



Original Article



Identification and Correlation of Novel Genes Associated with Progression of Alzheimer's Disease

Tania Arora^{1#}, Puneet Jain^{2#}, Harshita Sharma¹, Vikash Prashar¹, Randeep Singh¹, Arti Sharma³, Harish Changotra⁴ and Jyoti Parkash^{1*}

¹Department of Zoology, School of Basic Sciences, Central University of Punjab, Ghudda, Bathinda, Punjab, India; ²Department of Biochemistry, School of Basic Sciences, Central University of Punjab, Ghudda, Bathinda, Punjab, India; ³Department of Computational Biology, School of Basic Sciences, Central University of Punjab, Ghudda, Bathinda, Punjab, India; ⁴Department of Molecular Biology and Biochemistry, Guru Nanak Dev University, Amritsar, Punjab, India

Received: October 21, 2023 | Revised: February 29, 2024 | Accepted: March 20, 2024 | Published online: June 24, 2024

Abstract

Background and objectives: Alzheimer's disease (AD), an enduring neurodegenerative malady, contributes significantly to dementia cases, with late-onset AD being more common than early-onset AD. Despite extensive research to diagnose and treat AD, the intricate protein network impedes the development of efficacious drugs or targets. This study endeavored to identify previously undiscovered genetic reservoirs associated with AD progression, which could be targeted as therapeutic markers.

Methods: Employing the robust tools of R-language, we dissected vast RNA sequence datasets comprising numerous samples and thousands of genes, pinpointing potential candidates implicated in AD's trajectory. Thus, we selected the GSE203206 dataset, which includes AD patients and non-dementia controls, based on our criteria. After normalization, RNA-Seq data was compared, and log₂ fold change was calculated to determine the highly dysregulated genes. Further network analysis of genes and their associated miRNA was performed to determine a characteristic change in control and patient groups.

Results: Differential expression analysis revealed 13 dysregulated genes in AD, wherein 12 were upregulated, and one was down-regulated. Furthermore, we identified hsa-miR-30-5p as a significant miRNA associated with AD, aligning with previous studies and highlighting its high involvement.

Conclusions: This investigation has unveiled four novel genes and a paramount miRNA implicated in AD, thus furnishing potential targets for therapeutic interventions. These discoveries pave the way for further exploration into the intricate functions and implications of these genetic entities in AD.

Introduction

Alzheimer's disease (AD), a type of dementia, is a progressive neurodegenerative disease characterized by the presence of β -amyloid plaques and neurofibrillary tangles.¹ Among all neurodegenerative

disorders, AD claims unrivaled pre-eminence as the foremost prevailing affliction, which leads to severe deterioration in cognitive functions that impede everyday life activities. Apart from cognitive decline, patients with AD develop various neuropsychiatric symptoms such as depression, agitation, apathy, anxiety, and psychosis.² As the disease progresses, individuals may experience disorientation, visuospatial abnormalities, navigation difficulties, muscular dystrophy, language disturbance, and restrictions in various daily activities. Aging, head injuries, vascular illness, infections, and environmental variables such as heavy metals, trace metals, and others are identified among the risk factors for AD.³

According to the Institute of Health Metrics and Evaluation, cases of AD have shown the quickest rates of growth and continue to be one of the significant causes of mortality. AD alone accounts for 60% to 80% of all dementia cases and affects roughly 10% of individuals aged 65 and older.^{4,5} It has been estimated that more than 55

Keywords: Alzheimer's disease; Cognitive; Neurodegeneration; RNA-Seq; GSE203206; hsa-miR-30-5p.

***Correspondence to:** Jyoti Parkash, Department of Zoology, School of Basic Sciences, Central University of Punjab, Ghudda 151 401, Bathinda, Punjab, India. ORCID: <https://orcid.org/0000-0002-6399-700X>. Tel: +91-79-86980164, E-mail: jyoti.parkash@cup.edu.in

[#]These authors contributed equally.

How to cite this article: Arora T, Jain P, Sharma H, Prashar V, Singh R, Sharma A, et al. Identification and Correlation of Novel Genes Associated with Progression of Alzheimer's Disease. *Gene Expr* 2024;000(000):000–000. doi: 10.14218/GE.2023.00143.

million AD cases are currently present and are expected to quadruple by 2050, affecting more than 152.8 million people with AD and related dementia. Correlating to this data, the health-related burden of AD and related dementia will continue to increase, thus affecting the global economy.⁶ Since AD remains incurable, the World Health Organization has labeled it a “global public health priority”.⁷

According to anatomic pathology, AD can be recognized by two prototypical lesions: (1) Senile plaques, which are extracellular lesions of β -amyloid protein ($A\beta$ -42), and (2) Neurofibrillary tangles that comprise phosphorylated Tau protein in neuronal cytoplasm.⁸ Accumulation of $A\beta$ -42 plaques in the hippocampus, amygdala, and cerebral cortex activates microglia and astrocytes, mitochondria disturbance, synaptic loss, oxidative stress, and neurovascular dysfunction, thus ultimately triggering neuronal death.⁹ $A\beta$ -42 accumulation initiates a cascade during the pathogenesis of AD which further triggers the downstream processes leading to hyper-phosphorylation of tau-forming neurofibrillary tangles, ultimately causing neuronal toxicity and degeneration of cholinergic neurons, a main characteristic found in the brains of AD patients, which is believed to cause alterations in cognitive functions, sensory information, and memory loss.³ However, recent studies also highlight the strong association of $A\beta$ -42/40 ratio with tau and pTau181 in cerebrospinal fluid (CSF) of clinical patients compared to $A\beta$ 1-42 alone,¹⁰ favoring $A\beta$ -42/40 ratio as a better marker for disease pathology. Moreover, a low $A\beta$ -42/40 ratio in plasma samples of patients with mild cognitive impairment has also been shown to be negatively correlated with disease progression, indicating its role in determining early-stage dementia and its underlying AD pathology.¹¹

Genetically, AD is broadly categorized into two types, namely Sporadic AD and Familial AD (FAD). Achieving an early clinical diagnosis and gaining insights into the subsequent occurrence of AD can significantly benefit from comprehending both Familial AD and Sporadic AD pathogenesis. Such understanding could influence the trajectory of the disease and contribute to the development of more effective treatments for this devastating neurological disorder.¹² ENREF_10 Despite years of research and various clinical trials, there is still much to learn about the underlying mechanisms of pathological conditions. The development of new therapies for this appalling disease will be facilitated by knowledge of the genetic mechanisms underpinning AD pathogenesis and potentially implicated pathways.

Bioinformatics has revolutionized the biological sciences, enabling us to handle and analyze massive amounts of data. Numerous research studies emphasize the significance of analyzing extensive amounts of data and identifying new genes or pathways that could contribute to addressing the growing number of AD cases.¹³ To accomplish this objective, researchers are employing various tools to analyze RNA sequence data from many samples and investigate the potential involvement of multiple genes in AD progression.¹⁴ R programming is widely utilized in bioinformatics for statistical analysis of large data sets. It provides a wide range of tools and packages that enable efficient data exploration and analysis. These packages constitute an extensive collection of functions, datasets, and documentation designed to tackle particular problems. An understanding of the complex mechanisms of AD, including the role of genetic factors, protein accumulation, and neurodegenerative processes, is crucial for developing effective therapy.

Using bioinformatics and advanced data analysis tools, such as R programming, allows for exploring extensive datasets and identifying new genes or pathways implicated in AD progression.¹⁵ The datasets are derived from the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>), which is an open interna-

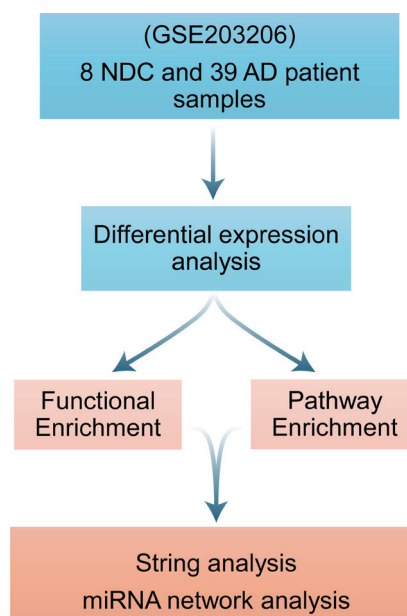


Fig. 1. The dataset GSE203206 consists of a total of 48 samples, including 8 healthy non-dementia control (NDC) and 39 Alzheimer's disease (AD) patients, composed of two groups: Early onset (EOAD) and Late onset (LOAD). The tissue source of all the samples in this dataset was the occipital lobe of the brain. Thus, three groups were created and compared: Control vs EOAD, Control vs LOAD, and EOAD vs LOAD. Data were processed using R programming and analyzed using various packages such as *GDCRNATools*, *edgeR*, and *clusterProfiler* to derive differentially expressed genes between control and AD patients.

tional database repository of high-throughput gene expression, microarray, and other functional genomics datasets administered by the National Center for Biotechnology Information. GEO allows researchers to deposit, search, and download high-throughput gene expression datasets. This study also utilized GEO to access and extract datasets on AD.

Materials and methods

Data retrieval

The GEO repository browser contains 86,871 series or databases generated by “Expression profiling by high throughput sequencing”. After applying inclusion filters, such as “Expression profiling by high throughput sequencing” as the series type, “Homo sapiens” as the sample organism, and “Alzheimer's disease” or “AD” as search keywords, a total of 91 series were obtained. Afterwards, exclusion criteria were applied, including “indirect tissue procurement”, “cell line-based studies”, and “studies based on certain drug treatment”. The dataset GSE203206 was finally chosen for further studies,¹⁶ which included eight healthy controls and 40 AD patients comprising early-onset AD (EOAD) and late-onset AD (LOAD). The analysis was conducted using various tools, and the pipeline is illustrated in Figure 1. The clinical delineation of patients is included in Supplementary Table 1.

Differential expression analysis

To analyze the differentially expressed genes of this dataset, the *GEOquery* package was used to extract the raw counts for all samples.¹⁷ It was further used to find relative expression between

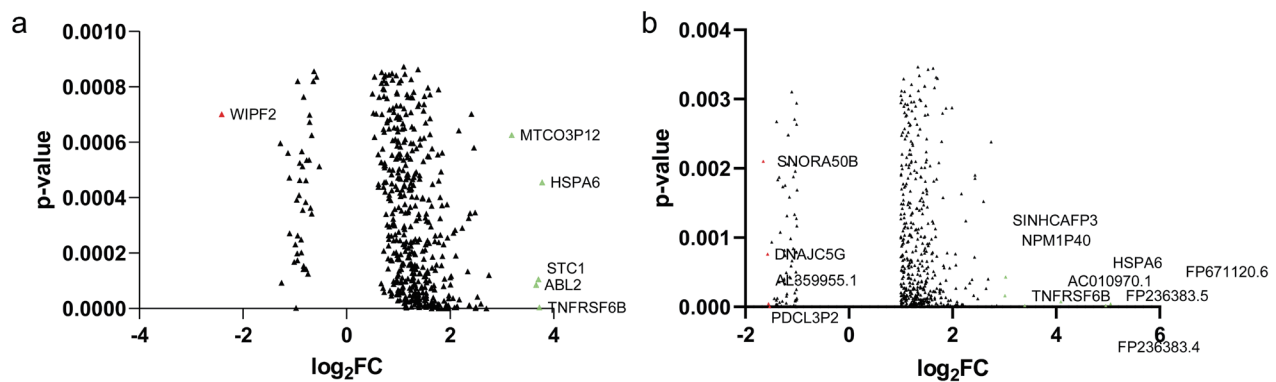


Fig. 2. Differentially expressed genes in Alzheimer's disease. (a) Volcano plot of comparison between subgroups "Control vs EOAD" genes of the total dysregulated 580 genes. (b) Volcano plot of comparison between subgroups "Control vs LOAD" of the total 405 genes. Red color represents the most downregulated genes, whereas green color indicates the upregulated genes. EOAD, early-onset Alzheimer's disease; LOAD, late-onset Alzheimer's disease.

three groups, Control vs EOAD, Control vs LOAD, and EOAD vs LOAD, using the *GDCRNATools* package.¹⁸ Initially, this package employed the *edgeR* package for count value normalization to remove gene length-based bias and library size differences among samples to enable accurate comparisons of expression levels.¹⁹ Differentially expressed genes were obtained by further screening based on a threshold of $|\log_2FC| > 1$ and false discovery rate (FDR)-adjusted p -value < 0.05 .

Gene set enrichment analysis (GSEA)

Post meticulous screening based on their ranking using log-fold changes, we employed the *clusterProfiler* package,²⁰ which offers a range of tools to obtain GSEA wherein we specifically opted for Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) database.²¹ KEGG unites gene information from the genomic database and functional information using a pathway database to systematically analyze gene functions. For gene enrichment, the pre-ranked gene list, sorted based on expression change was coupled with selected gene sets. The KEGG database ran certain phenotype-based permutations to obtain functional enrichment plots and Pathway enrichment plots. The specific gene sets were then further assessed based on their enrichment scores, with statistical significance determined through permutation testing, p -value, and FDR.

Protein-protein interactions (PPI)

To analyze protein-protein interactions, the String database (StringDB) and GeneMANIA were used.^{22,23} In the StringDB portal, the list of dysregulated genes was uploaded separately for upregulated and downregulated sets, and PPI data were obtained, which was further refined by applying a threshold filter to obtain more reliable interactions. These interactions were further explored in GeneMANIA software by uploading the combined list of genes, which highlighted various physical and predicted interactions based on protein structure.

miRNA network formation

To analyze the mRNA-miRNA interactions, we employed miRNET,²⁴ which is powerful bioinformatics software for extracting data from experimentally validated and computationally predicted interactions. Along with 13 dysregulated genes, the dataset also includes genes that play a direct role in the selected disease. Additionally, to obtain a more robust network, a degree filter of 2.0 and a betweenness filter of 1.0 were applied, providing 14 nodes

for upregulated genes and 19 nodes for downregulated genes. The resulting network was further arranged in a "linear bipartite/tripartite" graph layout.

Statistical analysis

Data analysis was handled using R programming. Statistical analysis was done using unpaired t-test and FDR-adjusted p -value < 0.05 was considered as significant.

Results

Differential expression analysis of the genes

A total of 24,448 genes included in the dataset were evaluated to determine differential expression in control individuals ($n = 8$) vs AD patients ($n = 39$), further categorized into EOAD and LOAD patients. Initially, the comparative analysis between control and EOAD samples provided 564 dysregulated genes (Supplementary Table 2) with statistically significant differences. These genes are shown in Figure 2a using a volcano plot, representing their changes in expressions ($\log_2$ -fold change) along with statistical significance (p -value). Notably, *SNORA50B*, *DNJC5G*, *PDCL3P2*, and *AL359955.1* are significantly downregulated, while *HSPA6*, *STC1*, *TNFRSF6B*, *SINHCAFP3*, *NPM1P40*, and *FP236383.5* are significantly upregulated.

Additionally, the comparison of control and LOAD samples conferred a set of 405 genes (Supplementary Table 3) with distinctive expression patterns. These findings are depicted in Figure 2b, highlighting genes such as *WIPF2*, *PDCL3P*, *ZPED9*, *LINC02068*, *REX01LIP*, *ATP5F1AP10*, and *CD44-AS1* as downregulated, and *MTCO3P12*, *HSPA6*, *STC1*, *ABL2*, and *TNFRSF6B* as upregulated. The comparative analyses provided a set of 13 genes (coding and pseudogenes), to be highly dysregulated (Table 1). Notably, the relative comparison of Control-EOAD and Control-LOAD fold changes did not expose any significant change in expression among the transcripts. This indicates that a unique genetic signature is associated with the EOAD and LOAD subtypes within the AD spectrum.

Functional enrichment analysis

Functional enrichment analysis of dysregulated genes was carried out with KEGG pathways using the *GDCRNATools* R package. This analysis helped in understanding the biological processes

Table 1. List of dysregulated genes associated with EOAD and LOAD

Genes (LOAD)	Fold change	Genes (EOAD)	Fold change
<i>HSPA6</i>	3.7777258	<i>FP671120.6</i>	5.05216
<i>TNFRSF6B</i>	3.7232833	<i>FP236383.5</i>	5.05216
<i>STC1</i>	3.7044567	<i>FP236383.4</i>	5.05216
<i>ABL2</i>	3.6604484	<i>HSPA6</i>	4.956657
<i>MTCO3P12</i>	3.1909349	<i>AC010970.1</i>	4.091295
<i>WIPF2</i>	-2.4091189	<i>TNFRSF6B</i>	3.394207
		<i>SINHCAFP3</i>	3.022924

EOAD, early-onset Alzheimer's disease; LOAD, late-onset Alzheimer's disease.

and pathways significantly enriched within the given gene list, shedding light on the functional implications of the input genes or proteins by identifying the most affected pathways. The phagosome pathway (hsa04145) was enriched by a set of three genes by 17-fold, while the collecting duct secretion pathway (hsa04966) was enriched by 30-fold but only by one gene (Fig. 3), highlighting their substantial role in AD. Since AD is recognized by the accumulation of A β -42 and τ -aggregates wherein disruptions in phagocytic clearance by microglia may lead to inefficient clearance of these aggregates, thus aiding in the progression of AD.²⁵ Consequently, the role of three novel genes (*TUBB6*, *TCIRG1*, and *MRC1*) connects to the pre-existing mechanism of AD, suggesting the development of therapeutics targeting the phagocytic mechanism. Further analysis of genes involved in each pathway was conducted using data mining from the KEGG database (Table 2).

Pathway enrichment analysis

The dysregulated genes were further assessed using gene set enrichment analysis. Pathway enrichment analysis using GSEA determines

whether a gene set is significantly enriched in a list of differentially expressed genes. Rather than focusing on specific gene-level alterations, GSEA evaluates the overall behavior of genes within a pathway. It aids researchers in determining if particular gene sets, such as those associated with particular biological pathways or functional groups, are perpetually upregulated or downregulated in a phenotype or pathology. The analysis showed that dysregulated genes were associated with multiple pathways, highlighting significant influences on organelle organization (Fig. 4). Previous studies also highlighted the relevance of the disrupted function of lysosomes,²⁶ mitochondria,²⁷ and Golgi apparatus with AD.^{28,29}

Correlation study of dysregulated genes

The 20 resultant dysregulated genes were further assessed for their correlation and affected pathways using the String database. This widely used database analyzes the PPI, predicting and creating a comprehensive network of protein interactions. The cluster with all possible PPIs (Fig. 5) and the pathways that are most affected or associated with these genes were identified. Upregulated genes

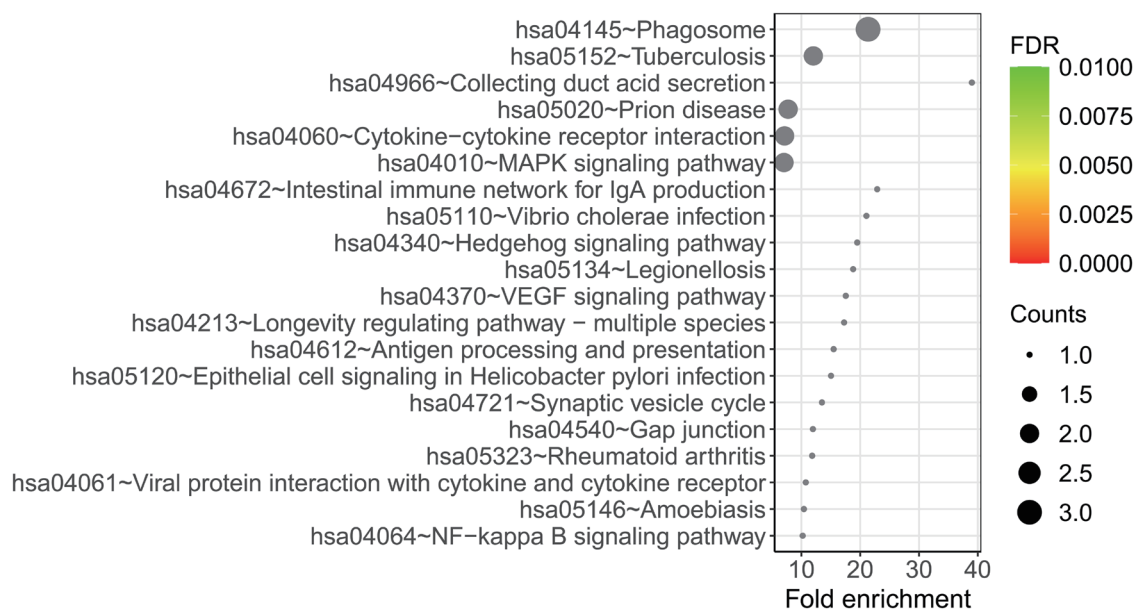


Fig. 3. Enrichment Analysis via KEGG: Functional annotation of dysregulated genes providing insights into the most relevant pathways or biological functions involved in AD. The plot typically displays functional categories on the y-axis, while the x-axis represents the fold change, where point represents the fold change and count represents a number of genes involved in a particular pathway. The plot also includes statistical indicators such as adjusted *p*-values (FDR) to assess the significance of enrichment. AD, Alzheimer's disease; FDR, false discovery rate; KEGG, Kyoto Encyclopaedia of Genes and Genomes.

Table 2. Tabular representation of genes involved in a multiple pathway, obtained through the KEGG database

Pathways	Genes involved
hsa04145-Phagosome	<i>TUBB6, TCIRG1, MRC1</i>
hsa05152-Tuberculosis	<i>MRC1, TCIRG1</i>
hsa04966-collecting duct acid secretion	<i>TCIRG1</i>
hsa05020-prion disease	<i>HSPA6, TUBB6</i>
hsa04060-cytokine –cytokine receptor interaction	<i>TNFRSF6B, LTBR</i>
hsa04010- MAPK signalling pathway	<i>HSPA6, HSPB1</i>
hsa04672-intestinal immune network for IgA production	<i>LTBR</i>
hsa05110-Vibrio cholera infection	<i>TCIRG1</i>
hsa04340-Hedgehog signaling pathway	<i>EVC2</i>
hsa05134-Legionellosis	<i>HSPA6</i>
hsa04370-VEGF signaling pathway	<i>HSPB1</i>
hsa04213-Longevity regulating pathway	<i>HSPA6</i>
hsa04612-Antigen processing and presentation	<i>HSPA6</i>
hsa05120-Epithelial cell signaling in Helicobacter pylori infection	<i>TCIRG1</i>
hsa04721-Synaptic vesicle cycle	<i>TCIRG1</i>
hsa04540-Gap junction	<i>TUBB6</i>
hsa05323- Rheumatoid arthritis	<i>TCIRG1</i>
hsa04061-Viral protein interaction with cytokine and cytokine receptor	<i>LTBR</i>
hsa05146-Amoebiasis	<i>HSPB1</i>
hsa04064-NF-kappa B signaling pathway	<i>LTBR</i>

IgA, immunoglobulin A; KEGG, Kyoto Encyclopaedia of Genes and Genomes; MAPK, mitogen-activated protein kinase; VEGF, vascular endothelial growth factor.

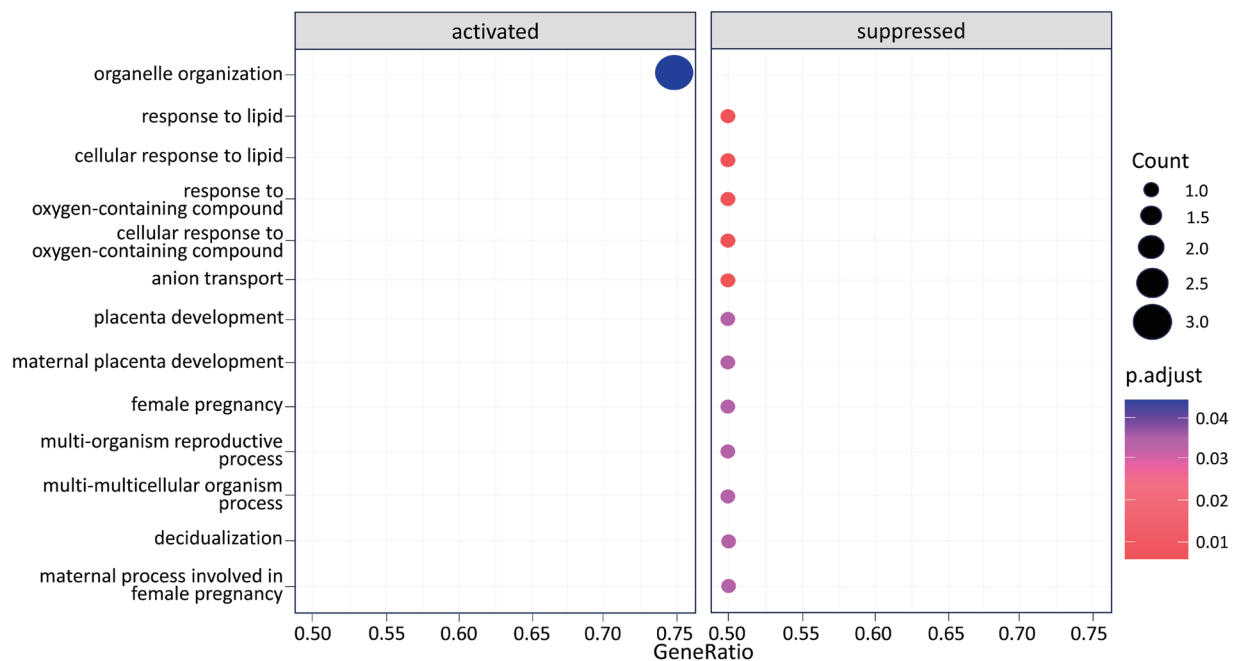


Fig. 4. Pathway Enrichment Plot where **organelle organization** is found to be highly influenced by the dysregulated genes. The plot typically displays functional categories on the y-axis, and the x-axis represents the gene ratio and underlying pathways with a lower *p*-value or a higher enrichment score indicating dark or intense color contrariwise less significant pathways represent lighter color.

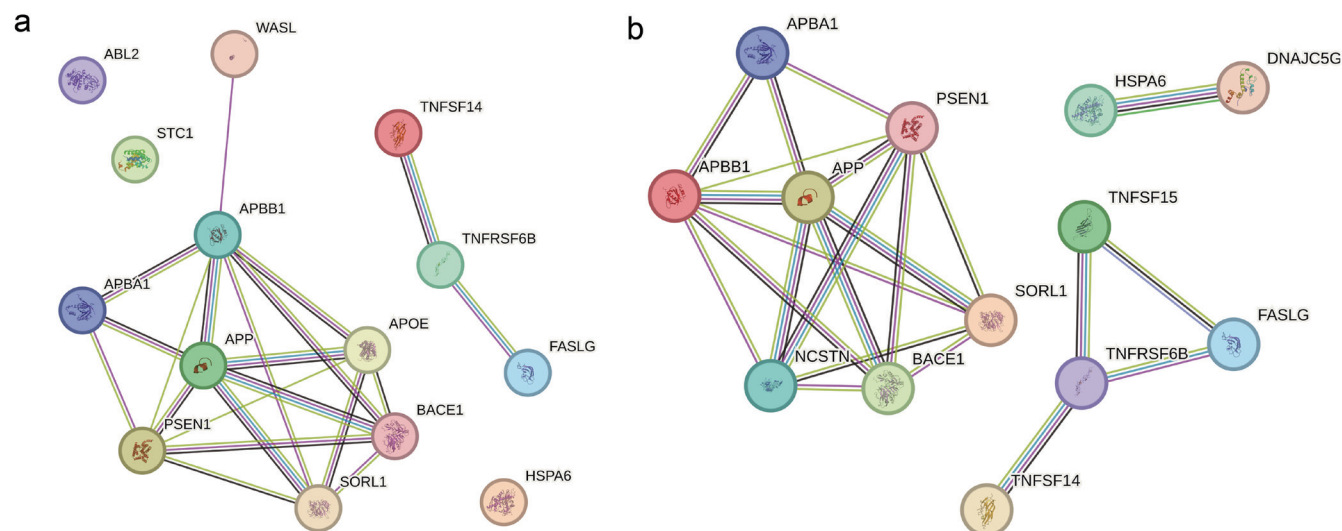


Fig. 5. String analysis of dysregulated genes showing protein-protein interaction plot. (a) Cluster demonstrating the correlation of upregulated genes involved in Alzheimer's disease (AD). The gene *WIPF2* and *TNFRSF6B* are indirectly associated with multiple biological pathways that affect AD. (b) Cluster demonstrating the correlation of downregulated genes involved in AD. The genes *HSPA6* and *TNFRSF6B* are indirectly associated with multiple biological pathways that affect AD.

affect the following pathways: (1) Circadian temperature homeostasis, (2) regulation of germinal center formation, (3) negative regulation of receptor biosynthetic processes, (4) p38MAPK cascade, and (5) type 2 immune response. Downregulated genes affect the following pathways: (1) Negative regulation of receptor biosynthetic processes, (2) regulation of germinal center formation, (3) phagosome acidification, (4) histone H4 deacetylation, and (5) regulation of T-helper 2 cell differentiation negative regulation of gene silencing by miRNA.

Since the PPIs network using STRING was unclear, we performed further interaction studies using GeneMANIA (Fig. 6). It revealed the physical interactions of ABL2 (Abelson-related gene) and HSPA6 (Heat Shock protein family A, member 6), which are further linked to Apolipoprotein E and A1. ApoE (apolipoprotein E) is mainly associated with lipid transport and plays a central role in the metabolism of cholesterol and triglycerides.³⁰ It is also involved in the clearance of A β -42 peptides, the aggregation of which is directly involved in AD progression.³¹

Correlation study of miRNA and genes

The association of dysregulated genes with miRNA was generated using miRNET – centric network analysis (Fig. 7). MiRNET software is widely used for studying and visualizing miRNA regulatory networks. This platform helps us explore miRNA-target interactions, functional enrichment analysis, and network visualization. It provides functional enrichment tools for insights into biological processes and molecular interactions using the KEGG or Gene Ontology database. It incorporates various modules to help understand miRNAs' role in gene regulation and biological processes. The miRNA interaction plot was constructed after applying filters, i.e., “betweenness filter = 1.0”, “degree filter = 1.0”, and using the “linear bipartite/tripartite” graph layout.

Discussion

Millions of people worldwide are afflicted by dreadful neurological ailments, including AD. AD is a prevalent and debilitating

neurodegeneration condition characterized by progressive cognitive decline, memory loss, and impaired daily activities.³ Despite decades of intensive research and substantial investment, the quest for successful treatments has encountered numerous setbacks. To address this serious concern, we conducted a comprehensive study utilizing the GSE203206 dataset obtained from the internationally recognized open repository, GEO. This dataset was derived from high-throughput sequencing focused on AD. Employing advanced bioinformatics tools such as GeneMania, StringDB, and R packages including *GDCRNATools*, *edgeR*, *clusterProfiler*, and *ggplot2*, we performed an in-depth analysis. In this study, we identified differentially expressed genes by comparing “control vs EOAD” and “control vs LOAD” groups, obtaining a total of 580 genes and 405 genes, respectively.

We found 13 uniquely highly dysregulated genes that might play a significant role in AD pathogenesis. These include 12 up-regulated genes and one downregulated gene based on their log₂-fold change (Fig. 2). We conducted a functional enrichment analysis to unravel the key pathways and biological processes associated with the dysregulated genes in AD. The hsa-04145-Phagosome pathway emerged as highly linked to the dysregulated genes involved in AD (Fig. 3). After manual filtering through the KEGG database, we identified that three dysregulated genes, i.e., *HSPA6*, *HSPB1*, and *TCIRG1*, are involved in multiple pathways obtained from functional enrichment analysis. Subsequently, pathway enrichment analysis revealed that organelle organization was the most significantly influenced pathway by the group of dysregulated genes associated with AD (Fig. 4).

Moreover, we conducted a PPI analysis using StringDB to gain further insights into the biological mechanisms underlying AD (Fig. 5) and GeneMANIA (Fig. 6). Additionally, we explored the association of dysregulated genes with miRNAs using the miRNET software (Fig. 7). Notably, B-cell lymphoma 6 (*BCL6*) is highly associated with several miRNAs among the upregulated genes, while JunB proto-oncogene (*JUNB*) exhibited significant associations among the downregulated genes. *HSPB1* and *HSPA6* were identified as the most enriched genes in the study (Fig. 3).

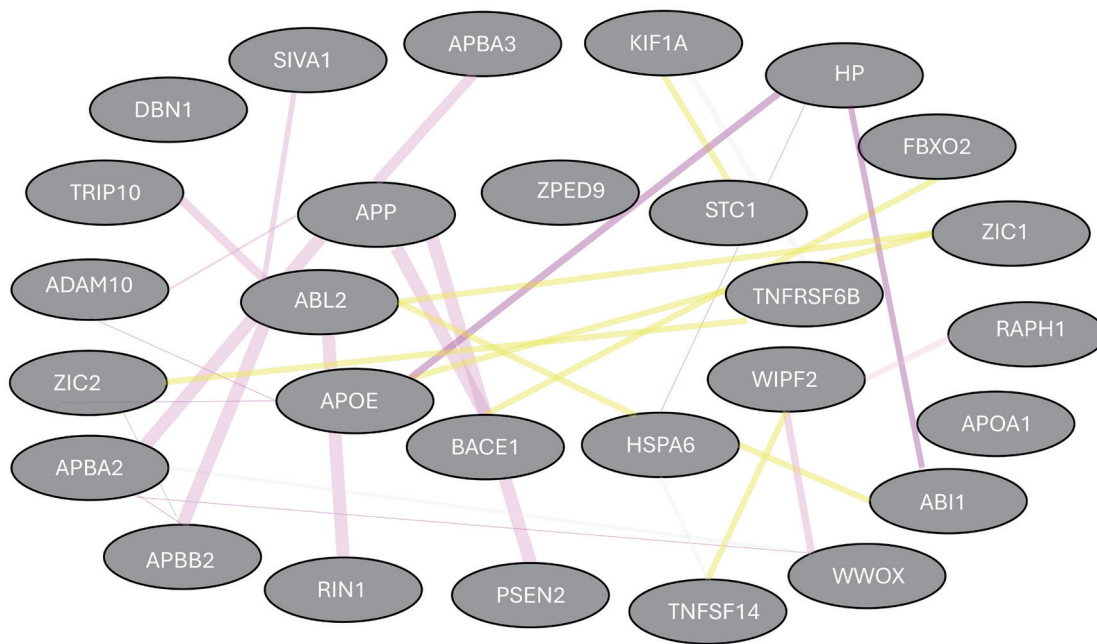


Fig. 6. GeneMANIA analysis shows the physical involvement of ABL2 and HSPA6 proteins, which are further connected to APOE and APOA1, responsible for amyloid-beta (A β) aggregation and deposition. Yellow lines represent predicted interactions, pink lines represent physical interactions, and grey lines represent genetic interactions. ABL2, Abelson-related gene; APOA1, apolipoprotein A-I; APOE, apolipoprotein E; HSPA6, heat shock protein family A, member 6.

Conclusively, according to our study, *HSPA6*, *HSPB1*, *TCIRG1*, and *BCL6* are identified as hub genes affecting multiple targets implicating in AD.

HSPA6 (70kDa family A member 6) is a protein-coding gene primarily located on chromosome 1 (1q23.3) and encodes a protein that comprises two important structural domains, namely the N-terminal nucleotide-binding protein and the C-terminal substrate-binding domain. *HSPA6* plays a crucial role in various biological

processes such as cellular heat acclimation, response to unfolded protein, protein refolding, neutrophil degranulation, cellular response to unfolded protein, and chaperone cofactor-dependent protein refolding. It plays a critical role in the protein quality control system, which includes the correct folding of proteins, refolding of misfolded proteins, and governing the control of the subsequent degradation of proteins.³² Co-chaperones mediate adenosine triphosphate (ATP) binding, hydrolysis, and ATP release cycles to

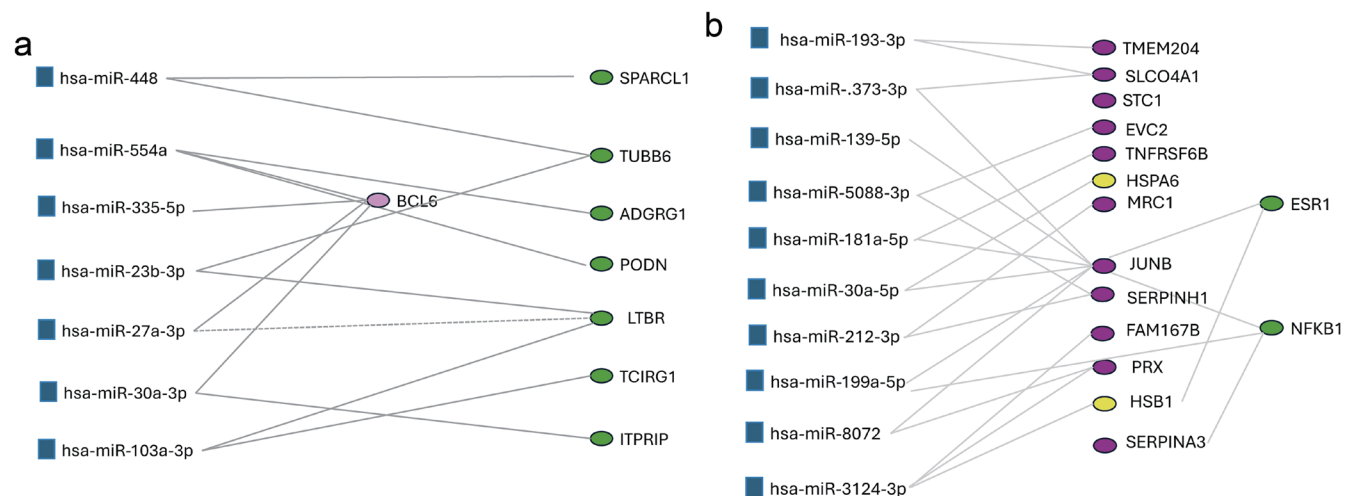


Fig. 7. miRNA-target interaction plot of the dysregulated genes. (a) miRNA target network plot of Upregulated genes shows a total of eight genes and seven associated miRNAs, with *BCL6* highly associated with most of the miRNAs. Blue color represents miRNAs, and the green color represents genes. (b) miRNA target network of downregulated genes shows a total of 13 genes and 10 genes were found. *JUNB* was found highly associated with most of miRNAs. With KEGG analysis via miRNET software, two highly enriched genes were identified. Pink color represents genes, the blue color represents miRNA, the green color represents transcription factor, and the yellow color represents enriched genes. KEGG, Kyoto Encyclopaedia of Genes and Genomes.

achieve this. According to recent studies, *HSPA6* contributes to the development and advancement of tumors and triggers disorders unrelated to cancer.³³ It is involved in the system that regulates the quality of the proteins, suppressing protein misfolding and aggregation, which are characteristics of neurodegeneration. Despite extensive research on *HSPA6*, its precise function and processes remain unknown. However, our study has highlighted its role in the progression of AD, suggesting it could be further studied and used as a prognostic marker.

HSPB1, i.e., heat shock protein family B (small) member 1, is a protein-coding gene located on chromosome 7 (7q11.23) and was first identified in 1987–1988.³⁴ According to earlier studies, *HSPB1* exhibited robust phosphorylation and the intriguing capacity to generate high molecular weight oligomers. After a long span of time, *HSPB1* is identified as a molecular chaperone primarily engaged in protein folding. It plays a crucial role in capturing and storing misfolded polypeptides brought on by stress to prevent their aggregation and subsequently encourage either refolding or proteolytic destruction.³⁴ It is abundantly expressed and can govern several cellular processes, including actin dynamics, oxidative stress regulation, and anti-apoptosis. Additionally, it serves a significant role in the regulation of cytoskeleton organization. In recent years, *HSPB1* has been studied intensively due to its constitutive expression in different tissues, particularly in pathological conditions.³⁴ *HSPB1* is associated with a protein quality control system and counteracts protein misfolding and aggregation, a characteristic of neurodegeneration. Disease-causing mutations in *HSPB1* cause disruption of cytoskeleton organization, apoptosis, and protein folding, and exclusively impact the peripheral nervous system, demonstrating an essential role in highly polarized motor and sensory neurons.³⁵ Cdk5 extensively phosphorylates neurofilaments in response to *HSPB1* mutations and disturbances in axonal transport.

BCL6 (B-cell lymphoma-6) is located on chromosome 3 (3q27.3) in humans and is a chief transcription factor that governs the growth of T follicular helper cells. Three conserved domains are present in *BCL6* (706 amino acids): an N-terminal BTB/POZ domain, a central region of about 400 amino acids, and a C-terminal C2H2-type zinc (Z.F.) DNA binding motifs. These domains interact with co-repressors to foster germinal center and B cell proliferation. *BCL6* contributes to the pathophysiology of several human lymphomas. It binds to the *ITM2B* protein and represses its transcription; when *BCL6* is removed, the transcription level of *ITM2B* demonstrates an inverse relationship between each other. Molecular modifications in the gene *ITM2B* were found to be associated with neurodegenerative disorders.³⁶ Recent studies have suggested a potential involvement of *BCL6* in AD pathogenesis, indicating its association with neuro-inflammation and amyloid-beta accumulation. However, additional investigations are required to fully understand the molecular mechanisms of *BCL6* in AD progression and to explore its potential as a target for intervention strategies.

TCIRG1 is a protein-coding gene located on chromosome 11 (11q13.2) in humans and encodes a component ($\alpha 3$ subunit) of a significant protein complex, vacuolar H⁺ ATPase (V-ATPase). V-ATPase comprises two domains: a cytosolic V1 domain and a transmembrane V0 domain. This protein complex operates as the pump aids in passing the proton across the membrane, hence regulating the pH of the cell and surrounding environment. Protein sorting, zymogen activation, and receptor-mediated endocytosis are examples of V-ATPase-dependent organelle acidification. *TCIRG1* is primarily involved in various biological pathways such

as lysosome acidification, insulin receptor signaling, neutrophil acidification, proton transmembrane transport, osteoclast proliferation, apoptosis process, inflammatory pathways, and macroautophagy. *TCIRG1* mutations are responsible for the rare genetic disorder called autosomal recessive osteopetrosis type-1. This genetic disorder is attributed to impaired bone resorption, resulting in an excessive build-up of thick and brittle bones.³⁷ Although the primary manifestation of *TCIRG1* mutation is related to bone disease, some indications may also play a role in neurodegeneration. However, according to several studies, *TCIRG1* mutation may disrupt lysosomal function in neurons, possibly resulting in impaired protein aggregation and toxic breakdown, a neurodegeneration characteristic.³⁸ Additionally, numerous neurodegenerative diseases, including AD, Parkinson's disease, and amyotrophic lateral sclerosis, have been linked to lysosomal dysfunction. It is conceivable that mutations in the *TCIRG1* gene could interfere with lysosomal function in neurons and contribute to the etiology of neurodegenerative studies since *TCIRG1* is involved in lysosomal acidification. However, additional studies are required to elucidate the specific role of *TCIRG1* and its implicated pathway.

hsa-miR-30a-5p is located on chromosome 6 (6q13). The miR-30a-5p molecule is a part of the miR-30 family, which consists of five closely related and evolutionarily conserved miRNAs. It plays a role in various cellular processes and has also been associated with the progression of different cancer type.³⁹ *In vivo* studies were conducted on mice, and it was found that elevated levels of miR-30a-5p resulted in neuronal damage and programmed cell death in cells associated with AD.^{39,40} Some studies have also indicated that miR-30a-5p may contribute to neurodegeneration by regulating key genes and pathways involved in processes such as neuronal survival, inflammation, oxidative stress, and protein aggregation. Overall, while evidence suggests the involvement of miR-30a-5p in neurodegeneration, further research is required to fully understand its specific role, target genes, and the underlying molecular mechanisms in different neurodegenerative diseases. Continued investigation into the functions of miR-30a-5p may provide valuable insights into the pathogenesis of neurodegenerative disorders and offer new therapeutic avenues for the intervention of AD.

Conclusions

The pursuit of a cure for AD has been marked by notable failures and substantial challenges. The complexity of the disease, limited understanding of underlying mechanisms, lack of early detection methods, and setbacks in clinical trial progress have impeded progress toward finding a definitive cure. The amalgamation of bioinformatics and AD research has revolutionized our understanding of the disease at a molecular level. By harnessing the power of computational analysis and data integration, bioinformatics has contributed significantly to identifying potential biomarkers, understanding disease mechanisms, and guiding the development of novel therapeutics. Our study delves into the intricate complexities of genetic mechanisms and understanding different dysregulated genes, namely, *HSPA6*, *HSPB1*, *TCIRG1*, *BCL6*, and miRNA, i.e., miR-30a-5p, and their associated pathways implicated in AD. Through an intricate dissection of molecular signatures, aberrant genetic profiles, perturbed signaling cascades have been elucidated, shedding light on the intricate molecular interplay underlying AD progression. These findings provide valuable insights into potential therapeutic targets and novel avenues for intervention to counteract this neurodegenerative disorder's devastating consequences. Notwithstanding the elucidating findings of our study, it

is imperative to underscore the necessity for subsequent investigations to validate our research outcomes and shape better treatments.

Acknowledgments

TA, RS and PJ are sincerely thankful for their fellowship from the Council of Scientific & Industrial Research.

Funding

The research was funded by the University Grants Commission (UGC) through start-up grant (F.30-312/2016 (BSR)), and the Science and Engineering Research Board through Early Career Research Grant, (ECR/2015/000240), and Core Research Grant (CRG/2020/003257).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Study conceptualization, data collection and analysis (TA, HS, PJ, JP), manuscript writing (TA, JP, VP, RA), final editing (PJ, AS), and proofreading (JP, AS, HC). All authors have made a significant contribution to this study and have approved the final manuscript.

Data sharing statement

Available on GEO, Accession: GSE203206.

References

- [1] Kumar A, Sidhu J, Goyal A, Tsao JW. Alzheimer disease. Treasure Island (FL): StatPearls Publishing; 2018.
- [2] Rosenberg PB, Nowrangi MA, Lyketsos CG. Neuropsychiatric symptoms in Alzheimer's disease: What might be associated brain circuits? *Mol Aspects Med* 2015;43-44:25–37. doi:10.1016/j.mam.2015.05.005, PMID:26049034.
- [3] Breijyeh Z, Karaman R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. *Molecules* 2020;25(24):5789. doi:10.3390/molecules25245789, PMID:33302541.
- [4] Calabrò M, Rinaldi C, Santoro G, Crisafulli C. The biological pathways of Alzheimer disease: a review. *AIMS Neurosci* 2021;8(1):86–132. doi:10.3934/Neuroscience.2021005, PMID:33490374.
- [5] GBD 2016 Dementia Collaborators. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* 2019;18(1):88–106. doi:10.1016/S1474-4422(18)30403-4, PMID:30497964.
- [6] Akushevich I, Kravchenko J, Yashkin A, Doraiswamy PM, Hill CV, Alzheimer's Disease and Related Dementia Health Disparities Collaborative Group. Expanding the scope of health disparities research in Alzheimer's disease and related dementias. *Alzheimer's Dement* 2023;15(1):e12415. doi:10.1002/dad2.12415, PMID:36935764.
- [7] Wortmann M. Dementia: a global health priority - highlights from an ADI and World Health Organization report. *Alzheimers Res Ther* 2012;4(5):40. doi:10.1186/alzrt143, PMID:22995353.
- [8] de Rijke TJ, Doting MHE, van Hemert S, De Deyn PP, van Munster BC, Harmsen HJM, *et al*. A Systematic Review on the Effects of Different Types of Probiotics in Animal Alzheimer's Disease Studies. *Front Psychiatry* 2022;13:879491. doi:10.3389/fpsy.2022.879491, PMID:35573324.
- [9] Silva MVF, Loures CMG, Alves LCV, de Souza LC, Borges KBG, Carvalho MDG. Alzheimer's disease: risk factors and potentially protective measures. *J Biomed Sci* 2019;26(1):33. doi:10.1186/s12929-019-0524-y, PMID:31072403.
- [10] Soto-Rojas LO, Pacheco-Herrero M, Martínez-Gómez PA, Campa-Córdoba BB, Apátiga-Pérez R, Villegas-Rojas MM, *et al*. The Neurovascular Unit Dysfunction in Alzheimer's Disease. *Int J Mol Sci* 2021;22(4):2022. doi:10.3390/ijms22042022, PMID:33670754.
- [11] Delaby C, Estellés T, Zhu N, Arranz J, Barroeta I, Carmona-Iragui M, *et al*. The Aβ1-42/Aβ1-40 ratio in CSF is more strongly associated to tau markers and clinical progression than Aβ1-42 alone. *Alzheimers Res Ther* 2022;14(1):20. doi:10.1186/s13195-022-00967-z, PMID:35105351.
- [12] Pérez-Grijalva V, Romero J, Pesini P, Sarasa L, Monleón I, San-José I, *et al*. Plasma Aβ42/40 Ratio Detects Early Stages of Alzheimer's Disease and Correlates with CSF and Neuroimaging Biomarkers in the AB25 Study. *J Prev Alzheimers Dis* 2019;6(1):34–41. doi:10.14283/jpad.2018.41, PMID:30569084.
- [13] Dorszewska J, Predecki M, Oczkowska A, Dezor M, Kozubski W. Molecular Basis of Familial and Sporadic Alzheimer's Disease. *Curr Alzheimer Res* 2016;13(9):952–963. doi:10.2174/1567205013666160314150501, PMID:26971934.
- [14] Arora T, Prashar V, Singh R, Barwal TS, Changotra H, Sharma A, *et al*. Dysregulated miRNAs in Progression and Pathogenesis of Alzheimer's Disease. *Mol Neurobiol* 2022;59(10):6107–6124. doi:10.1007/s12035-022-02950-z, PMID:35867206.
- [15] Guennewig B, Lim J, Marshall L, McCorkindale AN, Paasila PJ, Patrick E, *et al*. Defining early changes in Alzheimer's disease from RNA sequencing of brain regions differentially affected by pathology. *Sci Rep* 2021;11(1):4865. doi:10.1038/s41598-021-83872-z, PMID:33649380.
- [16] Caldwell AB, Anantharaman BG, Ramachandran S, Nguyen P, Liu Q, Trinh I, *et al*. Transcriptomic profiling of sporadic Alzheimer's disease patients. *Mol Brain* 2022;15(1):83. doi:10.1186/s13041-022-00963-2, PMID:36224601.
- [17] Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 2007;23(14):1846–1847. doi:10.1093/bioinformatics/btm254, PMID:17496320.
- [18] Li R, Qu H, Wang S, Wei J, Zhang L, Ma R, *et al*. GDCRNATools: an R/Bioconductor package for integrative analysis of lncRNA, miRNA and mRNA data in GDC. *Bioinformatics* 2018;34(14):2515–2517. doi:10.1093/bioinformatics/bty124, PMID:29509844.
- [19] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26(1):139–140. doi:10.1093/bioinformatics/btp616, PMID:19910308.
- [20] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16(5):284–287. doi:10.1089/omi.2011.0118, PMID:22455463.
- [21] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28(1):27–30. doi:10.1093/nar/28.1.27, PMID:10592173.
- [22] von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 2003;31(1):258–261. doi:10.1093/nar/gkg034, PMID:12519996.
- [23] Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD, *et al*. GeneMANIA update 2018. *Nucleic Acids Res* 2018;46(W1):W60–W64. doi:10.1093/nar/gky311, PMID:29912392.
- [24] Chang L, Zhou G, Soufan O, Xia J. miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology. *Nucleic Acids Res* 2020;48(W1):W244–W251. doi:10.1093/nar/gkaa467, PMID:32484539.
- [25] Perea JR, Bolós M, Avila J. Microglia in Alzheimer's Disease in the Context of Tau Pathology. *Biomolecules* 2020;10(10):1439. doi:10.3390/biom10101439, PMID:33066368.
- [26] Hung C, Livesey FJ. Endolysosome and autophagy dysfunction in Alzheimer disease. *Autophagy* 2021;17(11):3882–3883. doi:10.1080/1548627.2021.1963630, PMID:34429033.
- [27] Moreira PI, Carvalho C, Zhu X, Smith MA, Perry G. Mitochondrial dysfunction is a trigger of Alzheimer's disease pathophysiology. *Biochim*

- Biophys Acta 2010;1802(1):2–10. doi:10.1016/j.bbadis.2009.10.006, PMID:19853658.
- [28] Dal Canto MC. The Golgi apparatus and the pathogenesis of Alzheimer's disease. *Am J Pathol* 1996;148(2):355–360. PMID:8579097.
- [29] Baloyannis SJ, Mavroudis I, Baloyannis IS, Costa VG. Mammillary Bodies in Alzheimer's Disease: A Golgi and Electron Microscope Study. *Am J Alzheimers Dis Other Dement* 2016;31(3):247–256. doi:10.1177/1533317515602548, PMID:26399484.
- [30] Husain MA, Laurent B, Plourde M. APOE and Alzheimer's Disease: From Lipid Transport to Physiopathology and Therapeutics. *Front Neurosci* 2021;15:630502. doi:10.3389/fnins.2021.630502, PMID:33679311.
- [31] Verghese PB, Castellano JM, Garai K, Wang Y, Jiang H, Shah A, *et al*. ApoE influences amyloid- β (A β) clearance despite minimal apoE/A β association in physiological conditions. *Proc Natl Acad Sci U S A* 2013;110(19):E1807–E1816. doi:10.1073/pnas.1220484110, PMID:23620513.
- [32] Radons J. The human HSP70 family of chaperones: where do we stand? *Cell Stress Chaperones* 2016;21(3):379–404. doi:10.1007/s12192-016-0676-6, PMID:26865365.
- [33] Song B, Shen S, Fu S, Fu J. HSPA6 and its role in cancers and other diseases. *Mol Biol Rep* 2022;49(11):10565–10577. doi:10.1007/s11033-022-07641-5, PMID:35666422.
- [34] Arrigo AP. Mammalian HspB1 (Hsp27) is a molecular sensor linked to the physiology and environment of the cell. *Cell Stress Chaperones* 2017;22(4):517–529. doi:10.1007/s12192-017-0765-1, PMID:28144778.
- [35] Arrigo AP. Mammalian HspB1 (Hsp27) is a molecular sensor linked to the physiology and environment of the cell. *Cell Stress Chaperones* 2017;22(4):517–529. doi:10.1007/s12192-017-0765-1, PMID:28144778.
- [36] Baron BW, Baron RM, Baron JM. The ITM2B (BRI2) gene is a target of BCL6 repression: Implications for lymphomas and neurodegenerative diseases. *Biochim Biophys Acta* 2015;1852(5):742–748. doi:10.1016/j.bbadis.2014.12.018, PMID:25557390.
- [37] Capo V, Abinun M, Villa A. Osteoclast rich osteopetrosis due to defects in the TCIRG1 gene. *Bone* 2022;165:116519. doi:10.1016/j.bone.2022.116519, PMID:35981697.
- [38] Lie PPY, Nixon RA. Lysosome trafficking and signaling in health and neurodegenerative diseases. *Neurobiol Dis* 2019;122:94–105. doi:10.1016/j.nbd.2018.05.015, PMID:29859318.
- [39] Qi B, Wang Y, Chen ZJ, Li XN, Qi Y, Yang Y, *et al*. Down-regulation of miR-30a-3p/5p promotes esophageal squamous cell carcinoma cell proliferation by activating the Wnt signaling pathway. *World J Gastroenterol* 2017;23(45):7965–7977. doi:10.3748/wjg.v23.i45.7965, PMID:29259372.
- [40] Sun T, Zhao K, Liu M, Cai Z, Zeng L, Zhang J, *et al*. miR-30a-5p induces A β production via inhibiting the nonamyloidogenic pathway in Alzheimer's disease. *Pharmacol Res* 2022;178:106153. doi:10.1016/j.phrs.2022.106153, PMID:35257899.