allele, and 2 represents two APOE4 alleles). The clinical data of Fig. **A–C**, H is from GSE106241 (MCA, n = 16; MCB, n = 25; MCC, n = 19). The clinical data of Fig. **E**, **F** is from GSE8442201 (MCA, n = 7; MCB, n = 12; MCC, n = 15). The clinical data of Fig. **D** is from GSE106241 and GSE8442201 (MCA, n = 37; MCC, n = 34). The clinical data of Fig. **G**, **I** are from eight merged datasets (MCA, n = 173; MCB, n = 274; MCC, n = 350). Kruskal–Wallis test was used to compare statistical differences

When the soft-threshold was 4, the scale-free network and connectivity exhibited maximum efficiency (Fig. 6A). Using a hierarchical clustering algorithm, the clustering tree was classified into six gene modules, each of which assigned a unique color (Fig. 6B). Of these, the cyan module comprised 3284 genes and exhibited the strongest positive correlation with MCA (R=0.49) as well as a series of AD-related high-risk indicators, including NFTs (R=0.51), Braak (R=0.32), gamma-secretase activity (R=0.3), amyloid-beta 42 (R=0.28), and alpha-secretase activity (R=0.25) (Fig. 6C). Therefore, the cyan module was chosen as the hub module from which hub genes were extracted using the selection criteria cor.MM > 0.7 and cor.GS > 0.4 (Fig. 6D). In addition, we performed GO and KEGG enrichment analyses using the aforementioned hub genes (Fig. 6E, F). KEGG enrichment analysis revealed that various synapses, including GABAergic, glutamatergic, and dopaminergic synapses as well as synaptic transmission–related signaling pathways, including the calcium, adrenergic, and synaptic vesicle cycle



Fig. 4 Association between metabolism-associated signatures and the AD subclasses. A Heatmap of metabolism-associated signatures. B Boxplot of the signature score for the metabolism-associated signatures distinguished by different subclasses (ns indicates no significance, *P < 0.05, **P < 0.01, ***P < 0.005, ***P < 0.005, ***P < 0.0001)



Fig. 5 Association between AD subclasses and immune infiltration. **A** Heatmap describing the immune infiltration landscape in the three AD subclasses. Boxplots describing the distribution of expression for **B** the immune and stromal score, **B** immune checkpoint targets, **C** TME cells, and **D** fibroblast signatures in the three AD subclasses. (ns indicates no significance, *P < 0.05, **P < 0.01, ***P < 0.005, ***P < 0.001)



Fig. 6 Identification of AD feature genes and functional enrichment analysis A Analysis of the scale-free index and the mean connectivity for various soft-threshold powers. B Identification of coexpression gene modules. The branches of the dendrogram clustered into six modules, each labeled with a unique color. C Heatmap showing the correlation between modules and feature gene sets. The framed cyan module was the key module most relevant to the MCA subclass. Gene significance and corresponding p values were calculated and are shown in the heatmap. D Scatter diagram showing the correlation of gene significance in cyan module with gene significance for MCA subclass. E, F Dot-plots showing the KEGG E and GO F enrichment analysis of hub genes

signaling pathways, were closely associated with these hub genes (Fig. 6E, Additional file 3: Table S3). GO enrichment analysis revealed that these hub genes were predominantly enriched in cell morphogenesis regulation, actin filament organization, actin filament bundle assembly, actin filament bundle organization, and cell– matrix adhesion (Fig. 6F, Additional file 3: Table S4). These results indicate the important functions of these genes.

Selection of the AD feature genes based on hub genes of cyan module

We conducted three different machine-learning algorithms to screen for potential AD biomarkers. Using the LASSO regression algorithm, the hub genes were narrowed down to 51 variables (Fig. 7A, B). Using the SVM-REF algorithm, we identified a subset of 86 features among the hub genes (Fig. 7C, D). The RF algorithm revealed the top 20 feature genes (Fig. 7E, F). The overlapping genes among the LASSO, RF, and SVM-REF algorithms (*GFAP, CYB5R3, PMP2, DARS, KIAA0513, ITGB8, ENAH, EZR, RIN2, KCNC1, FOXO1, COLEC12, TST, AKR1C3, TSPO,* and *ANTXR2*) were selected for further study (Fig. 7G). Finally, we used logistic regression to out 8 feature genes (*GFAP, CYB5R3, DARS, KIAA0513, EZR, KCNC1, COLEC12,* and *TST*; p<0.05) from the above 16 overlapping genes.

Development and validation of the feature genes diagnostic signature for AD

A nomogram model was developed for AD diagnosis utilizing the eight feature genes (GFAP, CYB5R3, DARS, KIAA0513, EZR, KCNC1, COLEC12, and TST) (Fig. 8A). A calibration curve was used to assess the predictive capabilities of the nomogram model in the training and testing datasets. The calibration curve revealed a small error between the actual and predicted risk for AD, suggesting a high accuracy of the nomogram model for predicting AD (Fig. 8B). DCA revealed that the "nomogram" curve was higher than the curves representing "intervention for none," "intervention for all," and all single genes, suggesting that the patients may benefit from the nomogram model at a high-risk threshold from 0 to 1, and the clinical benefit of the nomogram model was higher compared with that of the single gene curve (Fig. 8C). Subsequently, the receiver operating characteristic (ROC) curve analysis was employed to evaluate the diagnostic capability of each feature gene for predicting AD progression in the internal datasets. The AUC values in the training dataset were 0.788 for the nomogram model, 0.729 for GFAP, 0.692 for EZR, 0.656 for COLEC12, 0.652 for KIAA0513, 0.698 for CYB5R3, 0.560 for DARS, 0.558 for KCNC1, and 0.557 for TST (Fig. 8D). The AUC values for the ROC curves in the testing set were 0.770 for nomogram model, 0.708 for GFAP, 0.698 for EZR, 0.677 for COLEC12, 0.692 for KIAA0513, 0.575 for CYB5R3, 0.566 for DARS, 0.566 for KCNC1, and 0.584 for TST (Fig. 8D). In addition, eight single validation datasets (GSE5281, GSE48350, GSE118553, GSE28146, GSE122063, GSE132903, GSE8442201, GSE28146, and GSE1297) were used to further confirm the diagnostic efficacy of these eight feature genes (Fig. 8E–L). To some extent, these results also suggest that the eight genes have a significant role in AD pathogenesis.

Validation of the feature genes expression

The differential expressions of the feature genes were verified in the aforementioned combined dataset (including GSE48350, GSE5281, GSE28146, GSE122063, GSE118553, GSE8442201, GSE132903, and GSE106241), which further demonstrated their diagnostic capacity for AD (Fig. 9B). In addition to the dataset, we further verified the expression of these eight feature genes by qRT-PCR analysis using tissues collected from AD mice or controls. Consistent with the bioinformatics analysis results, the expression of GFAP, CYB5R3, DARS, EZR, COLEC12, and TST were significantly higher in AD mice compared with controls, whereas KIAA0513 exhibited significant downregulation (Fig. 9A). In contrast, KCNC1 expression was not statistically different between the AD and control groups.

Discussion

AD is a neurodegenerative disease wherein A β and NFT aggregation causes the loss of synapses, neuronal death, and subsequent memory impairment. There is a large heterogeneity in AD pathogenesis among patients, and thus, AD progression biomarkers need to be further refined [27, 28]. Accordingly, suitable AD subtypes and more powerful biomarkers are necessary for improved diagnosis and therapy.

Accumulating evidence suggested that the occurrence and progression of AD is closely related to substance and energy metabolism. Glucose, lipids, and energy metabolism has an important impact on AD [29-31]. The energy of the brain is primarily dependent upon glucose, which is metabolized to ATP via glycolysis, tricarboxylic acid (TCA) cycle, and electron transport chain [32]. Glucose metabolism is markedly decreased in the AD brain. Attenuated ATP production due to inefficient glucose utilization is accompanied by signal transduction breakdown, ionic pump dysfunction, and neurotransmission failure, ultimately leading to neuronal degeneration and death [29]. Lipids are also involved in AD pathology [33]. Apolipoprotein E ε 4 (APOE4) is the strongest genetic risk factor for AD and



Fig. 7 Screen for potential AD biomarkers. **A**, **B** Genes identified by the LASSO algorithm. **C**, **D** Optimal biomarker selection using the SVM algorithm. **E**, **F** The top 20 genes selected by the RF algorithm. **G** Venn diagram showing the diagnostic markers intersected by LASSO, SVM, and RF algorithm

drives metabolic dysregulation in astrocytes and microglia, leading to cholesterol accumulation, decreased neuronal excitability, and neuroinflammation [34, 35]. Restoring metabolic homeostasis can exert a significant neuroprotective effect [36]. Despite evidence implicating disrupted metabolism as pathological mechanism underlying AD, the precise genes and biological functions are yet to be identified, particularly the role of metabolism in regulating AD immunity.



Fig. 8 Validation of the diagnostic efficacy of the feature genes. **A** Nomogram displaying the predicted risk for AD based on the feature genes. **B** Calibration curve showing the predicted performance of the nomogram in internal datasets including the training and testing datasets. **C** DCA showing the clinical benefits of the nomogram in the internal datasets, including the training and testing datasets. **D** ROC curves showing the diagnostic performance of the feature genes in the internal datasets, including the training and testing datasets. **E**–**L** ROC curves showing the diagnostic performance of genes identified in the eight datasets, including GSE5281 **E**, GSE48350 **F**, GSE118553 **G**, GSE122063 **H**, GSE132903 **I**, GSE8442201 **J**, GSE28146 **K**, and GSE1297 **L**. (ns indicates no significance, *P<0.05, **P<0.01, ***P<0.005, ***P<0.0001)



Fig. 9 Validation of the eight feature genes expression between control and AD groups in mice **A** and the GEO datasets **B**. **A** The relative expression of eight genes was measured via qRT-PCR using mouse tissues (n = 3 in the control mice; n = 3 in the AD mice). **B** Expression levels of the eight genes in the GEO datasets. (ns indicates no significance, *P < 0.05, **P < 0.01, ***P < 0.005, ****P < 0.0001)

In this study, to identify AD subclasses associated with metabolic processes, an AD classification was built based on metabolic genes from previous publications. Three distinct AD subclasses (MCA, MCB, and MCC) were identified. We explored the clinical features, metabolic signatures, and immune infiltration profile of each subclass. The results indicated that MCA exhibited specific metabolic signatures and was accompanied by high AD progression signatures (β -secretase activity, γ -secretase activity, NFT, Braak, and the AD-risk gene APOE4).

MCA was primarily associated with carbohydrate and lipid metabolism genes. The carbohydrate metabolism in MCA primarily involves glycolysis, fructose, mannose, and galactose metabolism, whereas the citrate cycle and pyruvate metabolism were decreased compared with the other two subclasses, indicating a reduction in the TCA cycle and glucose utilization (thereby reducing ATP production). Meanwhile, lipid metabolism in MCA mainly involves fatty acid degradation, probably due to low ATP production, which prompts a shift in energy metabolism to the ketogenic pathway. These metabolic disorders affect the energy supply of neurons in the brain. Furthermore, previous studies confirmed that mitochondrial ATP-synthase α subunit is lipoxidized and ATP-synthase activity was obviously reduced in the entorhinal cortex of patients with AD compared with the controls [37]. An analysis of the clinical features and metabolic signatures revealed that high APOE4 expression, NFT accumulation, and significant metabolic disorders were observed in the MCA subclass, thus presenting a poorer prognosis. Immune infiltration analysis suggested that MCA had an augmented immune score and a relatively higher abundance of immune cell infiltration compared with MCB and MCC. A significant change in the immune cell ratio was observed in the AD subclasses in which MCA exhibited higher levels of regulatory T cells (Tregs), CD4+T cells, memory CD4+T cells, B cells, activated dendritic cells, macrophages, and neutrophils compared with MCB and MCC, consistent with the findings of previous studies [38-40]. In addition, MCA exhibited a high stromal score and infiltration with endothelial cells and fibroblasts. Immune checkpoint genes that represent the potential targets for immunotherapy, such as CD274 (PDL1) and PDCD1 (PDL2), were primarily increased in the MCB.

To further elucidate the genomics characteristics of the AD subclasses, we used a combined dataset to the construct coexpression networks via WGCNA. The cyan module was positively correlated with MCA and the "A/T/N" system, such as NFTs, further supporting our hypothesis that the MCA subclass is a high-risk subclass for AD. Functional enrichment analysis revealed that the hub genes in the cyan module were primarily enriched in cellular morphological regulation and synapse-related functions and pathways. The impaired TCA cycle in the MCA is the main function of the mitochondria. These metabolic disorders may lead to mitochondrial dysfunction, inadequate energy supply, and massive reactive oxygen species release, inducing oxidative stress and calcium regulation imbalance, ultimately triggering neuronal apoptosis and synaptic loss [8].

Recently, various machine-learning algorithms have been used to identify new biomarkers and offer insights into disease pathogenesis, owing to an outstanding performance in diagnosis [41, 42]. Therefore, we used three machine-learning algorithms to further narrow down the number of hub genes. Eight feature genes were finally identified, including GFAP, CYB5R3, DARS, KIAA0513, EZR, KCNC1, COLEC12, and TST. GFAP is an astrogliosis marker. Recently, Shen et al. reported that plasma GFAP is significantly elevated from the preclinical stage of AD and is a promising diagnostic and predictive biomarker that distinguishes AD from the controls and non-AD dementia [43]. CYB5R3 encodes cytochrome b5 reductase 3, which is essential for reductive reactions, such as cholesterol biosynthesis, fatty acid elongation, methemoglobin reduction, and drug metabolism [44]. CYB5R3 expression was elevated in the human cortex in an AD proteomics study [45]. As an aspartyl-tRNA

synthetase, DARS missense mutations caused a significant pattern of hypomyelination, motor abnormalities, and cognitive impairment [46]. A bioinformatics analysis suggested that KIAA0513 reduction serves as a potential biomarker for early AD diagnosis [47]. EZR, which is a member of the ezrin-radixin-moesin protein family, has been recognized as a regulator of the adhesion signal pathways. EZR plays a key role in promoting the invasion and metastasis of malignant tumors [48]. KCNC1 encodes a subunit of the Kv3 voltage-gated potassium channels and is associated with various human diseases, including ataxia, epilepsy, and developmental delay [49]. COLEC12 encodes a member of the C-lectin family, which is a scavenger receptor that plays a crucial role in the binding and clearance of A β [50]. *TST* is an enzyme that is widely distributed in both prokaryotes and eukaryotes, which plays a crucial role in mitochondrial function [51]. These along with our findings are concordant and indicate that the overexpression of GFAP, CYB5R3, DARS, EZR, COLEC12, and TST as well as the downregulation of KIAA0513 and KCNC1 can predict poor AD prognosis. In addition, the nomogram model, calibration curves, DCA, and ROC curves verified the satisfactory diagnostic ability of these eight feature genes.

To the best of our knowledge, this was the first study to classify ADs from the perspective of metabolism. The screening and validation of the feature genes provided potential molecular targets for further exploring the metabolic mechanism of AD. However, this study had some limitations. First, the feature genes were only validated in AD mice and supporting human samples were lacking. Second, *KCNC1* showed inconsistent results in the AD datasets and AD mice, possibly due to the small mice sample size. Finally, the mechanism underlying metabolism regulation in AD warrants further investigated in vitro and in vivo, which will be our focus in future studies.

Conclusion

We found a strong relationship between the metabolic status and AD pathogenesis using a comprehensive bioinformatics analysis. Three AD subclasses from the perspective of metabolism were identified with substantial differences in clinical characteristics, metabolism signatures, and immune infiltration. The results can better elucidate the heterogeneity of patients with AD. In addition, we identified and verified eight feature genes, *GFAP, CYB5R3, DARS, EZR, COLEC12,* and *TST,* which showed high expression, whereas *KIAA0513* and *KCNC1* displayed showed low expression in AD. The diagnostic model built by these eight genes exhibited outstanding diagnostic value. These findings provide a basis for more accurate and early AD diagnosis.

Abbreviations

AD	Alzheimer's disease
Αβ	Amyloid-β
NFT	Neurofibrillary tangles
WGCNA	Weighted correlation network analysis
SVM	Support vector machines
RF	Random forest
GEO	Gene expression omnibus
GSVA	Gene set variation analysis
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes
AUC	Area under the curve
qRT-PCR	Quantitative reverse-transcription polymerase chain reaction
PCA	Principal component analysis
DCA	Decision curve analysis
TCA	Tricarboxylic acid
APOE4	Apolipoprotein Ε ε4

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-023-04324-y.

Additional file 1: Figure S1. Consensus clustering matrix for k = 2-5.

Additional file 2: Figure S2. Box-plot of tissue original of AD subclasses.

Additional file 3: Table S1. GEO datasets information. Table S2. Primer sequences of feature genes. Table S3. KEGG pathway enrichment analyses of hub genes in cyan module. Table S4. GO enrichment analyses of hub genes in cyan module.

Acknowledgements

We thank all contributors to the GEO database.

Author contributions

PL and XC were contributed to the conception and design of the study, processing of bioinformatics data, experimental validation and writing of the manuscript. CW was responsible for animal modeling. KL was responsible for the statistical work towards experimental data. YX and XC supervised the whole analysis and provided guidance and instructions. All authors contributed to the manuscript revision and approved the submitted version.

Funding

This study was supported by grant from the National Natural Science Foundation of China (NSFC Project, No. 81873734 and 81974200).

Availability of data and materials

The datasets analysed during the current study are available in the GEO database (https:// www. ncbi. nlm. nih. gov/ geo/), openly available for free download.

Declarations

Ethics approval and consent to participate

All animal experiments were reviewed and approved by the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology (ACUC Number:3121).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Neurology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. ²Department of Oncology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.

Received: 19 May 2023 Accepted: 1 July 2023 Published online: 15 September 2023

References

- 1. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022; 7:105–125.
- DeTure MA, Dickson DW. The neuropathological diagnosis of Alzheimer's disease. Mol Neurodegener. 2019;14:32.
- 3. Yang Y, Arseni D, Zhang W, Huang M, Lövestam S, Schweighauser M, Kotecha A, Murzin AG, Peak-Chew SY, Macdonald J, et al. Cryo-EM structures of amyloid- β 42 filaments from human brains. Science. 2022;375:167–72.
- 4. Hodson R. Alzheimer's disease. Nature. 2018;559:S1.
- 5. Poddar MK, Banerjee S, Chakraborty A, Dutta D. Metabolic disorder in Alzheimer's disease. Metab Brain Dis. 2021;36:781–813.
- Kuehn BM. In Alzheimer research, glucose metabolism moves to center stage. JAMA. 2020;323:297–9.
- Yu L, Jin J, Xu Y, Zhu X. Aberrant energy metabolism in Alzheimer's disease. J Transl Int Med. 2022;10:197–206.
- Peng Y, Gao P, Shi L, Chen L, Liu J, Long J. Central and peripheral metabolic defects contribute to the pathogenesis of Alzheimer's disease: targeting mitochondria for diagnosis and prevention. Antioxid Redox Signal. 2020;32:1188–236.
- Varma VR, Oommen AM, Varma S, Casanova R, An Y, Andrews RM, O'Brien R, Pletnikova O, Troncoso JC, Toledo J, et al. Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: a targeted metabolomics study. PLoS Med. 2018;15:e1002482.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41:D991-995.
- Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26:1572–3.
- Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, Sethumadhavan S, Woo HK, Jang HG, Jha AK, et al. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature. 2011;476:346–50.
- Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
- 14. Racle J, Gfeller D. EPIC: a tool to estimate the proportions of different cell types from bulk gene expression data. Methods Mol Biol. 2020;2120:233–48.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.
- Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, Krogsdam A, Loncova Z, Posch W, Wilflingseder D, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. Genome Med. 2019;11:34.
- 17. Li B, Liu JS, Liu XS. Revisit linear regression-based deconvolution methods for tumor gene expression data. Genome Biol. 2017;18:127.
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12:453–7.
- Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17:218.
- Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.
- 21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
- 22. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.

- 23. Motamedi F, Pérez-Sánchez H, Mehridehnavi A, Fassihi A, Ghasemi F. Accelerating big data analysis through LASSO-random forest algorithm in QSAR studies. Bioinformatics. 2022;38:469–75.
- Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak. 2019;19:281.
- Lai Y, Lin X, Lin C, Lin X, Chen Z, Zhang L. Identification of endoplasmic reticulum stress-associated genes and subtypes for prediction of Alzheimer's disease based on interpretable machine learning. Front Pharmacol. 2022;13:975774.
- Chang Y, Yao Y, Ma R, Wang Z, Hu J, Wu Y, Jiang X, Li L, Li G. Corrigendum: DI-3-n-butylphthalide reduces cognitive deficits and alleviates neuropathology in p301s tau transgenic mice. Front Neurosci. 2021;15:716049.
- Duara R, Barker W. Heterogeneity in Alzheimer's disease diagnosis and progression rates: implications for therapeutic trials. Neurotherapeutics. 2022;19:8–25.
- Cano A, Turowski P, Ettcheto M, Duskey JT, Tosi G, Sánchez-López E, García ML, Camins A, Souto EB, Ruiz A, et al. Nanomedicine-based technologies and novel biomarkers for the diagnosis and treatment of Alzheimer's disease: from current to future challenges. J Nanobiotechnology. 2021;19:122.
- Butterfield DA, Halliwell B. Oxidative stress, dysfunctional glucose metabolism and Alzheimer disease. Nat Rev Neurosci. 2019;20:148–60.
- Arnold M, Nho K, Kueider-Paisley A, Massaro T, Huynh K, Brauner B, MahmoudianDehkordi S, Louie G, Moseley MA, Thompson JW, et al. Sex and APOE ε4 genotype modify the Alzheimer's disease serum metabolome. Nat Commun. 2020;11:1148.
- Zhang X, Tong T, Chang A, Ang TFA, Tao Q, Auerbach S, Devine S, Qiu WQ, Mez J, Massaro J, et al. Midlife lipid and glucose levels are associated with Alzheimer's disease. Alzheimers Dement. 2023;19:181–93.
- Cunnane SC, Trushina E, Morland C, Prigione A, Casadesus G, Andrews ZB, Beal MF, Bergersen LH, Brinton RD, de la Monte S, et al. Brain energy rescue: an emerging therapeutic concept for neurodegenerative disorders of ageing. Nat Rev Drug Discov. 2020;19:609–33.
- Markesbery WR, Kryscio RJ, Lovell MA, Morrow JD. Lipid peroxidation is an early event in the brain in amnestic mild cognitive impairment. Ann Neurol. 2005;58:730–5.
- Tcw J, Qian L, Pipalia NH, Chao MJ, Liang SA, Shi Y, Jain BR, Bertelsen SE, Kapoor M, Marcora E, et al. Cholesterol and matrisome pathways dysregulated in astrocytes and microglia. Cell. 2022;185:2213-2233.e2225.
- Victor MB, Leary N, Luna X, Meharena HS, Scannail AN, Bozzelli PL, Samaan G, Murdock MH, von Maydell D, Effenberger AH, et al. Lipid accumulation induced by APOE4 impairs microglial surveillance of neuronalnetwork activity. Cell Stem Cell. 2022;29:1197-1212.e1198.
- Zheng J, Xie Y, Ren L, Qi L, Wu L, Pan X, Zhou J, Chen Z, Liu L. GLP-1 improves the supportive ability of astrocytes to neurons by promoting aerobic glycolysis in Alzheimer's disease. Mol Metab. 2021;47:101180.
- Terni B, Boada J, Portero-Otin M, Pamplona R, Ferrer I. Mitochondrial ATPsynthase in the entorhinal cortex is a target of oxidative stress at stages I/ II of Alzheimer's disease pathology. Brain Pathol. 2010;20:222–33.
- Saresella M, Calabrese E, Marventano I, Piancone F, Gatti A, Alberoni M, Nemni R, Clerici M. Increased activity of Th-17 and Th-9 lymphocytes and a skewing of the post-thymic differentiation pathway are seen in Alzheimer's disease. Brain Behav Immun. 2011;25:539–47.
- Song L, Yang YT, Guo Q, Zhao XM. Cellular transcriptional alterations of peripheral blood in Alzheimer's disease. BMC Med. 2022;20:266.
- Kim K, Wang X, Ragonnaud E, Bodogai M, Illouz T, DeLuca M, McDevitt RA, Gusev F, Okun E, Rogaev E, Biragyn A. Therapeutic B-cell depletion reverses progression of Alzheimer's disease. Nat Commun. 2021;12:2185.
- 41. Lai Y, Lin P, Lin F, Chen M, Lin C, Lin X, Wu L, Zheng M, Chen J. Identification of immune microenvironment subtypes and signature genes for Alzheimer's disease diagnosis and risk prediction based on explainable machine learning. Front Immunol. 2022;13:1046410.
- 42. Li J, Zhang Y, Lu T, Liang R, Wu Z, Liu M, Qin L, Chen H, Yan X, Deng S, et al. Identification of diagnostic genes for both Alzheimer's disease and Metabolic syndrome by the machine learning algorithm. Front Immunol. 2022;13:1037318.
- Shen XN, Huang SY, Cui M, Zhao QH, Guo Y, Huang YY, Zhang W, Ma YH, Chen SD, Zhang YR, et al. Plasma glial fibrillary acidic protein in the Alzheimer disease continuum: relationship to other biomarkers,

differential diagnosis, and prediction of clinical progression. Clin Chem. 2023;69:411–21.

- 44. Rahaman MM, Reinders FG, Koes D, Nguyen AT, Mutchler SM, Sparacino-Watkins C, Alvarez RA, Miller MP, Cheng D, Chen BB, et al. Structure guided chemical modifications of propylthiouracil reveal novel small molecule inhibitors of cytochrome b5 reductase 3 that increase nitric oxide bioavailability. J Biol Chem. 2015;290:16861–72.
- 45. Wang H, Dey KK, Chen PC, Li Y, Niu M, Cho JH, Wang X, Bai B, Jiao Y, Chepyala SR, et al. Integrated analysis of ultra-deep proteomes in cortex, cerebrospinal fluid and serum reveals a mitochondrial signature in Alzheimer's disease. Mol Neurodegener. 2020;15:43.
- 46. Fröhlich D, Suchowerska AK, Voss C, He R, Wolvetang E, von Jonquieres G, Simons C, Fath T, Housley GD, Klugmann M. Expression pattern of the aspartyl-tRNA synthetase DARS in the human brain. Front Mol Neurosci. 2018;11:81.
- 47. Zhu M, Jia L, Li F, Jia J. Identification of KIAA0513 and other hub genes associated with Alzheimer disease using weighted gene Coexpression network analysis. Front Genet. 2020;11:981.
- Xu J, Zhang W. EZR promotes pancreatic cancer proliferation and metastasis by activating FAK/AKT signaling pathway. Cancer Cell Int. 2021;21:521.
- Li X, Zheng Y, Li S, Nair U, Sun C, Zhao C, Lu J, Zhang VW, Maljevic S, Petrou S, Lin J. Kv3.1 Channelopathy: a novel loss-of-function variant and the mechanistic basis of its clinical phenotypes. Ann Transl Med. 2021;9:1397.
- Nakamura K, Ohya W, Funakoshi H, Sakaguchi G, Kato A, Takeda M, Kudo T, Nakamura T. Possible role of scavenger receptor SRCL in the clearance of amyloid-beta in Alzheimer's disease. J Neurosci Res. 2006;84:874–90.
- Buonvino S, Arciero I, Melino S. Thiosulfate-cyanide sulfurtransferase a mitochondrial essential enzyme: from cell metabolism to the biotechnological applications. Int J Mol Sci. 2022. https://doi.org/10.3390/ijms2 3158452.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

