Identification of biomarkers of Alzheimer's disease and their characterization of immune function

Keywords
Biomarker, Alzheimer's disease (AD), Machine learning algorithm, CD8 T cell, Trajectory analysis

Abstract
Introduction
Alzheimer's disease (AD) is a neurodegenerative disease with neurogenic fiber tangles caused by amyloid-β protein plaques and tau protein hyperphosphorylation as the pathological manifestations. This study was based on multi-omics to investigate the mechanisms and immune characterization of AD.

Material and methods
Based on bulk RNA-seq (GSE122063 and GSE97760), we screened potential biomarkers for AD by differential expression analysis and machine learning algorithms. Then, we elaborated the expression characteristics and immune functions of the above biomarkers by scRNA-seq (single-cell RNA sequencing) data analysis (GSM4996463 and GSM4996462) and immune infiltration analysis.

Results
Five biomarkers (RBM3, GOLGA8A, ALS2, FSD2, and LOC100287628) were identified using machine learning algorithms. Single-cell analysis revealed distinct expression patterns of these biomarkers in astrocytes from AD samples compared to normal samples. Additionally, three key biomarkers were selected based on interaction networks, and the diagnostic models indicated high diagnostic efficacy for these biomarkers. Based on immune infiltration and correlation analyses, RBM3, GOLGA8A, and ALS2 were all highly correlated with CD8 T cell content in the immune microenvironment of AD.

Conclusions
The biomarkers identified in this study demonstrate significant diagnostic potential for AD. Notably, the downregulation of RBM3 in astrocytes and the decreased presence of CD8 T cells infiltrating brain tissue are potential risk factors for AD.
Identification of biomarkers of Alzheimer's disease and their characterization of immune function

**Running title:** Identification of biomarkers of Alzheimer's disease
Abstract

Introduction: Alzheimer's disease (AD) is a neurodegenerative disease with neurogenic fiber tangles caused by amyloid-β protein plaques and tau protein hyperphosphorylation as the pathological manifestations. This study was based on multi-omics to investigate the mechanisms and immune characterization of AD.

Material and methods: Based on bulk RNA-seq (GSE122063 and GSE97760), we screened potential biomarkers for AD by differential expression analysis and machine learning algorithms. Then, we elaborated the expression characteristics and immune functions of the above biomarkers by scRNA-seq (single-cell RNA sequencing) data analysis (GSM4996463 and GSM4996462) and immune infiltration analysis.

Results: Five biomarkers (RBM3, GOLGA8A, ALS2, FSD2, and LOC100287628) were identified using machine learning algorithms. Single-cell analysis revealed distinct expression patterns of these biomarkers in astrocytes from AD samples compared to normal samples. Additionally, three key biomarkers were selected based on interaction networks, and the diagnostic models indicated high diagnostic efficacy for these biomarkers. Based on immune infiltration and correlation analyses, RBM3, GOLGA8A, and ALS2 were all highly correlated with CD8 T cell content in the immune microenvironment of AD.

Conclusions: The biomarkers identified in this study demonstrate significant diagnostic potential for AD. Notably, the downregulation of RBM3 in astrocytes and the decreased presence of CD8 T cells infiltrating brain tissue are potential risk factors for AD.

Keywords: Alzheimer's disease (AD), Biomarker, CD8 T cell, Machine learning algorithm, Trajectory analysis

Introduction

Alzheimer's disease (AD) is a neurodegenerative disease with amyloid beta protein plaques and tau protein hyperphosphorylation causing neurogenic fiber tangles as the main pathological manifestation [1]. It is divided into two main categories, “sporadic
AD (typical AD)” and “atypical AD”. Sporadic AD is common in the elderly and most typically presents clinically as a slowly progressive amnesic disorder that can eventually evolve into a multi-domain dementia in which amnesia predominates. The non-amnesic phenotype, on the other hand, is characterized mainly by dysfunction in the visual, language, executive, behavioral and motor domains and is most often seen in younger adults [2]. There are a number of hypotheses that explain the pathogenesis of AD, however Grobler et al. noted that the most dominant of these pathophysiologic hypotheses include the amyloid-β (Aβ), inflammatory, vascular, and infectious hypotheses [3]. Research supports that myelin damage may even predate Aβ lesions in AD [4]. Increasing evidence suggested that the peripheral anion site of AChE contributes to Aβ aggregation and neurotoxicity in AD, offering prospects for developing multifunctional anti-AD agents [5]. Concurrently, as confirmed by the studies, the cholinergic circuitry, a vital subcortical-cortical network in the brain, is progressively disrupted in AD, even in its early stages [6, 7]. These supports the "Cholinergic Hypothesis of AD" and has accelerated to the development of drugs, specifically acetylcholinesterase inhibitors. Several drugs in this class have been approved by the U.S. Food and Drug Administration as primary therapeutic agents for AD, which increase synaptic acetylcholine levels to relieve symptoms [8, 9]. Despite the progress that has been made in these studies, the exact pathogenesis is still in the exploratory stage. What is clearer are the risk factors that contribute to its development, including vascular dysfunction [10], sleep disturbances [11], and gender differences [12]. This neurodegenerative disease is the leading cause of dementia in the elderly population and affects approximately 240,000 people worldwide [13]. As it is characterized by progressively increasing cognitive deficits, language impairment, memory impairment, performance impairment, and personality-behavior changes, AD can lead to a decreased quality of life, increased dependency, and increased cost of care for patients [14-18]. In addition, AD patients often suffer from neuropsychiatric symptoms such as depression, apathy, and hallucinations, thus placing a heavier burden on the patients themselves and their families [19]. Therefore, an in-depth understanding
of the mechanisms, metabolic pathways, and gene targets associated with AD can help us treat the disease and alleviate the heavy burden it brings.

Through the analysis, we found a close correlation between AD and oxidative phosphorylation (OXPHOS) pathway. Mitochondrial OXPHOS is a crucial mechanism for energy production in eukaryotic cells and is central to cellular metabolism [20]. Based on its energy supply role and its involvement in cellular metabolism, it also plays a very important role in the development of diseases. Among cardiac diseases, cardiomyopathy and heart failure are common manifestations of mitochondrial diseases caused by defects in the mitochondrial OXPHOS system [21]. In neurodegenerative diseases, OXPHOS also plays a crucial role. On the one hand, oxidative stress involving OXPHOS is an essential factor for Aβ protein accumulation and tau protein hyperphosphorylation [22-24], and on the other hand, mitochondria are more susceptible to oxidative stress due to their role as sites of the ATP-generating electron transport chain and major sources of ROS [25, 26], and thus mitochondrial respiratory abnormalities have been seen in several neurodegenerative disorders, including AD and Parkinson's disease [27]. During the development of these diseases, brain glucose metabolism deteriorates in a region-specific and progressive manner, that is, the primary energy supply is converted from OXPHOS to glycolysis (metabolic reprogramming) [28]. Metabolic reprogramming leads to widespread defective impairment of cellular metabolism, ultimately causing cellular dysfunction. If this process occurs in immune cells of the brain (e.g. microglia), it leads to the development of neuroinflammation and ultimately promotes the development of AD [29].

The aim of this study is to reveal the potential driver genes for the occurrence of AD through multi-omics analysis and to develop unique biomarkers accordingly. **Figure 1** illustrates the workflow of this study.

**Material and methods**

**Transcriptome data and scRNA-seq (single-cell RNA sequencing) data acquisition**

Gene expression files for AD were obtained from the NCBI Gene Expression Omnibus
public database (GEO) (https://www.ncbi.nlm.nih.gov/). The GEO datasets GSE122063 and GSE97760 were chosen. Samples from GSE97760 were extracted from peripheral blood. Samples from GSE122063 were extracted from frontal and temporal cortex. After excluding 36 patients with vascular dementia, a total of 56 samples with AD and 44 healthy controls were retained in the GSE122063 dataset. A total of 9 samples with AD and ten healthy controls were retained in the GSE97760 dataset. The scRNA-seq data of 1 AD sample and 1 healthy control (GSM4996463 and GSM4996462) of GSE164089 were obtained from the GEO database.

**Differential expression analysis**

We replicated the DEG analysis with AD datasets (GSE122063 and GSE97760) to confirm the differentially expressed genes (DEGs). In R (version 4.2.2), the "limma" and "GSEABase" packages were used for DEGs identification and GSEA. Between the control group and AD group, the screening criteria for DEGs were $P\text{-value} < 0.05$ and $|\text{Log}_2(\text{fold change})| > 1$.

**Screening of co-expressed genes of AD-related pathways by weighted gene co-expression network analysis**

Using the GSEA database, we selected 10 potentially pathways relevant to AD, including GOBP\_RESPONSE\_TO\_OXIDATIVE\_STRESS (https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/GOBP\_RESPONSE\_TO\_OXIDATIVE\_STRESS) [30], GOBP\_INTRINSIC\_APOPTOTIC\_SIGNALLING\_PATHWAY (https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/GOBP\_INTRINSIC\_APOPTOTIC\_SIGNALLING\_PATHWAY) [31], HP\_NEUROINFLAMMATION (https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/HP\_NEUROINFLAMMATION) [32], HALLMARK\_OXIDATIVE\_PHOSPHORYLATION (http://www.gsea-msigdb.org/gsea/msigdb/human/geneset/HALLMARK\_OXIDATIVE\_PHOSPHORYLATION) [33], BIOCARTA\_HSP27\_PATHWAY (https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/BIOCARTA\_HSP27\_PATHWAY) [34]. The WGCNA package in R software was used to construct a gene co-expression network to
examine the co-expression relationships between genes and pathways [35]. We removed an abnormal sample based on cluster trees. To construct the similarity matrix, we calculated the correlation coefficient between each gene pair. We chose an appropriate soft threshold to transform the similarity matrix into an adjacency matrix, ensuring the construction of a scale-free network. Then, the average network connectivity for each gene was calculated using a topological overlap matrix (TOM). Genes with similar expression profiles were grouped into the same module using the "blockwiseModules" function and dynamic tree cutting. Different modules are characterized by different colors. Module eigengene (ME) is the first main component of each module's gene expression profile. We also calculated module membership (MM), gene significance (GS), and modules with AD-related pathways, and then visualized the network of feature genes. Finally, AD target genes were identified by intersecting WGCNA-derived significant module genes with DEGs from GSE122063 and GSE97760.

**GO, DO, KEGG analyses**

We used the "clusterProfiler" R package to perform GO (Gene Ontology), KEGG (Kyoto Encyclopedia of Genes and Genomes) and DO (Disease Ontology) analyses of the functions of the genes obtained from the intersection. A $P$-value $< 0.05$ was set as the filtering criterion for the three analyses mentioned above. GO analysis was conducted to identify target gene-related cellular components (CCs), molecular functions (MFs) and biological processes (BPs). DO analysis was used to identify diseases in which target genes are frequently involved. Additionally, we screened signaling pathways enriched by potential targets using KEGG enrichment analysis.

**Cerebral cortex marker screening in AD patients by machine learning algorithms**

Two machine learning algorithms were used for this study, including LASSO regression analysis and random forest analysis. Firstly, we used the "glmnet" package to perform the LASSO regression analysis. In addition, a random forest analysis was performed using the "randomForest" package and genes with a MeanDecreaseGini $> 2$ were selected. The intersection genes derived from two machine learning algorithms are
considered potential cortex biomarkers for patients with AD.

**Correlation analysis and GSEA of potential biomarkers**

Using the “corrplot” package, we analyzed the correlation between the expressions of potential biomarkers. Further, gene set enrichment analyses (GSEA) were performed on the biomarkers to better understand their functions.

**Single-cell data processing and cell annotation**

Seurat objects were created from scRNA-seq data using the R software package "Seurat 4.2.1" [36], which was also utilized for normalization, grouping, differential gene expression analysis, and visualization. Integration was performed using the Harmony package by removing batch effects for different samples. First, we conducted quality control on the scRNA-seq data by deleting clusters with fewer than three cells, cells with at least 20% mitochondrial genes, and cells with no more than 50 genes mapped. Then, the different cell clusters were obtained based on the first 2000 differential genes by descending and removing batch effects by principal component analysis (PCA) and “Harmony” algorithm. The clustering analysis was done based on the "FindClusters" and "FindNeighbors" functions in the Seurat package, and the results were visualized using the "RunUMAP" function. Different cell clusters were annotated using the SingleR package. By the subset() function, various cells from AD and control samples were extracted for differential gene expression analysis.

**Trajectory analysis of single cells**

The Monocle package (version 2.22.0) in R software (version 4.1.3) was used to create single-cell pseudotime trajectories of astrocytes. First, two samples of astrocytes were selected from the annotated cell subpopulations. Then, we applied the newCellDataSet() function to create a new object. Genes with mean expression <0.1 were removed and the remaining genes were used for trajectory analysis. The reduceDimension() function was applied to reduce the dimension and we set the appropriate parameters (reduction_method = "DDRTree" and max_components = 2). Immediately after, we used the plot_cell_trajectory() function to cells for sorting and visualization. We calculated the significance of changes in potential biomarkers with pseudotime by using
the differentialGeneTest() function with parameters set to qvalue<0.1 and visualized them with the plot_pseudotime_heatmap() function.

**Construction of the miRNA–mRNA and TF-mRNA network**

Investigating the relationship between miRNAs and potential biomarkers, transcription factors and potential biomarkers is essential for understanding the regulatory mechanisms and functions of potential biomarkers. In addition, we use this network to find the most crucial potential biomarkers that are interlinked among the biomarkers. Based on the NetworkAnalyst 3.0 platform (https://www.networkanalyst.ca/NetworkAnalyst/), we used the “ChEA” database to enrich potential biomarker-associated transcription factors (TFs), used the “TarBase” database to enrich potential biomarker-associated miRNAs.

**Construction of the potential biomarkers-based nomogram**

Nomogram was used to predict the relationship between gene expression and prevalence in a population. We selected potential biomarkers that were linked in both the TF-target gene network and the miRNA-target gene network for the construction of the nomogram. The “rms” package was used to construct a nomogram base on the potential biomarker. Subsequently, the expression of potential biomarkers was used to predict AD prevalence in the population.

**Immune infiltration and immune cell correlation analyses**

The "CIBERSORTx" website (https://cibersortx.stanford.edu/) was used to analyze the level of infiltration of 22 different immune cells in the cerebral cortex and blood tissue of patients. Then, we performed a correlation analysis of the 22 immune cells infiltrating the cerebral cortex of AD patients to better understand how the immune system changes.

**Correlation analysis of key potential biomarkers and immune cell infiltration**

The Spearman coefficient was used to investigate the correlation between key potential marker genes and immune cell infiltration further, to identify which immune cells they have a significant connection with, and to try to understand how key potential marker genes play an essential part in AD by regulating immune cell infiltration.
Results

Identification of DEGs in the cerebral cortex and peripheral blood of AD patients

First, the sample was standardized to obtain 307 and 2862 DEGs in the cerebral cortex (Figure 2A) and peripheral blood (Figure S1A) of patients with AD. Among the DEGs in the cerebral cortex, 118 genes were up-regulated and 189 genes were down-regulated. Meanwhile, among the DEGs in peripheral blood, 1283 genes were down-regulated and 1579 genes were up-regulated. This DEG information was displayed as a volcano plot (Figure 2B, S1B). Two heatmaps of the top 100 DEGs were plotted (Figure 2C, S1C).

Subsequently, GSEA of the 307 and 2862 DEGs respectively yielded nucleotide excision repair, alanine, aspartate and glutamate metabolism pathways, hippo signaling pathway – multiple species and cell cycle, RNA degradation pathways (Figure 2D, S2D).

Target gene screening using WGCNA

Weighted gene co-expression network analysis was performed on AD cortical samples and control samples. Outliers were removed in the process of sample clustering. Then, the scale-free fit index was adjusted to 0.9, and the minimum soft thresholds for constructing scale-free networks were set at 5 and 2, respectively (Figure 2E, S1E). Additionally, the minimum number of genes in each module was set at 50, resulting in 4 and 25 modules, respectively (Figure 2F, S1F). Gene significance (GS) denotes the relationship between samples and module genes. Module membership (MM) denotes the relationship between the module eigen gene and module gene expression values. By calculating GS and MM values, the modules were connected to the passage and clinical features. The MEturquoise module in AD brain cortical samples and the MEblue module in peripheral blood samples showed a strong correlation with both AD and OXPHOS (Figure 2G, S1G). Figure 2H, S1H demonstrated the correlation of target module genes with OXPHOS-related genes. Notably, the turquoise module genes correlated extremely well with OXPHOS-related genes (correlation = 0.9, P<1e-200)(Figure 2H). Similarly, the genes in the blue module showed a similar correlation.
Finally, the MEnturquoise module genes and MEnblue module genes were intersected with two DEG sets, yielding 29 target genes associated with both AD and OXPHOS (Figure 2I).

**GO, DO and KEGG enrichment analyses for the target genes**

To study the biological functions (BFs) and related diseases of the target genes, GO, DO, and KEGG enrichment analyses were performed. By GO analysis, we discovered that these genes are engaged in a range of functions, including regulation of macrophage activation. Furthermore, they are involved in many MFs, such as SH3 domain binding, adenylate cyclase binding, and are associated with several CCs, such as cytoplasmic side of plasma membrane, extrinsic component of cytoplasmic side of plasma membrane (Figure S2A, B, C). Correspondingly, according to the results of the KEGG enrichment analysis, these target genes were predominantly enriched by diseases and signaling pathways such as amyotrophic lateral sclerosis (ALS), pathways of neurodegeneration – multiple diseases and GABAergic synapse (Figure S2D). According to DO analysis, the target genes were linked to conditions like craniosynostosis and neurodegenerative diseases such as PD and ALS (Figure S2E).

**Machine learning algorithm-based biomarker screening in the cerebral cortex of AD patients**

This study used 2 machine learning techniques to further investigate putative AD biomarkers in 29 target genes. The LASSO regression model was created using both control and AD samples. The $\lambda$ analysis revealed that when $\lambda = 18$, the model could correctly forecast AD. Consequently, LASSO analysis produced 18 potential biomarkers (Figure 3A, Table 1). Additionally, in the random forest results, 8 genes with a “MeanDecreaseGini” greater than 3 were selected as candidate biomarkers (Figure 3B, Table 2). Finally, combining the results of the 2 algorithms yielded RBM3, GOLGA8A, ALS2, FSD2, and LOC100287628 in the cerebral cortex as potential biomarkers associated with AD (Figure 3C).

**Correlation between potential biomarkers and GSEA**

The correlation analysis of potential biomarkers showed a negative correlation between
FSD2 and ALS2, RBM3, GOLGA8A and LOC100287628 (Figure 4A). Then, these 5 potential biomarkers were subjected to GSEA functional analysis, and we identified several pathways, including alanine, aspartate and glutamate metabolism pathways, hippo signaling pathway – multiple species, which closely matched the outcomes of the DEG functional analysis. Subsequently, Figure 4B-F display the outcomes of the GSEA functional analysis of the 5 potential biomarkers. In addition, there are some biological functions that are highly associated with these biomarkers, such as apoptosis and the NF-kappa B signaling pathway, suggesting that these genes can be involved in the inflammatory response, cell proliferation, differentiation and survival[37], and play a significant role in AD through these pathways.

**Cell quality control, cell annotation and biomarker expression analysis**

In the current study, 3594 cells from the AD and normal control samples of GSE188545 were used to generate the scRNA-seq data. The range of the number of genes detected, the level of sequencing, and the proportion of mitochondria in each sample were all within acceptable bounds (Figure 5A). Following data normalization, the top 2000 highly variable genes (HVGs) were chosen (Figure 5B). The PCA method and Harmony software package were used to reduce dimensions and remove batch effects (Figure 5D), and the cells were finally classified into 14 clusters (Figure 5C). With the SingleR package to annotate the cells, we classified the cells into 4 subclusters, namely astrocytes, endothelial cells, macrophages, and tissue stem cells (Figure 5E). By separating AD samples from normal samples, we analyzed the differences in expression of four potential biomarkers ALS2, FSD2, GOLGA8A and RBM3 between the two samples (LOC100287628 was not analyzed as we did not have a gene symbol for it). The results showed that ALS2 and GOLGA8A were lower expressed in astrocytes (Figure 5F); ALS and RBM3 were higher expressed in endothelial cells (Figure 5G); GOLGA8A was lower expressed in macrophages (Figure 5H); and GOLGA8A and RBM3 were lower expressed in tissue stem cells (Figure 5I) of AD samples compared with the control samples. Differential expression of potential biomarkers in different cells may be a potential mechanism for causing AD.
Trajectory of astrocyte in AD

Due to the high content of astrocytes in this current single-cell data and the low risk of analysis bias, we selected astrocytes for trajectory analysis. Subsequently, after reducing the dimensionality of astrocytes according to different genes (Figure 5J), we visualized the trajectory of astrocyte distribution with pseudotime. Both trajectory analyses yielded 1 key time node, and we divided the astrocytes into 3 cell states by dividing the time node (Figure 5K, L). By analyzing the changes in expression of potential biomarkers across astrocyte states (Figure 5M, N) and calculating the significance of changes in marker expression with pseudotime (Figure 5O, P), we found significant changes in the expression of GOLGA8A and FSD2 in AD samples and ALS2, RBM3 and FSD2 in control samples.

Regulatory mechanisms and relevance of potential biomarkers

TFs are engaged in the transcriptional process of genes, while one or more miRNAs have fine regulation of gene expression. By constructing the miRNA-mRNA network and TF-mRNA network, we clarify the regulatory relevance of potential biomarkers and search for the most key potential biomarkers. The results of network construction suggest that RBM3, GOLGA8A and ALS2 may be regulated by the same miRNAs, such as has-mir-101-3p, hsa-mir-23b-3p, hsa-mir-30b-5p, hsa-mir-30c-5p, hsa-mir-124-3p, hsa-mir-16-5p and hsa-mir-129-2-3p (Figure S3A). Meanwhile, RBM3, GOLGA8A, ALS2 and FSD2 were regulated by the same TFs, such as CREM, STAT4, MTF2, TBP and GATA1, while E2F1 may be involved in the regulation of these four genes (Figure S3B).

Clinical diagnostic evaluation of key potential biomarkers

The diagnostic bar graph shows that low expression of 3 key biomarkers contributes to the clinical diagnosis of AD. We calculated the likelihood that a patient would be given an AD diagnosis by combining the gene scores (Figure 6A). Also, the column line model exhibited great capacity to forecast AD, according to calibration curves (Figure 6B). Additionally, DCA (Figure 6C) and CIC (Figure 6D) visually illustrated that the column line diagram had superior overall net benefit over a practical and wide range of
threshold probabilities, indicating that the column line diagram acquired in our current research could assist clinicians in more accurately determining patient prognosis. Taken together, these analyses' findings imply that RBM3, GOLGA8A, and ALS2 have the potential to function as reliable diagnostic biomarkers for AD.

**Immune cell infiltration results**

The microenvironment in the brain contains a variety of growth factors, inflammatory factors, extracellular matrix and immune cells that have an important impact on the development of brain diseases. With the CIBERSORT algorithm, we estimated the proportion of 22 types of immune cells in 44 control samples and 56 AD samples in the cerebral cortex (Figure 7A) and performed the same analysis on control and AD samples of blood and visualized the results (Figure S4A). Also, in cortical samples from AD brains, we examined the relationship between immune cell infiltration (Figure 7B). Then, immune cell infiltration in AD cerebral cortex and blood samples and respective control samples were compared in two boxplots (Figure 7C, S4B). The results showed that in the cerebral cortex, the proportions of T cells CD8 ($P < 0.05$) were significantly lower in the AD group than in the control group. In peripheral blood, the proportions of B cells memory ($P < 0.05$) and T cells memory resting ($P < 0.05$) were significantly higher in the AD group than in the control group, while the proportions of T cells CD4 naïve ($P < 0.05$), monocytes ($P < 0.05$) and macrophages M0 ($P < 0.01$) were lower.

**Correlation analysis between key potential biomarkers and immune cells**

Correlation analysis of biomarker genes and 22 immune cells allows us to speculate how these genes are involved in the process of AD by regulating the infiltration of immune cells. We explored the relevance of 3 key potential biomarkers to immune cells in cortical samples of AD brains (Figure 8). The marker gene RBM3 in AD was positively correlated with mast cells resting, eosinophils and T cells CD8, and negatively correlated with mast cells activated (Figure 8A). The gene GOLGA8A was positively correlated with T cells CD8, whereas it was negatively correlated T cells follicular helper and macrophages M0 (Figure 8B). The gene ALS2 was positively
correlated with eosinophils and T cells CD8, and negatively correlated with NK cells resting, macrophages M0 and neutrophils (Figure 8C).

Discussion

This study conducted AD target gene screening and functional analysis. It also integrated machine learning algorithms for biomarker screening and performed regulatory mechanism prediction of the screened biomarkers. These results have theoretical value for the diagnosis and prediction of AD as well as insight into the immune characterization of AD. First, we screened cortical samples and blood samples from AD for 29 common genes that may be closely related to the OXPHOS pathway. Eukaryotic OXPHOS, crucial for ATP production, significantly influences AD development [38]. On the other hand, OXPHOS facilitates the maintenance of central immune system homeostasis. An essential role for neuroinflammation in the etiology of AD may be played by microglia, according to the discovery of higher levels of inflammatory markers in AD patients and the finding of AD risk genes related to innate immune function [39]. The 29 target genes were found to be highly co-expressed with genes involved in the OXPHOS pathway, implying that these target genes may affect the development of neuroinflammation and AD directly or indirectly through the OXPHOS pathway. Resident macrophages of the central nervous system (CNS), microglia, play a dual role in brain function by secreting both pro-inflammatory cytokines and neurotrophic factors. They interact bidirectionally with other brain cells and peripheral immune cells, influencing diseases like neuroinflammation and maintaining healthy brain tissue [40, 41]. Some of the above target genes regulate macrophage activation, potentially affecting AD progression by altering microglial activation. In the pathogenesis of AD, reduced estrogen signaling leads to diminished estrogen-mediated neuroprotection. Estrogen, a neuroactive steroid, plays a neuroprotective role in AD by reducing glutamate toxicity, Aβ, brain inflammation, and tau hyperphosphorylation [42-46]. Reduced ERα expression in AD patients' hippocampal neurons demonstrates estrogen's protective effects, involving pathways
like Ca2+ mobilization and adenylate cyclase-modulated G protein-coupled receptor
signaling [47]. Some of the target genes may also mediate similar protective functions.

We screened five potential biomarkers based on two machine learning algorithms,
namely FSD2, ALS2, RBM3, GOLGA8A and LOC100287628. Autophagy, a
lysosome-driven self-degradation process, plays a key role in cellular homeostasis by
removing dysfunctional organelles and proteins [48, 49]. Mitophagy specifically targets
mitochondria [50]. Autophagy dysfunction, often due to DNA damage, is linked to
aging and diseases like AD and PD [49]. Our study's GSEA analysis shows that
biomarker FSD2, upregulated in AD patients' cortical samples, might inhibit autophagy
in the CNS, leading to protein and organelle accumulation. New research now suggests
that microglia autophagy plays an integral role in the development of AD. Pyroptosis,
autophagy, and ferroptosis of microglia have a significant impact on AD [51].

Astrocytes, abundant in the CNS, are key to regulating neuroinflammation and
maintaining CNS health [52, 53]. In AD, reactive astrocytes accumulate around
amyloid plaques, participating in phagocytosis and secreting inflammatory cytokines,
exacerbating neurodegeneration. Furthermore, dysregulated interaction between
astrocytes, microglia, and neurons contributes to AD's neuroinflammation [54, 55].

RBM3 (RNA-binding motif protein 3) is a cold shock protein that is rapidly increased
in cold conditions to maintain cellular and internal environmental equilibrium and aid
survival [56]. Various cellular physiological processes, including tumorigenesis [57]
and neuroprotection [58], are likewise regulated by RBM3. Furthermore, RBM3 is
expressed in a variety of immune cells, including eosinophils, T cells, and macrophages
[59]. Our analysis showed that in normal samples, there was a trend of elevated RBM3
expression as astrocytes developed and matured, while there was no such trend in AD
samples. In AD patients, downregulation of RBM3 in astrocytes affects various RBM3-
mediated physiological functions, including cell cycle inhibition and neuroprotective
effects [58, 60], and may also interfere with cellular communication [55], thereby
causing abnormal activation of microglia, which in turn promotes the development of
neuroinflammation and the progression of AD. Jana H Badrani et al. also demonstrated
that RBM3 intrinsically inhibits innate lymphocyte activation and inflammation partially through CysLT1R, suggesting that RBM3 may contribute to the suppression of central neuroinflammation [59], which coincides with our results.

MicroRNAs (miRNA), produced by RNA polymerase II, are crucial for post-transcriptional gene regulation in eukaryotes [61, 62]. They're linked to AD progression, influencing neuronal functions and gene expression [63]. A bioinformatics analysis also pointed to miRNAs as biomarkers and therapeutic targets for AD [64]. Based on five potential biomarkers, we explored the correlation between them by constructing miRNA-mRNA networks. The results showed that mRNAs of three genes could constitute miRNA-mRNA networks with each other. These three genes are GOLGA8A, ALS2, and RBM3. The mRNAs produced by transcription of the three genes are regulated by the same miRNAs, such as hsa-mir-101-3p, hsa-mir-23b-3p, hsa-mir-30b-5p, hsa-mir-30c-5p, hsa-mir-124-3p, hsa-mir-16-5p, and hsa-mir-129-2-3p. In our study, we found that GOLGA8A, ALS2, and RBM3 are correlated with the development of AD by bioinformatic methods, and the above seven miRNAs are jointly involved in the expression of these genes, suggesting that these miRNAs may influence the CNS through the ROS accumulation, mitochondrial function, and Aβ deposition to affect the occurrence and development of AD. Interestingly, one study showed that the role of hsa-miR-101-3p in AD increases with NFT stage and regulates RAC1 [65], while hsa-mir-124-3p emerged as a predictive biomarker for AD [66, 67].

We constructed TF-mRNA networks for 5 potential biomarkers to explore the relevance of transcriptional regulation of these potential biomarkers. The results showed that 4 genes are known to possess transcriptional regulatory relevance. We found that the transcription factor E2F1 co-regulates the transcription of four genes: GOLGA8A, ALS2, RBM3, and FSD2; similarly, SPI1 co-regulates the transcription of three genes: ALS2, RBM3, and FSD2. KLF4 merits consideration as well. KLF4, a zinc finger nuclear protein [68], binds GC-rich DNA sequences and CACCC elements to control transcription [69, 70]. It's linked to inflammation regulation in endothelial cells and macrophages [71, 72] and is upregulated by Aβ42-induced neuroinflammation and
activated P53 in microglial cells [73]. This inflammation, worsened by activated astrocytes producing IL-1, IL-6, TNF-α [74, 75], contributes to neuronal apoptosis and dysfunction. We selected three genes that are linked in the miRNA-mRNA network among potential biomarkers, namely RBM3, GOLGA8A and ALS2, to construct a nomogram to estimate the accuracy of these three genes for predicting the occurrence of AD, and this prediction accuracy was over 90%. Based on the results of our study, the RBM3, GOLGA8A and ALS2 may possess potential as diagnostic biomarkers of AD. Recent data show that innate immune cells crucially impact CNS immune homeostasis, influencing the neuropathogenesis of AD either positively or negatively [76]. By immune infiltration analysis, we found a significantly lower percentage of CD8 T cell infiltration in cortical samples from AD patients and these immune cells may be derived from peripheral blood. However, the functional relevance of CD8 T lymphocytes in the AD brain remains largely unknown. The cellular component "postsynaptic" and biological process "regulation of neuronal synaptic plasticity" were over-represented after CD8 T cell ablation, and the pathways "long-term dementia", "calcium signaling", and "axon guidance" were upregulated [77]. Hence, we conclude that there is an decreased proportion of CD8 T cell infiltration in the AD brain, and the direct involvement of these lymphocytes in controlling synaptic plasticity may lead to neuronal dysfunction. Of note to us are the results of the correlation analysis between the expression of key potential biomarkers and immune cells showing that RBM3, GOLGA8A and ALS2 all have a significant positive correlation with CD8 T cell infiltration. As such, these three genes may play an irreplaceable influence in CD8 T cells affecting neuronal dysfunction in the AD brain. In conclusion, our study identified the biomarkers RBM3, GOLGA8A, and ALS2, which may mediate the onset and progression of AD by modulating the OXPHOS pathway and influencing the proportion of CD8 T cell infiltration. The above results provide new target genes and research ideas for studying the pathogenesis and development of AD, as well as potential biomarkers for clinical prediction of AD.
Declarations

Funding
This research was supported by grants from Guangzhou basic and applied basic research project (2023A04J2337).

Conflicts of interest
The authors have no conflict of interest to report.

Consent to participate
All patients signed written informed consent.

Consent for publication
Consent for publication was obtained from the participants.

Availability of data and material
Data supporting the reported results may be found in the article and supplementary information.

Authors' contributions
Mingkai Lin and Pei Shang were responsible for the design of the project. Mingkai Lin and Peixian Liang was responsible for the analysis and organization of the data. Mingkai Lin, Pei Shang, Yue Zhou, Peixian Liang and Ruoyi Zheng were responsible for writing the manuscript. Minwei Du, Xintong Ke, and Yue Zhou were responsible for the manuscript revision. Wenjing Zhang and Ruoyi Zheng were responsible for the review of the manuscript.

Acknowledgements
None
References


72. Yang HJ, Xi XX, Zhao B, Su ZX, Wang ZY. KLF4 protects brain microvascular endothelial cells from ischemic stroke induced apoptosis by transcriptionally activating MALATI. Biochemical and Biophysical Research Communications 2018; 495: 2376-82.


Legends

Graphical Abstract

Our studies have screened for key biomarkers that influence the occurrence of Alzheimer's disease (AD). The changes in the expression of these genes in astrocytes may promote the development of AD. For example, down-regulation of RBM3 in
astrocytes can promote the release of inflammatory factors from microglia, and cause a decrease in the level of CD8 T cell infiltration in the cerebral cortex, leading to neuronal dysfunction.

**Figure 1** The flow chart for multi-omics method of the current study.

**Figure 2** Identification of DEGs in frontal and temporal cortex samples and results of WGCNA.

(A) Cortex samples display after standardization. (B) The volcano Plot showing the expression characteristics of DEGs, where red represents gene upregulation in the AD group, and blue represents gene upregulation in the normal control group. (C) The heat map showing the expression of the top 50 DEGs in the cortex samples. (D) GSEA (gene set enrichment analysis) of DEGs. (E) The left panel shows the scale-free fit index (scale independence, y-axis) as a function of soft threshold power (x-axis); the right panel shows the average connectivity (degree, y-axis) as a function of soft threshold power (x-axis). (F) AD module clustering dendrogram based on a dissimilarity measure (1-TOM). The colored squares below the dendrogram represent the module assignment identified by dynamic tree shearing. (G) Feature-module correlations. Each column corresponds to a clinical feature, whereas individual rows in the heat map correspond to an ME. Individual cells contain the corresponding P-value and correlation coefficient. (H) Scatter plot of OXPHOS pathway correlation with genes in the turquoise module. (I) WGCNA module genes mapped to DEGs.

**Figure 3** Screening of potential biomarkers by 2 types of machine learning algorithms.

(A) LASSO regression model. (B) The left panel shows the random forest model; the right panel shows the 7 genes in order of importance. (D) The intersection of the results of the 2 types of machine learning algorithms.
**Figure 4 GSEA of potential biomarkers.**
(A) Correlation analysis between five potential biomarkers. (B–F) GSEA functional analysis of five potential biomarkers (B, RBM3, C, GOLGA8A, D, LOC100287628, E, FSD2, F, ALS2).

**Figure 5 Results of single-cell analysis and trajectory analysis.**
(A) Quality control of scRNA-seq data from AD samples (human brain middle temporal gyrus). (B) Variance plot showing 20,432 genes in all cells, with red dots representing the top 2,000 highly variable genes. The 8 most highly variable genes were labeled. (C) Visualization of 14 clusters using the UMAP algorithm. (D) Visualization of cells after removal of batch effects. (E) Cell subclusters identification. (F–I) Differential expression of potential biomarkers in 4 types of cells in AD versus normal control. (F) Astrocyte, (G) endothelial cells, (H) macrophage, (I) tissue stem cells. (J) Dimensionality reduction in trajectory analysis of cells from AD and control samples (above, AD; below, control). (K–L) The trajectory plot of Monocle 2 shows the dynamics of the astrocyte of AD (K) and control (L) samples and their pseudotime curve. (M, N) Variation in expression of potential biomarkers in different astrocyte states (M, AD; N, control). (O, P) Significance analysis of potential biomarkers changed along the pseudotime ($q$ value $< 0.1$) (O, AD; P, control).

**Figure 6 Construction of diagnostic models.**
(A) Nomogram; we calculated the corresponding score for each genetic variable, then draw a straight line above the "point axis" and add the calculated scores for all the variables to find the final value on the "total score axis" and then draw a vertical line on the "probability axis" to identify the patient's risk of developing AD. (B) Calibration curve showing the relationship between predicted and actual probability. (C) Decision Curve Analysis (DCA); horizontal line indicated no AD, gray diagonal line indicated no AD, and column diagram showed more net benefit within the threshold probability range. (D) Clinical impact curves; red curves showed the number of individuals
classified as positive by the model at each threshold probability; green curves showed the number of true positives at each threshold probability.

**Figure 7** Results of immune cell infiltration of AD brain cortex samples.
(A) The relative percentage of 22 immune cells in each AD brain cortex sample. (B) Heatmap of correlation in 18 types of immune cells in cortex samples (4 types of immune cells were eliminated due to the absence of infiltration). Brown represented a positive correlation, while blue represented a negative correlation. Darker color indicated stronger association. (C) The box plot for immune cell infiltration analysis in AD cortex. Red represented the AD group and blue represented the control group. ns: no significant, \(*P < 0.05; **P < 0.01; ***P < 0.001.\)

**Figure 8** Relationship between 3 potential diagnostic biomarkers and immune infiltration in AD (cortex sample).
(A) RBM3, (B) GOLGA8A, (C) ALS2.

**Figure S1** Identification of DEGs in blood samples and results of WGCNA.
(A) Blood samples display after standardization. (B) The volcano Plot showing the expression characteristics of DEGs, where red represents gene upregulation in the AD group, and blue represents gene upregulation in the normal control group. (C) The heat map showing the expression of the top 50 DEGs in the blood samples. (D) GSEA (gene set enrichment analysis) of DEGs. (E) The left panel shows the scale-free fit index (scale independence, y-axis) as a function of soft threshold power (x-axis); the right panel shows the average connectivity (degree, y-axis) as a function of soft threshold power (x-axis). (F) AD module clustering dendrogram based on a dissimilarity measure (1-TOM). The colored squares below the dendrogram represent the module assignment identified by dynamic tree shearing. (G) Feature-module correlations. Each column corresponds to a clinical feature, whereas individual rows in the heat map correspond to an ME. Individual cells contain the corresponding P-value and correlation coefficient.
(H) Scatter plot of OXPHOS pathway correlation with genes in the blue module.

Figure S2 Enrichment analysis of function and disease of potential target genes.
(A) The barplot showed the top 8 enriched results of target genes from 3 aspects of BP, CC, and MF. (B) The net plot showed the connection of all the pathways from the 3 aspects of BP, CC, and MF. (C) The circle chart showed the pathways and target genes from the aspect of BP. (D) The bubble plot showed the top 5 enriched KEGG pathways of target genes. (E) The barplot showed the most enriched results of target genes from disease.

Figure S3 Regulatory network of potential biomarker markers.
(A) The miRNA–potential biomarker network, where the red nodes symbolize potential biomarkers, the green nodes symbolize microRNA (miRNA). (B) The TF–potential biomarker network, where the red nodes symbolize potential biomarkers, the blue nodes symbolize transcription factors (TFs).

Figure S4 Results of immune cell infiltration of AD blood samples.
(A) The relative percentage of 22 immune cells in each AD blood sample. (B) The box plot for immune cell infiltration analysis in AD blood. Red represented the AD group and blue represented the control group. ns: no significant, *P < 0.05; **P < 0.01.

Table 1 Results of the LASSO analysis.

Table 2 Results of random forest analysis.
<table>
<thead>
<tr>
<th>No.</th>
<th>Gene name</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CCDC66</td>
<td>-0.023330326</td>
</tr>
<tr>
<td>2</td>
<td>FSD2</td>
<td>0.144008264</td>
</tr>
<tr>
<td>3</td>
<td>PVALB</td>
<td>0.033221064</td>
</tr>
<tr>
<td>4</td>
<td>TNS1</td>
<td>0.052256082</td>
</tr>
<tr>
<td>5</td>
<td>KHDRBS2</td>
<td>0.037955948</td>
</tr>
<tr>
<td>6</td>
<td>LOC100505971</td>
<td>0.000779881</td>
</tr>
<tr>
<td>7</td>
<td>EFTUD1</td>
<td>-0.062305877</td>
</tr>
<tr>
<td>8</td>
<td>LOC100287628</td>
<td>-0.002449124</td>
</tr>
<tr>
<td>9</td>
<td>FAM19A1</td>
<td>0.013057823</td>
</tr>
<tr>
<td>10</td>
<td>RBM3</td>
<td>-0.21634374</td>
</tr>
<tr>
<td>11</td>
<td>GNG2</td>
<td>0.016065787</td>
</tr>
<tr>
<td>12</td>
<td>RGS1</td>
<td>0.074340248</td>
</tr>
<tr>
<td>13</td>
<td>ALS2</td>
<td>0.037761796</td>
</tr>
<tr>
<td>14</td>
<td>XIST</td>
<td>0.013727952</td>
</tr>
<tr>
<td>15</td>
<td>LOC100288814</td>
<td>-0.01847535</td>
</tr>
<tr>
<td>16</td>
<td>GOLGA8A</td>
<td>-0.198714156</td>
</tr>
<tr>
<td>17</td>
<td>XLOC_013896</td>
<td>0.038040005</td>
</tr>
<tr>
<td>18</td>
<td>ZDHHC23</td>
<td>-0.061786556</td>
</tr>
</tbody>
</table>
Table 2 Results of random forest analysis.

<table>
<thead>
<tr>
<th>No.</th>
<th>Gene name</th>
<th>MeanDecreaseGini</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RBM3</td>
<td>5.677111625</td>
</tr>
<tr>
<td>2</td>
<td>GOLGA8A</td>
<td>5.481686647</td>
</tr>
<tr>
<td>3</td>
<td>CCT6B</td>
<td>4.316600027</td>
</tr>
<tr>
<td>4</td>
<td>ENPP5</td>
<td>3.992431065</td>
</tr>
<tr>
<td>5</td>
<td>ALS2</td>
<td>3.932123715</td>
</tr>
<tr>
<td>6</td>
<td>FSD2</td>
<td>3.670724866</td>
</tr>
<tr>
<td>7</td>
<td>LOC100287628</td>
<td>3.183618109</td>
</tr>
<tr>
<td>8</td>
<td>IL1RL1</td>
<td>3.101110037</td>
</tr>
<tr>
<td>9</td>
<td>LOC100505971</td>
<td>1.837803457</td>
</tr>
<tr>
<td>10</td>
<td>LOC100288814</td>
<td>1.509276095</td>
</tr>
<tr>
<td>11</td>
<td>TNS1</td>
<td>1.333639869</td>
</tr>
<tr>
<td>12</td>
<td>RGS1</td>
<td>1.17357217</td>
</tr>
<tr>
<td>13</td>
<td>TASP1</td>
<td>1.09041616</td>
</tr>
<tr>
<td>14</td>
<td>ZDHHHC23</td>
<td>1.014779372</td>
</tr>
<tr>
<td>15</td>
<td>XIST</td>
<td>0.729755195</td>
</tr>
<tr>
<td>16</td>
<td>NELL1</td>
<td>0.707489867</td>
</tr>
<tr>
<td>17</td>
<td>ZNF215</td>
<td>0.623515798</td>
</tr>
<tr>
<td>18</td>
<td>FAM19A1</td>
<td>0.598571032</td>
</tr>
<tr>
<td>19</td>
<td>AKAP5</td>
<td>0.537364545</td>
</tr>
<tr>
<td>20</td>
<td>NEFM</td>
<td>0.507005845</td>
</tr>
<tr>
<td>21</td>
<td>XLOC_013896</td>
<td>0.506096223</td>
</tr>
<tr>
<td>22</td>
<td>CCDC66</td>
<td>0.493729086</td>
</tr>
<tr>
<td>23</td>
<td>PVALB</td>
<td>0.472640711</td>
</tr>
<tr>
<td>24</td>
<td>GNG2</td>
<td>0.437003367</td>
</tr>
<tr>
<td>25</td>
<td>KHDRBS2</td>
<td>0.418461147</td>
</tr>
<tr>
<td>26</td>
<td>SRRM2</td>
<td>0.407525122</td>
</tr>
<tr>
<td>27</td>
<td>CD200</td>
<td>0.406332736</td>
</tr>
<tr>
<td>28</td>
<td>SULT4A1</td>
<td>0.360802487</td>
</tr>
<tr>
<td>29</td>
<td>EFTUD1</td>
<td>0.30306696</td>
</tr>
</tbody>
</table>
Figure 1 The flow chart for multi-omics method of the current study.
Figure 2 Identification of DEGs in frontal and temporal cortex samples and results of WGCNA.

(A) Cortex samples display after standardization. (B) The volcano Plot showing the expression characteristics of DEGs, where red represents gene upregulation in the AD group, and blue represents gene upregulation in the normal control group. (C) The heat map showing the expression of the top 50 DEGs in the cortex samples. (D) GSEA (gene set enrichment analysis) of DEGs. (E) The left panel shows the scale-free fit index (scale independence, y-axis) as a function of soft threshold power (x-axis); the right panel shows the average connectivity (degree, y-axis) as a function of soft threshold power (x-axis). (F) AD module clustering dendrogram based on a dissimilarity measure (1-TOM). The colored
squares below the dendrogram represent the module assignment identified by dynamic tree shearing. (G) Feature-module correlations. Each column corresponds to a clinical feature, whereas individual rows in the heat map correspond to an ME. Individual cells contain the corresponding P-value and correlation coefficient. (H) Scatter plot of OXPHOS pathway correlation with genes in the turquoise module. (I) WGCNA module genes mapped to DEGs.
Figure 3 Screening of potential biomarkers by 2 types of machine learning algorithms. (A) LASSO regression model. (B) The left panel shows the random forest model; the right panel shows the 7 genes in order of importance. (C) The intersection of the results of the 2 types of machine learning algorithms.
Figure 4 GSEA of potential biomarkers. (A) Correlation analysis between five potential biomarkers. (B–F) GSEA functional analysis of five potential biomarkers (B, RBM3, C, GOLGA8A, D, LOC100287628, E, FSD2, F, ALS2).
Figure 5 Results of single-cell analysis and trajectory analysis.

(A) Quality control of scRNA-seq data from AD samples (human brain middle temporal gyrus). (B) Variance plot showing 20,432 genes in all cells, with red dots representing the top 2,000 highly variable genes. The 8 most highly variable genes were labeled. (C) Visualization of 14 clusters using the UMAP algorithm. (D) Visualization of cells after removal of batch effects. (E) Cell subclusters identification. (F-I) Differential expression of potential biomarkers in 4 types of cells in AD versus normal control. (F) Astrocyte, (G) endothelial cells, (H) macrophage, (I) tissue stem cells. (J) Dimensionality reduction in trajectory analysis of cells from AD and control samples (above, AD; below, control). (K-L) The trajectory plot of Monocle 2 shows the dynamics of the astrocyte of AD (K) and control (L) samples and their pseudotime curve. (M, N) Variation in expression of potential biomarkers in different astrocyte states (M, AD; N, control). (O, P) Significance analysis of potential biomarkers changed along the pseudotime (q value < 0.1) (O, AD; P, control).
Figure 6 Construction of diagnostic models.
(A) Nomogram; we calculated the corresponding score for each genetic variable, then draw a straight line above the "point axis" and add the calculated scores for all the variables to find the final value on the "total score axis" and then draw a vertical line on the "probability axis" to identify the patient's risk of developing AD. (B) Calibration curve showing the relationship between predicted and actual probability. (C) Decision Curve Analysis (DCA); horizontal line indicated no AD, gray diagonal line indicated no AD, and column diagram showed more net benefit within the threshold probability range. (D) Clinical impact curves; red curves showed the number of individuals classified as positive by the model at each threshold probability; green curves showed the number of true positives at each threshold probability.
Figure 7 Results of immune cell infiltration of AD brain cortex samples. 
(A) The relative percentage of 22 immune cells in each AD brain cortex sample. 
(B) Heatmap of correlation in 18 types of immune cells in cortex samples (4 types of immune cells were eliminated due to the absence of infiltration). Brown represented a positive correlation, while blue represented a negative correlation. Darker color indicated stronger association. 
(C) The box plot for immune cell infiltration analysis in AD cortex. Red represented the AD group and blue represented the control group. ns: no significant, *P < 0.05; **P < 0.01; ***P < 0.001.
Figure 8 Relationship between 3 potential diagnostic biomarkers and immune infiltration in AD (cortex sample).
(A)RBM3, (B) GOLGA8A, (C) ALS2.
Figure S1 Identification of DEGs in blood samples and results of WGCNA.
(A) Blood samples display after standardization. (B) The volcano Plot showing the
expression characteristics of DEGs, where red represents gene upregulation in the AD
group, and blue represents gene upregulation in the normal control group. (C) The heat map
showing the expression of the top 50 DEGs in the blood samples. (D) GSEA (gene set
enrichment analysis) of DEGs. (E) The left panel shows the scale-free fit index (scale
independence, y-axis) as a function of soft threshold power (x-axis); the right panel shows
the average connectivity (degree, y-axis) as a function of soft threshold power (x-axis). (F)
AD module clustering dendrogram based on a dissimilarity measure (1-TOM). The colored
squares below the dendrogram represent the module assignment identified by dynamic tree
shearing. (G) Feature-module correlations. Each column corresponds to a clinical feature, whereas individual rows in the heat map correspond to an ME. Individual cells contain the corresponding P-value and correlation coefficient. (H) Scatter plot of OXPHOS pathway correlation with genes in the blue module.
Figure S2 Enrichment analysis of function and disease of potential target genes.

(A) The barplot showed the top 8 enriched results of target genes from 3 aspects of BP, CC, and MF. (B) The net plot showed the connection of all the pathways from the 3 aspects of BP, CC, and MF. (C) The circle chart showed the pathways and target genes from the aspect of BP. (D) The bubble plot showed the top 5 enriched KEGG pathways of target genes. (E) The barplot showed the most enriched results of target genes from disease.
Figure S3 Regulatory network of potential biomarker markers.
(A) The miRNA–potential biomarker network, where the red nodes symbolize potential biomarkers, the green nodes symbolize microRNA (miRNA). (B) The TF–potential biomarker network, where the red nodes symbolize potential biomarkers, the blue nodes symbolize transcription factors (TFs).
Figure S4 Results of immune cell infiltration of AD blood samples. (A) The relative percentage of 22 immune cells in each AD blood sample. (B) The box plot for immune cell infiltration analysis in AD blood. Red represented the AD group and blue represented the control group. ns: no significant, *P < 0.05; **P < 0.01.