



OPEN

# Alzheimer's disease may develop from changes in the immune system, cell cycle, and protein processing following alterations in ribosome function

Akiko Yamakawa<sup>1</sup>, Mutsumi Saganuma<sup>1</sup>, Risa Mitsumori<sup>1</sup>, Shumpei Niida<sup>2</sup>, Kouichi Ozaki<sup>1,3,4</sup> & Daichi Shigemizu<sup>1,3</sup>✉

The prevalence of Alzheimer's disease (AD) is increasing as society ages. The details of AD pathogenesis have not been fully elucidated, and a comprehensive gene expression analysis of the process leading up to the onset of AD would be helpful for understanding the mechanism. We performed an RNA sequencing analysis on a cohort of 1227 Japanese blood samples, representing 424 AD patients, 543 individuals with mild cognitive impairment (MCI), and 260 cognitively normal (CN) individuals. A total of 883 and 1169 statistically significant differentially expressed genes (DEGs) were identified between CN and MCI (CN-MCI) and between MCI and AD (MCI-AD), respectively. Pathway analyses using these DEGs, followed by protein–protein interaction network analysis, revealed key roles of ribosomal function in MCI progression, whereas immune responses, cell cycle, and protein processing in endoplasmic reticulum were involved in AD progression. Our findings indicate that the onset of AD might be associated with gene expression changes in the immune system, cell cycle, and protein processing following alterations in the expression of ribosomal protein genes during the MCI stage, although validation using brain tissue samples will be necessary in the future. Given the known effectiveness of delaying MCI progression in preventing AD, the genes related to ribosomal function might emerge as biomarkers for early diagnosis.

**Keywords** Alzheimer's disease, Mild cognitive impairment, RNA sequencing

With the aging of the global population, the prevalence of dementia is expected to reach an estimated 152.8 million individuals worldwide by 2050<sup>1</sup>. Alzheimer's disease (AD) is the leading cause of dementia, accounting for 60–80% of all cases<sup>2</sup>. The majority of AD cases are sporadic and are typically diagnosed in people over 65 years of age (late-onset AD). AD is characterized by its heterogeneity, with intricate interactions among various genetic and environmental risk factors. The heritability of AD is estimated to be in the range of 58–79%<sup>3</sup>. The ε4 polymorphism in the protein encoded by the apolipoprotein E (*APOE*) gene, located on chromosome 19, is recognized as the most potent genetic risk factor for AD<sup>4,5</sup>. Recently, additional genetic risk factors influencing the development of AD have been identified by genome-wide association studies (GWAS) using a large number of samples<sup>6</sup>. However, a substantial proportion of the heritability remains unexplained. Genetic variants can affect gene expression, and perturbations in gene expression are frequently involved in the onset of various diseases. Additionally, perturbations in gene expression due to aging can be factors that ultimately lead to disease.

One of the most powerful tools for the comprehensive analysis of all gene expression in an organism is whole RNA sequencing (RNA-seq) using a next-generation sequencer<sup>7</sup>. The identification of differentially expressed genes (DEGs) between patients and healthy individuals in RNA-seq data, and subsequent gene set enrichment analysis (GSEA) of the expression signatures, contributes to our understanding of the biological mechanisms underlying disease pathogenesis and the identification of potential disease biomarkers. Recently, RNA-seq

<sup>1</sup>Medical Genome Center, Research Institute, National Center for Geriatrics and Gerontology, 7-430 Morioka-cho, Obu 474-8511, Aichi, Japan. <sup>2</sup>Research Institute, National Center for Geriatrics and Gerontology, Obu 474-8511, Aichi, Japan. <sup>3</sup>Department of Cardiovascular Medicine, Hiroshima University Graduate School of Biomedical and Health Sciences, Hiroshima 734-8551, Japan. <sup>4</sup>RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Kanagawa, Japan. ✉email: daichi@ncgg.go.jp

data from human blood have offered promising biomarkers for the diagnosis and treatment of a wide variety of diseases, including cancers<sup>8</sup>, circadian sleep-wake disorders<sup>9</sup>, and depression<sup>10</sup>. For AD, some potential biomarkers (e.g., *PHGDH*, *BASE-AS1*, *NEAT1*, and *GAS5*) have been identified from blood-based RNA-seq data<sup>11,12</sup>. Our previous study also reported several candidate AD biomarkers (e.g., *EEF2*, *RPL7*, and *WDR37*) through analyses of blood-based RNA-seq data<sup>13,14</sup> and miRNA expression data<sup>15</sup>. However, AD progresses through three stages—cognitively normal (CN), mild cognitive impairment (MCI), and dementia—and stage-specific gene expression during the transitions from CN to MCI (CN-MCI) and from MCI to AD (MCI-AD) would shed light on what causes the progression from one stage to the next and have not yet been sufficiently investigated.

Here, we explored the stage-specific gene expression patterns in CN-MCI and MCI-AD by using a large-scale retrospective Japanese cohort dataset. We observed the involvement of the ribosomal pathway related genes in CN-MCI and genes associated with the immune system, cell cycle, and protein processing in MCI-AD. Our findings indicate that the onset of AD may be associated with gene expression changes in the immune system, cell cycle, and protein processing, following alterations in the expression of ribosomal protein genes during the MCI stage. Given the effectiveness of delaying MCI progression in preventing AD<sup>16</sup>, the expression of ribosome-related genes may serve as potential biomarkers. We believe that these characteristics will help to elucidate the mechanism of AD onset.

## Results

### Sample collection

A total of 1227 samples were included in this study, from 424 individuals with AD, 543 with MCI, and 260 CN individuals. The mean age and *APOE* ε4 allele frequency (AF) were highest for AD (age = 79.5 years, AF = 0.23), followed by MCI (age = 76.8 years, AF = 0.19), and CN (age = 72.6 years, AF = 0.085). Each category had more samples from females than from males, with the largest difference in the MCI group (Table 1). Similar allele frequencies were reported by Farrer et al.<sup>17</sup>.

### Distribution of white blood cell types

We investigated the differences in cell type distribution between CN and MCI samples and between MCI and AD samples. By using CIBERSORT (see Methods), we estimated the relative proportions (as TPM) of 12 major leukocyte cell types in each sample (Fig. 1a). Of those, only γδ T cells showed a statistically significant difference between CN and MCI samples ( $FDR = 7.88 \times 10^{-3}$ ). Four cell types (B cells,  $2.78 \times 10^{-7}$ ; plasma cells,  $6.28 \times 10^{-3}$ ; γδ T cells,  $1.64 \times 10^{-7}$ ; neutrophils,  $3.62 \times 10^{-2}$ ) showed statistically significant differences between MCI and AD samples.

### Detection of DEGs

To elucidate differences in DEGs between CN and MCI, and between MCI and AD, we conducted two-group comparisons. A total of 883 and 1169 statistically significant DEGs with Entrez Gene IDs were identified in CN-MCI and MCI-AD, respectively, with statistical significance defined as  $FDR < 0.05$  and  $|FC| > 1.2$  (Supplementary Table 1). In CN-MCI, 293 DEGs were upregulated and 590 were downregulated (Fig. 1b). In MCI-AD, 1027 DEGs were upregulated and 142 were downregulated (Fig. 1c). Notably, a substantial number of upregulated genes were obtained in MCI-AD samples, whereas CN-MCI samples exhibited predominantly downregulated genes.

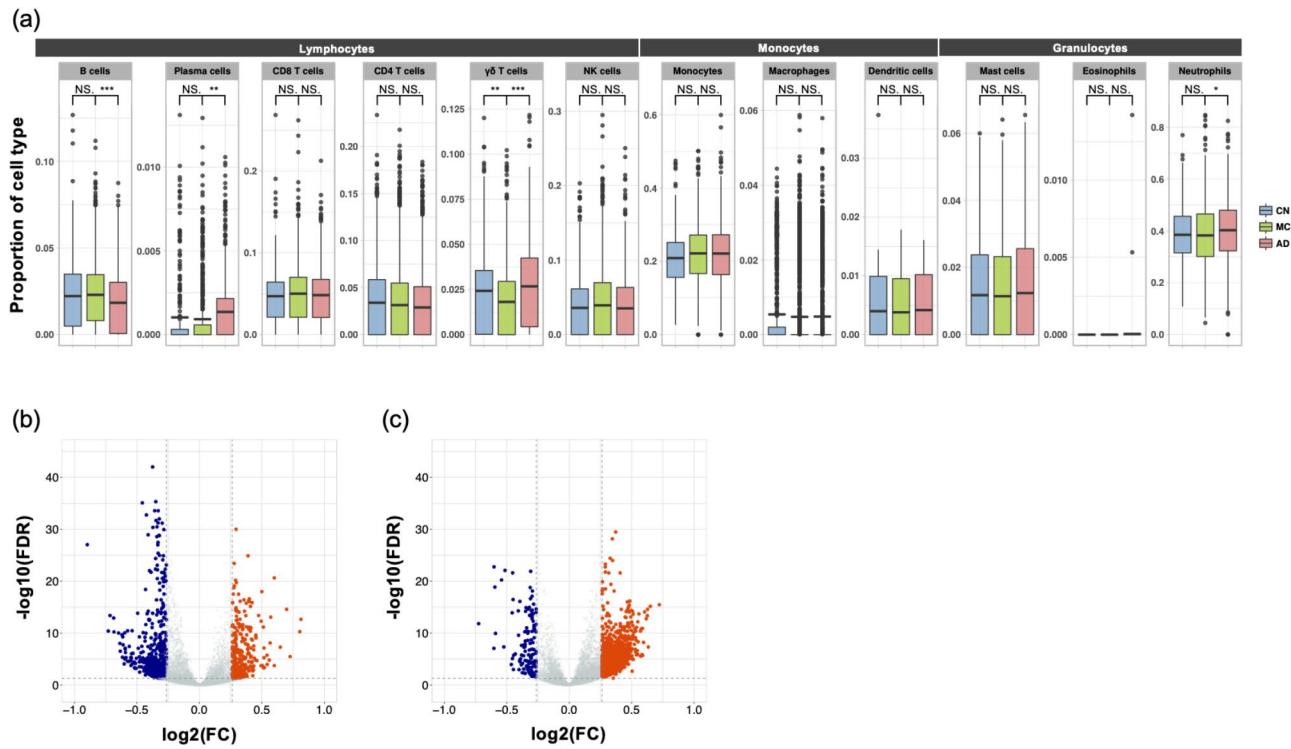
### Gene set enrichment analysis of DEGs

To obtain further insight into the biological functions of the DEGs, we performed GSEA by using the DAVID gene functional classification tool. Our analysis revealed that DEGs were enriched in two KEGG biological pathways (coronavirus disease – COVID-19 and ribosome) for CN-MCI and in five pathways (TCR signaling pathway, oocyte meiosis, protein processing in endoplasmic reticulum, salmonella infection, and NF-κB signaling pathways) for MCI-AD. The significance level was set at  $FDR < 0.05$  (Table 2). The number of unique genes was 38 in CN-MCI and 101 in MCI-AD.

In CN-MCI, a considerable number of genes were involved in the COVID-19 pathway, with more than 70% (26/36) overlapping with those in the ribosomal pathway (Supplementary Fig. 1a). In contrast, in MCI-AD, there were fewer gene overlaps among the identified pathways. However, more than 57% of genes (58/101) were associated with immune-related pathways, including the TCR signaling, salmonella infection, and NF-κB

	AD	MCI	CN
Number of samples	424	543	260
Male:Female	1:1.04	1:1.41	1:1.08
Age (mean ± 1 SD)	$79.5 \pm 5.4$	$76.8 \pm 6.4$	$72.6 \pm 6.0$
Number of <i>APOE</i> ε4 alleles (ratio)			
2	25 (0.06)	22 (0.04)	4 (0.01)
1	149 (0.35)	164 (0.30)	36 (0.14)
0	250 (0.59)	357 (0.66)	220 (0.85)

**Table 1.** Summary of characteristics of AD, MCI, and CN samples



**Fig. 1.** Proportions of immune cell types and distribution of differentially expressed genes.

Proportions of the 12 major immune cell types among CN, MCI, and AD samples (a). Statistical significance was set at a false discovery rate  $< 0.05$ . \*:  $< 0.05$ ; \*\*:  $< 0.01$ ; \*\*\*:  $< 0.001$ , NS: not significant. Data are represented as box and whisker plots, depicting minimum, lower quartile (Q1), mean (Q2), upper quartile (Q3), and maximum values. Distribution of differentially expressed genes in CN-MCI (b) and MCI-AD (c) comparisons. Each point represents a gene. Red and blue dots represent upregulated and downregulated genes, respectively. FC, fold change; FDR, false discovery rate.

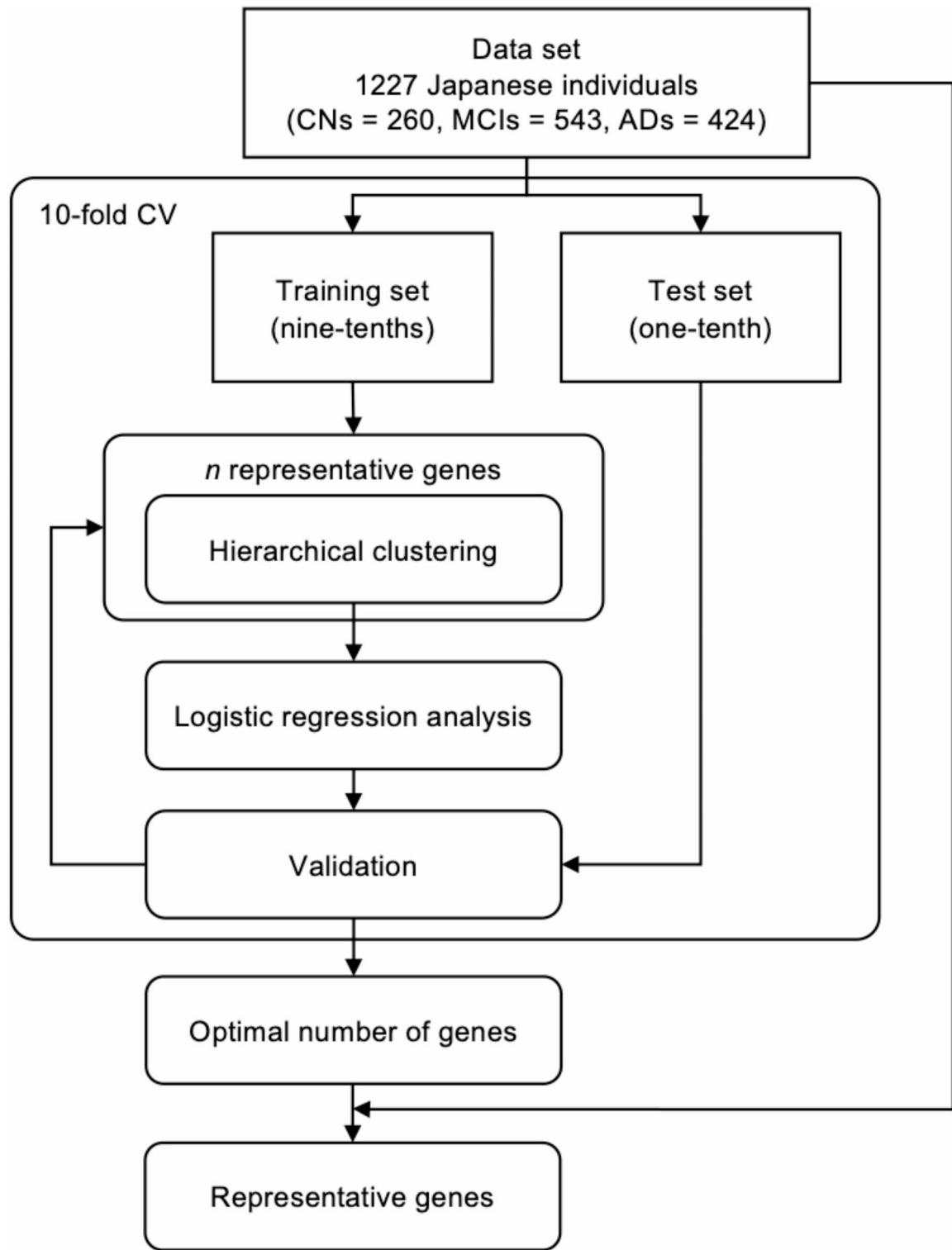
Case	KEGG pathway	Description	Number of genes (n)	FDR (BH)	Unique genes (n)
CN-MCI	hsa05171	Coronavirus disease – COVID-19	36	$6.05 \times 10^{-8}$	38
	hsa03010	Ribosome	28	$8.49 \times 10^{-7}$	
MCI-AD	hsa04660	T cell receptor signaling pathway	20	$2.76 \times 10^{-2}$	101
	hsa04114	Oocyte meiosis	22	$4.12 \times 10^{-2}$	
	hsa04141	Protein processing in endoplasmic reticulum	26	$4.19 \times 10^{-2}$	
	hsa05132	Salmonella infection	33	$4.19 \times 10^{-2}$	
	hsa04064	NF-κB signaling pathway	18	$4.95 \times 10^{-2}$	

**Table 2.** Pathways and numbers of genes in pathways from gene set enrichment analysis. BH: Benjamini-Hochberg.

signaling pathways (Supplementary Fig. 1b). Given the involvement of many immune-related genes in the progression to AD, we further investigated potential differences in the diversities of TCR and BCR repertoires. However, no statistically significant difference was observed in either MCI-AD or CN-MCI (Supplementary Fig. 2).

### Extraction of representative genes

We detected 38 candidate genes in CN-MCI samples and 101 in MCI-AD samples through GSEA. To determine the optimal  $n$  representative genes from these candidates ( $n \leq 38$  for CN-MCI;  $n \leq 101$  for MCI-AD, see the Methods), we applied a logistic regression method. Gene selection was based on hierarchical clustering, with nine-tenths of the dataset for clustering and model fitting in each cross-validation step. The adjusted model was then evaluated on the remaining one-tenth of the dataset, with the process repeated 10 times (10-fold cross-validation, Fig. 2). Consequently, we identified 8 representative genes for CN-MCI and 22 for MCI-AD (Supplementary Fig. 3).

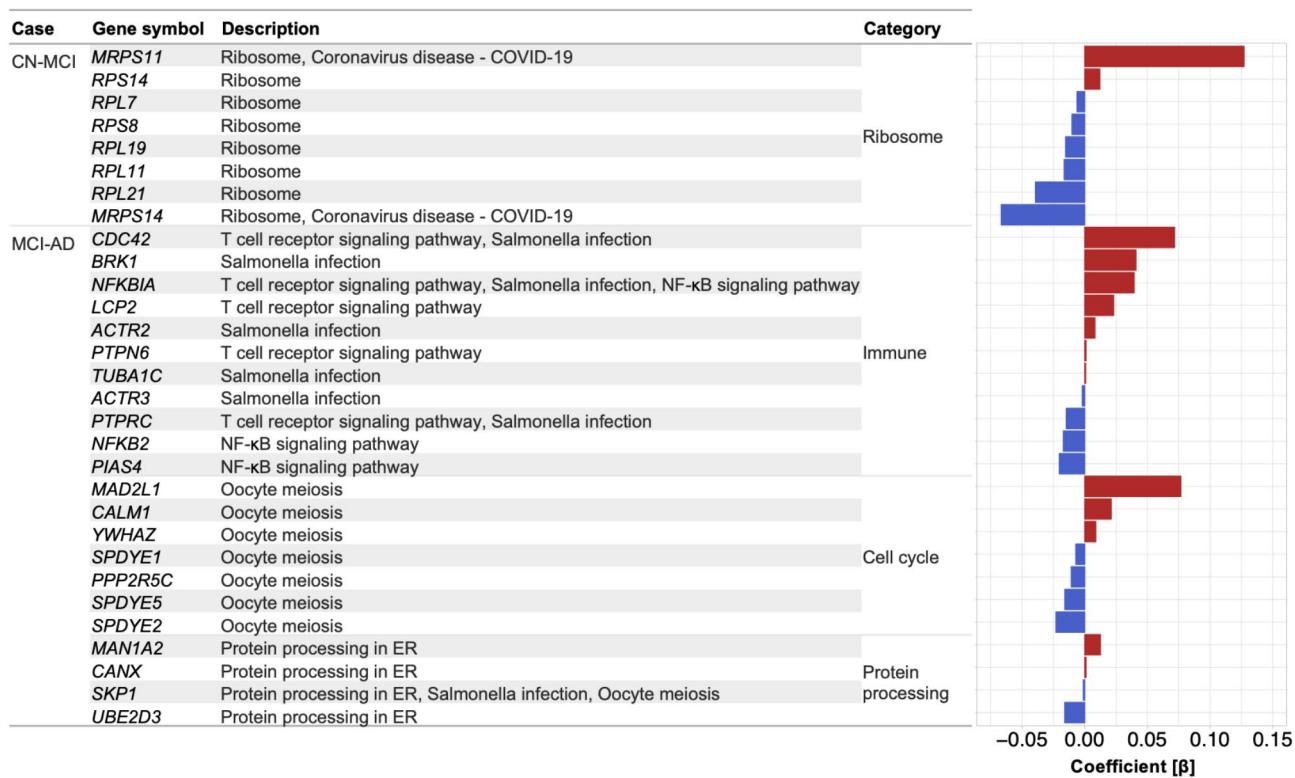


**Fig. 2.** Outline of the representative genes' extraction.

AD, Alzheimer's disease; CN, cognitively normal; CV, cross-validation; MCI, mild cognitive impairment.

#### Functional modules in MCI and AD samples

All eight representative genes in CN-MCI were associated with ribosome functions, while among the 22 representative genes in MCI-AD, 11 were linked to the immune system, 7 were related to the cell cycle, and 4 were involved in the protein processing (Fig. 3 and **Supplementary Table 2**). There were no overlapping genes between CN-MCI and MCI-AD, suggesting distinct molecular mechanisms in the progression of MCI and AD.



**Fig. 3.** Functional classification of representative genes identified.

The descriptions obtained from the KEGG pathway were further classified into four categories (Ribosome, Immune, Cell cycle, and Protein processing). Red and blue bars represent beta coefficients obtained from a logistic regression method.

To further investigate these molecular mechanisms, we performed a PPI network analysis by using NetworkAnalyst 3.0<sup>18</sup> with the STRING Interactome database<sup>19</sup>. The resulting PPI network comprised 929 nodes and 2056 edges, identifying 18 hub genes (node degrees  $\geq 25$ ) (Fig. 4a). These hub genes were categorized into four groups: CN-MCI (ribosome) and MCI-AD (immune, cell cycle, and protein processing). Alterations in CN-MCI were observed in eight ribosomal protein genes, while the remaining 10 genes were involved in NF- $\kappa$ B mediated immune response, cell cycle, and protein processing in MCI-AD. Interestingly, seven of the eight hub genes were down-regulated in CN-MCI, whereas nine of the ten were up-regulated in MCI-AD (Fig. 4a). Additionally, several ribosomal protein genes (*RPL7*, *RPL11*, *RPS14*, and *RPL21*) and immune-related genes (*ACTR2* and *CDC42*) showed opposing responses between CN-MCI and MCI-AD (Fig. 4a). These findings suggest that ribosome and immune functions play a crucial role in the pathogenesis of AD.

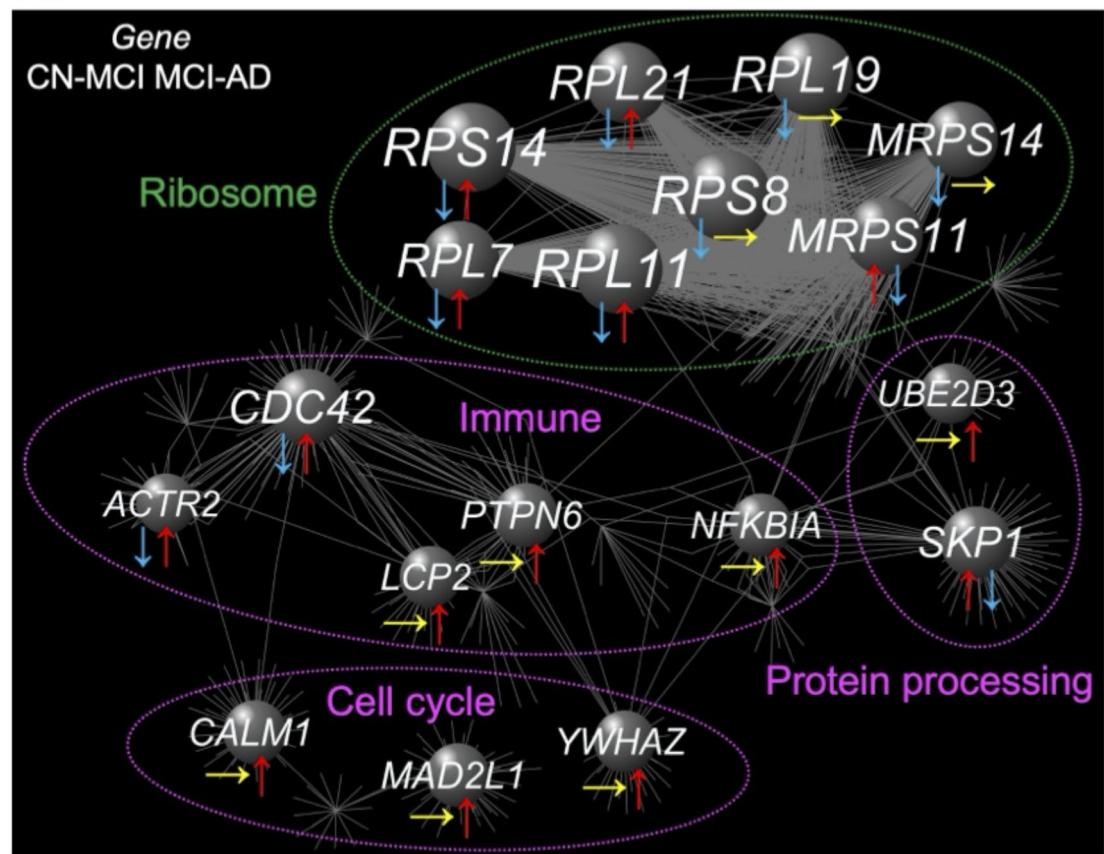
To validate the RNA-seq results, we performed qPCR analysis for *RPL11* and *RPS14* from the ribosomal group and for *CDC42* from the immune group. The qPCR results (Fig. 4b) were overall consistent with the trend observed in the RNA-seq results (Supplementary Table 2). Although there was no significant difference in either *RPS14* or *CDC42* between CN and MCI, this may have been due to the small number of samples ( $n=10$  each).

## Discussion

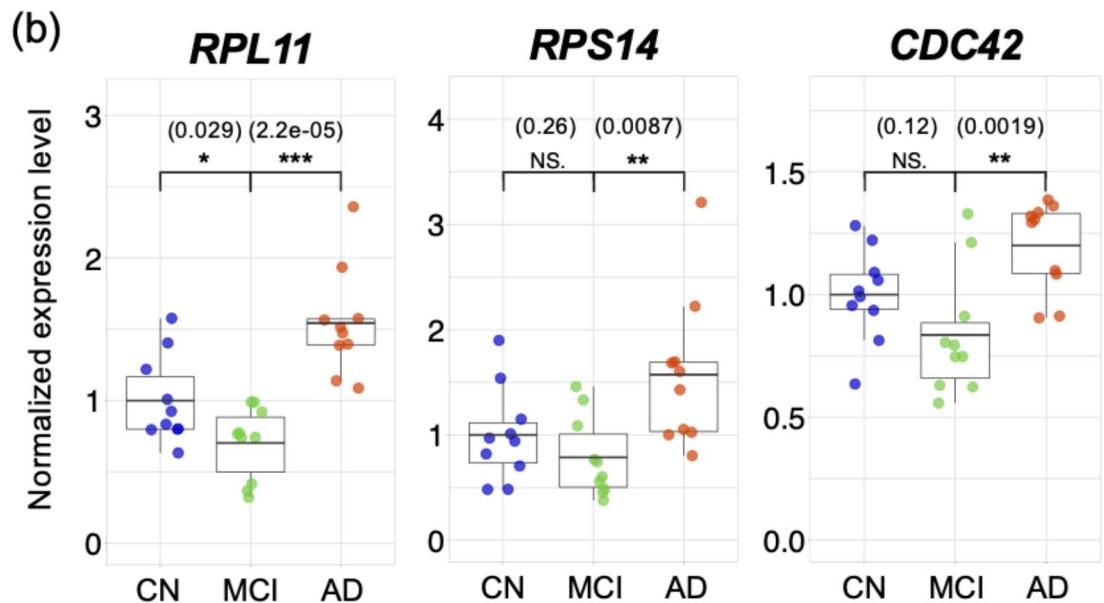
Through transcriptome analysis and subsequent pathway and PPI network analyses, we revealed distinct molecular mechanisms differentiating MCI from CN and AD from MCI. The involvement of genes associated with the ribosome function was observed in the CN-MCI comparison, whereas the MCI-AD comparison highlighted genes associated with the immune system, cell cycle, and protein processing, although validation using brain tissue samples will be necessary in the future.

Most of ribosomal protein genes detected in this study (i.e., *RPL7*, *RPL11*, *RPL19*, *RPL21*, *RPS8*, and *RPS14*) were also identified in our previous study, where we observed their downregulation in AD through transcriptomic comparison of CN with AD<sup>13</sup>. This finding is consistent with previous reports indicating changes in ribosomal protein abundance not only in AD<sup>20,21</sup>, but also in Parkinson's disease<sup>22</sup>, spinal muscular atrophy<sup>23</sup>, and recently in tauopathy<sup>24,25</sup>. These observations suggest an early and widespread involvement of ribosomal proteins in neurodegenerative diseases. However, our stage-specific transcriptome analyses revealed that the downregulations of these ribosomal protein genes have occurred in only CN-MCI, whereas a slight upregulation was observed in MCI-AD. Also, ribosome-associated quality control deposition at the core of amyloid plaques in AD brains may result from translational stalling due to ribosomal defects in AD pathogenesis<sup>26</sup>. Moreover, recent proteome analysis of AD patients identified the upregulation of ribosomal proteins, including *PRL7*, *RPL11*,

(a)



(b)

**Fig. 4.** Functional modules obtained from PPI network analysis.

(a) PPI network diagram including representative genes. Gene names are shown for nodes with  $\geq 25$  edges. The size of the circle representing each gene reflects the number of edges. Green and magenta dashed circles represent the modules in CN-MCI and MCI-AD, respectively. Pairs of arrows indicate genes that are upregulated ( $\uparrow$  in red), unchanged ( $\rightarrow$  in yellow), or downregulated ( $\downarrow$  in blue) in CN-MCI and MCI-AD, respectively.

(b) Validation by quantitative PCR analysis. Each dot represents one CN (blue), MCI (green), or AD (red) sample. The results were consistent with the RNA-seq results for all genes. Data are represented as box and whisker plots, depicting minimum, lower quartile (Q1), mean (Q2), upper quartile (Q3), and maximum values. The data are normalized such that the mean in CNs is 1 for each gene.  $P$  values are shown in parentheses, and statistical significance was set at  $P < 0.05$ . \*:  $< 0.05$ ; \*\*:  $< 0.005$ ; \*\*\*:  $< 0.0005$ , NS: not significant.

PRL14, and RPS8—consistent with our results, but specifically in brain capillaries, not in brain parenchyma<sup>27</sup>. These results suggest that the expression of ribosomal protein genes, especially *RPL7*, *RPL11*, *RPL14*, and *RPS8*, might be a potential target for AD prevention. Furthermore, considering the reported association between mitochondrial dysfunction and the pathogenesis of AD<sup>28,29</sup>, mitochondrial ribosomal protein genes differentially expressed between CN and MCI, such as *MRPS11* and *MRPS14*, could also serve as potential targets for AD prevention.

Many genes involved in the immune system, cell cycle, and protein processing were upregulated in MCI-AD. Abnormal activation of NF-κB in the immune system is involved in synaptic dysfunction<sup>30</sup> and is currently attracting attention as a therapeutic target for AD<sup>31,32</sup>. Cdc42, a member of the small G protein family, plays a crucial role in regulating synaptic plasticity. Activation of the Cdc42 signaling pathway is a key factor in mediating the progression of AD-like phenotypes<sup>33</sup>. Although ACTR2 protein expression has decreased in the postmortem parietal cortex of AD patients<sup>34</sup>, our stage-specific investigations indicated that expression of the ACTR2 decreased at the MCI stage. Cell-cycle reactivation is involved in apoptosis in neurons of AD patients<sup>35,36</sup>. Protein processing, including by ubiquitin ligases such as *SKP1* and *UBE2D3*, has a key role in the pathogenesis of AD<sup>37,38</sup>. Overall, our findings are consistent with previous reports despite blood-based transcriptome analysis.

The primary limitation of this study lies in the small sample size and the sample variation between the groups (AD,  $n=424$ ; MCI,  $n=543$ ; CN,  $n=260$ ). Collecting samples from healthy individuals  $\geq 60$  years old has been challenging owing to the need to verify their cognitive normality. Increasing the number of samples could improve resolution and provide other important factors in the CN-MCI-AD process. Furthermore, our findings were identified from blood samples and should be validated using brain tissues to elucidate AD pathology. However, they do align with a study<sup>27</sup> demonstrating similar expression trends among certain ribosomal proteins in brain or brain capillaries, adding strength to their usefulness.

Finally, our transcriptome analyses among three phenotypes (CN, MCI, and AD) provided detailed insights into the complicated changes in gene expression from CN to AD through MCI. Unlike the ribosomal protein genes observed in CN-MCI, none of the genes involved in immune, cell cycle, and protein processing in MCI-AD were detected in our previous CN-AD comparison<sup>13</sup>. This suggests that changes in MCI-AD may have been masked by the large variations seen in CN-MCI, highlighting the importance of conducting stepwise investigations in both CN-MCI and MCI-AD. Moreover, the genes identified in this study could serve as valuable biomarkers, as they are associated with the progression of MCI and/or AD. Notably, several ribosomal protein genes and immune-related genes showed significant responses in both CN-MCI and MCI-AD comparisons. These findings suggest that the ribosomal protein genes (*RPL7*, *RPL11*, *RPS14*, and *RPL21*) and immune-related genes (*ACTR2* and *CDC42*) may present promising targets for early intervention not only for AD but also for MCI.

## Methods

### Clinical samples

All blood samples and their associated clinical data were obtained from the National Center for Geriatrics and Gerontology (NCGG) Biobank. Of the 1227 samples, 424 were from patients with AD, 543 from individuals with MCI, and 260 from CN donors. The AD and MCI subjects were diagnosed with probable or possible AD according to the criteria of the National Institute on Aging and the Alzheimer's Association workgroups<sup>39,40</sup>. For this study, only patients with probable AD were used as AD subjects. The CN samples were obtained from individuals who had subjective cognitive complaints but normal cognition on a neuropsychological assessment, including a score  $> 23$  on the Mini-Mental State Examination, a comprehensive neuropsychological test. All samples were obtained from adults aged 60 years or older.

This study protocol was approved by the ethics committee of the NCGG of Japan. All participants were volunteers and completed informed consent in writing before registering with the NCGG Biobank.

### RNA sequencing

Isolation of buffy coat from whole blood and of total RNAs from the buffy coat was conducted in accordance with the standard operating procedure of the NCGG Biobank<sup>13</sup>. Only high-quality samples with an RNA integrity number (RIN)  $\geq 6.0$  were used to construct the sequencing library. Sequencing libraries were prepared by using 500 ng of total RNA for each sample; Illumina TruSeq Stranded Total RNA with Ribo-Zero Globin and IDT for Illumina TruSeq UD Indexes (Illumina, San Diego, CA) were used in accordance with the manufacturer's instructions. The libraries were subsequently sequenced on the Illumina NovaSeq6000 platform with paired-end reads of 151 bp in accordance with the manufacturer's instructions.

### RNA sequencing data analysis

All RNA-seq data were downloaded from the NCGG Biobank database<sup>13</sup>. The quality of read sequences (fastq files) was assessed by using FastQC (version 0.11.7). Low-quality reads ( $< Q20$ ) and trimmed reads with adaptor sequences (shorter than 50 bp) were discarded by using Cutadapt (version 1.16). The remaining clean, sequenced reads were mapped to the human reference genome (GRCh37) by using STAR<sup>41</sup> (two-pass option, version 2.5.2b). Read counts for each gene were calculated with the featureCounts program<sup>42</sup> from the subread package (version 1.6.6) to generate expression levels. Outlier read counts, defined as the top and bottom 5% of read counts for each gene, were replaced with the maximum and minimum, respectively, of the remaining effective counts.

An average of 39.1 million raw read sequences were obtained from AD samples, with 99.4% being high-quality (i.e.,  $> Q20$ ). For MCI this was 99.4% of an average of 39.4 million raw read sequences, and for CN it was 99.3% of an average of 35.6 million raw read sequences. After we had discarded low-quality read sequences and trimmed reads with adaptor sequences, 38.6 million reads of cleaned data remained for the AD samples, with

82.5% uniquely mapping to the human reference genome (GCh37). For MCI, this was 79.1% of 39.1 million reads of cleaned data, and for CN, it was 82.1% of 35.1 million reads of cleaned data (**Supplementary Table 3**).

### Proportion of immune cell types

After the alignment of RNA-seq reads to the human reference genome by using STAR, RSEM<sup>43</sup> (version 1.3.0) was employed to quantify transcripts per million (TPM) that were suitable for use with CIBERSORT<sup>44</sup> (version 1.0.1). We reclassified 22 immune cell types with CIBERSORT into 12 major cell types<sup>13</sup>: (1) B cells (naive and memory), (2) plasma cells, (3) CD8<sup>+</sup> T cells, (4) CD4<sup>+</sup> T cells (CD4<sup>+</sup> T cells naive, memory resting, and memory activated; T cells follicular helper; and T cells regulatory), (5)  $\gamma\delta$  T cells, (6) NK cells (resting and activated), (7) monocytes, (8) macrophages (M0, M1, and M2), (9) dendritic cells (resting and activated), (10) mast cells (resting and activated), (11) eosinophils, and (12) neutrophils. Differences in the proportions of each cell type between CN and MCI samples and between MCI and AD samples were evaluated with Welch's *t* test. False discovery rate (FDR) values were calculated with the Benjamini–Hochberg method. An FDR value less than 0.05 was considered statistically significant.

### Differential gene expression analysis

The read counts from each sample were aggregated into a count file for subsequent differential expression analysis by using edgeR<sup>45</sup> (version 3.18.1). Genes with a threshold CPM (counts per million reads mapped)  $> 1$  in more than one-fourth of all sequenced samples were used for subsequent analysis. The calcNormFactors function in edgeR<sup>45</sup> was employed to obtain a trimmed mean of *M* value normalization factors (TMM) to account for library sizes. Dispersion was calculated with the estimateCommonDisp and estimateTagwiseDisp functions in edgeR<sup>45</sup>. The exactTest function in edgeR<sup>45</sup> was applied to obtain DEGs between CN and MCI samples (CN-MCI, MCI compared with CN) and between MCI and AD samples (MCI-AD, AD compared with MCI). DEGs with  $|\log_2(\text{fold change: FC})| > \log_2(1.2)$  and an FDR  $< 0.05$  were defined as statistically significant. This definition was based on the small sample size, following the approach used in the previous reports<sup>13,14</sup>.

### Gene set enrichment analysis of DEGs

A GSEA was implemented by using the Database for Annotation, Visualization and Integrated Discovery (DAVID)<sup>46,47</sup> (version 2021, <http://david.ncifcrf.gov>). Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>48,49</sup> biological pathways with an FDR  $< 0.05$  were identified as statistically significant.

### Detection of immune receptor repertoires

Immune receptor repertoires in T cells and B cells were detected from RNA-seq data by using TRUST4 software<sup>50</sup> (v1.0.5). The estimated CDR3 clonotypes included  $\alpha\beta/\gamma\delta$  T cell receptors (TCRs: TRA, TRB, TRG, and TRD) and B cell receptors (BCRs: IGH, IGK, and IGL). Clonal diversity of TCRs and BCRs was estimated by using an inverse Simpson index calculated with VDJtools<sup>51</sup> (v1.2.1). A linear regression model was employed to identify significant increases or decreases in the proportion of TCR and BCR repertoires in CN-MCI and MCI-AD. An FDR value  $< 0.05$  was considered statistically significant.

### Extraction of representative genes

All datasets were classified for each category (MCI from CN-MCI and AD from MCI-AD). The CN-MCI and MCI-AD datasets were split into nine-tenths for a training set and one-tenth for a test set, respectively (Fig. 2). For GSEA-detected candidate genes, we performed logistic regression analysis, adjusting for clinical information (age, sex, and number of *APOE* ε4 alleles) and immune cell types obtained from the CIBERSORT results in the training set. *P* values and beta coefficients of the candidate genes were obtained. Subsequently, a hierarchical clustering analysis was conducted on the candidate genes within the training set. The number of clusters (*n*) was selected stepwise (*n* = 1, 2, ..., maximum number of candidate genes), and a representative gene with the minimum *P* value was selected from each cluster.

Logistic regression analysis, incorporating clinical information, cell types, and the *n* representative genes, was applied in the training set and evaluated on one-tenth of the test set. This process was repeated 10 times. The average area under the curve (AUC) values from these iterations were used to determine the optimal *n* representative genes. Representative genes were finally defined using logistic regression analysis on the entire dataset, including both the training and test sets. The logistic regression method was implemented using the R statistical software packages stats (version 3.4.3) and ROCR (version 1.0.7).

### Network-based meta-analysis

A network-based analysis was performed by using NetworkAnalyst 3.0 (<https://www.networkanalyst.ca>)<sup>18</sup> with the STRING Interactome database<sup>19</sup>. The PPI network was constructed by using first-order interaction network analysis with a default confidence cutoff score of 900 and was visually represented by using Cytoscape v3.10.1 (<https://cytoscape.org>)<sup>52</sup>.

### Quantitative PCR validation

cDNA was synthesized by using a PrimeScriptII 1st Strand cDNA Synthesis Kit (Takara Bio, Shiga, Japan). Quantitative PCR (qPCR) was performed by using TaqMan Fast Advanced Master Mix (Thermo Fisher Scientific, Waltham, MA), TaqMan Probes (Thermo Fisher Scientific), and the Quantstudio7 Flex Real-Time PCR System (Thermo Fisher Scientific) in accordance with the manufacturer's instructions. The TaqMan gene expression assays used were commercially available: *RPL11* (Hs00831112\_s1), *RPS14* (Hs00735285\_m1), *CDC42* (Hs00918044\_g1), and *ACTB* (Hs99999903\_m1). The qPCR conditions comprised one cycle of 50 °C for 2 min and 95 °C for 20 s, followed by 42 cycles of 95 °C for 1 s and 60 °C for 20 s. Each gene was assayed at least

in duplicate. *ACTB* was preselected as a reference gene for normalization of target gene expression levels, and relative expression levels were calculated by using the  $\Delta\Delta Ct$  method. Gene expression was obtained for 10 AD patients, 10 individuals with MCI, and 10 CN individuals, all randomly selected.

## Data availability

The data supporting the findings of this study are available under controlled access from the NBDC (National Bioscience Database Center) website (<https://biosciencedbc.jp/en/>). The accession number is hum0481.

Received: 9 November 2024; Accepted: 29 January 2025

Published online: 30 January 2025

## References

1. Collaborators, G. F. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the global burden of Disease Study 2019. *Lancet Public. Health.* **7**, e105–e125 (2022).
2. Gauthier, S., Servaes, W. C., Morais, S. & Rosa-Neto, J. A. P. Alzheimer Report 2022: Life after Diagnosis: Navigating Treatment, Care and Support. *Alzheimer's Disease International: London, UK.* (2022).
3. Gatz, M. et al. Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry.* **63**, 168–174 (2006).
4. Harold, D. et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.* **41**, 1088–1093 (2009).
5. Lambert, J. C. et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat. Genet.* **41**, 1094–1099 (2009).
6. Bellenguez, C. et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
7. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
8. Best, M. G. et al. RNA-Seq of Tumor-educated platelets enables blood-based Pan-cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. *Cancer Cell.* **28**, 666–676 (2015).
9. Laing, E. E. et al. Blood transcriptome based biomarkers for human circadian phase. *Elife* **6**, e20214 (2017).
10. Bhak, Y. et al. Depression and suicide risk prediction models using blood-derived multi-omics data. *Transl Psychiatry.* **9**, 262 (2019).
11. Yan, Z. et al. Presymptomatic Increase of an Extracellular RNA in Blood Plasma Associates with the Development of Alzheimer's Disease. *Curr Biol* **30**, 1771–1782 e3 (2020).
12. Shobeiri, P. et al. Circulating long non-coding RNAs as novel diagnostic biomarkers for Alzheimer's disease (AD): a systematic review and meta-analysis. *PLoS One.* **18**, e0281784 (2023).
13. Shigemizu, D. et al. Identification of potential blood biomarkers for early diagnosis of Alzheimer's disease through RNA sequencing analysis. *Alzheimers Res. Ther.* **12**, 87 (2020).
14. Shigemizu, D., Akiyama, S., Mitsuhashi, R., Niida, S. & Ozaki, K. Identification of potential blood biomarkers for early diagnosis of Alzheimer's disease through immune landscape analysis. *NPJ Aging.* **8**, 15 (2022).
15. Shigemizu, D. et al. A comparison of machine learning classifiers for dementia with Lewy bodies using miRNA expression data. *BMC Med. Genomics.* **12**, 150 (2019).
16. Crous-Bou, M., Minguillon, C., Gramunt, N. & Molinuevo, J. L. Alzheimer's disease prevention: from risk factors to early intervention. *Alzheimers Res. Ther.* **9**, 71 (2017).
17. Farrer, L. A. et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* **278**, 1349–1356 (1997).
18. Zhou, G. et al. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* **47**, W234–W241 (2019).
19. Szklarczyk, D. et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
20. Ding, Q., Markesberry, W. R., Chen, Q., Li, F. & Keller, J. N. Ribosome dysfunction is an early event in Alzheimer's disease. *J. Neurosci.* **25**, 9171–9175 (2005).
21. Hernandez-Ortega, K., Garcia-Esparcia, P., Gil, L., Lucas, J. J. & Ferrer, I. Altered Machinery of Protein Synthesis in Alzheimer's: from the Nucleolus to the Ribosome. *Brain Pathol.* **26**, 593–605 (2016).
22. Garcia-Esparcia, P. et al. Altered machinery of protein synthesis is region- and stage-dependent and is associated with alpha-synuclein oligomers in Parkinson's disease. *Acta Neuropathol. Commun.* **3**, 76 (2015).
23. Bernabo, P. et al. In vivo translatome profiling in spinal muscular atrophy reveals a role for SMN protein in Ribosome Biology. *Cell. Rep.* **21**, 953–965 (2017).
24. Evans, H. T., Benetatos, J., van Roijen, M., Bodea, L. G. & Gotz, J. Decreased synthesis of ribosomal proteins in tauopathy revealed by non-canonical amino acid labelling. *EMBO J.* **38**, e101174 (2019).
25. Evans, H. T., Taylor, D., Kneynsberg, A., Bodea, L. G. & Gotz, J. Altered ribosomal function and protein synthesis caused by tau. *Acta Neuropathol. Commun.* **9**, 110 (2021).
26. Rimal, S. et al. Inefficient quality control of ribosome stalling during APP synthesis generates CAT-tailed species that precipitate hallmarks of Alzheimer's disease. *Acta Neuropathol. Commun.* **9**, 169 (2021).
27. Suzuki, M. et al. Upregulation of ribosome complexes at the blood-brain barrier in Alzheimer's disease patients. *J. Cereb. Blood Flow. Metab.* **42**, 2134–2150 (2022).
28. Wang, W., Zhao, F., Ma, X., Perry, G. & Zhu, X. Mitochondria dysfunction in the pathogenesis of Alzheimer's disease: recent advances. *Mol. Neurodegener.* **15**, 30 (2020).
29. Apajai, N. et al. Cognitive impairment is associated with mitochondrial dysfunction in peripheral blood mononuclear cells of elderly population. *Sci. Rep.* **10**, 21400 (2020).
30. Lian, H. et al. NFκB-activated astroglial release of complement C3 compromises neuronal morphology and function associated with Alzheimer's disease. *Neuron* **85**, 101–115 (2015).
31. Jha, N. K. et al. Nuclear factor-κappa beta as a therapeutic target for Alzheimer's disease. *J. Neurochem.* **150**, 113–137 (2019).
32. Sun, E., Motolani, A., Campos, L. & Lu, T. The Pivotal Role of NF-κB in the Pathogenesis and Therapeutics of Alzheimer's Disease. *Int. J. Mol. Sci.* **23**, 8972 (2022).
33. Zhu, M. et al. Cdc42GAP deficiency contributes to the Alzheimer's disease phenotype. *Brain* **146**, 4350–4365 (2023).
34. Lauterborn, J. C. et al. Synaptic actin stabilization protein loss in Down syndrome and Alzheimer disease. *Brain Pathol.* **30**, 319–331 (2020).
35. Herrup, K. The involvement of cell cycle events in the pathogenesis of Alzheimer's disease. *Alzheimers Res. Ther.* **2**, 13 (2010).
36. Moh, C. et al. Cell cycle deregulation in the neurons of Alzheimer's disease. *Results Probl. Cell. Differ.* **53**, 565–576 (2011).
37. Oddo, S. The ubiquitin-proteasome system in Alzheimer's disease. *J. Cell. Mol. Med.* **12**, 363–373 (2008).

38. Hegde, A. N., Smith, S. G., Duke, L. M., Pourquoi, A. & Vaz, S. Perturbations of ubiquitin-proteasome-mediated Proteolysis in Aging and Alzheimer's Disease. *Front. Aging Neurosci.* **11**, 324 (2019).
39. Albert, M. S. et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* **7**, 270–279 (2011).
40. McKhann, G. M. et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* **7**, 263–269 (2011).
41. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
42. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
43. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011).
44. Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. Profiling Tumor infiltrating Immune cells with CIBERSORT. *Methods Mol. Biol.* **1711**, 243–259 (2018).
45. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
46. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
47. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
48. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
49. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2024).
50. Song, L. et al. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat. Methods.* **18**, 627–630 (2021).
51. Shugay, M. et al. VDJtools: unifying post-analysis of T cell receptor repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015).
52. Su, G., Morris, J. H., Demchak, B. & Bader, G. D. Biological network exploration with Cytoscape 3. *Curr. Protoc. Bioinf.* **47** (8 13), 1–24 (2014).

## Acknowledgements

We thank the NCGG Biobank and the NCGG Integrated Database for Dementia Research (iDDR) for providing the study materials, clinical information, and technical support. We also thank ELSS editors for English corrections. This study was supported by grants from Research Funding for Longevity Sciences from the NCGG (21-23 to K.O., and 21-24 and 23-7 to D.S.); JSPS KAKENHI (Grant Number JP21H02470 to D.S.); The Hori Science and Arts Foundation (to A.Y.); and a grant from the Japanese Ministry of Health, Labour, and Welfare for Research on Dementia (to K.O.).

## Author contributions

A.Y. and D.S. designed the study and the main conceptual ideas and performed the analyses; M.S. and R.M. performed the experiments on mRNA expression. S.N. and K.O. contributed to data acquisition and analyses. A.Y. and D.S. wrote the manuscript. D.S. organized this work. All authors contributed to and approved the final manuscript.

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-88526-y>.

**Correspondence** and requests for materials should be addressed to D.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025, corrected publication 2025