

# SCIENTIFIC REPORTS



OPEN

## Identification of molecular alterations in leukocytes from gene expression profiles of peripheral whole blood of Alzheimer's disease

Hongdong Li<sup>1</sup>, Guini Hong<sup>1</sup>, Mengna Lin<sup>1</sup>, Yidan Shi<sup>1</sup>, Lili Wang<sup>1</sup>, Fengle Jiang<sup>1</sup>, Fan Zhang<sup>1</sup>, Yuhang Wang<sup>1</sup> & Zheng Guo<sup>1,2</sup>

Blood-based test has been considered as a promising way to diagnose and study Alzheimer's disease (AD). However, the changed proportions of the leukocytes under disease states could confound the aberrant expression signals observed in mixed-cell blood samples. We have previously proposed a method, Ref-REO, to detect the leukocyte specific expression alterations from mixed-cell blood samples. In this study, by applying Ref-REO, we detect 42 and 45 differentially expressed genes (DEGs) between AD and normal peripheral whole blood (PWB) samples in two datasets, respectively. These DEGs are mainly associated with AD-associated functions such as Wnt signaling pathways and mitochondrion dysfunctions. They are also reproducible in AD brain tissue, and tend to interact with the reported AD-associated biomarkers and overlap with targets of AD-associated PWB miRNAs. Moreover, they are closely associated with aging and have severer expression alterations in the younger adults with AD. Finally, diagnostic signatures are constructed from these leukocyte specific alterations, whose area under the curve (AUC) for predicting AD is higher than 0.73 in the two AD PWB datasets. In conclusion, gene expression alterations in leukocytes could be extracted from AD PWB samples, which are closely associated with AD progression, and used as a diagnostic signature of AD.

Alzheimer's disease (AD) is the predominant form of dementia. The pathological features of AD include the presence of amyloid plaques, neurofibrillary tangles, synaptic loss, soluble amyloid- $\beta$  (A $\beta$ ) oligomers, neuritic dystrophy, and eventual neurodegeneration<sup>1</sup>. In clinical practice, AD diagnosis is mainly based on PET imaging or cerebral spinal fluid biomarkers. The major disadvantages of these diagnostic approaches are the high cost, the low patient compliance, and most importantly, the difficulty in diagnosing AD at an early stage<sup>2</sup>.

The natural role of blood cells in immune response to physiologic and pathologic changes has made blood an important source for investigation of disease-associated molecular biomarkers<sup>3</sup>. Recent studies have also demonstrated a significant degree of covariability in gene expression between brain tissue and peripheral blood cells<sup>4-7</sup>. Therefore, a diagnostic blood biomarker for AD would be valuable and convenient for the early diagnosis of patients presenting at clinics with memory complaints. Actually, the potential use of peripheral whole blood (PWB) or peripheral blood mononuclear cell (PBMC) gene expression profiling in the diagnosis of brain disorders has been described<sup>6-10</sup>. These studies detected and analyzed the significantly altered genes<sup>6,9</sup> or modules<sup>8,10</sup> by directly comparing the expression measurements between AD and normal blood samples. It's noted that relative proportions of the blood cells may shift under disease states<sup>11,12</sup>, which may confound the aberrant disease signals originated from leukocytes when directly comparing expression values of genes between disease and normal blood samples<sup>13,14</sup>. Consequently, the changed proportions of leukocyte subtypes could introduce some differentially expressed genes (DEGs) between disease samples and normal controls which actually have no expression changes in any leukocyte subtypes<sup>15</sup>. Therefore, it is necessary to exclude alteration signals originating from leukocyte subtype proportion changes when trying to detect AD-associated cellular molecular changes from mixed-cell blood samples.

<sup>1</sup>Department of bioinformatics, Key Laboratory of Ministry of Education for Gastrointestinal Cancer, Fujian Medical University, Fuzhou, China. <sup>2</sup>Fujian Key Laboratory of Tumor Microbiology, Fujian Medical University, Fuzhou, China. Hongdong Li and Guini Hong contributed equally to this work. Correspondence and requests for materials should be addressed to G.H. (email: [gnhong@fjmu.edu.cn](mailto:gnhong@fjmu.edu.cn))

Dataset*	Characteristic of Sample	Age (Years)	Platform (#Gene)	GEO Accession ID	Ref
<b>Purified leukocyte subtypes (PLS)</b>					
PLS-47	All: 47 Monocytes: 10, B cells:5, CD4+ T cell:5, NK cells:5, CD8+ T cell:5, Eosinophils:4, mDCs:5; Neutrophils:3, pDCs:5	—	GPL570 (11,241)	GSE28490	21
PLC-33	All: 33 CD19+ B cells: 5, CD14+ monocytes:5, CD4+ T cells:5, CD8+ T cells:5, Eosinophils:3, NK cells:5; Neutrophils:5	—	GPL570 (10,689)	GSE28491	21
<b>peripheral whole blood (PWB)</b>					
PWB-AD-249	ALL:249 Control: 104; AD:145	52 ~ 90	GPL1122 (21067)	GSE63060	22
PWB-AD-275	ALL:275 Control: 135;AD:140	57 ~ 100	GPL1122 (18327)	GSE63061	22
PWB-Normal-61	61	18 ~ 56	GPL571 (12432)	GSE19151	23
<b>Brain tissue</b>					
Brain-AD-161	Brain region (Control: AD) Entorhinal Cortex (13:10) Hippocampus (13:10) Middle temporal gyrus (12:16) Posterior cingulate cortex (13:9) Superior frontal gyrus (11:23) Primary visual cortex Control (12:19)	63 ~ 102	GPL570 (20848)	GSE5281	24

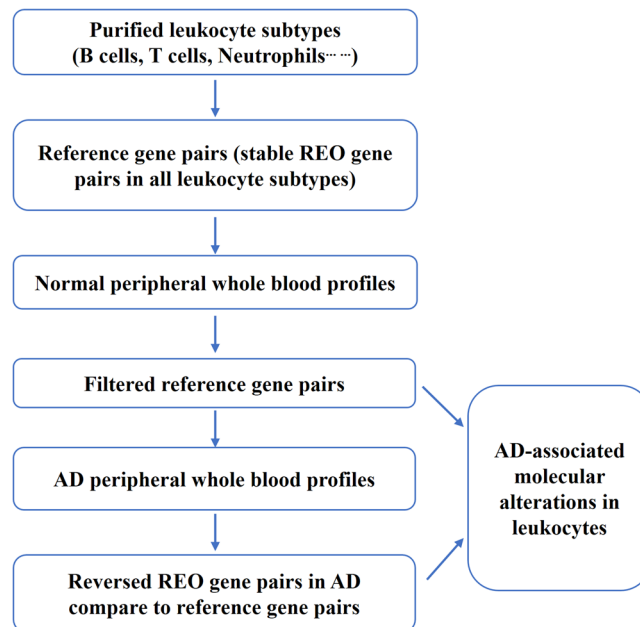
**Table 1.** Datasets analyzed in this study. \*The abbreviation of each dataset is denoted by the phenotype followed by sample size.

In recent years, researchers have developed methods based on deconvolution<sup>16</sup> or surrogate variable analysis algorithms<sup>17,18</sup> to avoid the influence of relative leukocyte subtype proportion changes on the overall signals of PWB or PBMCs. Methods based on deconvolution algorithms aim to estimate and adjust the proportion of each leukocyte subtype in blood samples using the expression profiles of purified leukocyte subtypes<sup>16</sup>. However, the absolute quantitative gene expression level measurements used in these methods could be sensitive to systematic biases of microarray measurements especially examined in different microarray platforms<sup>19</sup>. Methods based on surrogate variable analysis aim to find true disease-associated alterations by estimating and adjusting the confounding factors that could have effects on gene expression levels<sup>17,18</sup>. However, it's difficult for them to avoid the influence of cell proportion changes that are indeed associated with disease progression<sup>15</sup>. More recently, we proposed a method, *Ref-REO*, to detect leukocyte-specific molecular alterations from mixed-cell blood samples of patients through analyzing the disrupted patterns of the pre-determined within-sample relative expression orderings (REOs) of genes which are consistent in purified normal leukocyte subtypes<sup>15</sup>. This method is based on the fundamental that if the REOs of any two genes have consistent patterns (eg.  $E_A > E_B$ ) in all normal leukocyte subtypes, these consistent patterns could be stable in PWB or PBMCs, no matter how the proportion of the constituent cells changes when no expression alterations occur in leukocytes. If inconsistent patterns are observed in disease samples, at least one of these two genes has altered gene expression in certain leukocyte subtypes. The *Ref-REO* method has been shown to have higher precision and recall than the previous methods<sup>15</sup>. Most importantly, the REOs of genes have been reported to be more robust than the absolute measured levels as REOs are invariant to monotonic data transformation (normalization) and rather resistant to batch effects<sup>19,20</sup>, indicating the disease-associated biomarkers detected by this method could be easily validated and transferred.

Therefore, in this study, we apply the *Ref-REO* method to detect and analyze the AD-associated cellular expression alterations from two independent AD PWB datasets. The results showed that these AD-associated molecular alterations detected from PWB by *Ref-REO* were significantly enriched in AD-associated pathways. They were reproducible in brain tissue of AD-patients, and had interactions with reported AD biomarkers and overlaps with the targets of AD-associated miRNAs.

## Materials and Methods

**Datasets.** The gene expression data were downloaded from the Gene Expression Omnibus database (GEO, <http://www.ncbi.nlm.nih.gov/geo/>). Detailed information for each dataset was described in Table 1. The PLS-47 (GSE28490) dataset examined 47 expression profiles for nine leukocyte subtypes, which were isolated from healthy human blood and assessed for cell type purity by flow cytometry<sup>21</sup>. The PLS-33 (GSE28491) dataset examined 33 expression profiles for seven leukocyte subtypes, which were obtained from a separate panel of healthy donors at the University Hospital of Geneva. These two datasets were used to detect the gene pairs with stable REOs in each purified leukocytes<sup>21</sup>. The PWB-AD-249 (GSE63060) and PWB-AD-275 (GSE63061) datasets examined the PWB expression profiles for AD and normal control samples which were obtained from the AddNeuroMed consortium, a large cross-European AD biomarker study and a follow-on Dementia Case Register (DCR) cohort in London<sup>22</sup>. These two datasets were used to detect the AD-associated molecular alterations in leukocytes. The PWB-Normal-61 dataset included the PWB expression profiles for 61 healthy controls with age ranging from 18 to 56, which were obtained from the GEO dataset (GSE19151)<sup>23</sup>. In GSE19151, control samples of unknown age were excluded, thus only the 61 samples with age information were collected in PWB-Normal-61, which was used to detect aging-associated genes. The Brain-AD-161 (GSE5281) dataset examined 161 expression profiles of six brain regions for AD and normal control samples<sup>24</sup>. This dataset was used to evaluate whether the AD-associated PWB molecular alterations had expression changes in brain tissue. For each dataset, the normalized data were downloaded from GEO. The original platform annotation file obtained from GEO for each dataset



**Figure 1.** The flow chart of detecting AD-associated cellular alterations by the *Ref-REO* method.

was used to annotate the CloneIDs to GeneIDs. The number of genes measured in each dataset was shown in Table 1. Totally, 8,708 genes commonly measured in all datasets were analyzed in the study.

**Detecting disease-associated cellular alterations from AD PWB samples.** The *Ref-REO* method was employed to detect the AD-associated cellular alterations from AD PWB samples<sup>15</sup>. Briefly, the algorithm detected the disease-associated cellular alterations according to the following steps as shown in Fig. 1: (1) Select reference gene pairs. Reference gene pairs are gene pairs whose REOs are stable and consistent across different purified normal leukocytes. The REOs of these gene pairs could be stable under normal condition or disease state, no matter how the proportion of the constituent cells changes when no expression alterations occur in leukocytes. (2) Filter the reference gene pairs by the control samples in the dataset under study to exclude the gene pairs whose REOs are easily affected by age, sex, experimental batch effects<sup>25</sup> and other factors. (3) Detect reversed gene pairs. Reversed gene pairs are the gene pairs that have inconsistent REO patterns with reference gene pairs in the disease samples. The reversed REO patterns of these gene pairs should be caused by expression alterations occurred in leukocyte subtypes, given that the changed proportion of the constituent cells could not affect REOs of the reference gene pairs when no expression alterations occur in leukocytes. (4) Detect DEGs. Based on the filtered reference gene pairs and reversed gene pairs, whether a gene could be observed in the reversed gene pairs by random chance was evaluated by the hypergeometric distribution model. The DEGs were detected as significant if the adjusted *P*-value was less than 0.05<sup>15</sup>.

**Detecting DEGs in various brain regions of AD patients.** The student's *t*-test was used to detect DEGs in various brain regions of AD patients comparing to normal controls. *P*-values were adjusted for multiple testing using the Benjamini-Hochberg procedure to control the FDR level. If the adjusted *p*-value for a gene was less than 0.05, this gene was considered as a DEG.

**Detecting aging-associated genes from normal PWB samples.** Because the relative proportions of the blood cells in older adults shifted weakly as age increases<sup>12</sup>, the linear regression model was employed to detect the aging-associated genes in the normal PWB controls of the datasets analyzed in this study. *P*-values were adjusted for multiple testing using the Benjamini-Hochberg procedure<sup>26</sup> to control the FDR level at 0.05. If the adjusted *p*-value for a gene was less than 0.05, this gene was considered as an aging-associated gene.

**Random experiments.** Two random experiments were performed in this study.

The first random experiment was performed to evaluate whether the number of reversed gene pairs detected in a dataset was significantly more than expected by chance. Supposed there are *n* reversed gene pairs observed in a study dataset. The random experiment consists of the following steps: (1) Randomly disturb sample labels. The sample sizes of normal controls and AD samples are kept the same in randomized data and in the study dataset. (2) Calculate the number of reversed gene pairs detected in the randomized data, denoted as *m*. (3) Repeat step 2 for 1,000 times and calculate the percentage of the random experiments in which *m* is larger than *n*, defined as the probability of observing *n* reversed gene pairs by random chance. A *p*-value < 0.05 was considered as significant.

The second random experiment was performed to evaluate whether the number of observed AD-DEGs having interactions with the AD-associated biomarkers was significantly more than expected by chance. Suppose *k* out of *n* AD-DEGs interact with at least one of the AD-associated biomarkers, the random experiments were done

with the following steps: (1) Randomly select  $n$  genes from the background genes as the randomized AD-DEGs. (2) Calculate the number of the randomly defined AD-DEGs that interact with at least one of the AD-associated biomarkers, denoted as  $m$ . (3) Repeat step 2 for 1,000 times and calculate the percentage of the random experiments in which  $m$  is larger than  $k$ , defined as the probability of observing  $k$  AD-DEGs interacting with at least one of AD-associated biomarkers by random chance. A  $p$ -value  $<0.05$  was considered as significant.

**Enrichment analysis.** The KEGG (Kyoto Encyclopedia of Genes and Genomes) and Gene Ontology databases were used to evaluate the AD-associated cellular alterations in PWB by the functional annotation tool DAVID (<https://david.ncifcrf.gov/>, version 6.8)<sup>27</sup>. For a given dataset, all of the measured genes annotated in the KEGG or Gene Ontology database were considered as the background genes.

**String, AlzGene, and miRTarBase databases.** The protein-protein interactions were downloaded from STRING database (<https://string-db.org/>, Version 10) that collected known and predicted 82,160 protein-protein interactions involving 7,638 proteins<sup>28</sup>. The AD-associated biomarkers were download from AlzGene database which is a collection of published AD genetic association studies<sup>29</sup> including 618 genes (<http://alsgene.org/>, download at April, 2017). These two databases were used to evaluate the interactions between AD-associated cellular alterations in PWB and AD-associated alterations collected in AlzGene database. MiRNA-mRNA interactions were downloaded from miRTarBase database at April 2017 (<http://mirtarbase.mbc.nctu.edu.tw/>). In this study, only the experimentally validated microRNA-target interactions were downloaded, which included 322,161 miRNA-mRNA interactions involving 2,618 miRNAs and 14,831 mRNAs<sup>30</sup>.

## Results

**AD-associated cellular alterations in PWB.** First, gene pairs with stable REOs in different normal leukocytes were detected in purified leukocyte expression datasets using the *Ref-REO* method. Totally, 9,638,173 and 8,832,824 gene pairs were detected in the gene expression profiles of purified leukocyte subtypes examined in PLS-47 and PLS-33, respectively. The two lists shared 6,133,414 gene pairs, and 99.9% of them had consistent REO patterns, which was unlikely to happen by chance (binomial distribution test,  $p$ -value  $<2.2 \times 10^{-16}$ ). These consistent gene pairs (totally 6,124,866) were used as the reference for recognition of abnormal disease signals in leukocytes.

Then, the reference gene pairs were evaluated in AD PWB expression profiles examined in PWB-AD-249 and PWB-AD-275, respectively. For dataset PWB-AD-249, 4,524,607 of the reference gene pairs retained the REO patterns in at least 95% of the 104 normal PWB profiles. Among them, 1,145 gene pairs had significantly reversed REO patterns in the 145 AD PWB profiles (Fisher exact test, adjusted  $p$ -value  $<0.05$ ). For dataset PWB-AD-275, 4,673,704 genes pairs had consistent REO patterns with the reference in at least 95% of the 135 normal PWB profiles. Among them, 1,249 gene pairs showed significantly reversed REO patterns in the 140 AD PWB profiles (Fisher exact test, adjusted  $p$ -value  $<0.05$ ). Though relatively little in quantity, these reversed gene pairs may reflect the true AD-associated information, as no reversed genes pairs could be detected in the 1,000 random experiments by randomly disturbing the sample labels of controls and cases ( $p$ -value  $<0.001$ , see Materials and Methods).

Based on the reference gene pairs and reversed gene pairs, DEGs were detected from PWB-AD-249 and PWB-AD-275 using the *Ref-REO* method<sup>15</sup>. Totally, with FDR  $<5\%$ , 42 and 45 DEGs were detected respectively (Table 2), which shared 21 genes. All of the 21 shared DEGs had consistent expression dysregulated directions (up- or down-regulation) in AD samples compared to normal samples in the two datasets, which was unlikely to happen by chance (binomial distribution test,  $p$ -value  $<2.2 \times 10^{-16}$ ). This result indicated the cellular alterations detected from AD PWB could be reproducible. In the following study, genes detected as significant in at least one of these two AD datasets were analyzed, denoted as AD-DEGs, which included 66 genes.

For the AD-DEGs, the enrichment analysis was conducted based on KEGG and Gene Ontology databases by DAVID<sup>27</sup>. The result showed that these DEGs were significantly enriched in oxidative phosphorylation, proteasome and ribosome pathways in KEGG, and Wnt signaling pathway and translation pathway in Gene Ontology (Table 3). These enriched pathways have been demonstrated to be closely associated with AD development and progression. For example, oxidative phosphorylation has been reported to be down-regulated in AD patients<sup>31,32</sup>. Ribosome dysfunction has been considered as an early event in AD progression<sup>33</sup>. The inhibition of proteasome activity has been reported to be sufficient to induce neuron degeneration in AD<sup>34</sup>. Specially, 10 AD-DEGs were enriched in mitochondrial inner membrane, suggesting that mitochondrial dysfunctions could play an important role in AD development. In fact, mitochondrial dysfunctions have been found in PWB lymphocytes of AD patients, which have been considered as a trigger of AD pathophysiology<sup>35,36</sup>.

### Expression changes of AD-associated PWB cellular alterations in various brain regions of AD patients.

The Brain-AD-161 dataset, which examined expression profiles of six different brain regions in AD and normal controls, was used to evaluate the expression changes of PWB AD-DEGs in brain tissue of AD patients. By using the Student's  $t$ -test with FDR  $<5\%$ , the DEGs were detected from various brain regions of AD patients (Table 4). The AD-DEGs were found to significantly overlap with the DEGs detected from hippocampus, middle temporal gyrus, superior frontal gyrus and primary visual cortex of AD patients (Table 4). For example, among the 66 AD-DEGs, 53 DEGs were detected as DEGs in superior frontal gyrus of AD patients (hypergeometric distribution test,  $p$ -value =  $7.30 \times 10^{-4}$ ) and 49 out of them had consistent alteration directions in AD patients compared to normal controls. Totally, 60 out of the 66 AD-DEGs had expression changes in at least one of the six brain regions between AD and normal control samples, among which 52 genes had consistent alteration directions in PWB and brain tissue. The results further suggested that the cellular alterations observed in PWB could be closely associated with AD.

Gene ID	Symbol	Direction	P-value	Gene ID	Symbol	Direction	P-value
DEGs detected from PWB-AD-249				DEGs detected from PWB-AD-275			
5684	<i>PSMA3</i>	down	$<2.2 \times 10^{-16}$	2079	<i>ERH</i>	down	$<2.2 \times 10^{-16}$
9360	<i>PPIG</i>	down	$<2.2 \times 10^{-16}$	6160	<i>RPL31</i>	down	$<2.2 \times 10^{-16}$
51574	<i>LARP7</i>	down	$<2.2 \times 10^{-16}$	9552	<i>SPAG7</i>	down	$<2.2 \times 10^{-16}$
51611	<i>DPH5</i>	down	$<2.2 \times 10^{-16}$	51188	<i>SS18L2</i>	down	$<2.2 \times 10^{-16}$
84987	<i>COX14</i>	down	$<2.2 \times 10^{-16}$	5716	<i>PSMD10</i>	down	$<2.2 \times 10^{-16}$
7381	<i>UQCRB</i>	down	$7.00 \times 10^{-9}$	7155	<i>TOP2B</i>	down	$1.35 \times 10^{-11}$
6741	<i>SSB</i>	down	$1.98 \times 10^{-8}$	522	<i>ATP5J</i>	down	$4.74 \times 10^{-11}$
219927	<i>MRPL21</i>	down	$3.83 \times 10^{-6}$	20	<i>ABCA2</i>	up	$2.32 \times 10^{-8}$
57396	<i>CLK4</i>	down	$4.04 \times 10^{-6}$	5204	<i>PFDN5</i>	down	$5.59 \times 10^{-8}$
10600	<i>USP16</i>	down	$1.83 \times 10^{-5}$	79746	<i>ECHDC3</i>	up	$5.69 \times 10^{-8}$
51637	<i>C14orf166</i>	down	$1.03 \times 10^{-4}$	10632	<i>ATP5L</i>	down	$6.84 \times 10^{-8}$
11168	<i>PSIP1</i>	down	$3.31 \times 10^{-4}$	23500	<i>DAAM2</i>	up	$1.85 \times 10^{-6}$
126208	<i>ZNF787</i>	up	$5.17 \times 10^{-4}$	27089	<i>UQCRQ</i>	down	$3.35 \times 10^{-4}$
762	<i>CA4</i>	up	$6.52 \times 10^{-4}$	8664	<i>EIF3D</i>	down	$6.21 \times 10^{-4}$
23478	<i>SEC. 11 A</i>	down	$3.00 \times 10^{-3}$	5683	<i>PSMA2</i>	down	$1.69 \times 10^{-3}$
65056	<i>GPBP1</i>	down	$3.32 \times 10^{-3}$	51019	<i>CCDC53</i>	down	$1.73 \times 10^{-3}$
5685	<i>PSMA4</i>	down	$1.20 \times 10^{-2}$	29093	<i>MRPL22</i>	down	$1.89 \times 10^{-3}$
202018	<i>TAPT1</i>	down	$1.51 \times 10^{-2}$	9550	<i>ATP6V1G1</i>	down	$2.56 \times 10^{-3}$
84343	<i>HPS3</i>	up	$1.55 \times 10^{-2}$	6256	<i>RXRA</i>	up	$4.15 \times 10^{-3}$
6038	<i>RNASE4</i>	up	$1.49 \times 10^{-2}$	23294	<i>ANKS1A</i>	up	$4.65 \times 10^{-3}$
54556	<i>ING3</i>	down	$4.01 \times 10^{-2}$	4082	<i>MARCKS</i>	up	$1.18 \times 10^{-3}$
				85464	<i>SSH2</i>	up	$1.66 \times 10^{-2}$
				286006	<i>LSMEM1</i>	up	$3.46 \times 10^{-2}$
				22900	<i>CARD8</i>	up	$4.26 \times 10^{-2}$
<b>Overlapped DEGs between PWB-AD-249 and PWB-AD-275</b>							
521	<i>ATP5I</i>	down	$<2.2 \times 10^{-16}$	51371	<i>POMP</i>	down	$<2.2 \times 10^{-16}$
2959	<i>GTF2B</i>	down	$<2.2 \times 10^{-16}$	80135	<i>RPF1</i>	down	$<2.2 \times 10^{-16}$
3301	<i>DNAJA1</i>	down	$<2.2 \times 10^{-16}$	3476	<i>IGBP1</i>	down	$9.67 \times 10^{-13}$
4694	<i>NDUFA1</i>	down	$<2.2 \times 10^{-16}$	153527	<i>ZMAT2</i>	down	$1.06 \times 10^{-11}$
6119	<i>RPA3</i>	down	$<2.2 \times 10^{-16}$	8813	<i>DPM1</i>	down	$2.22 \times 10^{-11}$
6154	<i>RPL26</i>	down	$<2.2 \times 10^{-16}$	25847	<i>ANAPC13</i>	down	$6.66 \times 10^{-9}$
6233	<i>RPS27A</i>	down	$<2.2 \times 10^{-16}$	388789	<i>LINC00493</i>	down	$1.12 \times 10^{-6}$
9553	<i>MRPL33</i>	down	$<2.2 \times 10^{-16}$	1622	<i>DBI</i>	down	$7.53 \times 10^{-6}$
51258	<i>MRPL51</i>	down	$<2.2 \times 10^{-16}$	6645	<i>SNTB2</i>	up	$3.35 \times 10^{-5}$
51503	<i>CWC15</i>	down	$<2.2 \times 10^{-16}$	55505	<i>NOP10</i>	down	$3.98 \times 10^{-3}$
55854	<i>ZC3H15</i>	down	$<2.2 \times 10^{-16}$				

**Table 2.** DEGs detected from PWB-AD-249 and PWB-AD-275.

**AD-associated cellular alterations in PWB tend to interact with reported AD-biomarkers.** The AD-DEGs were compared with the 618 AD-associated biomarkers collected from AlzGene database<sup>29</sup>. The result showed that only three AD-DEGs (*ABCA2*, *CARD8* and *RXRA*) overlapped with the AD-associated biomarkers. With the protein-protein interaction data from STRING, we found 23 AD-DEGs each interacting with at least one of the AD-associated biomarkers (Fig. 2). The number was significantly more than expected by chance (p-value = 0.027): when randomly choosing the same number of genes as AD-DEGs, the mean number of random DEGs having interactions with the AD-associated biomarkers was  $16.37 \pm 3.6878$ , which was estimated in the 1,000 random experiments (see Materials and Methods). With FDR < 0.05, the KEGG pathway enrichment analysis showed the interaction partners of AD-DEGs also tended to be enriched in the AD-associated oxidative phosphorylation (p-value =  $5.2 \times 10^{-4}$ ) and RRAR signaling pathways (p-value =  $1.30 \times 10^{-6}$ )<sup>37</sup>.

AD-associated aberrant miRNA alterations observed in PWB were also analyzed for the AD-DEGs. The PWB aberrant miRNAs were collected from a study, which detected 12 diagnostic miRNAs (brain-miR-112, brain-miR-161, hsa-let-7d-3p, hsa-miR-5010-3p, hsa-miR-26a-5p, hsa-miR-1285-5p, hsa-miR-151a-3p, hsa-miR-103a-3p, hsa-miR-107, hsa-miR-532-5p, hsa-miR-26b-5p and hsa-let-7f-5p) between 48 AD patients and 22 non-AD control PWB samples<sup>38</sup>. From the miRTarBase database<sup>30</sup>, 1620 targets of the 10 miRNAs (two miRNAs weren't collected by miRTarbase) were downloaded. Among the 66 AD-DEGs, 19 were included in the 1620 targets, which was unlikely to happen by chance (hypergeometric distribution test, p-value = 0.028) (Fig. 3).

The above results further indicated that the AD-DEGs could be closely associated with AD development and progression.

ID	Name	Hit Genes	P-value
<b>KEGG</b>			
hsa00190	Oxidative phosphorylation	<i>ATP5I, ATP5J, NDUFA1, UQCRB, ATP6V1G1, ATP5L, UQCRQ</i>	$6.63 \times 10^{-5}$
hsa03010	Ribosome	<i>RPL26, RPL31, RPS27A, MRPL21, MRPL22, MRPL33</i>	$6.62 \times 10^{-4}$
hsa03050	Proteasome	<i>PSMA2, PSMA3, PSMA4, POMP</i>	$5.15 \times 10^{-3}$
<b>Gene ontology</b>			
molecular function			
GO:0003735	structural constituent of ribosome	<i>RPL26, RPL31, RPS27A, MRPL21, MRPL22, MRPL33, MRPL51</i>	$2.43 \times 10^{-4}$
O:0003723	RNA binding	<i>RPF1, LARP7, RPL31, C14orf166, ZMAT2, RPL26, MRPL21, SSB, CWC15, EIF3D</i>	$1.18 \times 10^{-3}$
cellular component			
GO:0005743	mitochondrial inner membrane	<i>ATP5I, ATP5J, MRPL22, UQCRB, ATP5L, MRPL21, MRPL51, NDUFA1, UQCRQ</i>	$2.25 \times 10^{-4}$
GO:0019773	proteasome core complex,	<i>PSMA2, PSMA3, PSMA4, POMP</i>	$8.89 \times 10^{-4}$
biological process			
GO:0060071	Wnt signaling pathway	<i>PSMA2, PSMA3, PSMA4, RPS27A, PSMD10</i>	$9.50 \times 10^{-4}$
GO:0006412	translation	<i>RPL26, RPL31, RPS27A, MRPL21, MRPL22, MRPL33, MRPL51</i>	$3.57 \times 10^{-3}$

**Table 3.** Pathways and functional terms enriched for AD-DEGs.

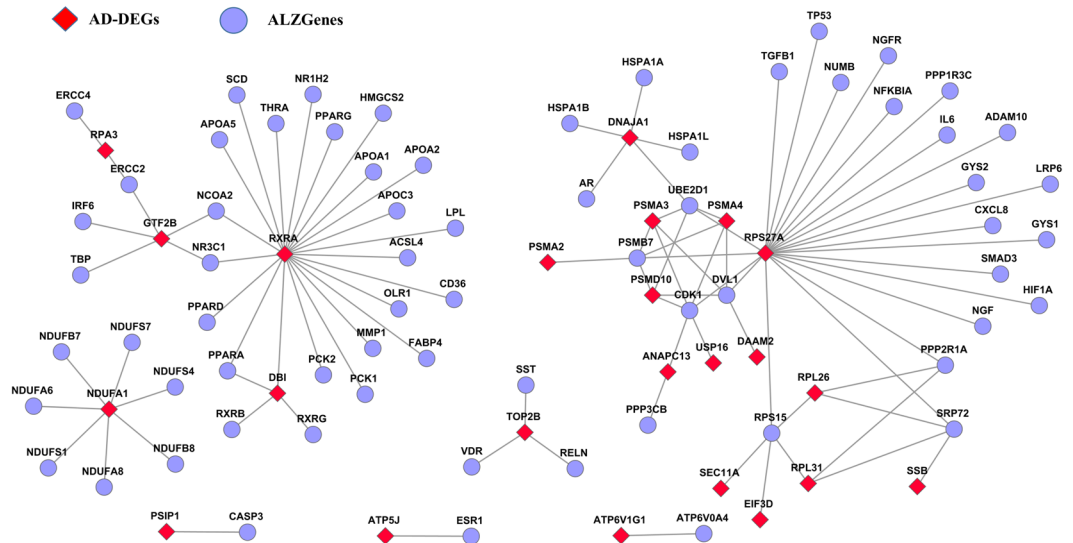
Brain region	DEGs	Overlapped with AD-DEG	P-value*
Entorhinal Cortex	1334	5 (5) <sup>#</sup>	0.98
hippocampus	2346	26 (21)	$1.90 \times 10^{-2}$
Middle temporal gyrus	2253	28 (26)	$2.52 \times 10^{-3}$
Posterior cingulate cortex	3273	24 (24)	0.64
Superior frontal gyrus	5344	53 (49)	$7.3 \times 10^{-4}$
Primary visual cortex	373	9 (9)	$1.91 \times 10^{-3}$

**Table 4.** Comparison of PWB AD-DEGs with DEGs detected from various brain regions of AD patients.

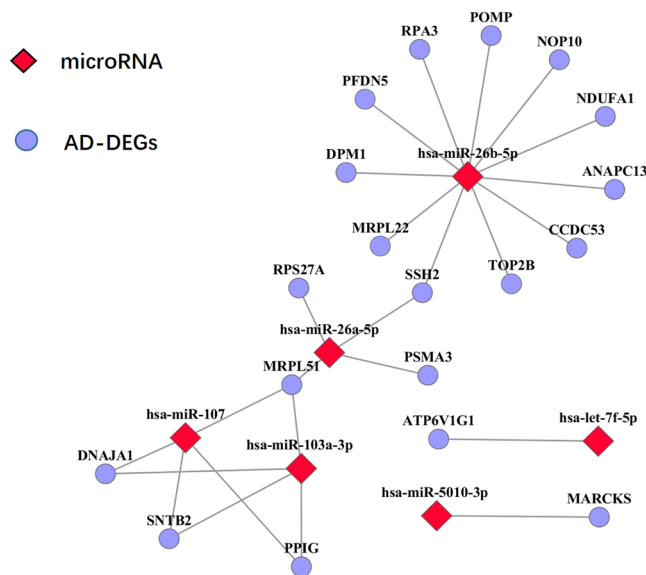
<sup>#</sup>The number outside the parentheses indicates the number of overlapped genes between AD-DEGs and brain tissue DEGs and the number inside the parentheses indicates the number of overlapped genes with consistent expression alterations in both PWB and brain tissue of AD patients. \*Represent the probability of observing the number of overlapped genes by chance calculated by the hypergeometric distribution model.

**AD-associated cellular alterations in PWB and aging-associated genes.** Aging-associated genes were detected from normal control samples in PWB-AD-249, PWB-AD-275 and PWB-Normal-61 using the linear regression model (see Materials and Methods). With FDR < 5%, 1,332 and 3,146 aging-associated genes were identified in PWB-AD-249 and PWB-Normal-61, respectively, while no aging-associated genes were identified in PWB-AD-275. The numbers of detected aging-associated genes differed significantly in the three datasets, which could be explained by the different age distribution patterns: there were 104 normal controls with ages ranging from 52 to 87 in PWB-AD-249, while in PWB-AD-275, there were 135 normal controls with ages ranging from 63 to 91; and the PWB-AD-249 dataset had more samples with age  $\leq 65$  (9.62%) than the PWB-AD-275 dataset (2.22%, Fisher's exact test, p-value = 0.019). In PWB-Normal-61, the age distribution was wider: there was 61 samples with ages ranging from 18 to 56. Compared the aging-associated genes identified in PWB-AD-249 and PWB-Normal-61, the results showed that these two lists shared 515 aging-associated genes (hypergeometric distribution test, p-value =  $1.90 \times 10^{-2}$ ), and 479 out of them had the same alteration directions with age in both datasets, which was unlikely to happen by chance (binomial distribution test, p-value <  $2.2 \times 10^{-16}$ ). Thus, only the 479 genes were analyzed in the following study, which were denoted as aging-DEGs.

The aging-DEGs and the AD-DEGs shared 21 genes, and all of them had consistent directions of correlations with age and expression changes (hypergeometric distribution test, p-value =  $2.02 \times 10^{-11}$ ). The pathway enrichment analysis showed that these aging- and AD-associated genes were also enriched in proteasome (p-value =  $1.82 \times 10^{-4}$ ) and RRAR signaling pathways (p-value =  $3.70 \times 10^{-3}$ ). Interestingly, more reserved gene pairs were found from AD patients with age  $\leq 65$  than AD patients with age > 65. In PWB-AD-249, the average numbers of reversed gene pairs identified from AD patients with age  $\leq 65$  and age > 65 were  $529.64 \pm 286.45$  and  $269.18 \pm 228.31$  (Wilcoxon rank sum test, p-value =  $4.5 \times 10^{-3}$ ), respectively. In PWB-AD-275, there were only three AD patients with age  $\leq 65$ . Therefore, the phenomenon that the number of reserved gene pairs differed between the two groups was unobvious: the average numbers of reversed gene pairs identified from AD patients with age  $\leq 65$  and age > 65 were  $282.52 \pm 176.77$  and  $228.67 \pm 163.95$  (Wilcoxon rank sum test, p-value = 0.57), respectively. Actually, previous studies have compared the prevalence of a number of clinical features occurring in patients with early- and late-onset primary degenerative dementia of the Alzheimer type. The early-onset group



**Figure 2.** Interactions between AD-DEGs and AD-associated biomarkers.

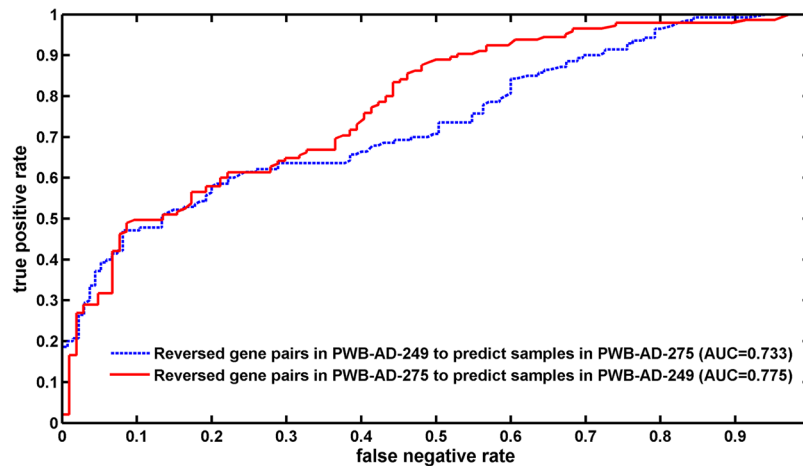


**Figure 3.** Interactions between AD-DEGs and PWB aberrant miRNAs.

demonstrated a greater prevalence of language disturbance, a disproportionate number of left-handers and a much shorter relative survival time than the late-onset group<sup>39</sup>.

These results showed that the occurrence of AD is closely related to aging, further suggesting that AD could be an excessive aging disease.

**Prediction performance of AD-associated cellular alterations in PWB.** One task of the blood-based test is to detect biomarkers for disease diagnostic workup. In this study, the detections of DEGs were based on gene pairs that have reversed REO patterns in AD patients compared with normal controls. These reversed gene pairs could be naturally used as diagnostic biomarkers for AD. The result showed that these reversed gene pairs had good prediction performance for classifying AD and normal samples: the 1,145 reversed gene pairs detected in PWB-AD-249 discriminated AD from the normal samples in PWB-AD-275 with an area under the curve (AUC) of 0.733, while the 1,249 reversed gene pairs detected in PWB-AD-275 discriminated AD from the normal samples in PWB-AD-249 with an AUC of 0.775 (Fig. 4). The prediction performances were better than the reported prediction performance by Nicola *et al.* (AUC 0.729,<sup>10</sup>), which used a pathway based classification method based on the same data.



**Figure 4.** The prediction performance of reversed gene pairs.

## Discussion

The concept of using blood such as peripheral blood cells as the source of information to detect disease-associated molecular alterations relies on the natural role of these cells in immune response to physiologic and pathologic changes<sup>40</sup>. In this study, we tried to detect AD-associated leukocyte cellular alterations from AD peripheral whole blood samples using the *Ref-REO* method<sup>15</sup>, which is based on within-sample REOs in purified leukocytes.

In the study, the number of PWB AD-DEGs identified by *Ref-REO* was relatively low, suggesting that the *Ref-REO* method had its disadvantages to some extent. Firstly, this method may fail to detect some disease-associated cellular alterations, as the reference pairs were required to have stable REO patterns in all purified leukocytes and in more than 95% of normal PWB samples, which could miss some AD-associated alterations. Secondly, the method of *Ref-REO* itself tended to detect DEGs with larger magnitude of cellular expression alterations, as greater expression alterations will have greater REO changes. Thus, genes with weak changes may be missed. Lastly, the *Ref-REO* method may not be able to detect those alterations occurring in leukocyte subtypes with limited proportions, as the signal could be easily covered by other leukocyte subtypes. However, it's important to note that the PWB alterations identified by *Ref-REO* were closely associated with the development and progression of disease: the AD-DEGs were significantly enriched in AD-associated pathways, reproducible in brain tissue of AD patients and tended to interact with reported AD-associated biomarkers. Therefore, though low in quantity, the AD-DEGs identified by *Ref-REO* could be real expression alterations occurring in PWB leukocytes<sup>15</sup>.

Interestingly, our results showed that few AD-associated cellular alterations were enriched in PWB immune associated pathways or functional terms. In contrast, most of them were involved in aging-associated pathways or functional terms. For example, oxidative phosphorylation and mitochondrial have been reported to be closely associated with aging<sup>41</sup>. In AD blood samples, an increased neutrophil-lymphocyte ratio has also been reported as a function of age<sup>42</sup>. These results suggested that AD-associated cellular alterations in PWB might mainly reflect aging-associated alterations. Fortunately, such alterations could also be detected in the diagnostic workup of Alzheimer's disease (AD), as AD is an excessive aging disease.

## References

- Peng, S. *et al.* Decreased brain-derived neurotrophic factor depends on amyloid aggregation state in transgenic mouse models of Alzheimer's disease. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **29**, 9321–9329, <https://doi.org/10.1523/JNEUROSCI.4736-08.2009> (2009).
- Dubois, B. *et al.* Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *The Lancet. Neurology* **6**, 734–746, [https://doi.org/10.1016/S1474-4422\(07\)70178-3](https://doi.org/10.1016/S1474-4422(07)70178-3) (2007).
- Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature biotechnology* **31**, 142–147, <https://doi.org/10.1038/nbt.2487> (2013).
- Sullivan, P. F., Fan, C. & Perou, C. M. Evaluating the comparability of gene expression in blood and brain. *American journal of medical genetics. Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics* **141B**, 261–268, <https://doi.org/10.1002/ajmg.b.30272> (2006).
- Liew, C. C., Ma, J., Tang, H. C., Zheng, R. & Dempsey, A. A. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *The Journal of laboratory and clinical medicine* **147**, 126–132, <https://doi.org/10.1016/j.lab.2005.10.005> (2006).
- Maes, O. C. *et al.* Transcriptional profiling of Alzheimer blood mononuclear cells by microarray. *Neurobiology of aging* **28**, 1795–1809, <https://doi.org/10.1016/j.neurobiolaging.2006.08.004> (2007).
- Grunblatt, E. *et al.* Gene expression as peripheral biomarkers for sporadic Alzheimer's disease. *Journal of Alzheimer's disease: JAD* **16**, 627–634, <https://doi.org/10.3233/JAD-2009-0996> (2009).
- Fehlbaum-Beurdeley, P. *et al.* Toward an Alzheimer's disease diagnosis via high-resolution blood gene expression. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **6**, 25–38, <https://doi.org/10.1016/j.jalz.2009.07.001> (2010).
- Booij, B. B. *et al.* A gene expression pattern in blood for the early detection of Alzheimer's disease. *Journal of Alzheimer's disease: JAD* **23**, 109–119, <https://doi.org/10.3233/JAD-2010-101518> (2011).
- Voyle, N. *et al.* A Pathway Based Classification Method for Analyzing Gene Expression for Alzheimer's Disease Diagnosis. *Journal of Alzheimer's disease: JAD* **49**, 659–669, <https://doi.org/10.3233/JAD-150440> (2016).



11. Reinius, L. E. *et al.* Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS one* **7**, e41361, <https://doi.org/10.1371/journal.pone.0041361> (2012).
12. Li, H. *et al.* Common DNA methylation alterations of Alzheimer's disease and aging in peripheral whole blood. *Oncotarget*, <https://doi.org/10.18632/oncotarget.7862> (2016).
13. Hong, G., Zhang, W., Li, H., Shen, X. & Guo, Z. Separate enrichment analysis of pathways for up- and downregulated genes. *Journal of the Royal Society, Interface* **11**, 20130950, <https://doi.org/10.1098/rsif.2013.0950> (2014).
14. Li, H. *et al.* Similar blood-borne DNA methylation alterations in cancer and inflammatory diseases determined by subpopulation shifts in peripheral leukocytes. *British journal of cancer* **111**, 525–531, <https://doi.org/10.1038/bjc.2014.347> (2014).
15. Hong, G. *et al.* A simple way to detect disease-associated cellular molecular alterations from mixed-cell blood samples. *Briefings in bioinformatics*, <https://doi.org/10.1093/bib/bbx009> (2017).
16. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics* **13**, 86, <https://doi.org/10.1186/1471-2105-13-86> (2012).
17. Chakraborty, S., Datta, S. & Datta, S. Surrogate variable analysis using partial least squares (SVA-PLS) in gene expression studies. *Bioinformatics* **28**, 799–806, <https://doi.org/10.1093/bioinformatics/bts022> (2012).
18. Houseman, E. A., Molitor, J. & Marsit, C. J. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* **30**, 1431–1439, <https://doi.org/10.1093/bioinformatics/btu029> (2014).
19. Qi, L. *et al.* Critical limitations of prognostic signatures based on risk scores summarized from gene expression levels: a case study for resected stage I non-small-cell lung cancer. *Briefings in bioinformatics* **17**, 233–242, <https://doi.org/10.1093/bib/bbv064> (2016).
20. Wang, H. *et al.* Individual-level analysis of differential expression of genes and pathways for personalized medicine. *Bioinformatics* **31**, 62–68, <https://doi.org/10.1093/bioinformatics/btu522> (2015).
21. Allantaz, F. *et al.* Expression profiling of human immune cell subsets identifies miRNA-mRNA regulatory relationships correlated with cell type specific expression. *PLoS one* **7**, e29979, <https://doi.org/10.1371/journal.pone.0029979> (2012).
22. Sood, S. *et al.* A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. *Genome biology* **16**, 185, <https://doi.org/10.1186/s13059-015-0750-x> (2015).
23. Lewis, D. A. *et al.* Whole blood gene expression analyses in patients with single versus recurrent venous thromboembolism. *Thrombosis research* **128**, 536–540, <https://doi.org/10.1016/j.thromres.2011.06.003> (2011).
24. Liang, W. S. *et al.* Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain. *Physiological genomics* **28**, 311–322, <https://doi.org/10.1152/physiolgenomics.00208.2006> (2007).
25. Leek, J. T. *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* **11**, 733–739, <https://doi.org/10.1038/nrg2825> (2010).
26. Benjamini Y, H. Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *journal of the royal statistical society series B* **57** (1995).
27. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**, 1–13, <https://doi.org/10.1093/nar/gkn923> (2009).
28. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research* **45**, D362–D368, <https://doi.org/10.1093/nar/gkw937> (2017).
29. Bertram, L., McQueen, M. B., Mullin, K., Blacker, D. & Tanzi, R. E. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nature genetics* **39**, 17–23, <https://doi.org/10.1038/ng1934> (2007).
30. Chou, C. H. *et al.* miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic acids research* **44**, D239–247, <https://doi.org/10.1093/nar/gkv1258> (2016).
31. Chandrasekaran, K., Hatanpaa, K., Brady, D. R., Stoll, J. & Rapoport, S. I. Downregulation of oxidative phosphorylation in Alzheimer disease: loss of cytochrome oxidase subunit mRNA in the hippocampus and entorhinal cortex. *Brain research* **796**, 13–19 (1998).
32. Biffi, A. *et al.* Genetic variation of oxidative phosphorylation genes in stroke and Alzheimer's disease. *Neurobiology of aging* **35**(1956), e1951–1958, <https://doi.org/10.1016/j.neurobiolaging.2014.01.141> (2014).
33. Ding, Q., Markesbery, W. R., Chen, Q., Li, F. & Keller, J. N. Ribosome dysfunction is an early event in Alzheimer's disease. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **25**, 9171–9175, <https://doi.org/10.1523/JNEUROSCI.3040-05.2005> (2005).
34. Keller, J. N., Hanni, K. B. & Markesbery, W. R. Impaired proteasome function in Alzheimer's disease. *Journal of neurochemistry* **75**, 436–439 (2000).
35. Sultana, R. *et al.* Lymphocyte mitochondria: toward identification of peripheral biomarkers in the progression of Alzheimer disease. *Free radical biology & medicine* **65**, 595–606, <https://doi.org/10.1016/j.freeradbiomed.2013.08.001> (2013).
36. Moreira, P. I., Carvalho, C., Zhu, X., Smith, M. A. & Perry, G. Mitochondrial dysfunction is a trigger of Alzheimer's disease pathophysiology. *Biochimica et biophysica acta* **1802**, 2–10, <https://doi.org/10.1016/j.bbadis.2009.10.006> (2010).
37. Zolezzi, J. M. *et al.* PPARs in the central nervous system: roles in neurodegeneration and neuroinflammation. *Biological reviews of the Cambridge Philosophical Society*. <https://doi.org/10.1111/brv.12320> (2017).
38. Leidinger, P. *et al.* A blood based 12-miRNA signature of Alzheimer disease patients. *Genome biology* **14**, R78, <https://doi.org/10.1186/gb-2013-14-7-r78> (2013).
39. Seltzer, B. & Sherwin, I. A comparison of clinical features in early- and late-onset primary degenerative dementia. *One entity or two?* *Archives of neurology* **40**, 143–146 (1983).
40. LaBrecche, H. G., Nevins, J. R. & Huang, E. Integrating factor analysis and a transgenic mouse model to reveal a peripheral blood predictor of breast tumors. *BMC medical genomics* **4**, 61, <https://doi.org/10.1186/1755-8794-4-61> (2011).
41. Li, H. *et al.* Aging-associated mitochondrial DNA mutations alter oxidative phosphorylation machinery and cause mitochondrial dysfunctions. *Biochimica et biophysica acta*. <https://doi.org/10.1016/j.bbadis.2017.05.022> (2017).
42. Kuyumcu, M. E. *et al.* The evaluation of neutrophil-lymphocyte ratio in Alzheimer's disease. *Dementia and geriatric cognitive disorders* **34**, 69–74, <https://doi.org/10.1159/000341583> (2012).

## Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (grant numbers 81501215, 81501829, 81372213 and 81572935) and Natural Science Foundation of Fujian Province, China (grant number 2016J01706) and Joint Fund for Program of Science and Technology innovation of Fujian Province, China (grant number 2016Y9102, 2016Y9044).

## Author Contributions

All authors meet the authorship requirements. G.N.H. and H.D.L. designed the study. M.N.L., Y.D.S. and L.L.W. analyzed the data. F.L.J., F.Z. and Y.H.W. collected the gene expression datasets. G.N.H. and H.D.L. drafted the manuscript. Z.G. revised the manuscript critically for important intellectual content. G.N.H., H.D.L. agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors read and approved the final manuscript.

## Additional Information

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2017