Macrothink Institute™

Journal of Biology and Life Science
ISSN 2157-6076
2019, Vol. 10, No. 2

Whole Blood mRNA Expression Pattern Differentiates AD Patients and Healthy Controls Through Bioinformatics Analysis

Mengjia Zhu

College of Pharmaceutical Engineering and Life Sciences, Changzhou University 213000, Changzhou, China, Email: mengjiazhu0729@163.com

Liqun Wang(Corresponding author)

College of Pharmaceutical Engineering and Life Sciences, Changzhou University 213000, Changzhou, China, Email: wlq@cczu.edu.cn

Received: March 13, 2019 Accepted: April 25, 2019

doi:10.5296/jbls.v10i2.14491 URL: https://doi.org/10.5296/jbls.v10i2.14491

Abstract

Background: Gene chip has a wide range of applications in screening disease markers.

Methods: GSE63063 dataset including 238 healthy controls and 285 patients with Alzheimer's disease (AD) was downloaded to investigate the whole blood mRNA expression pattern. Lumi and LIMMA packages of R software were used to screening differential-expressed genes (DEGs). We functionally annotate DEGs through DAVID database. Then STRING database and Cytoscape software were used to construct protein-protein interaction models for hub genes.

Results: Our results indicated that 51 DEGs altered in AD patients compared with healthy controls. These DEGs was associated with transcription (BP), RNA binding (MF) and ribosome (CC) terms and the ribosome signaling pathway. In addition, Ribosomal protein S17 (RPS17) was identified as the top 1 in hub genes using maximal clique centrality. RPS17 mutations reduced erythrocyte production and impaired brain development. Finally, the expression levels of the three genes (NDUFA1, RPL36AL, and NDUFS5) showed a good predictive effect.

Conclusion: In conclusion, we explored the expression of genes in the AD blood and NDUFA1 may be a potential biomarker for predicting AD.

Keywords: Alzheimer's disease, whole blood, differentially-expressed genes, bioinformatics analysis

1 Introduction

Alzheimer's disease (AD) is a common form of dementia in the elderly, which is characterized by amyloid plaques and neurofibrillary tangles (Henderson, 2014). There is an evidence that 24 million dementia patients and 4.6 million new cases are estimated to occur in the worldwide every year. Among them, 70% of dementia patients will become AD (Reitz & Mayeux, 2014). Notably, large affected population and long-term disability pose a huge challenge to public health (Sibener et al., 2014). Therefore, people have attached greatly importance to early diagnosis and treatment of AD.

Clinically, the diagnosis of AD is based on neuropsychological tests such as hippocampal volumetric magnetic resonance imaging, micro mental status examination and clinical evaluation (Sui, Liu, & Yang, 2014). However, we only determine whether the patient was an AD through autopsy. Cerebrospinal fluid (CSF) have been demonstrated to be a vector reflecting pathological features in AD brain (Blennow, 2005). Generally, we diagnosed probable AD through β-amyloid 1-42 ($A\beta_{1\text{-}42}$), total tau and phospho-tau-181 in CSF, which has high specificity and sensitivity (Humpel, 2011). However, patient screening is difficult to achieve because of the side effects of CSF collection. Blood-based tests would be widely accessible, non-invasive and economic (O'Bryant et al., 2011).

With AD being studied most extensively, gene chip has attracted people's attention. Public microarray data (GSE63063) was downloaded and processed to obtain differentially expressed genes (DEGs) altered in AD patients. Bioinformatics analyses were used to analyze the function and interactions of these DEGs. In addition, the diagnostic value of mRNA level of DEGs in distinguishing between AD patients and controls were also examined.

2 Patients and Methods

2.1 Public Microarray Data

We downloaded the microarray data (GSE63063) from the gene expression omnibus (GEO) (www.ncbi.nlm.nih.gov/geo). Jamie et al. uploaded the dataset and examined the blood genome-wide transcription profile of AD and controls. Totally 516 samples were analyzed (healthy controls with 238 replicates, and AD patients with 285 replicates).

2.2 Identification of DEGs

Lumi package and Linear Models for Microarray Data package (LIMMA, www.biocondutor.org/packages/release/bioc/html/limma.html) of Bioconductor (www.bioconductor.org/) were used to screening DEGs altered in AD patients. The screening criteria were adjusted $P < 0.001$ (Z. Lin & Lin, 2017; L. Zhang et al., 2017) and $|\text{fold change}| \geq 1.5$.

2.3 GO and KEGG Pathway Enrichment Analyses of DEGs

Database for annotation, Visualization, and integrated Discovery (DAVID https://david.ncifcrf.gov/) provided four web-based modules (Annotation tool, GoCharts, KEGG charts and DomainCharts). We analyze the function and pathway enrichment of DEGs

by using DAVID database (Huang da, Sherman, & Lempicki, 2009). The $P < 0.05$ was considered as a cut-off value.

2.4 PPI Network Establishment and Receiver Operating Characteristic (ROC) Curves

STRING (www.string-db.org/) can display physical and/or functional associations (score (median confidence) > 0.4) of proteins encoded by DEGs (Szklarczyk et al., 2015). Hub Object Analyzer (Hubba), an APP of Cytoscape (www.cytoscape.org/), visualized explores important nodes in an interactome network (C. Y. Lin et al., 2008). The diagnostic value of DEGs in distinguishing AD or healthy controls were analyzed, and areas under the curve (AUCs) were calculated.

3 Results

3.1 Identification of DEGs

After data processing, we identified 51 DEGs (50 downregulated and 1 upregulated DEGs) altered in AD patients. Table 1 displays the top 5 downregulated and upregulated DEGs. The hierarchical clustering of expression data (AD vs. control: GPL6947) was shown in Figure 1.

Table 1. The top 5 significant up-regulated and down-regulated differential expressed genes in the whole blood of AD patients

GSE63063(cohort 1)				GSE63063(cohort 2)			
ID_REF	Gene symbol	Log2(FC)	Adj. P value	ID_REF	Gene symbol	Log2(FC)	Adj. P value
Down-regulated				Down-regulated			
ILMN_1784286	NDUFA1	-1.1	4.24E-22	ILMN_1784286	NDUFA1	-1.1	4.24E-22
ILMN_1656625	RPS24	-1.08	1.24E-11	ILMN_1656625	RPS24	-1.08	1.24E-11
ILMN_1658283	RPL17	-1.05	5.99E-09	ILMN_1658283	RPL17	-1.05	5.99E-09
ILMN_1731546	RPL26	-1.03	2.81E-11	ILMN_1731546	RPL26	-1.03	2.81E-11
ILMN_2128128	SHFM1	-0.994	4.16E-18	ILMN_2128128	SHFM1	-0.994	4.16E-18
Up-regulated				Up-regulated			
ILMN_1733559	RNA28S5	1.13	1.55E-07	ILMN_1733559	RNA28S5	1.13	1.55E-07

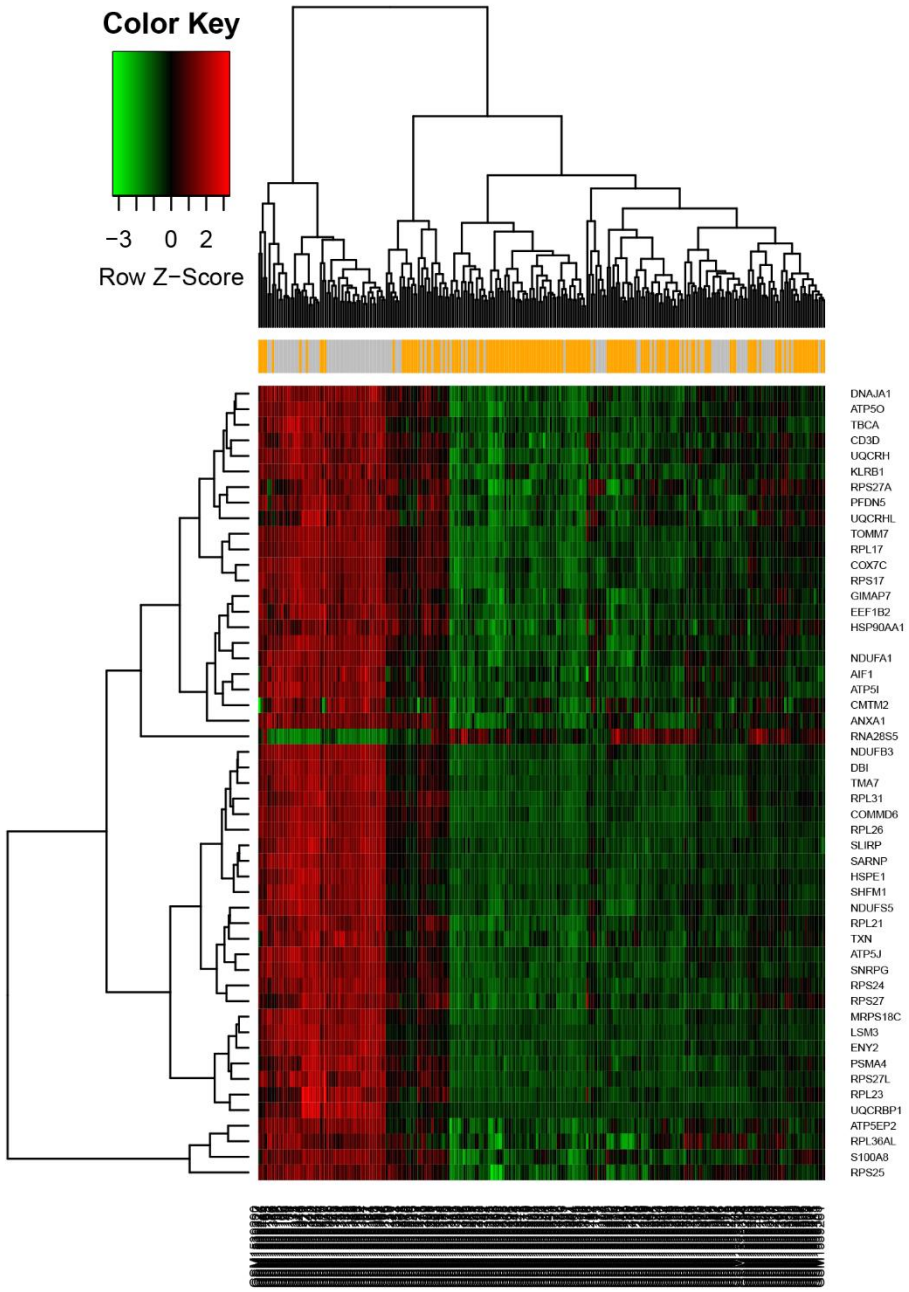

Figure 1. Heatmap for differentially expressed genes (DEGs). Grey and yellow color represents controls and AD patients respectively. The color bar denotes z-score adjusted expression values, green used for down-regulation and red for up-regulation

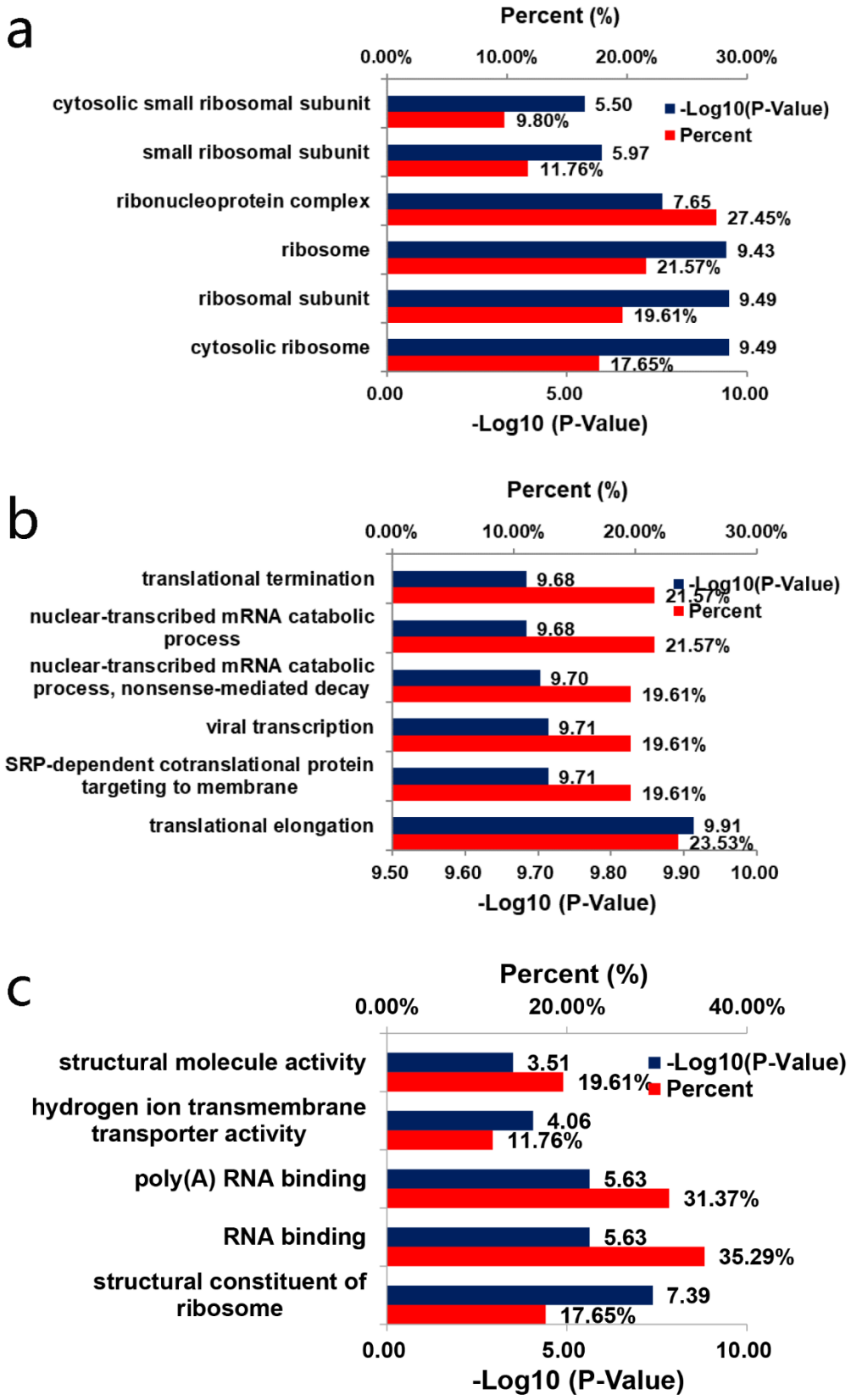

Figure 2. GO enrichment of DEGs. a, cellular component (CC); b, biological process (BP); c, molecular function (MF)

3.2 GO Enrichment and KEGG Pathway Analysis of DEGs and PPI Network Analysis

The enriched GO terms consist of cellular component (CC), biological process (BP) and molecular function (MF). As is shown in Table 2, In the CC ontology, DEGs were correlated with the ribosome, e.g. cytosolic ribosome (9 genes), ribosomal subunit (10 genes) and ribosome (11 genes). In the BP ontology, these DEGs is mainly involved in the transcription-related terms, such as translational elongation (12 genes). viral transcription (10 genes), nuclear-transcribed mRNA catabolic process, and nonsense-mediated decay (10 genes). In the MF ontology, these DEGs play the function of the binding terms, e.g. structural constituent of ribosome (9 genes), RNA binding (18 genes) and poly(A) RNA binding (16 genes). KEGG pathway enrichment analysis suggested the DEGs were mainly enriched in pathways of the ribosome (e.g. MRPS18C, RPL17 and RPL21), oxidative phosphorylation (e.g. ATP5I, ATP5J, and ATP5O) and Alzheimer's disease (e.g. ATP5J, ATP5O, COX7C) (Table 3).

Table 2. The enriched KEGG pathway of differential expression genes

Term	Count	P value	Genes
hsa3010: Ribosome	11	1.53E-12	MRPS18C, RPL17, RPL21, RPL23, RPL31, RPL36AL, RPS24, RPS25, RPS27, RPS27A, RPS27L
hsa190: Oxidative phosphorylation	8	7.84E-08	ATP5I, ATP5J, ATP5O, COX7C, NDUFA1, NDUFB3, NDUFS5, UQCRH
hsa5012: Parkinson s disease	7	2.75E-06	ATP5J, ATP5O, COX7C, NDUFA1, NDUFB3, NDUFS5, UQCRH
hsa5010: Alzheimer s disease	7	6.45E-06	ATP5J, ATP5O, COX7C, NDUFA1, NDUFB3, NDUFS5, UQCRH
hsa5016: Huntington s disease	7	1.01E-05	ATP5J, ATP5O, COX7C, NDUFA1, NDUFB3, NDUFS5, UQCRH
hsa4932: Non-alcoholic fatty liver disease (NAFLD)	4	0.00103	COX7C, NDUFA1, NDUFB3, UQCRH

Totally 51 DEGs were submitted to the STRING website to construct PPI networks (combined scores > 0.4) (Figure 3). We identified the ribosomal protein S17 (RPS17) as the most important hub gene.

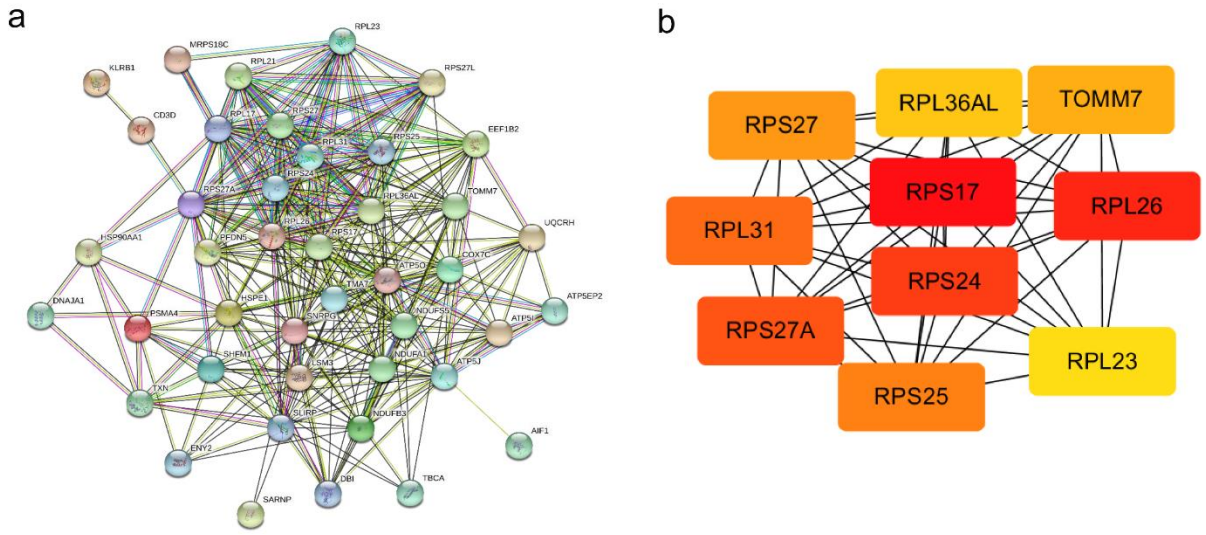

Figure 3. Protein-protein network of DEGs altered in AD patients. a, string database; b, the hub genes identified from protein-protein network

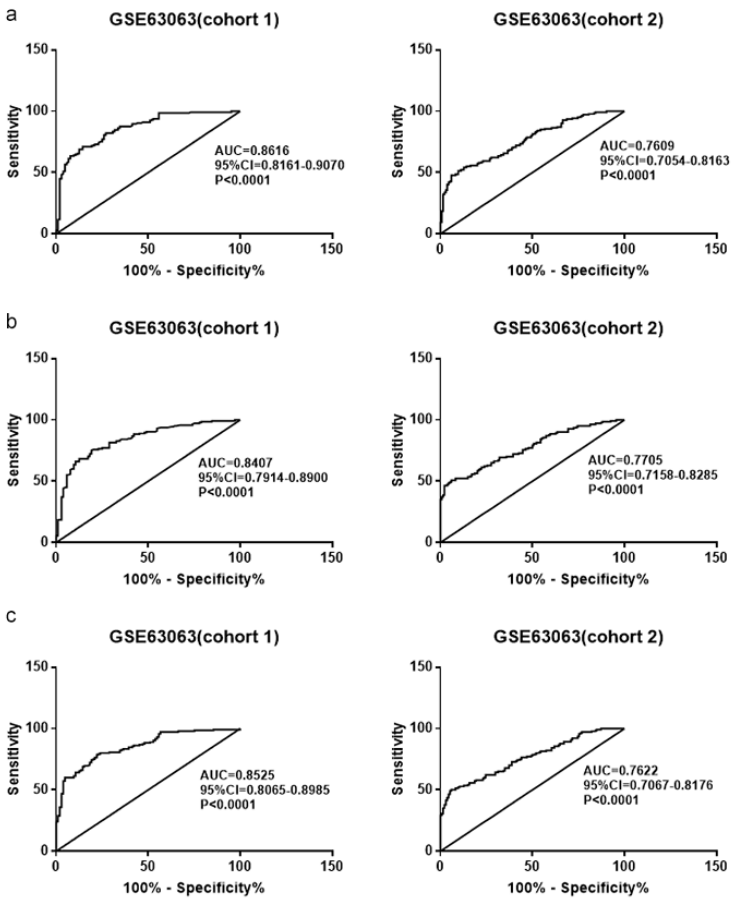

Figure 4. Receiver operating characteristic curve of the expression of three DEGs (a. NDUFA1; b, RPL36AL; c, NDUFS5)

3.3 Performance of These DEGs as a Biomarker

ROC analysis was performed to predict probable AD patients. Three genes show better predictive effect. The AUCs based on NDUFA1 expression was 0.8616 (GSE63063 cohort 1) and 0.7609 (GSE63063 cohort 2). For RPL36AL, the AD patients and healthy controls were classified with an area under the curve of 0.8407 and 0.7705 respectively. Further, we determined the sensitivity and specificity of NDUFS5 expression. AD patients showed AUCs of 0.8525 and 0.7622 respectively compared with healthy controls. This finding indicated that the expression level of these genes may be a new biomarker predicting AD patients.

4. Discussion

Our group conducted bioinformatics analyses of gene expression profiles from healthy controls and AD patients to explore the pathogenesis of AD. Our results suggested that these DEGs (50 down-regulated DEGs and 1 up-regulated DEGs) was associated with ribosome (CC), transcription (BP) and binding (MF) terms and the pathway of ribosome. RPS17 was considered as the most important gene of PPU models.

Ribosomal RNA genes modulate gene expression through cell transcription and translation processes and can. Payao et al. found that mature rRNA 28S and 18S ratio decrease significantly in the elderly groups compared to the young (Payao, Smith, Winter, & Bertolucci, 1998). The lowest 28S/18S ratio was observed in the AD patients (Payao et al., 1998). Changes in rRNA and rDNA expression were associated with cellular and organism aging and AD pathogenesis (Rasmussen et al., 2015). RNA oxidation is a prominent feature of vulnerable neurons in AD (Nunomura et al., 1999). Mouse experiment demonstrated that the deficiency of ribosomal protein S6 kinase 1 expression is beneficial to spatial memory and synaptic plasticity. In this study, we found that these DEGs (such as RPL17, RPS17, and RPS27) are mainly enriched in ribosome-related terms. The lack of the mouse RPL17 alters the diversity of mature ribosomes by enhancing production of shortened 5.8S rRNA (Wang, Parshin, Shcherbik, & Pestov, 2015). The function of the remaining genes in AD has not been reported. Previous studies reported that RPS17 mutation would cause the reduction of red blood cells production contributing to anemia (Kenney & Meng, 2015). In vitro knockdown of gene expression disturbed pre-ribosomal RNA processing, zebrafish models of RPL27 and RPL27 mutations showed impairments of erythrocyte production and tail and/or brain development (R. Wang et al., 2015). Some studies demonstrated that lower hemoglobin correlated with cognitive impairment and AD (Faux et al., 2014). In other words, individuals with anemia are more likely to get AD.

In addition, these DEGs (AD vs. control) were mainly involved in pathways of the ribosome (e.g. MRPS18C, RPL17 and RPL21), oxidative phosphorylation (e.g. ATP5I, ATP5J, and ATP5O) and Alzheimer's disease (e.g. ATP5J, ATP5O, COX7C). Ribosomal RNA is one of the most abundant molecules in most cells and is affected by oxidative stress in the human brain (Ding et al., 2012). Oxidative stress may augment the production and aggregation of Abeta and facilitate the phosphorylation and polymerization of tau, thus promoting the initiation and progression of AD (Zhao & Zhao, 2013). Genetic variants within oxidative phosphorylation genes increase the risk of AD (Biffi et al., 2014). Decreased expression of

COX7C occurs in total homogenates of the entorhinal cortex in AD stages V-VI when compared with stage I-II (Armand-Ugon, Ansoleaga, Berjaoui, & Ferrer, 2017). ATP5J may involve in the neurodegeneration and pathogenesis of Parkinson's disease by regulating oxidative phosphorylation (Kong et al., 2018).

Furthermore, 10 hub genes (RPS17, RPS24, RPL26, RPL31, RPS27A, RPS27, RPL36AL, TOMM7, RPS25, RPL23) were identified through MCC centrality. The human large subunit RPL36AL contacts the CCA end of P-site bound tRNA (Baouz et al., 2009). ROC analysis indicated that RPL36AL expression showed a better predictive effect. RPS25 was predicted as candidate biomarkers in peripheral blood for monitoring cardiac allograft rejection (Shen & Gong, 2015). Furthermore, RPS25 was identified to be a transcriptional target of p53; p53 directly bounded to RPS25 promoter and suppressed RPS25 expression (X. Zhang et al., 2013).

Ma et al. analyzed four GEO datasets (Hippocampus: GSE1297 and GSE5281; PBMCs: GSE18309 and GSE4226) in the study (Ma et al., 2019). GSE18309 dataset included 3 elderly samples and 6 AD patients, then sequencing with Affymetrix Human Genome U133 Plus 2.0 Array (Ma et al., 2019). GSE4226 dataset included 14 elderly samples and 14 AD patients, then sequencing with Affymetrix Human Genome U133 Plus 2.0 Array (Ma et al., 2019). In this study, GSE63063 dataset including 238 healthy controls and 285 AD patients was analyzed with Illumina HumanHT-12 V3.0 expression beadchip. The explanations for different DEGs (differential-expressed genes) among three datasets were as follows. First, the gender, age, lifestyle and disease severity of AD patients and sequencing platform were significantly different among three above datasets. Second, the sample size among three datasets was significantly different. The sample size of GSE18039 or GSE4226 is not enough to reach a convincing result compared to the GSE63063. Three, clinical heterogeneity and ethnic differences may also explain the different findings. Therefore, the data among three datasets is difficult to compare. Generally, the data from larger, well-designed dataset is more convincing.

Several limitations need to be addressed in this study. To begin with, we did not validate these findings using experiments due to limited experimental conditions. In addition, the confounding factors such as cigarette smoking, comorbid medical disease and socioeconomic status would affect the results. Three, the dataset was conducted in UK, which might be not applicable to other ethnicities.

In conclusion, this study explored the whole blood mRNA expression pattern of AD patients by bioinformatics methods. Furthermore, the expression levels of several genes were selected as potential biomarkers for predicting AD. However, larger studies are needed to confirm these results.

References

Armand-Ugon, M., Ansoleaga, B., Berjaoui, S., & Ferrer, I. (2017). Reduced Mitochondrial Activity is Early and Steady in the Entorhinal Cortex but it is Mainly Unmodified in the Frontal Cortex in Alzheimer's Disease. *Curr Alzheimer Res, 14*(12), 1327-1334.

https://doi.org/10.2174/1567205014666170505095921

Baouz, S., Woisard, A., Sinapah, S., Le Caer, J. P., Argentini, M., Bulygin, K., ... Hountondji, C. (2009). The human large subunit ribosomal protein L36A-like contacts the CCA end of P-site bound tRNA. *Biochimie, 91*(11-12), 1420-1425. https://doi.org/10.1016/j.biochi.2009.07.013

Biffi, A., Sabuncu, M. R., Desikan, R. S., Schmansky, N., Salat, D. H., Rosand, J., ... Alzheimer's disease Neuroimaging, I. (2014). Genetic variation of oxidative phosphorylation genes in stroke and Alzheimer's disease. *Neurobiol Aging, 35*(8), 1956, e1951-1958. https://doi.org/10.1016/j.neurobiolaging.2014.01.141

Blennow, K. (2005). CSF biomarkers for Alzheimer's disease: use in early diagnosis and evaluation of drug treatment. *Expert Rev Mol Diagn, 5*(5), 661-672. https://doi.org/10.1586/14737159.5.5.661

Ding, Q., Zhu, H., Zhang, B., Soriano, A., Burns, R., & Markesbery, W. R. (2012). Increased 5S rRNA oxidation in Alzheimer's disease. *J Alzheimers Dis, 29*(1), 201-209. https://doi.org/10.3233/JAD-2012-111058

Faux, N. G., Rembach, A., Wiley, J., Ellis, K. A., Ames, D., Fowler, C. J., ... Bush, A. I. (2014). An anemia of Alzheimer's disease. *Mol Psychiatry, 19*(11), 1227-1234. https://doi.org/10.1038/mp.2013.178

Henderson, V. W. (2014). Alzheimer's disease: review of hormone therapy trials and implications for treatment and prevention after menopause. *J Steroid Biochem Mol Biol, 142*, 99-106. https://doi.org/10.1016/j.jsbmb.2013.05.010

Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc, 4*(1), 44-57. https://doi.org/10.1038/nprot.2008.211

Humpel, C. (2011). Identifying and validating biomarkers for Alzheimer's disease. *Trends Biotechnol, 29*(1), 26-32. https://doi.org/10.1016/j.tibtech.2010.09.007

Kenney, S. P., & Meng, X. J. (2015). Identification and fine mapping of nuclear and nucleolar localization signals within the human ribosomal protein S17. *PLoS One, 10*(4), e0124396. https://doi.org/10.1371/journal.pone.0124396

Kong, P., Lei, P., Zhang, S., Li, D., Zhao, J., & Zhang, B. (2018). Integrated microarray analysis provided a new insight of the pathogenesis of Parkinson's disease. *Neurosci Lett, 662*, 51-58. https://doi.org/10.1016/j.neulet.2017.09.051

Lin, C. Y., Chin, C. H., Wu, H. H., Chen, S. H., Ho, C. W., & Ko, M. T. (2008). Hubba: hub objects analyzer--a framework of interactome hubs identification for network biology. *Nucleic Acids Res, 36*(Web Server issue), W438-443. https://doi.org/10.1093/nar/gkn257

Lin, Z., & Lin, Y. (2017). Identification of potential crucial genes associated with steroid-induced necrosis of femoral head based on gene expression profile. *Gene, 627*,

322-326. https://doi.org/10.1016/j.gene.2017.05.026

Ma, G., Liu, M., Du, K., Zhong, X., Gong, S., Jiao, L., & Wei, M. (2019). Differential Expression of mRNAs in the Brain Tissues of Patients with Alzheimer's Disease Based on GEO Expression Profile and Its Clinical Significance. *Biomed Res Int, 2019*, 8179145. https://doi.org/10.1155/2019/8179145

Nunomura, A., Perry, G., Pappolla, M. A., Wade, R., Hirai, K., Chiba, S., & Smith, M. A. (1999). RNA oxidation is a prominent feature of vulnerable neurons in Alzheimer's disease. *J Neurosci, 19*(6), 1959-1964. https://doi.org/10.1523/JNEUROSCI.19-06-01959.1999

O'Bryant, S. E., Xiao, G., Barber, R., Huebinger, R., Wilhelmsen, K., Edwards, M., ... Alzheimer's Disease Neuroimaging, I. (2011). A blood-based screening tool for Alzheimer's disease that spans serum and plasma: findings from TARC and ADNI. *PLoS One, 6*(12), e28092. https://doi.org/10.1371/journal.pone.0028092

Payao, S. L., Smith, M. A., Winter, L. M., & Bertolucci, P. H. (1998). Ribosomal RNA in Alzheimer's disease and aging. *Mech Ageing Dev, 105*(3), 265-272. https://doi.org/10.1016/S0047-6374(98)00095-5

Rasmussen, L., de Labio, R. W., Viani, G. A., Chen, E., Villares, J., Bertolucci, P. H., ... Payao, S. L. (2015). Differential Expression of Ribosomal Genes in Brain and Blood of Alzheimer's Disease Patients. *Curr Alzheimer Res, 12*(10), 984-989. https://doi.org/10.2174/1567205012666151027124017

Reitz, C., & Mayeux, R. (2014). Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochem Pharmacol, 88*(4), 640-651. https://doi.org/10.1016/j.bcp.2013.12.024

Shen, Z., & Gong, W. (2015). Identification of Candidate Biomarkers in Peripheral Blood for Cardiac Allograft Rejection based on Bioinformatics Analysis. *Ann Transplant, 20*, 312-319. https://doi.org/10.12659/AOT.893029

Sibener, L., Zaganjor, I., Snyder, H. M., Bain, L. J., Egge, R., & Carrillo, M. C. (2014). Alzheimer's Disease prevalence, costs, and prevention for military personnel and veterans. *Alzheimers Dement, 10*(3 Suppl), S105-110. https://doi.org/10.1016/j.jalz.2014.04.011

Sui, X., Liu, J., & Yang, X. (2014). Cerebrospinal fluid biomarkers of Alzheimer's disease. *Neurosci Bull, 30*(2), 233-242. https://doi.org/10.1007/s12264-013-1412-1

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., ... von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res, 43*(Database issue), D447-452. https://doi.org/10.1093/nar/gku1003

Wang, M., Parshin, A. V., Shcherbik, N., & Pestov, D. G. (2015). Reduced expression of the mouse ribosomal protein Rpl17 alters the diversity of mature ribosomes by enhancing production of shortened 5.8S rRNA. *RNA, 21*(7), 1240-1248. https://doi.org/10.1261/rna.051169.115

Wang, R., Yoshida, K., Toki, T., Sawada, T., Uechi, T., Okuno, Y., ... Ito, E. (2015). Loss of function mutations in RPL27 and RPS27 identified by whole-exome sequencing in Diamond-Blackfan anaemia. *Br J Haematol, 168*(6), 854-864. https://doi.org/10.1111/bjh.13229

Zhang, L., Huang, Y., Zhuo, W., Zhu, Y., Zhu, B., & Chen, Z. (2017). Identification and characterization of biomarkers and their functions for Lapatinib-resistant breast cancer. *Med Oncol, 34*(5), 89. https://doi.org/10.1007/s12032-017-0953-y

Zhang, X., Wang, W., Wang, H., Wang, M. H., Xu, W., & Zhang, R. (2013). Identification of ribosomal protein S25 (RPS25)-MDM2-p53 regulatory feedback loop. *Oncogene, 32*(22), 2782-2791. https://doi.org/10.1038/onc.2012.289

Zhao, Y., & Zhao, B. (2013). Oxidative stress and the pathogenesis of Alzheimer's disease. *Oxid Med Cell Longev, 2013*, 316523. https://doi.org/10.1155/2013/316523