Am J Transl Res 2023;15(4):2498-2514
www.ajtr.org /ISSN:1943-8141/AJTR0148473

Original Article

Ribosomal proteins are blood biomarkers and associated with CD4+ T cell activation in Alzheimer's disease: a study based on machine learning strategies and scRNA-Seq data validation

Yuejun Wang, Dehua Zhan, Lifang Wang

Department of Geriatrics, Zhejiang Aged Care Hospital, Hangzhou Normal University, Hangzhou, Zhejiang, P. R. China

Received December 13, 2022; Accepted March 16, 2023; Epub April 15, 2023; Published April 30, 2023

Abstract: Background: Alzheimer's disease (AD) is a widespread neurodegenerative disease that primarily affects the elderly. Unfortunately, the lack of convenient early diagnostic tools makes it difficult to intervene and treat the disease during its initial stages. Methods: We obtained four bulk and single-cell RNA-sequencing peripheral blood samples related to AD from public databases. Using Boruta and LASSO machine learning algorithms, we screened the signature genes and constructed a diagnostic model using lightGBM. The model was further validated in a test cohort. Additionally, we extracted hub biomarkers using the protein-protein interactions method and validated them in a single-cell RNA-seq dataset. Results: Our analysis revealed the identification of 37 AD-related peripheral blood signature genes, with their main enrichment in ribosome-related biological functions. Four core biomarkers, RPL24, RPL5, RPS27A, and RPS4X, were identified and exhibited good diagnostic power in the testing cohort. Immune infiltration analysis revealed a higher proportion of CD4+ T cells in AD patients' peripheral blood compared to healthy controls, with a negative correlation with the four ribosome-associated core genes. Validation in a single-cell RNA-seq dataset confirmed these findings. Conclusions: Ribosomal family proteins have the potential to serve as biomarkers for the diagnosis and treatment of AD, and are associated with CD4+ T cell activation.

Keywords: Neurodegenerative diseases, machine learning algorithms, bioinformatics, serum biomarker, single-cell sequencing

Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder that is characterized by progressive cognitive decline and memory impairment. The typical pathological hallmarks of AD include extracellular plaques of amyloid β (Aβ) depositions and neurofibrillary tangles, which are formed due to hyperphosphorylation of microtubule-associated tau protein. These pathological changes are accompanied by neuronal loss, synaptic degeneration, glial cell activation, and neuroinflammation [1]. The burden of AD on patients and their families is significant, and it has become a significant social problem. Unfortunately, there is currently no cure for AD, and available drugs can only delay the progression of the disease. Therefore, early diagnosis and intervention are critical. Imaging techniques, such as positron emission tomography (PET) and magnetic resonance imaging (MRI), are commonly used for the early diagnosis of AD. However, the high cost of PET and the low specificity of MRI present significant challenges to their practical application [2].

Recently, molecular biomarkers that reflect changes in the structure and function of neurons have been discovered, which has greatly improved the efficiency of AD diagnosis. For example, cerebrospinal fluid biomarkers, such as Aβ40, Aβ42, and tau, can directly reflect AD neuropathology [3]. However, the invasive nature of cerebrospinal fluid collection and the associated risks limit their use for screening healthy individuals. As a result, there is a grow-

ing need to identify biomarkers that are specific, sensitive, and minimally invasive.

Recent research has examined the use of peripheral blood biomarkers for AD diagnosis [4]. However, the abundance of biomarkers in the blood that reflect neuropathological damage to brain tissue is extremely low due to the blood-brain barrier. Traditional techniques, such as enzyme-linked immunosorbent assays and electrochemiluminescence, are not sensitive enough for quantitative and efficient detection. Therefore, the value of blood indicators in AD diagnosis is limited. However, the development of advanced detection techniques, such as genomic, proteomic, and metabolomic analysis, is anticipated to lead to the identification of reliable and widely applicable peripheral blood biomarkers for early screening of AD.

AD is a complex disease that involves interactions between multiple factors, and it is challenging to understand using simplified models. Machine learning has emerged as a powerful tool for identifying patterns in complex data and can be used to diagnose and determine prognoses in AD. Previous studies have demonstrated the application of machine learning in identifying potential biomarkers for AD by analyzing gene expression profiles and constructing differential co-expression networks [5]. Other studies have explored the role of long non-coding RNA expression in AD pathogenesis [6]. Despite these promising findings, the application of machine learning in AD diagnosis and prognosis remains in its early stages.

In this study, we used transcriptome and single-cell transcriptome data from public databases to identify serum biomarkers for AD using a range of bioinformatics methods and machine learning algorithms. We also investigated immune infiltration in AD. **Figure 1** provides an overview of our research methodology.

Methods

Datasets preprocessing in training, testing, and validation cohorts

Four datasets related to Alzheimer's disease were acquired through the Gene Expression Omnibus (GEO). All datasets were collected from blood samples taken from patients with AD or healthy controls. Two of these datasets, GSE63060 and GSE63061, were selected for use as the training cohort. Both datasets were based on the GPL6947 chip and were derived from peripheral blood samples. GSE63060 includes 145 AD samples and 104 control samples, while GSE63061 has 139 AD samples and 134 control samples [7]. The RNA-Seq dataset GSE140829 was used for the testing cohort, consisting of 204 AD blood samples and 249 healthy control samples [8]. Additionally, we used a single-cell RNA-Seq dataset GSE181279, which contains 36,849 peripheral blood mononuclear cells from three AD and two healthy control samples, as the validation cohort [9]. Detailed information about the datasets used in this study is shown in **Table 1**. The clinical baseline characteristics of the AD and control groups from the training and testing cohorts are displayed in **Table 2**.

To begin the analysis, the microarray gene datasets GSE63060 and GSE63061 were downloaded, normalized, and log2-transformed using the "limma" package in R software (version 4.2). The raw RNA-Seq dataset GSE1408-29 was annotated using the human genome to build 38 references and formatted to transcripts per million (TPM) data for subsequent analysis. Finally, samples in the single-cell RNA-Seq (scRNA-Seq) dataset GSE181279 were normalized, and batch effects were eliminated using the canonical correlation analysis in the "Seurat" R package (version 4.0) [10].

Filtering for differentially expressed genes

To identify differentially expressed genes (DEGs) from the GSE63060 and GSE63061 datasets, we applied the "sva" package [11] for data integration and batch correction. The "limma" package [12] was then utilized to analyze the expression data for DEG identification. Our criteria for DEG screening were set at $|log2Foldchange| > 1$ and adjusted $P < 0.05$.

Features selection using Boruta and LASSO

In this section, we describe the methods used to select relevant features for further analysis. Firstly, we applied the Boruta algorithm to exclude genes that have low correlation with AD. To accomplish this, we reshuffled the original features to obtain new features and trained a random forest model based on the input features. The importance of each feature was then

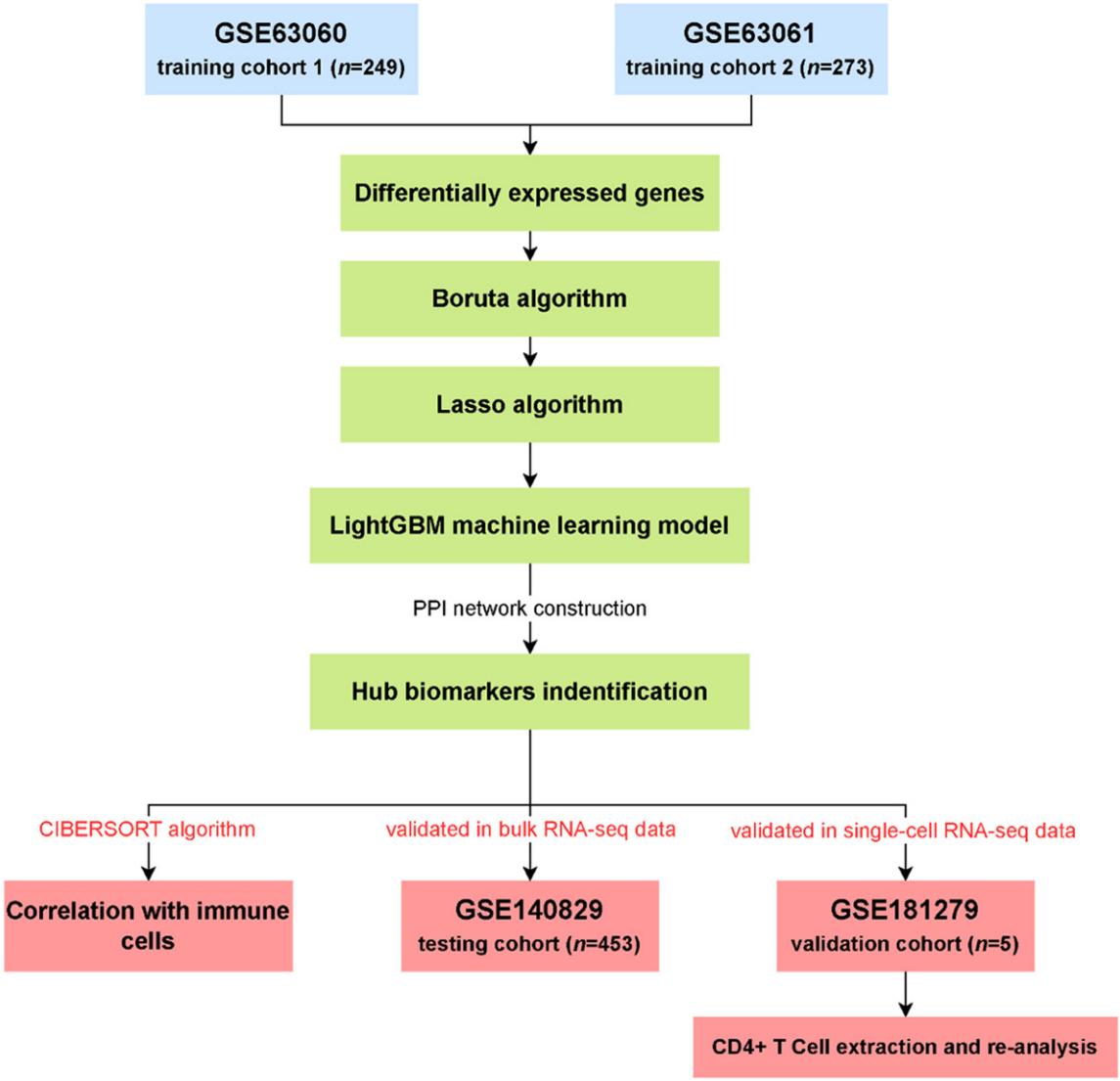

Figure 1. The flow chart of this study.

calculated, and relevant features were selected from the original features using the feature importance in the new features as a reference. The iterations were stopped after all features were judged to be important or useless. This process helped to identify signature genes that may be associated with the onset and development of AD.

Next, we continued the screening of signature genes using the LASSO algorithm, which builds a regularized linear model and filters important features for classification by eliminating features that are worthless or redundant. To assess the discriminative ability of the hub genes, we applied principal component analysis (PCA).

To implement these methods, we utilized the "Boruta" (v7.0) [13] and "glmnet" (v4.1) [14] R packages for filtering and analyzing genetic data. By conducting this analysis, we aimed to select genes that are most relevant to AD, and that may be used as potential biomarkers or therapeutic targets.

Construction and validation of the diagnostic model in the testing cohort

After selecting signature genes using the Boruta and LASSO algorithms, we constructed an Alzheimer's disease diagnosis model using the lightGBM algorithm. To evaluate the model's performance, we utilized the testing cohort GSE140829 and conducted PCA to estimate

Table 1. The detailed information for gene datasets used in this study

GEO accession	Platform; Data type	Sample (number)	Tissue	Attribute
GSE63060	GPL6947; bulk RNA-seq	AD (145); Control (104)	Blood	Training cohort
GSE63061	GPL10558; bulk RNA-seq	AD (139); Control (134)	Blood	Training cohort
GSE140829	GPL15988; bulk RNA-seq	AD (204); Control (249)	Blood	Testing cohort
GSE181279	GPL24676; scRNA-seq	AD (3); Control (2)	Blood	Validation cohort

AD: Alzheimer's disease; scRNA: single-cell RNA; GPL6947 Illumina HumanHT-12 V3.0 expression beadchip; GPL10558 Illumina HumanHT-12 V4.0 expression beadchip; GPL15988 HumanHT-12 v4 Expression beadchip; GPL24676 Illumina NovaSeq 6000 (Homo sapiens).

Table 2. The baseline characteristics of AD and control groups from training and test cohorts

		Training cohort			Testing cohort		
		AD	Control	*P* value	AD	Control	*P* value
n		284	238		204	249	
Gender (%)	Female	184 (64.8)	143 (60.1)	0.31	104 (51.0)	139 (55.8)	0.35
	Male	100 (35.2)	95 (39.9)		100 (49.0)	110 (44.2)	
Age (Year, mean (SD))		76.62 (6.73)	74.02 (6.32)	0.112	73.00 (7.09)	73.66 (6.25)	0.288

AD: Alzheimer's disease; SD: standard deviation.

the model's ability to distinguish Alzheimer's disease samples from healthy controls. We used the "lightGBM" R package (v.3.3.2) [15] for model construction and receiver operating characteristic (ROC) analysis to evaluate the model's performance.

Hub biomarker identification

In order to identify the core biomarkers associated with AD, we conducted a protein-protein interaction (PPI) network analysis using genes from the diagnostic model. The String online tool (https://www.string-db.org/, v11.5) was employed with a high confidence score of 0.7 to construct the PPI network. Subsequently, we performed enrichment analysis to identify genes that play pivotal roles in the network. To identify the hub biomarkers, we conducted molecular complex detection (MCODE) analysis on the PPI network. The predictive power of these hub biomarkers was then analyzed and demonstrated in the training and testing cohorts. Additionally, we presented the expression profile of each core gene in different cohorts for both AD and control data using box plots.

Immune cell infiltration and association with hub biomarkers

To further explore the relationship between hub biomarkers and immune cell infiltration in AD, we utilized the CIBERSORT algorithm [16]. To analyze immune cell fraction in the merged gene expression matrix of GSE63060 and GSE63061. The gene expression matrix of the training dataset was uploaded to the official online platform (https://cibersortx.stanford.edu/) to estimate the immune cell infiltration in each sample. Immune cells with *P*-values less than 0.05 were selected for further analysis. We then calculated and demonstrated the correlation between immune cell infiltration and the expression of hub biomarkers.

Validation in sing-cell RNA seq dataset

To validate the results obtained from bulk RNA-seq data, we further analyzed scRNA-Seq data. To begin with, we used the "Seurat" package to normalize the scRNA-Seq data and identify highly variable features for each sample using the "FindVariableFeatures" function. We then employed the "FindIntegrationAnchors" method to find anchors based on variable features, and used the "IntegrateData" function to integrate the five scRNA-Seq samples. The data were scaled, and the cells were clustered and analyzed using the uniform manifold approximation and projection for dimension reduction (UMAP) algorithm. Cell types for cell clustering were identified using the CellMarker database [17] and the "singleR" package (v1.10) [18].

In addition, we calculated the proportions of various cell types in different samples and demonstrated expression differences of hub

biomarkers across samples and different cell types.

Biological enrichment investigation in CD4+ T cells

To explore the potential mechanisms in CD4+ T cells, we extracted this subset cell type from the scRNA-Seq dataset and performed gene set enrichment analysis (GSEA). Unlike focusing on the enrichment of a few significantly different genes, GSEA sorts all the different genes between the two groups of samples by logFC value and examines the overall trends of enrichment in the gene set. This approach helps to prevent important but non-significant biological traits from being overlooked. Terms with adjusted *P*-values < 0.05 were considered significant in our analysis.

Results

Important feature genes associated with AD

We initially identified 267 differentially expressed genes (DEGs) from the gene expression matrix in the two combined training cohorts, which consisted of 131 upregulated and 136 downregulated DEGs (Table S1; **Figure 2A**).

To refine the selection of important feature genes, we applied the Boruta algorithm and identified 39 genes as significant feature genes in the blood of AD (Table S2; **Figure 2B**). To further narrow down the selection, we employed the LASSO algorithm and selected 21 hub genes, which were then used to construct a gene model (Table S3; **Figure 2C**, **2D**). The PCA plots demonstrated that these hub genes effectively discriminated between samples from AD patients and healthy controls (**Figure 2E**).

Construction of gene model and validation in testing cohort

The hub genes identified from the LASSO analysis were utilized to construct a diagnostic model for AD using the lightGBM machine learning algorithm (Table S4). Subsequently, we evaluated the model's diagnostic performance. In the training cohort, the area under the ROC curve (AUC) and precision-recall curves were 0.994 and 0.997, respectively (**Figure 3A**, **3B**). In the testing cohort, the AUC and precision-recall curves were 0.76 and 0.79, respectively (**Figure 3C**, **3D**). These results indicate that the diagnostic model possessed satisfactory diagnostic capability.

Hub biomarkers identification

The gene module was input into the STRING platform to construct a protein-protein interaction (PPI) network. The biological pathways mainly enriched in the PPI network were ribosomal pathways (Figure S1; **Figure 3E**). Subsequently, the MCODE algorithm identified four hub biomarkers (RPL24, RPL5, RPS27A, and RPS4X) (**Figure 3F**). The ROC curves showed that these ribosomal biomarkers had good diagnostic abilities for AD, with AUC values of 0.745 for RPS27A, 0.683 for RPL5, 0.705 for RPL24, and 0.661 for RPS4X (**Figure 3G**, **3H**).

To further investigate the role of these ribosomal biomarkers in AD, we compared the expression levels of these biomarkers between healthy controls and patients with AD in the training cohort. The results showed that the expression levels of these ribosomal biomarkers were significantly higher in healthy controls than in patients with AD (P < 2.2e-16 for RPS27A, P = 5.4e-13 for RPL5, P = 8.1e-16 for RPL24, and P = 2.5e-10 for RPS4X) (**Figure 3I-L**). A similar trend was observed in the testing cohort (P = 0.028 for RPS27A, P = 0.0027 for RPL5, P = 0.03 for RPL24, and P = 0.0041 for RPS4X) (**Figure 3M-P**).

Immune cell infiltration in the training cohort

Using the CIBERSORT analysis, we examined the proportion of seven different immune cells and found statistically significant differences between the control and AD groups. Our results revealed that dendritic cells, M2 macrophages, and CD8+ T cells had higher proportions in control samples than in AD samples, while mast cells, natural killer cells, and CD4+ T cells had a higher proportion in the AD samples (**Figure 4A**).

Furthermore, we evaluated the relationships between the hub biomarkers and immune cells using correlation heatmaps. Our analysis showed that RPL24, RPL5, RPS27A, and RPS4X were positively correlated with dendritic cells,

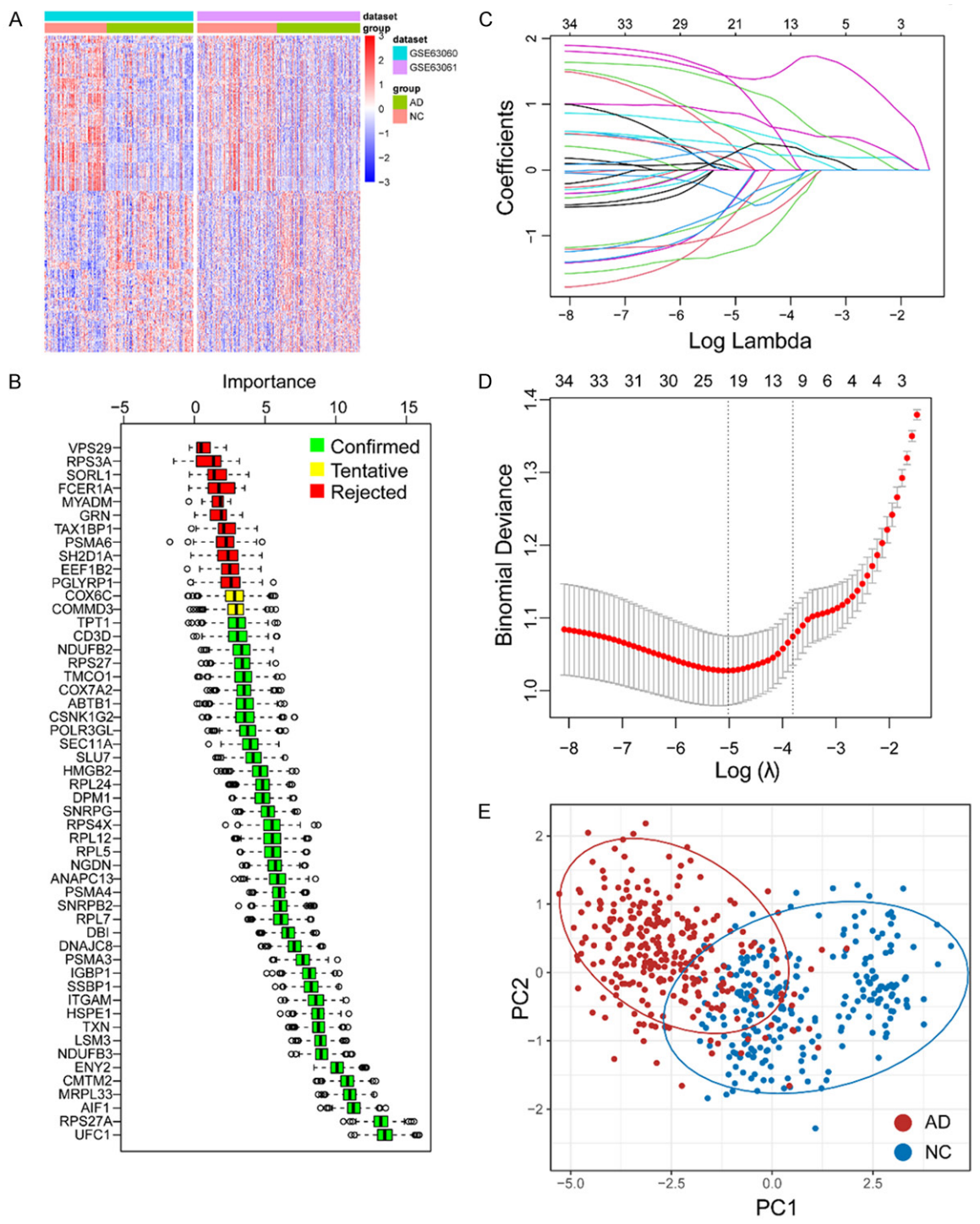

Figure 2. Identification of important genes associated with AD. A. The heatmap of differentially expressed genes in combined training cohort (GSE63060 and GSE63061). B. Confirmed important genes associated with AD by using Boruta algorithm. C, D. LASSO algorithm was applied for further narrow down the important genes related to AD. E. PCA plot has shown the ability of the hub genes to distinguish between AD and normal groups. Abbreviation: AD: Alzheimer's disease; NC: normal control.

M2 macrophages, and CD8+ T cells, whereas they all showed a negative correlation with mast cells, natural killer cells, and CD4+ T cells (**Figure 4B**).

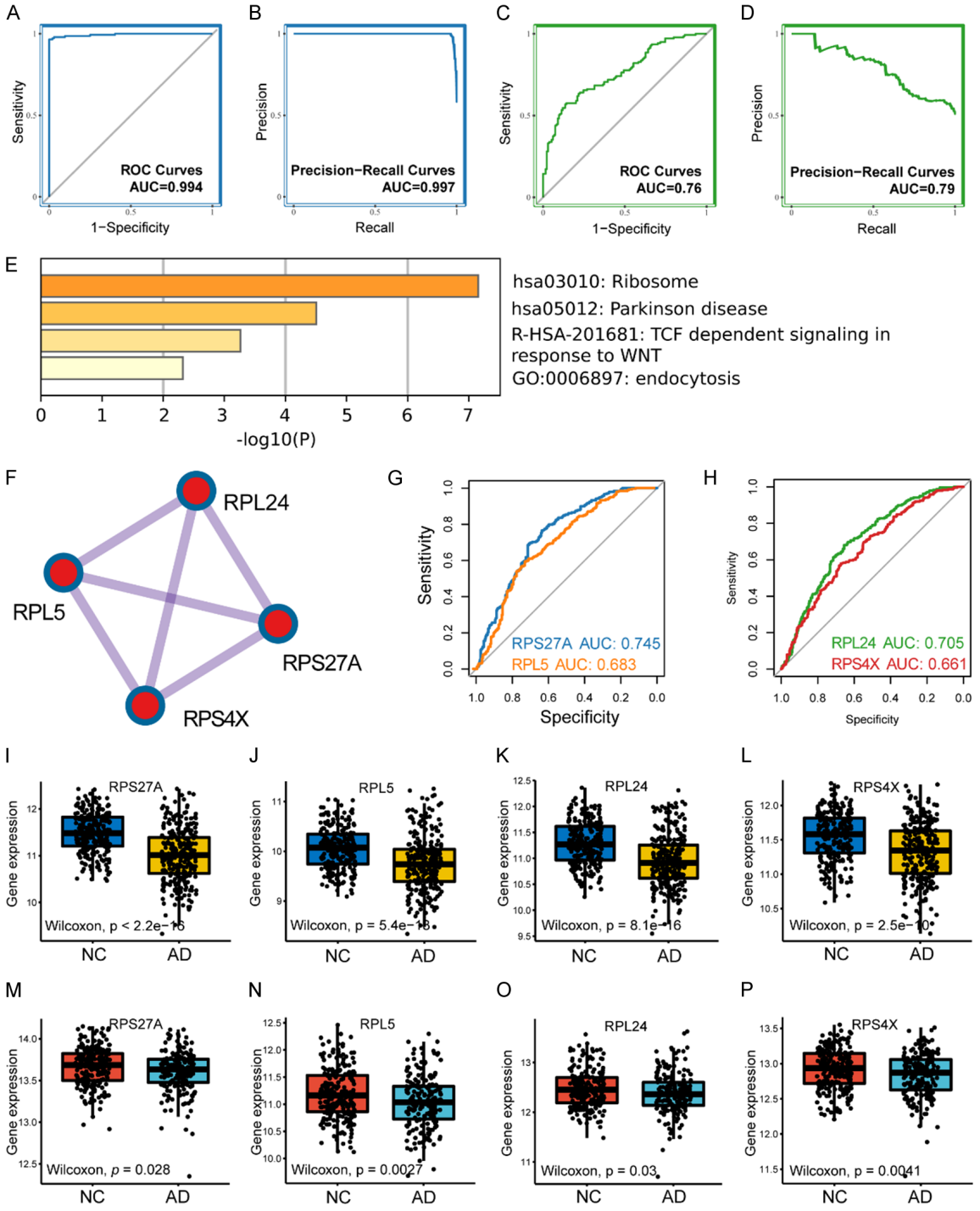

Figure 3. Gene model validation and hub biomarkers identification. Receiver operator characteristic curve of gene model in training cohort (A) and testing cohort (C). Precision-recall curve of gene model in training cohort (B) and testing cohort (D). (E) The bar plot has displayed the results of functional enrichment analysis of the gene model. (F) Hub biomarkers identified by molecular complex detection algorithm. (G) ROC analysis of RPS27A and RPL5 in testing cohort. (H) ROC analysis of RPL24 and RPS4X in testing cohort. The gene expression profile of RPS27A (I), RPL5 (J), RPL24 (K) and RPS4X (L) in training cohort. The boxplots have demonstrated the expression of RPS27A (M), RPL5 (N), RPL24 (O) and RPS4X (P) in testing cohort. Abbreviation: ROC: receiver operator characteristic; AUC: area under curve; AD: Alzheimer's disease; NC: normal control.

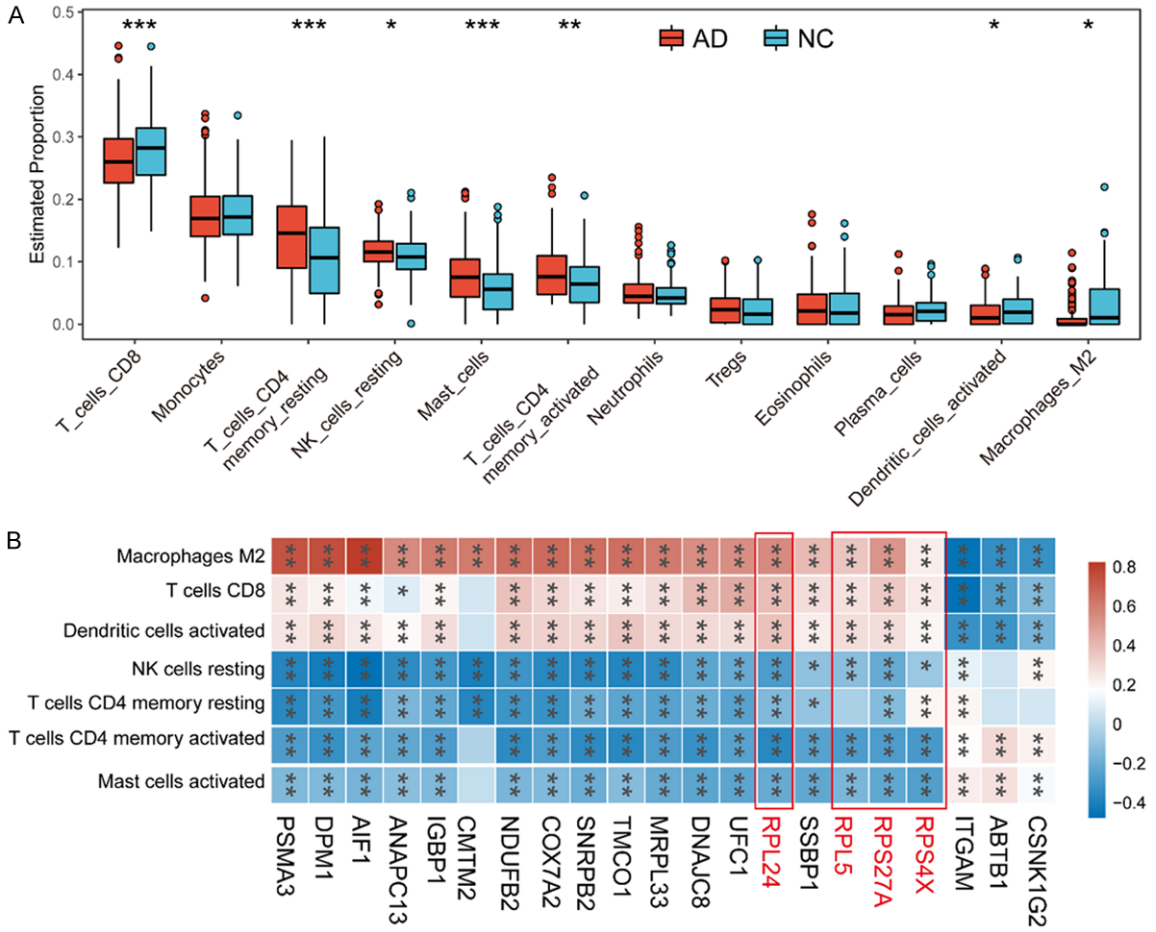

Figure 4. Immune cell infiltration in training cohort. A. The boxplot has shown the proportions of different types of immune cells in AD and normal control samples. B. Correlation heatmap has displayed the association between the expression of hub biomarkers and immune cell infiltration. Abbreviation: AD: Alzheimer's disease; NC: normal control.

Validation using scRNA-Seq data

Following pre-processing and application of the UMAP algorithm, we identified 19 cell clusters (Figure S2), and ultimately identified five different cell types, including monocytes, B cells, natural killer cells, CD8+ T cells, and CD4+ T cells (**Figure 5A**). Notably, CD4+ T cells were the most abundant cell type and had a higher proportion in AD samples (63.08%) compared to healthy controls (49.34%), suggesting a potentially significant role in AD progression (**Figure 5B**).

Subsequent analysis revealed that the expression levels of RPL24, RPL5, RPS27A, and RPS4X were higher in healthy controls compared to patients with AD (**Figure 5C**). Furthermore, violin plots demonstrated that these hub biomarkers were primarily expressed in CD4+ T cells, indicating the important role of ribosomal genes in CD4+ T cells (**Figure 5D**).

Specifically, all four biomarkers showed higher expression levels in the control group in CD4+ and CD8+ T cells and B cells, with RPL5, RPL24, and RPS4X significantly more highly expressed in natural killer cells of the control group. The expression pattern of these four gene markers in monocytes varied, with RPS27A and RPL24 being highly expressed in the AD group and RPS4X showing lower expression (**Figure 6A-D**).

Biological functions and pathway enrichment in CD4+ T cells

Based on the results of GSEA the top five GO and KEGG terms were selected and presented. As shown in **Figure 7A**, in patients with AD, several upregulated biological processes (BP) were

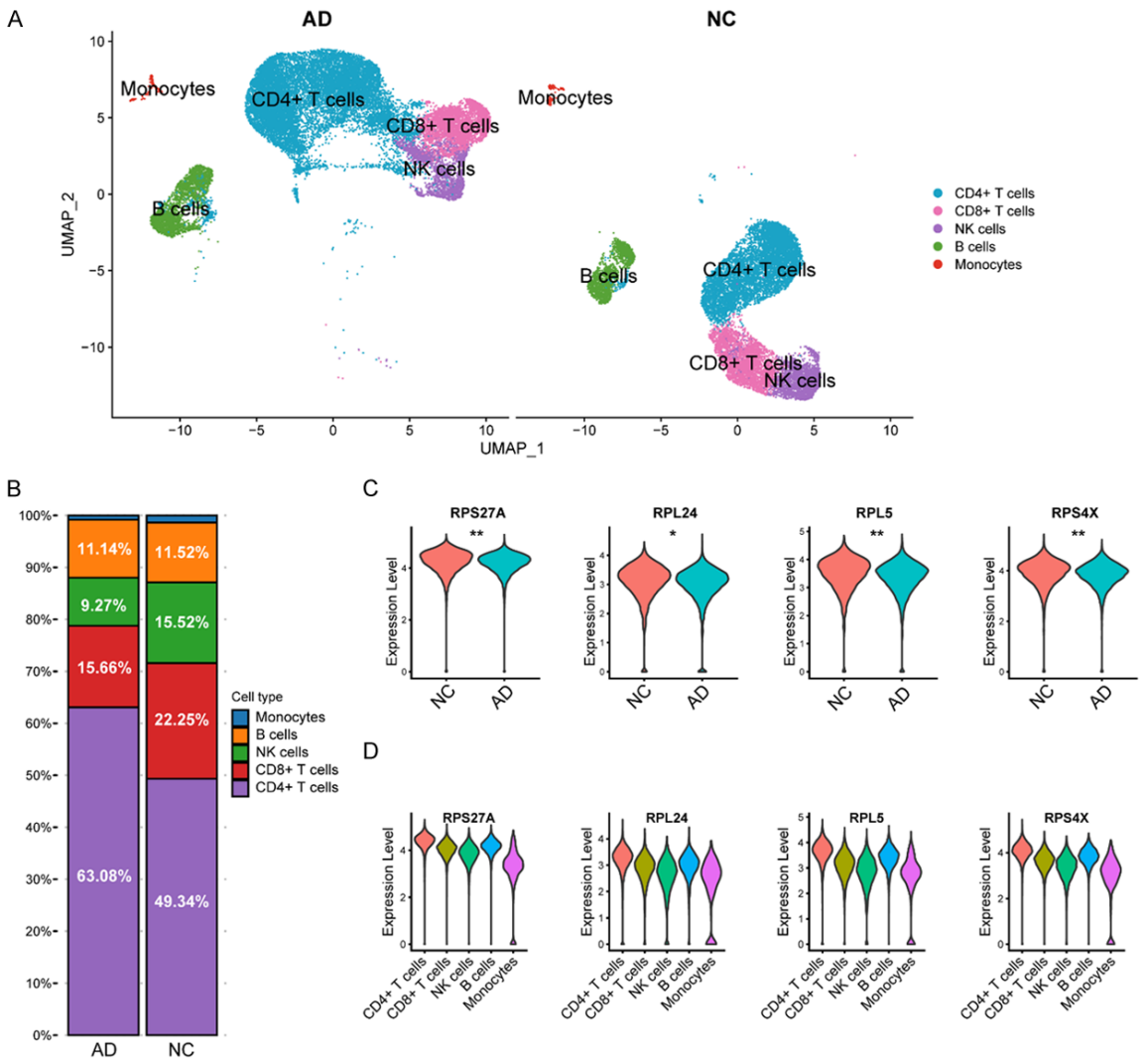

Figure 5. Validation in single-cell RNA-Seq dataset. A. Identification of varied cell types in the single-cell RNA-Seq dataset by using UMAP algorithm. B. Proportions of different cell types in AD and normal control samples. C. The violin diagram shows the differences in the expression of RPS27A, RPL24, RPL5 and RPS4X in AD and normal samples (Wilcoxon t-test, $^{*}P < 0.05$, $^{**}P < 0.01$). D. The expression profile of RPS27A, RPL24, RPL5 and RPS4X in different types of cells in scRNA-Seq dataset. Abbreviation: AD: Alzheimer's disease; NC: normal control.

identified, including "regulation of leukocyte chemotaxis", "regulation of cation transmembrane transport", and "regulation of transmembrane transport". In contrast, "cytoplasmic translation" and "ribosomal large subunit biogenesis" were significantly downregulated. In terms of cellular components (CC), terms with significant statistical differences in the AD group were mainly related to ribosome function and included "cytosolic large ribosomal subunit" and "cytosolic ribosome" (**Figure 7B**). Patients with AD showed upregulated "protein serine threonine phosphatase activity" in molecular functions (MF), whereas RNA synthesis functions were downregulated, including "translation regulator activity nucleic acid binding" and "translation regulator activity" (**Figure 7C**). Additionally, pathway analysis revealed that the MAPK signaling and T-cell differentiation pathways were upregulated in the AD group, while the "ribosome" pathway was downregulated, consistent with the results of GO analysis (**Figure 7D**).

Discussion

Alzheimer's disease (AD) is a multifaceted and intricate process that involves multiple sys-

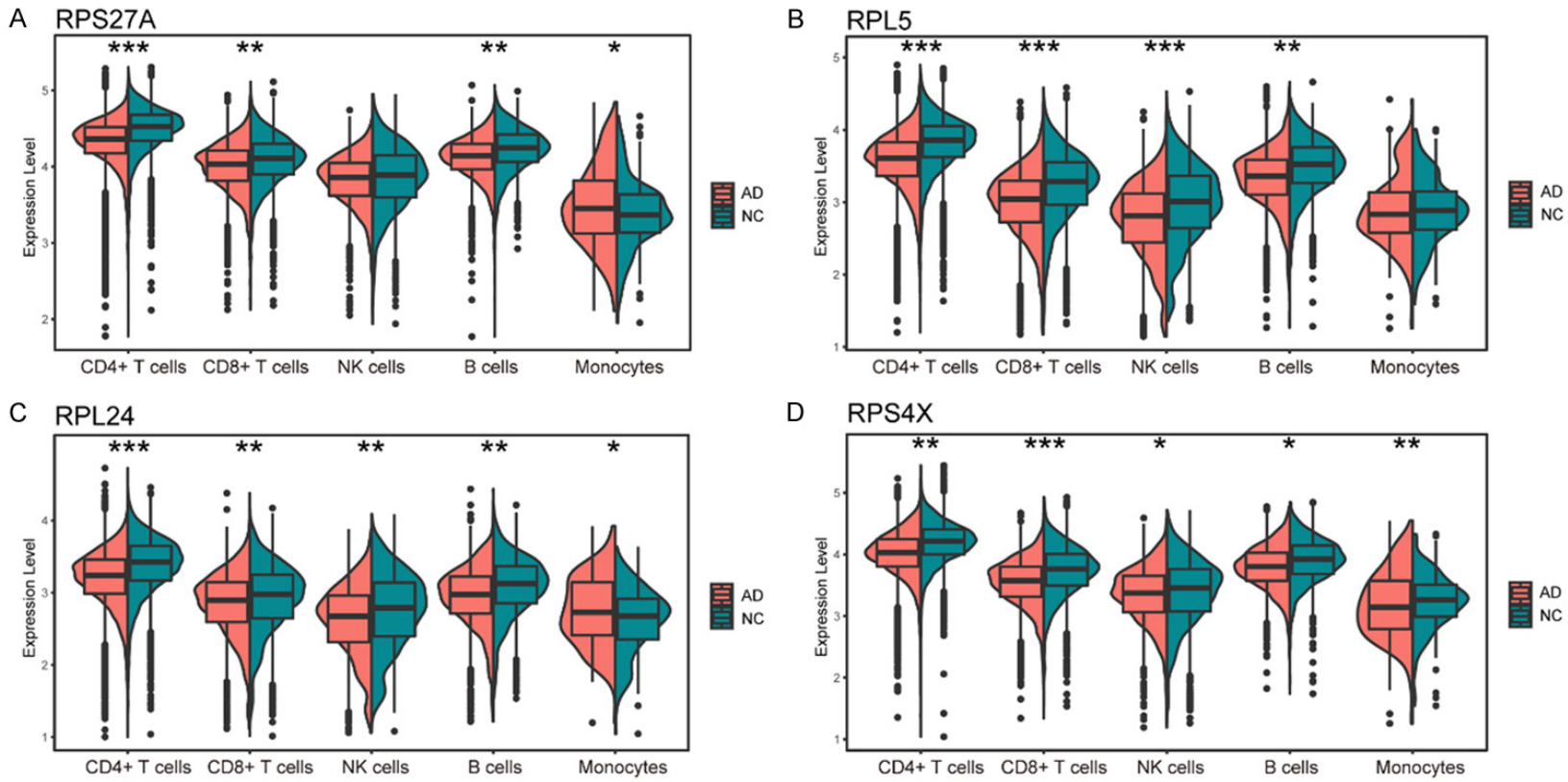

Figure 6. Violin diagrams show the differential expression of RPS27A (A), RPL5 (B), RPL24 (C) and RPS4X (D) in various cell types in AD patients and normal controls, respectively. (Wilcoxon t-test, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$). Abbreviation: AD: Alzheimer's disease; NC: normal control.

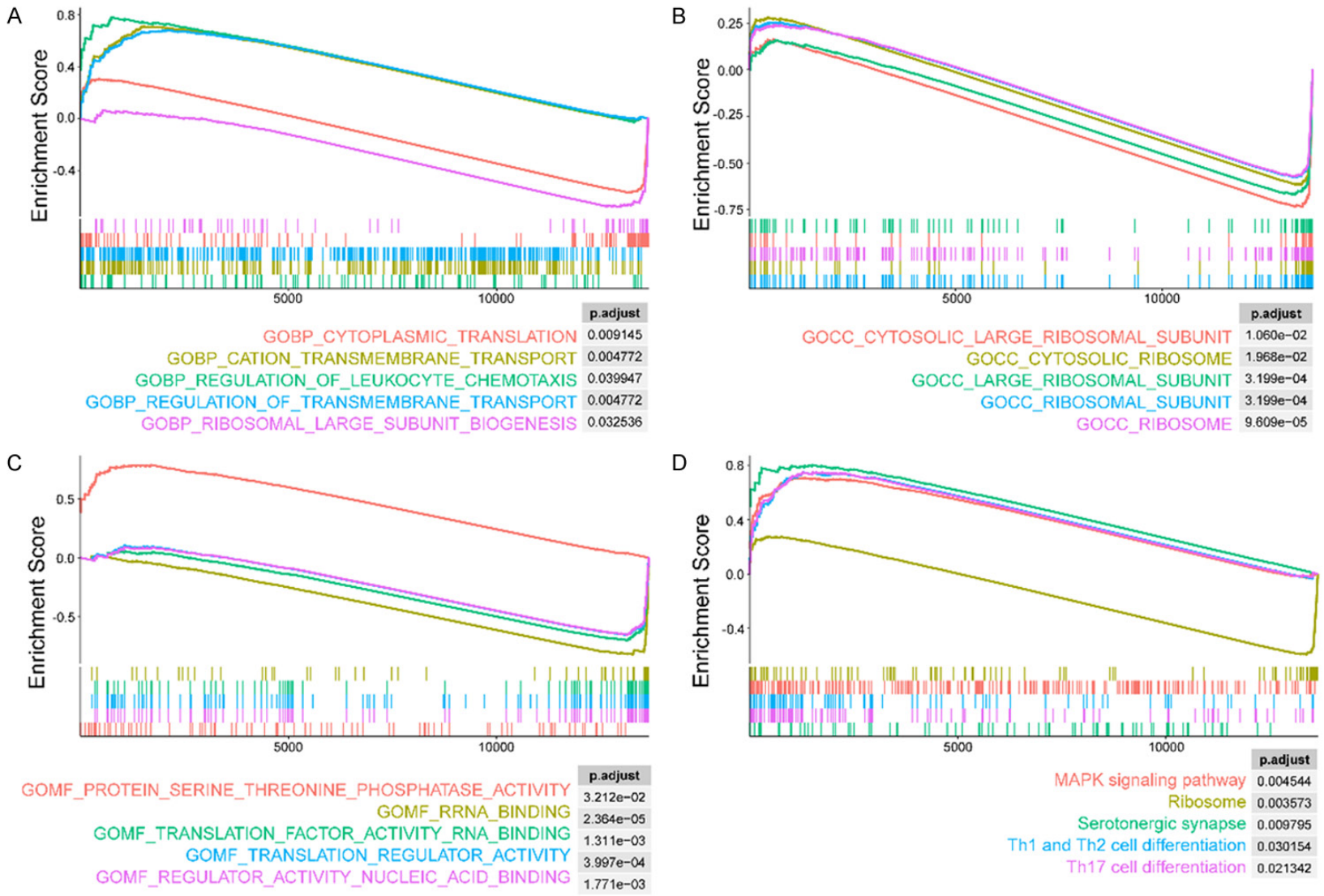

Figure 7. The GSEA plots have shown the biological function and pathway enrichment of DEGs in CD4+ T cells in the peripheral blood of AD patients and normal control. By using gene ontology gene sets, we conducted enrichment analysis of BP (A), CC (B) and MF (C) for DEGs and only the top 5 terms were displayed. For explore the molecular pathway enrichment for DEGs, the KEGG analysis was introduced, and the top 5 enriched pathway were demonstrated (D). Abbreviation: BP: biological process; CC: cellular component; MF: molecular function; DEGs: differentially expressed genes.

tems, making it challenging for a single biomarker to capture the entire pathological process. Several studies have highlighted the importance of combining the detection of multiple blood biomarkers to significantly enhance diagnostic efficiency for AD [3, 9, 19]. High-throughput sequencing technologies offer tremendous opportunities for investigating AD. Transcriptome sequencing can be employed to identify differences in mRNA expression at the transcriptome level, and follow-up studies can subsequently validate the findings using protein analyses. However, given the high-dimensional gene expression profile data, an increasing number of researchers are turning to machine learning rather than traditional statistical techniques for data analysis to effectively reveal the biological properties of AD [20].

In this study, we employed the Boruta algorithm to identify characteristic genes associated with AD. Compared to traditional feature-selection algorithms, the Boruta algorithm is known to provide superior results in identifying the importance of variables. In the field of AD research, the Boruta algorithm has been widely used for feature selection in multi-omics data. For example, it can be used to screen biomarkers of various neurodegenerative diseases at the microRNA level to determine the characteristics of these diseases [21]. Additionally, the Boruta feature filtering method has been employed in a study to analyze the characteristics of each brain region according to the systematic methylation map in patients with AD [22]. The study revealed important methylation signatures that contribute to the development of the disease. Moreover, the Boruta algorithm has also been applied to single nucleotide polymorphism data to aid in the early detection of AD [23]. The above studies demonstrate the versatility and wide applicability of the Boruta algorithm in the context of AD research.

Next, we utilized the LASSO algorithm to further refine characteristic genes and employed the lightGBM algorithm to construct a gene model with potential relevance to AD. The performance of this model was assessed using ROC analysis, which demonstrated favorable predictive ability in both the training and testing sets, suggesting that it may be a valuable tool for identifying patients with AD. Furthermore, we utilized the MCODE algorithm to identify the core genes of the model, and found that these genes were primarily involved in the regulation of ribosomal activity. In cells, ribosomes play a central role in the translation of proteins from mRNA, but with age, ribosomal function deteriorates, leading to an increase in defective proteins [24]. Synthesis of new proteins is a critical process for neuronal activity-dependent learning and memory, with ribosome biosynthesis being the key rate-limiting step in intracellular protein translation [25]. Multiple studies have reported a reduction in ribosome numbers and increased oxidation of ribosomal RNA in the brains of AD patients, resulting in reduced protein synthesis viability due to ribosomal dysfunction, which is a key feature of metabolic disorders in the AD brain [26]. A decrease in total ribosomal RNA and total RNA can be observed in the cerebrum of AD patients [27]. Oxidative stress in the cerebrum of AD patients has also been found to significantly impair the synthesis of 5S ribosomal RNA, which is responsible for ribosome stability, and specifically impairs ribosome function [28]. Animal studies have shown that Aβ injections lead to protein synthesis impairment by heavy polyribosomes formation in the hippocampus of the AD rat model [29]. Similarly, pathological tau proteins can bind to ribosomes and damage RNA translation [30]. Impaired protein synthesis due to ribosome dysfunction has been observed not only in patients with AD, but also in various other neurodegenerative pathologies [31].

Furthermore, we identified four core genes, namely RPL24, RPS27A, RPL5, and RPS4X, that exhibited lower expression levels in AD. These genes showed a strong ability for the diagnosis of AD as suggested by the ROC analysis. RPL24 is located in the cytoplasm and is an essential constituent of the 60S subunit of ribosomes. Its primary biological functions include RNA binding and forming the structure of the ribosomal protein complex. Previous studies have demonstrated that RPL24 can inhibit translation elongation and improve protein synthesis homeostasis, thereby inhibiting protein synthesis in tumor cells [32, 33]. Nevertheless, the understanding of the role of RPL24 in AD is still limited. RPS27A, a member of the 40S subunit of the ribosome, is one of the core genes identified in multiple bioinformatics studies investigating mild cogni-

tive impairment, and its levels are significantly altered in the blood of AD patients [34, 35]. Recent studies have shown that RPS27A, as part of a triplet with interleukin (IL)-18 and CX3CL1, acts as a potential upstream regulator in microglia cells to reduce IL-18 and alleviate neurodegenerative diseases [36]. As one of the components of the 60S subunit of the ribosome, RPL5 has been shown to inhibit tumorigenesis by activating downstream tumor suppressor factors and downregulating oncoprotein expression [37]. RPL5 can also suppress breast cancer cell growth by regulating E2F transcription factor 1 and endoplasmic reticulum stress of tumor cells [38]. However, the role of RPL5 in AD remains unclear. Previous research has shown that numerous ribosomal genes, including RPS4X, are highly assembled and translated in axons far from the cellular body of neurons to maintain local ribosomal function [39]. Interestingly, proteomic analysis of high-purity cerebral capillaries isolated from the gray and white matter of four donors with AD and three controls showed that RPS4X was upregulated in AD brain vessels, but no significant difference was observed in brain parenchyma [40]. This finding may explain the low expression of RPS4X in the peripheral blood of AD patients.

Further immune infiltration analysis was conducted on the training set and demonstrated that the level of CD4+ T cells was significantly higher in patients with AD as compared to the control group, and this showed a negative correlation with the previously identified ribosomal biomarkers. These findings suggest that the low expression of key ribosomal genes in AD is associated with the accumulation of CD4+ T cells. To validate this observation, peripheral blood scRNA-Seq data from the AD and control groups were utilized as a validation set, which also showed significant enrichment of CD4+ T cells and reduction of ribosomal key genes in AD, thereby corroborating the results obtained from the bulk RNA-Seq data. Previous studies have reported that the increased levels of activated CD4+ T cells and CD8+ T cells in the peripheral blood of AD patients are closely linked to cognitive deficits and magnetic resonance imaging changes in specific brain regions [41]. CD4+ T cells can aggravate or alleviate AD symptoms based on their infiltrating subgroups and constitute a major source of pro-inflammatory cytokines that decrease endothelial integrity and stimulate astrocytes, leading to Aβ production [42]. Furthermore, *in vitro* studies have demonstrated that β-secretase 1 levels in 5xFAD-transgenic mice were higher in CD4+ T cells, and its activation was enhanced [43, 44]. Recent studies have demonstrated the significance of ribosomal proteins in the regulation of the immune system. Immune cells, such as T cells and B cells, require high levels of protein synthesis to support their proliferation and activation during immune responses, and ribosomal proteins play a vital role in this process. Dysregulation of ribosomal proteins can affect the immune response, and certain ribosomal proteins, such as RPL5 and RPL11, can interact directly with the p53 tumor suppressor protein to regulate the expression of genes involved in cell growth and proliferation [45]. Moreover, ribosomal stress caused by ribosomal protein dysfunction can activate the p53 pathway and induce cell cycle arrest or apoptosis, resulting in the elimination of potentially harmful cells.

Evidence suggests that the RPS4X gene is associated with immune cell infiltration. For instance, the RPS4X gene is upregulated in peripheral blood mononuclear cells (PBMCs) from patients with systemic lupus erythematosus (SLE), a disease characterized by chronic immune cell infiltration and inflammation [46]. Additionally, RPS4X expression is positively correlated with disease activity and the number of infiltrating immune cells. Another study found that RPS4X expression was significantly downregulated in CD4+ T cells from multiple sclerosis (MS) patients compared with healthy controls. MS is a chronic inflammatory disease of the central nervous system characterized by immune cell infiltration, and RPS4X expression was inversely correlated with disease severity and the frequency of infiltrating T cells in the brain [47]. Furthermore, RPS27A has been shown to modulate immune responses in a variety of disease settings by regulating the recruitment and activity of tumor-associated macrophages, which play a crucial role in shaping the tumor microenvironment and promoting cancer progression [48]. Studies have shown that RPL5 gene expression is related to immune cell infiltration in different types of cancer. For example, in colorectal cancer, high RPL5 expression is associated with increased

infiltration of CD8+ T cells and natural killer cells, as well as improved patient survival [49]. However, the role played by the interaction between ribosomes and CD4 cells in the pathogenesis of AD remains unclear. A possible explanation is that as the degree of ageing and inflammation increases in AD patients, abnormalities in the translation of proteins by ribosomes occur, leading to the activation of CD4 cells [50].

In this study, we utilized the GSEA algorithm to examine the functional enrichment of CD4+ T cells from patients with Alzheimer's disease (AD). Our analysis revealed that ribosomes and their synthesis-related functions and pathways were inhibited in these cells. Previous studies have highlighted the critical role of ribosomes in the protein production of effector CD4+ T cells following TCR stimulation [51]. Moreover, ribosomal proteins (RPs) have been linked to a variety of physiological and pathological processes, including the regulation of T-cell development and immune-related diseases [52]. For instance, Noc4L-mediated ribosome biogenesis controls the activation of regulatory T cells (Tregs) and maintains immune homeostasis [53]. Additionally, post-transcriptional mechanisms have been shown to regulate ribosomes in murine CD4+ T cells after 24 hours of activation [54]. Nonetheless, further studies are needed to comprehensively elucidate the role of RPs in AD and their potential as a therapeutic target.

Furthermore, we observed that leukocyte chemoattraction-related pathways were highly expressed in CD4+ T cells from AD patients. Leukocyte chemotaxis refers to the process by which white blood cells migrate to specific sites of inflammation or infection in response to chemical signals. In AD, leukocyte chemotaxis enhances the recruitment of activated immune cells into the brain via chemotaxis [55]. Aβ has been found to be chemotactic for monocytes and it induces the secretion of proinflammatory cytokines and chemokines in the periphery as well as the brain [56]. Additionally, microglia and monocyte-derived cells have been shown to play an important role in promoting proinflammatory and neurotoxic pathways [57]. Notably, blood samples from AD patients with dementia have revealed neutrophil hyperactivation associated with increased reactive oxygen species production [58].

Overall, ribosome family proteins have been shown to play a role in Alzheimer's disease (AD) pathogenesis. Studies have reported that alterations in ribosomal function and biogenesis are associated with AD [59, 60]. Ribosomal proteins have been identified as potential biomarkers for AD, and changes in the expression levels of certain ribosomal proteins have been observed in AD patients [26]. Additionally, several studies have suggested that ribosomal dysfunction can lead to the accumulation of misfolded proteins, including amyloid-beta (Aβ) and tau, which are hallmarks of AD pathology [29, 61]. Therefore, ribosome family proteins may play an important role in AD pathogenesis, and further research is needed to fully elucidate the mechanisms involved.

However, the present study has certain limitations that must be considered. Firstly, although bioinformatics methods are useful for generating hypotheses, it is essential to verify the findings using *in vivo* or *in vitro* experiments. Secondly, a larger sample size and more extensive clinical data collection could enhance the validity and reliability of the conclusions. Therefore, future studies are warranted to confirm the results of the present investigation and to expand upon the findings reported here.

Conclusions

In conclusion, our study utilized machine learning and bioinformatics approaches to analyze multiple datasets of bulk RNA-Seq and scRNA data, identifying four under expressed ribosome family protein biomarkers (RPL24, RPS27A, RPL5, RPS4X) in AD peripheral blood. Further, our findings suggest that these biomarkers are negatively correlated with CD4+ T cells activation in AD patients.

Acknowledgements

This work was supported by Hangzhou Medical Health Science and Technology Projects (B20210660, B20200287). Additionally, we wish to acknowledge to the patients and researchers who contributed to the Gene Expression Omnibus project.

Disclosure of conflict of interest

None.

Address correspondence to: Lifang Wang, Department of Geriatrics, Zhejiang Aged Care Hospital, Hangzhou Normal University, 21 Jinhua Road, Gongshu District, Hangzhou, Zhejiang, P. R. China. Tel: +86-17600380182; E-mail: Prof.wanglifang@outlook.com

References

[1] Scheltens P, De Strooper B, Kivipelto M, Holstege H, Chetelat G, Teunissen CE, Cummings J and van der Flier WM. Alzheimer's disease. Lancet 2021; 397: 1577-1590.

[2] Dubois B, Villain N, Frisoni GB, Rabinovici GD, Sabbagh M, Cappa S, Bejanin A, Bombois S, Epelbaum S, Teichmann M, Habert MO, Nordberg A, Blennow K, Galasko D, Stern Y, Rowe CC, Salloway S, Schneider LS, Cummings JL and Feldman HH. Clinical diagnosis of Alzheimer's disease: recommendations of the International Working Group. Lancet Neurol 2021; 20: 484-496.

[3] Ossenkoppele R, van der Kant R and Hansson O. Tau biomarkers in Alzheimer's disease: towards implementation in clinical practice and trials. Lancet Neurol 2022; 21: 726-734.

[4] Hansson O. Biomarkers for neurodegenerative diseases. Nat Med 2021; 27: 954-963.

[5] Wang L and Liu ZP. Detecting diagnostic biomarkers of Alzheimer's disease by integrating gene expression data in six brain regions. Front Genet 2019; 10: 157.

[6] Zhou M, Zhao H, Wang X, Sun J and Su J. Analysis of long noncoding RNAs highlights region-specific altered expression patterns and diagnostic roles in Alzheimer's disease. Brief Bioinform 2019; 20: 598-608.

[7] Sood S, Gallagher IJ, Lunnon K, Rullman E, Keohane A, Crossland H, Phillips BE, Cederholm T, Jensen T, van Loon LJ, Lannfelt L, Kraus WE, Atherton PJ, Howard R, Gustafsson T, Hodges A and Timmons JA. A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. Genome Biol 2015; 16: 185.

[8] Nachun D, Ramos E, Karydas A, Dokuru D, Gao F, Yang Z, Van Berlo V, Sears R, Kramer J, Boxer A, Rosen H, Miller B and Coppola G. Systems-level analysis of peripheral blood gene expression in dementia patients reveals an innate immune response shared across multiple disorders. bioRxiv 2019; 2019.2012.2013.875112.

[9] Xu H and Jia J. Single-Cell RNA sequencing of peripheral blood reveals immune cell signatures in Alzheimer's disease. Front Immunol 2021; 12: 645666.

[10] Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P and Satija R. Comprehensive integration of single-cell data. Cell 2019; 177: 1888-1902, e1821.

[11] Leek JT, Johnson WE, Parker HS, Jaffe AE and Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012; 28: 882-883.

[12] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43: e47.

[13] Kursa MB and Rudnicki WR. Feature selection with the boruta package. Journal of Statistical Software 2010; 36: 1-13.

[14] Simon N, Friedman J, Hastie T and Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. J Stat Softw 2011; 39: 1-13.

[15] Yan J, Xu Y, Cheng Q, Jiang S, Wang Q, Xiao Y, Ma C, Yan J and Wang X. LightGBM: accelerated genomically designed crop breeding through ensemble learning. Genome Biol 2021; 22: 271.

[16] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M and Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015; 12: 453-457.

[17] Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, Luo T, Xu L, Liao G, Yan M, Ping Y, Li F, Shi A, Bai J, Zhao T, Li X and Xiao Y. CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res 2019; 47: D721-D728.

[18] Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ and Bhattacharya M. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 2019; 20: 163-172.

[19] Huang Z, Li M, Zhang L and Liu Y. Electrochemical immunosensor based on superwettable microdroplet array for detecting multiple Alzheimer's disease biomarkers. Front Bioeng Biotechnol 2022; 10: 1029428.

[20] Xie C, Zhuang XX, Niu Z, Ai R, Lautrup S, Zheng S, Jiang Y, Han R, Gupta TS, Cao S, Lagartos-Donate MJ, Cai CZ, Xie LM, Caponio D, Wang WW, Schmauck-Medina T, Zhang J, Wang HL, Lou G, Xiao X, Zheng W, Palikaras K, Yang G, Caldwell KA, Caldwell GA, Shen HM, Nilsen H, Lu JH and Fang EF. Amelioration of Alzheimer's disease pathology by mitophagy inducers identified via machine learning and a cross-species workflow. Nat Biomed Eng 2022; 6: 76-93.

[21] Li Z, Guo W, Ding S, Chen L, Feng K, Huang T and Cai YD. Identifying key MicroRNA signatures for neurodegenerative diseases with machine learning methods. Front Genet 2022; 13: 880997.
[22] Li Z, Guo W, Zeng T, Yin J, Feng K, Huang T and Cai YD. Detecting brain structure-specific methylation signatures and rules for Alzheimer's disease. Front Neurosci 2022; 16: 895181.
[23] Ahmed H, Soliman H and Elmogy M. Early detection of Alzheimer's disease using single nucleotide polymorphisms analysis based on gradient boosting tree. Comput Biol Med 2022; 146: 105622.
[24] Stein KC, Morales-Polanco F, van der Lienden J, Rainbolt TK and Frydman J. Ageing exacerbates ribosome pausing to disrupt cotranslational proteostasis. Nature 2022; 601: 637-642.
[25] Yordanova MM, Loughran G, Zhdanov AV, Mariotti M, Kiniry SJ, O'Connor PBF, Andreev DE, Tzani I, Saffert P, Michel AM, Gladyshev VN, Papkovsky DB, Atkins JF and Baranov PV. AMD1 mRNA employs ribosome stalling as a mechanism for molecular memory formation. Nature 2018; 553: 356-360.
[26] Ding Q, Markesbery WR, Chen Q, Li F and Keller JN. Ribosome dysfunction is an early event in Alzheimer's disease. J Neurosci 2005; 25: 9171-9175.
[27] Ding Q, Markesbery WR, Cecarini V and Keller JN. Decreased RNA, and increased RNA oxidation, in ribosomes from early Alzheimer's disease. Neurochem Res 2006; 31: 705-710.
[28] Ding Q, Zhu H, Zhang B, Soriano A, Burns R and Markesbery WR. Increased 5S rRNA oxidation in Alzheimer's disease. J Alzheimers Dis 2012; 29: 201-209.
[29] Maina MB, Bailey LJ, Doherty AJ and Serpell LC. The involvement of Abeta42 and Tau in nucleolar and protein synthesis machinery dysfunction. Front Cell Neurosci 2018; 12: 220.
[30] Meier S, Bell M, Lyons DN, Rodriguez-Rivera J, Ingram A, Fontaine SN, Mechas E, Chen J, Wolozin B, LeVine H 3rd, Zhu H and Abisambra JF. Pathological Tau promotes neuronal damage by impairing ribosomal function and decreasing protein synthesis. J Neurosci 2016; 36: 1001-1007.
[31] Turi Z, Lacey M, Mistrik M and Moudry P. Impaired ribosome biogenesis: mechanisms and relevance to cancer and aging. Aging (Albany NY) 2019; 11: 2512-2540.
[32] Challa S, Khulpateea BR, Nandu T, Camacho CV, Ryu KW, Chen H, Peng Y, Lea JS and Kraus WL. Ribosome ADP-ribosylation inhibits translation and maintains proteostasis in cancers. Cell 2021; 184: 4531-4546, e4526.
[33] Knight JR, Vlahov N, Gay DM, Ridgway RA, Faller WJ, Proud C, Mallucci GR, von der Haar T, Smales CM, Willis AE and Sansom OJ. Rpl24(Bst) mutation suppresses colorectal cancer by promoting eEF2 phosphorylation via eEF2K. Elife 2021; 10: e69729.
[34] Ji W, An K, Wang C and Wang S. Bioinformatics analysis of diagnostic biomarkers for Alzheimer's disease in peripheral blood based on sex differences and support vector machine algorithm. Hereditas 2022; 159: 38.
[35] Tao Y, Han Y, Yu L, Wang Q, Leng SX and Zhang H. The predicted key molecules, functions, and pathways that bridge Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD). Front Neurol 2020; 11: 233.
[36] Khayer N, Mirzaie M, Marashi SA and Jalessi M. Rps27a might act as a controller of microglia activation in triggering neurodegenerative diseases. PLoS One 2020; 15: e0239219.
[37] Xie J, Zhang W, Liang X, Shuai C, Zhou Y, Pan H, Yang Y and Han W. RPL32 promotes lung cancer progression by facilitating p53 degradation. Mol Ther Nucleic Acids 2020; 21: 75-85.
[38] Ma X, Li Y and Zhao B. Ribosomal protein L5 (RPL5)/E2F transcription factor 1 (E2F1) signaling suppresses breast cancer progression via regulating endoplasmic reticulum stress and autophagy. Bioengineered 2022; 13: 8076-8086.
[39] Shigeoka T, Koppers M, Wong HH, Lin JQ, Cagnetta R, Dwivedy A, de Freitas Nascimento J, van Tartwijk FW, Strohl F, Cioni JM, Schaeffer J, Carrington M, Kaminski CF, Jung H, Harris WA and Holt CE. On-site ribosome remodeling by locally synthesized ribosomal proteins in axons. Cell Rep 2019; 29: 3605-3619, e3610.
[40] Suzuki M, Tezuka K, Handa T, Sato R, Takeuchi H, Takao M, Tano M and Uchida Y. Upregulation of ribosome complexes at the blood-brain barrier in Alzheimer's disease patients. J Cereb Blood Flow Metab 2022; 42: 2134-2150.
[41] Fu J, Duan J, Mo J, Xiao H, Huang Y, Chen W, Xiang S, Yang F, Chen Y and Xu S. Mild cognitive impairment patients have higher regulatory T-cell proportions compared with Alzheimer's disease-related dementia patients. Front Aging Neurosci 2021; 12: 624304.
[42] Kubick N, Flournoy PCH, Enciu AM, Manda G and Mickael ME. Drugs modulating CD4+ T cells blood-brain barrier interaction in Alzheimer's disease. Pharmaceutics 2020; 12: 880.
[43] Hernandez-Mir G, Raphael I, Revu S, Poholek CH, Avery L, Hawse WF, Kane LP and McGeachy MJ. The Alzheimer's Disease-associated protein BACE1 modulates T cell activation

and Th17 function. J Immunol 2019; 203: 665-675.
[44] Dai L, Wang Q, Lv X, Gao F, Chen Z and Shen Y. Elevated beta-secretase 1 expression mediates CD4(+) T cell dysfunction via PGE2 signalling in Alzheimer's disease. Brain Behav Immun 2021; 98: 337-348.
[45] Lindstrom MS, Bartek J and Maya-Mendoza A. p53 at the crossroad of DNA replication and ribosome biogenesis stress pathways. Cell Death Differ 2022; 29: 972-982.
[46] Shi ZR, Han YF, Yin J, Zhang YP, Jiang ZX, Zheng L, Tan GZ and Wang L. The diagnostic benefit of antibodies against ribosomal proteins in systemic lupus erythematosus. Adv Rheumatol 2020; 60: 45.
[47] Tajouri L, Mellick AS, Ashton KJ, Tannenberg AE, Nagra RM, Tourtellotte WW and Griffiths LR. Quantitative and qualitative changes in gene expression patterns characterize the activity of plaques in multiple sclerosis. Brain Res Mol Brain Res 2003; 119: 170-183.
[48] Luo J, Zhao H, Chen L and Liu M. Multifaceted functions of RPS27a: an unconventional ribosomal protein. J Cell Physiol 2023; 238: 485-497.
[49] Zhang H, Liu J, Dang Q, Wang X, Chen J, Lin X, Yang N, Du J, Shi H, Liu Y and Han J. Ribosomal protein RPL5 regulates colon cancer cell proliferation and migration through MAPK/ERK signaling pathway. BMC Mol Cell Biol 2022; 23: 48.
[50] Hu B, Jadhav RR, Gustafson CE, Le Saux S, Ye Z, Li X, Tian L, Weyand CM and Goronzy JJ. Distinct age-related epigenetic signatures in CD4 and CD8 T cells. Front Immunol 2020; 11: 585168.
[51] Asmal M, Colgan J, Naef F, Yu B, Lee Y, Magnasco M and Luban J. Production of ribosome components in effector CD4+ T cells is accelerated by TCR stimulation and coordinated by ERK-MAPK. Immunity 2003; 19: 535-548.
[52] Chen C, Peng J, Ma S, Ding Y, Huang T, Zhao S, Gao L, Liang X, Li C and Ma C. Ribosomal protein S26 serves as a checkpoint of T-cell survival and homeostasis in a p53-dependent manner. Cell Mol Immunol 2021; 18: 1844-1846.
[53] Zhu X, Zhang W, Guo J, Zhang X, Li L, Wang T, Yan J, Zhang F, Hou B, Gao N, Gao GF and Zhou X. Noc4L-mediated ribosome biogenesis controls activation of regulatory and conventional T cells. Cell Rep 2019; 27: 1205-1220, e1204.
[54] Jurgens AP, Popovic B and Wolkers MC. T cells at work: How post-transcriptional mechanisms control T cell homeostasis and activation. Eur J Immunol 2021; 51: 2178-2187.
[55] Goldeck D, Larbi A, Pellicano M, Alam I, Zerr I, Schmidt C, Fulop T and Pawelec G. Enhanced chemokine receptor expression on leukocytes of patients with Alzheimer's disease. PLoS One 2013; 8: e66664.
[56] Pietronigro E, Zenaro E and Constantin G. Imaging of leukocyte trafficking in Alzheimer's disease. Front Immunol 2016; 7: 33.
[57] Hohsfield LA and Humpel C. Migration of blood cells to beta-amyloid plaques in Alzheimer's disease. Exp Gerontol 2015; 65: 8-15.
[58] Dong Y, Lagarde J, Xicota L, Corne H, Chantran Y, Chaigneau T, Crestani B, Bottlaender M, Potier MC, Aucouturier P, Dorothee G, Sarazin M and Elbim C. Neutrophil hyperactivation correlates with Alzheimer's disease progression. Ann Neurol 2018; 83: 387-405.
[59] Hernandez-Ortega K, Garcia-Esparcia P, Gil L, Lucas JJ and Ferrer I. Altered machinery of protein synthesis in Alzheimer's: from the nucleolus to the ribosome. Brain Pathol 2016; 26: 593-605.
[60] Evans HT, Taylor D, Kneynsberg A, Bodea LG and Gotz J. Altered ribosomal function and protein synthesis caused by tau. Acta Neuropathol Commun 2021; 9: 110.
[61] Koren SA, Hamm MJ, Meier SE, Weiss BE, Nation GK, Chishti EA, Arango JP, Chen J, Zhu H, Blalock EM and Abisambra JF. Tau drives translational selectivity by interacting with ribosomal proteins. Acta Neuropathol 2019; 137: 571-583.

Table S1. Differentially expressed genes in training cohorts (|logFC| >1 and adjust *P*-value < 0.05)

Genes	logFC	AveExpr	t	*P*.Value	adj.P.Val	B
RPS27	1.609555	9.360017	8.004628	7.82E-15	6.39E-13	22.93928
SNRPG	1.560314	8.624294	8.761071	2.71E-17	4.33E-15	28.49838
NDUFB3	1.525468	8.209974	9.3203	3.25E-19	7.97E-17	32.84635
LSM3	1.518026	7.449509	9.399124	1.72E-19	5.26E-17	33.47474
TXN	1.505974	9.029185	9.330874	2.99E-19	7.84E-17	32.93042
RPS3A	1.488942	9.137976	5.789404	1.22E-08	2.88E-07	9.031426
AIF1	1.482472	9.583373	10.57976	7.85E-24	6.36E-21	43.31744
RPS27A	1.471432	11.22784	10.56869	8.65E-24	6.36E-21	43.22158
ENY2	1.46671	7.751628	10.50344	1.53E-23	9.39E-21	42.65787
HSPE1	1.442115	7.878766	10.48004	1.88E-23	9.87E-21	42.45623
CMTM2	1.423501	9.803581	8.642466	6.76E-17	9.55E-15	27.60172
CD3D	1.412729	9.713352	8.18586	2.08E-15	2.19E-13	24.23616
FCER1A	1.374636	8.465744	6.248154	8.63E-10	2.64E-08	11.60099
COX7A2	1.367872	9.666367	7.089347	4.40E-12	2.28E-10	16.74402
RPL7	1.364625	7.449686	8.400067	4.24E-16	4.73E-14	25.79776
NDUFB2	1.348251	9.117247	8.001095	8.03E-15	6.41E-13	22.91422
PSMA4	1.345087	7.687818	8.417991	3.71E-16	4.27E-14	25.92982
DBI	1.342793	7.573618	8.653271	6.22E-17	9.15E-15	27.68303
RPL24	1.340774	11.10152	8.444713	3.03E-16	3.72E-14	26.12711
TMCO1	1.319965	8.386782	6.489204	2.01E-10	7.28E-09	13.01867
EEF1B2	1.308908	8.298641	8.059392	5.26E-15	4.72E-13	23.32879
DPM1	1.30758	7.926542	8.417659	3.72E-16	4.27E-14	25.92737
RPL5	1.307361	9.893617	7.151992	2.91E-12	1.55E-10	17.14862
UFC1	1.289101	8.789978	11.6999	3.04E-28	1.12E-24	53.33741
PSMA6	1.27603	7.944563	6.722693	4.69E-11	2.04E-09	14.43524
MRPL33	1.275817	9.322367	10.76414	1.54E-24	2.83E-21	44.92362
TAX1BP1	1.275691	9.0186	6.564644	1.26E-10	4.78E-09	13.47173
SNRPB2	1.274217	8.620786	9.411019	1.56E-19	5.21E-17	33.56989
PSMA3	1.273224	7.272025	10.0806	5.90E-22	2.71E-19	39.06033
COX6C	1.272149	9.396508	7.237844	1.64E-12	9.13E-11	17.70784
SEC11A	1.268193	8.960055	8.296291	9.21E-16	9.96E-14	25.03737
TPT1	1.260235	12.64568	4.951095	9.98E-07	1.37E-05	4.786157
POLR3GL	1.247474	8.724454	8.53491	1.53E-16	2.01E-14	26.7965
DNAJC8	1.243341	9.662491	9.276359	4.64E-19	1.00E-16	32.49769
COMMD3	1.24251	8.778826	6.71466	4.93E-11	2.08E-09	14.3858
SLU7	1.238382	7.618838	8.062144	5.15E-15	4.72E-13	23.34842
IGBP1	1.234193	8.058435	9.349145	2.58E-19	7.29E-17	33.07587
HMGB2	1.233574	7.338737	6.721438	4.73E-11	2.04E-09	14.42751
VPS29	1.230121	8.767872	6.389813	3.68E-10	1.25E-08	12.42857
ANAPC13	1.229039	9.062079	8.075493	4.68E-15	4.41E-13	23.4437
RPS4X	1.226629	11.42672	6.715426	4.91E-11	2.08E-09	14.39051
SH2D1A	1.211206	7.805405	6.922	1.31E-11	6.26E-10	15.67762
NGDN	1.207658	7.404584	8.963076	5.61E-18	1.09E-15	30.04633
RPL12	1.206323	11.00729	7.574583	1.65E-13	1.21E-11	19.95344
SSBP1	1.202602	8.844253	8.864241	1.22E-17	2.03E-15	29.28572
S100P	1.201722	8.534098	3.066349	0.002279	0.011524	-2.49927
CLEC4A	1.199682	8.853891	6.010665	3.47E-09	9.59E-08	10.2495

ACAT1	1.196367	7.5686	7.43027	4.46E-13	2.83E-11	18.98096
SEC11C	1.191696	7.671402	7.301401	1.07E-12	6.05E-11	18.12534
VBP1	1.189889	7.767461	5.933711	5.40E-09	1.42E-07	9.821328
VPREB3	1.189533	7.766085	4.820647	1.88E-06	2.45E-05	4.179462
MTIF3	1.189342	7.695863	8.904146	8.91E-18	1.64E-15	29.59207
ZNHIT3	1.188955	7.537992	6.884199	1.67E-11	7.79E-10	15.43966
SUB1	1.188161	7.189655	6.886199	1.65E-11	7.79E-10	15.45222
BOLA2	1.186158	8.837085	6.038841	2.95E-09	8.28E-08	10.40747
FRG1	1.18499	7.705225	8.009453	7.56E-15	6.31E-13	22.97352
GNL2	1.184044	7.7993	9.305633	3.66E-19	8.42E-17	32.72984
UCHL3	1.183686	7.269827	9.80055	6.27E-21	2.56E-18	36.73236
LARP7	1.180143	7.184561	8.13943	2.93E-15	2.91E-13	23.90177
MPHOSPH10	1.17912	7.854866	8.152241	2.67E-15	2.73E-13	23.99389
FGL2	1.177594	11.29	5.278012	1.92E-07	3.37E-06	6.371032
ZC3H15	1.173928	7.188201	7.556873	1.87E-13	1.34E-11	19.83329
TMSB10	1.172418	12.6362	8.656653	6.06E-17	9.15E-15	27.70849
POLE4	1.170632	9.105497	5.80595	1.11E-08	2.69E-07	9.121129
EIF4A2	1.168289	10.43317	4.268268	2.34E-05	0.000225	1.775971
TRAPPC4	1.165273	8.540057	8.864463	1.21E-17	2.03E-15	29.28741
KLRF1	1.162783	7.794414	3.765954	0.000185	0.001395	-0.17241
GNL3	1.161637	7.27984	6.625284	8.65E-11	3.38E-09	13.83912
SET	1.160755	9.058105	6.374144	4.05E-10	1.37E-08	12.33625
MRPL36	1.160749	7.854119	7.424341	4.64E-13	2.89E-11	18.94133
RPL26L1	1.160499	7.067921	7.410133	5.12E-13	3.13E-11	18.84646
TRAT1	1.158254	7.308237	5.689111	2.13E-08	4.75E-07	8.492512
ZBED5	1.156992	8.186188	4.486661	8.91E-06	9.95E-05	2.693956
NDUFB6	1.156392	8.41702	7.350937	7.65E-13	4.61E-11	18.45279
VAMP7	1.156114	8.100145	5.563607	4.23E-08	9.14E-07	7.829854
SNRK	1.154602	9.019402	4.457393	1.02E-05	0.000111	2.568467
PTRH2	1.154589	7.223705	7.544441	2.03E-13	1.41E-11	19.74908
ZCCHC17	1.15268	7.743866	8.078528	4.57E-15	4.41E-13	23.46538
DYNLT3	1.148081	7.147284	6.627953	8.50E-11	3.36E-09	13.85536
MRPS17	1.147402	7.228969	10.60016	6.56E-24	6.36E-21	43.49432
NUP88	1.146683	8.088509	6.642555	7.76E-11	3.17E-09	13.94428
NAP1L1	1.145913	8.556691	4.568925	6.12E-06	7.12E-05	3.050733
BTF3	1.144319	9.098358	7.319806	9.45E-13	5.51E-11	18.24679
NDUFAF2	1.144043	7.201889	8.022276	6.89E-15	6.03E-13	23.06458
SNURF	1.143706	8.549658	5.452381	7.68E-08	1.53E-06	7.253555
CUTC	1.142512	7.552621	7.22285	1.81E-12	9.95E-11	17.60978
SNRPF	1.14176	8.824739	5.126121	4.17E-07	6.54E-06	5.623288
LYRM2	1.139303	7.546508	6.631655	8.31E-11	3.36E-09	13.87788
CD79A	1.135425	7.862534	3.5857	0.000368	0.00249	-0.8154
TBC1D15	1.134592	7.623807	5.790084	1.22E-08	2.88E-07	9.035108
ATG3	1.134195	7.976037	5.849834	8.69E-09	2.17E-07	9.36013
PTPN4	1.134169	8.050844	4.605276	5.18E-06	6.20E-05	3.210292
TCEAL8	1.132485	7.034816	8.534871	1.53E-16	2.01E-14	26.79621
CD58	1.131889	7.699419	4.419536	1.20E-05	0.000129	2.407279
RPL15	1.131746	10.51133	3.697884	0.00024	0.001721	-0.41874
AGTPBP1	1.130733	8.64499	5.246539	2.26E-07	3.86E-06	6.214478

TAF12	1.1285	7.811007	7.796496	3.48E-14	2.66E-12	21.47794
HLA-DOB	1.126125	7.86112	4.136751	4.11E-05	0.000372	1.243759
HMGN1	1.125817	8.84021	5.26283	2.08E-07	3.62E-06	6.295406
PLAC8	1.121874	9.908433	4.613638	4.98E-06	5.99E-05	3.247165
SPAG7	1.121814	7.449676	7.317478	9.60E-13	5.51E-11	18.23142
PPP1R2	1.12169	8.009864	4.642495	4.36E-06	5.29E-05	3.374868
BTBD10	1.120736	7.515313	7.107501	3.90E-12	2.05E-10	16.86097
CCT8	1.120637	9.397083	5.537508	4.87E-08	1.04E-06	7.693697
COMMD1	1.116948	8.456085	7.040907	6.05E-12	3.00E-10	16.43318
ZNF22	1.116835	7.533323	4.277447	2.25E-05	0.000219	1.813695
LYPLAL1	1.116631	7.465392	6.70958	5.09E-11	2.13E-09	14.35456
SDAD1	1.115534	8.990701	5.392528	1.05E-07	2.00E-06	6.947732
MTIF2	1.115318	7.553229	7.551995	1.93E-13	1.36E-11	19.80023
SLC30A9	1.113518	7.482694	5.394483	1.04E-07	2.00E-06	6.957674
SYF2	1.112482	7.847031	4.554967	6.53E-06	7.52E-05	2.989777
RPL34	1.11233	6.832642	7.5304	2.24E-13	1.53E-11	19.65411
PPP1CC	1.111531	9.496883	3.643695	0.000296	0.002066	-0.61179
MRPS28	1.110568	6.91372	7.489131	2.98E-13	1.95E-11	19.37579
RABEP1	1.109975	7.866664	4.174793	3.49E-05	0.000322	1.396103
WBP4	1.108676	7.171659	5.945118	5.06E-09	1.34E-07	9.884495
PDCD10	1.108059	7.25675	4.273146	2.29E-05	0.000222	1.79601
CFDP1	1.107623	7.312833	7.344341	8.00E-13	4.74E-11	18.40909
BTLA	1.107028	7.065203	5.772089	1.34E-08	3.15E-07	8.93779
NFU1	1.106481	7.336419	6.318371	5.67E-10	1.85E-08	12.00921
RPS21	1.105676	7.44293	4.736828	2.80E-06	3.55E-05	3.797446
RGS18	1.105546	9.012759	2.129215	0.033704	0.108111	-4.89122
CD160	1.104222	7.316553	4.049876	5.90E-05	0.000518	0.900752
PRMT1	1.103759	8.093079	6.290136	6.72E-10	2.13E-08	11.84459
LTV1	1.103603	7.513847	6.780627	3.25E-11	1.47E-09	14.79324
BCCIP	1.103004	7.281884	5.994611	3.81E-09	1.01E-07	10.15977
CLK1	1.101738	7.987373	3.06231	0.00231	0.011634	-2.51137
ARPC2	1.101513	12.05039	5.697407	2.03E-08	4.59E-07	8.536773
USP1	1.101417	7.262528	4.349162	1.64E-05	0.000168	2.111035
NDUFA9	1.100453	7.974848	6.299781	6.34E-10	2.03E-08	11.90076
MRFAP1L1	1.100101	7.745686	4.571153	6.06E-06	7.08E-05	3.060478
CLN3	-1.10002	7.767752	-5.92669	5.62E-09	1.47E-07	9.782523
TBC1D3B	-1.10012	7.34544	-5.87844	7.39E-09	1.89E-07	9.516801
ACSL1	-1.10017	9.274877	-2.23205	0.026036	0.087244	-4.66997
S100A10	-1.10052	11.6047	-3.60725	0.000339	0.002318	-0.7401
JAK1	-1.10083	9.0193	-3.16823	0.001624	0.008738	-2.18885
ADAM15	-1.10124	8.334721	-4.22926	2.77E-05	0.000261	1.616476
DUSP18	-1.10134	8.860939	-4.5835	5.73E-06	6.79E-05	3.114588
PUM1	-1.10155	9.615751	-4.82956	1.80E-06	2.36E-05	4.220467
NINJ1	-1.1018	10.89604	-3.1977	0.001469	0.008002	-2.09725
NKG7	-1.10223	12.03308	-2.36091	0.018597	0.066113	-4.37826
SLC16A5	-1.10227	7.820328	-4.95594	9.74E-07	1.35E-05	4.808981
RHBDD2	-1.10232	9.357268	-4.4807	9.15E-06	0.000102	2.668354
SLC11A1	-1.10322	8.70159	-3.57067	0.000389	0.002614	-0.86767
IGFBP7	-1.10322	7.62935	-4.74774	2.66E-06	3.38E-05	3.846841

PIAS4	-1.10328	9.048995	-4.93074	1.10E-06	1.50E-05	4.690526
HIP1	-1.10352	7.398625	-3.66296	0.000275	0.001931	-0.54346
YIPF6	-1.10378	8.802682	-4.49781	8.47E-06	9.55E-05	2.741962
SIGLEC5	-1.104	7.699953	-3.37482	0.000794	0.004822	-1.52951
WWC3	-1.10431	8.062959	-5.16029	3.51E-07	5.76E-06	5.789793
TST	-1.10446	9.516769	-3.75656	0.000192	0.001435	-0.20668
OSBPL5	-1.10474	7.821929	-4.25601	2.47E-05	0.000236	1.725716
KIAA0513	-1.10541	8.734343	-3.99801	7.31E-05	0.000622	0.699215
CNNM3	-1.10546	7.572529	-5.49771	6.03E-08	1.25E-06	7.487175
GATAD2B	-1.1057	8.130054	-5.13473	4.00E-07	6.41E-06	5.66514
ADAR	-1.10595	10.12459	-3.54899	0.000422	0.002793	-0.94269
CHFR	-1.10679	9.43106	-6.51774	1.68E-10	6.32E-09	13.18953
ZAP70	-1.10717	8.671486	-4.01221	6.89E-05	0.000594	0.754179
TSPAN32	-1.10725	8.688197	-4.9849	8.45E-07	1.19E-05	4.945777
SYT11	-1.10737	8.848883	-4.25068	2.52E-05	0.00024	1.703903
IER3	-1.10768	7.744146	-4.91147	1.21E-06	1.63E-05	4.600326
CEP350	-1.1088	8.390328	-6.43685	2.77E-10	9.69E-09	12.70685
PLCG2	-1.1092	9.852003	-3.99613	7.37E-05	0.000625	0.691989
IFNGR2	-1.11019	11.30619	-4.098	4.83E-05	0.000431	1.089898
TMEM154	-1.11123	10.23983	-2.54028	0.011365	0.044023	-3.9456
NT5C2	-1.11126	9.282328	-4.33735	1.73E-05	0.000176	2.061762
SPSB3	-1.11127	9.693225	-4.94708	1.02E-06	1.40E-05	4.76727
CTSW	-1.11138	8.613774	-3.19809	0.001467	0.008002	-2.09604
ATP6V1F	-1.11202	10.32295	-5.12589	4.18E-07	6.54E-06	5.622147
AKT1	-1.11208	10.03643	-5.01411	7.31E-07	1.06E-05	5.084526
SLC22A4	-1.11216	7.819931	-3.97878	7.91E-05	0.000665	0.625151
E4F1	-1.1127	8.505191	-6.25489	8.29E-10	2.56E-08	11.64
ZNF746	-1.11306	9.233585	-4.41972	1.20E-05	0.000129	2.40805
NISCH	-1.11355	8.497056	-6.26327	7.89E-10	2.46E-08	11.68852
RBCK1	-1.1136	8.314714	-5.22366	2.54E-07	4.30E-06	6.101184
PSTPIP1	-1.11369	8.534335	-5.32501	1.50E-07	2.71E-06	6.606381
KIR3DL1	-1.11373	7.80358	-3.43652	0.000636	0.003964	-1.32483
NFKBIZ	-1.11378	9.466229	-4.08126	5.18E-05	0.000458	1.023871
MGAT1	-1.1142	9.938249	-5.30701	1.65E-07	2.94E-06	6.516016
GDPD3	-1.11422	7.550261	-6.07218	2.43E-09	6.87E-08	10.59522
PNPLA6	-1.11482	9.193207	-5.12637	4.17E-07	6.54E-06	5.624499
AKAP13	-1.11497	8.456562	-6.2278	9.74E-10	2.93E-08	11.4834
SLC26A8	-1.11525	7.19938	-5.00279	7.73E-07	1.11E-05	5.030684
ANKS1A	-1.11575	7.871144	-7.50219	2.72E-13	1.82E-11	19.46376
UBR4	-1.11628	8.897887	-5.16306	3.46E-07	5.73E-06	5.803303
PPP1R14B	-1.11665	8.347555	-7.20978	1.98E-12	1.07E-10	17.52444
TRIM38	-1.11799	8.946014	-5.25446	2.17E-07	3.76E-06	6.253794
SPI1	-1.11803	10.21408	-3.33	0.00093	0.005504	-1.67592
RAB24	-1.11832	7.765331	-6.31773	5.69E-10	1.85E-08	12.0055
SLC27A3	-1.1187	8.59676	-5.4655	7.16E-08	1.44E-06	7.320988
STAT5B	-1.11871	8.834655	-5.38548	1.09E-07	2.05E-06	6.911907
TNFRSF1A	-1.11975	10.59215	-4.37724	1.45E-05	0.000152	2.228716
TRPC4AP	-1.12011	9.039122	-5.01913	7.13E-07	1.04E-05	5.108408
PISD	-1.1212	8.973175	-4.50975	8.02E-06	9.13E-05	2.793476

TNFSF13B	-1.12137	10.32359	-3.78362	0.000172	0.001313	-0.1078
CTNNA1	-1.12212	9.058431	-4.83048	1.79E-06	2.35E-05	4.224664
IDS	-1.12244	8.948136	-5.67116	2.35E-08	5.21E-07	8.396951
SLC9A8	-1.12261	7.814915	-6.08298	2.28E-09	6.51E-08	10.6562
ZC3H3	-1.12305	8.417697	-6.09685	2.11E-09	6.09E-08	10.73471
KLF6	-1.12327	8.515199	-5.75912	1.45E-08	3.36E-07	8.867834
STAT3	-1.12348	9.236638	-4.11882	4.43E-05	0.000397	1.172399
B4GALT5	-1.12378	9.570983	-3.76106	0.000188	0.001419	-0.19026
CREBBP	-1.12418	9.131086	-4.63759	4.46E-06	5.39E-05	3.35313
TNFRSF14	-1.12494	10.963	-6.27253	7.46E-10	2.34E-08	11.74229
PPTC7	-1.12519	8.792531	-6.00353	3.62E-09	9.71E-08	10.20959
BCKDK	-1.12746	9.185912	-5.46243	7.28E-08	1.45E-06	7.30519
NCSTN	-1.12991	9.644502	-5.07555	5.38E-07	8.11E-06	5.3787
KLF2	-1.13077	12.49163	-5.46534	7.17E-08	1.44E-06	7.320148
EMILIN2	-1.13123	8.785816	-5.13854	3.92E-07	6.35E-06	5.683661
MAPK8IP3	-1.1313	8.707327	-6.75584	3.80E-11	1.70E-09	14.63974
HBQ1	-1.13162	10.26462	-2.29827	0.021941	0.075848	-4.52206
GAK	-1.13179	9.788639	-5.48058	6.61E-08	1.36E-06	7.398673
PGS1	-1.13396	9.113736	-5.47493	6.81E-08	1.39E-06	7.369535
SLC7A7	-1.13443	10.39408	-5.26299	2.07E-07	3.62E-06	6.296205
STAT1	-1.13457	9.402702	-3.57413	0.000384	0.002585	-0.85565
MKNK2	-1.13559	9.475694	-4.92667	1.12E-06	1.52E-05	4.671436
FCGRT	-1.1357	10.75588	-5.16873	3.36E-07	5.59E-06	5.831082
FRAT1	-1.13618	7.940727	-5.39263	1.05E-07	2.00E-06	6.948265
AGTRAP	-1.13664	8.520648	-6.3319	5.23E-10	1.73E-08	12.0883
NOL12	-1.137	7.944724	-8.01711	7.15E-15	6.11E-13	23.02788
KLHDC8B	-1.13802	7.697377	-4.99393	8.08E-07	1.15E-05	4.988618
MEGF9	-1.13833	7.964413	-5.75611	1.47E-08	3.38E-07	8.851622
HSPA6	-1.13856	9.246179	-4.16615	3.63E-05	0.000332	1.361375
RXRA	-1.14058	10.79991	-5.04347	6.32E-07	9.29E-06	5.224698
ALPL	-1.14151	10.85446	-2.53757	0.011452	0.044314	-3.95237
HELZ	-1.14229	8.122759	-9.04475	2.95E-18	6.02E-16	30.67949
ARHGEF2	-1.14303	11.04351	-7.04596	5.85E-12	2.95E-10	16.46554
EIF2AK2	-1.14442	9.050618	-3.70077	0.000238	0.001717	-0.4084
APBB3	-1.14472	8.112898	-7.48108	3.15E-13	2.03E-11	19.32161
PIK3CD	-1.14619	9.466606	-5.36662	1.21E-07	2.24E-06	6.816288
NXF1	-1.15088	9.277887	-7.91847	1.46E-14	1.14E-12	22.33068
CSF1R	-1.15111	10.10815	-4.46709	9.73E-06	0.000107	2.609944
ZNF281	-1.15415	8.601829	-5.02938	6.78E-07	9.89E-06	5.157305
CA4	-1.15434	8.585433	-3.72413	0.000217	0.00159	-0.32425
SMAP2	-1.15472	12.43063	-4.88769	1.36E-06	1.81E-05	4.489434
DCUN1D1	-1.15771	7.871169	-5.79905	1.16E-08	2.76E-07	9.083691
EFHD2	-1.15826	11.92824	-6.73603	4.31E-11	1.91E-09	14.51744
FPR2	-1.15936	8.952108	-3.78242	0.000173	0.001316	-0.1122
TBXAS1	-1.16107	9.847507	-6.08574	2.25E-09	6.45E-08	10.67185
ITGA5	-1.1626	8.763511	-5.88828	6.99E-09	1.80E-07	9.570841
MYO1G	-1.16508	9.949531	-6.41311	3.20E-10	1.11E-08	12.56618
ITGB2	-1.16527	13.21236	-6.40996	3.26E-10	1.12E-08	12.54757
PLOD1	-1.16601	8.902714	-6.23176	9.51E-10	2.89E-08	11.50626

FOXO3	-1.16664	10.06672	-4.64507	4.31E-06	5.26E-05	3.386318
ACADVL	-1.168	9.851089	-7.7459	4.97E-14	3.73E-12	21.12727
PPM1F	-1.16989	10.37548	-5.34873	1.33E-07	2.43E-06	6.725851
SLC15A4	-1.17323	9.615326	-6.20753	1.10E-09	3.28E-08	11.36663
TLN1	-1.1741	9.189393	-5.10243	4.70E-07	7.17E-06	5.508425
BEST1	-1.17693	8.18837	-6.19241	1.20E-09	3.53E-08	11.27969
RHBDF2	-1.17751	9.006029	-6.87514	1.77E-11	8.15E-10	15.38281
CEBPB	-1.18182	12.08324	-5.84151	9.11E-09	2.25E-07	9.314671
SRRM2	-1.18189	10.02406	-6.57623	1.17E-10	4.49E-09	13.5417
CSAD	-1.18473	8.247861	-6.99147	8.35E-12	4.04E-10	16.11772
ULK1	-1.1859	10.03443	-6.4738	2.21E-10	7.80E-09	12.92668
FPR1	-1.1872	12.42206	-5.11369	4.44E-07	6.83E-06	5.562954
OSCAR	-1.19048	9.26419	-6.47432	2.20E-10	7.80E-09	12.92983
APBB1IP	-1.19128	10.68937	-6.0041	3.61E-09	9.71E-08	10.2128
RNF24	-1.19134	9.954112	-4.84265	1.69E-06	2.23E-05	4.280765
FES	-1.19348	8.875001	-6.99583	8.12E-12	3.98E-10	16.14549
LPP	-1.19569	9.726372	-9.57503	4.08E-20	1.50E-17	34.8905
MYADM	-1.20489	11.64442	-4.99868	7.89E-07	1.12E-05	5.011149
GRN	-1.21056	9.844343	-5.82237	1.01E-08	2.49E-07	9.210354
SORL1	-1.21059	10.85277	-5.2221	2.56E-07	4.32E-06	6.093503
ABTB1	-1.2221	10.46012	-6.62955	8.42E-11	3.36E-09	13.8651
PGLYRP1	-1.22315	9.569679	-4.26985	2.32E-05	0.000224	1.782453
CSNK1G2	-1.22986	10.26575	-5.73617	1.64E-08	3.74E-07	8.744359
ITGAM	-1.26164	9.471993	-8.47597	2.40E-16	3.04E-14	26.35848

Table S2. Feature genes identified by Boruta algorithm

Gene	Important	decision
VPS29	1.099922	Rejected
RPS3A	1.205005	Rejected
GRN	1.576741	Rejected
SORL1	1.706963	Rejected
MYADM	2.126641	Rejected
TAX1BP1	2.147483	Rejected
PSMA6	2.148816	Rejected
FCER1A	2.179526	Rejected
SH2D1A	2.318225	Rejected
PGLYRP1	2.546298	Rejected
EEF1B2	2.588323	Rejected
COX6C	2.808093	Tentative
COMMD3	2.935152	Tentative
CD3D	2.975145	Confirmed
TPT1	3.01185	Confirmed
NDUFB2	3.369418	Confirmed
RPS27	3.397104	Confirmed
COX7A2	3.459972	Confirmed
TMCO1	3.467814	Confirmed
ABTB1	3.490556	Confirmed
CSNK1G2	3.527216	Confirmed

POLR3GL	3.807013	Confirmed
SEC11A	3.928164	Confirmed
SLU7	4.142269	Confirmed
HMGB2	4.597455	Confirmed
DPM1	4.789581	Confirmed
RPL24	4.80265	Confirmed
SNRPG	5.258554	Confirmed
RPL12	5.447277	Confirmed
RPL5	5.554934	Confirmed
RPS4X	5.565047	Confirmed
NGDN	5.639681	Confirmed
ANAPC13	5.906456	Confirmed
PSMA4	5.97657	Confirmed
SNRPB2	6.095489	Confirmed
RPL7	6.140573	Confirmed
DBI	6.630868	Confirmed
DNAJC8	7.058027	Confirmed
PSMA3	7.641101	Confirmed
IGBP1	8.153762	Confirmed
SSBP1	8.222691	Confirmed
ITGAM	8.553201	Confirmed
HSPE1	8.776746	Confirmed
TXN	8.798499	Confirmed
LSM3	8.901268	Confirmed
NDUFB3	9.049477	Confirmed
ENY2	10.15341	Confirmed
CMTM2	10.84303	Confirmed
MRPL33	11.0729	Confirmed
AIF1	11.28228	Confirmed
RPS27A	13.19591	Confirmed
UFC1	13.50049	Confirmed

Table S3. Hub genes identified by LASSO algorithm

AIF1
RPS27A
CMTM2
COX7A2
NDUFB2
RPL24
TMCO1
DPM1
RPL5
UFC1
MRPL33
SNRPB2
PSMA3
DNAJC8
IGBP1
ANAPC13
RPS4X
SSBP1
ABTB1
CSNK1G2
ITGAM

Table S4. Gene model constructed by lightGBM

Gene	Coefficient
AIF1	0.653944
RPS27A	0.863563
CMTM2	1.284774
COX7A2	-0.40885
NDUFB2	-0.44523
RPL24	-0.66977
TMCO1	-1.23248
DPM1	-0.42652
RPL5	0.053078
UFC1	1.409454
MRPL33	0.02332
SNRPB2	0.338794
PSMA3	0.621297
DNAJC8	1.316004
IGBP1	0.144825
ANAPC13	0.269325
RPS4X	0.36695
SSBP1	0.263203
ABTB1	-0.91195
CSNK1G2	-0.64088
ITGAM	-0.53259

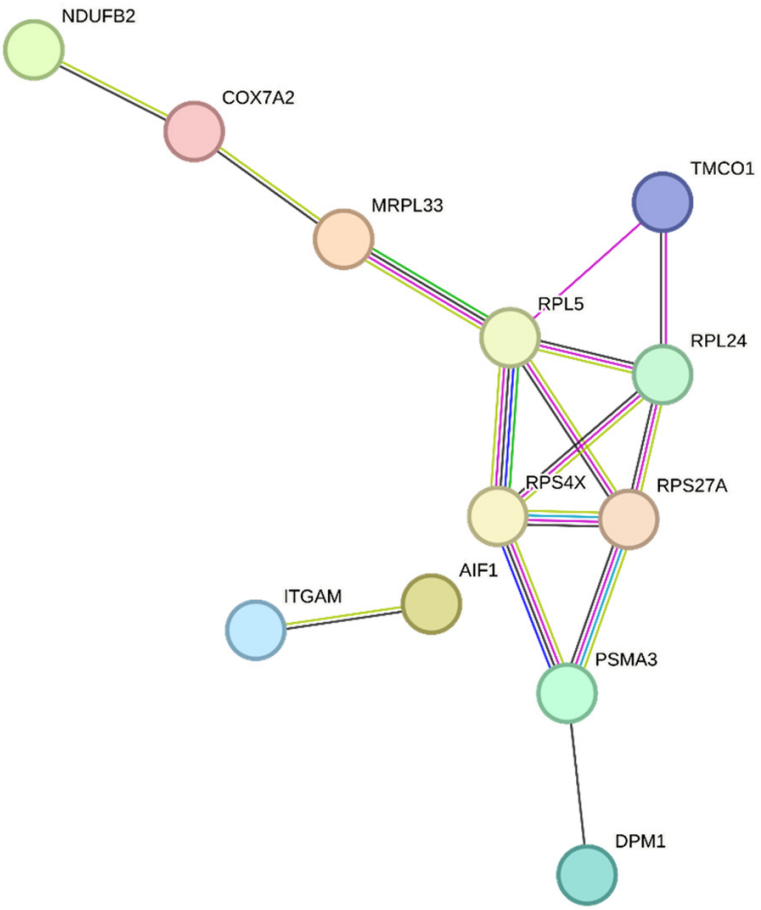

Figure S1. Protein-protein interaction network based on genes obtained from gene module constructed by lightGBM algorithm.

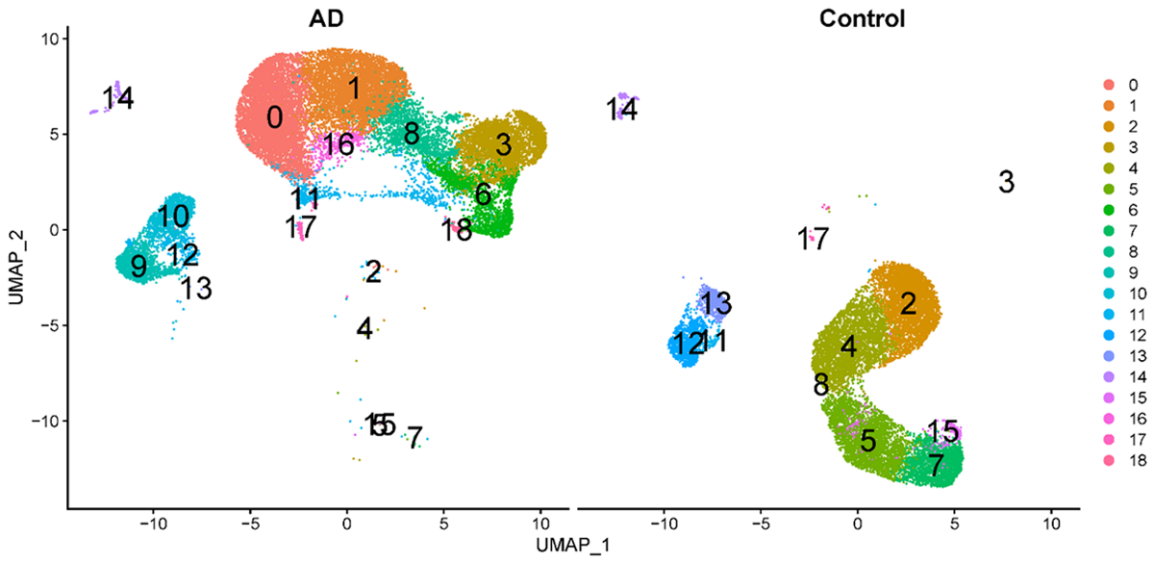

Figure S2. Nineteen clusters identified by using the uniform manifold approximation and projection for dimension reduction (UMAP) algorithm in blood single-cell RNA-seq collected from patients with Alzheimer's disease.