MEDICAL SCIENCE MONITOR

MOLECULAR BIOLOGY

e-ISSN 1643-3750
© Med Sci Monit, 2020; 26: e919249
DOI: 10.12659/MSM.919249

Received: 2019.08.08
Accepted: 2019.11.04
Published: 2020.01.27

Microarray Analysis of Differential Gene Expression in Alzheimer's Disease Identifies Potential Biomarkers with Diagnostic Value

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABCDEFG 1 Liping Liu*
BCEF 2 Qin Wu*
BCDEF 3 Weiwei Zhong
BCFG 4 Yuping Chen
BCE 5 Wenying Zhang
BCD 1 Huiling Ren
BC 1 Ling Sun
ABCDEFG 6 Jihu Sun

1 Pharmaceutical College, Jiangsu Vocational College of Medicine, Yancheng, Jiangsu, P.R. China
2 Medical Technology College, Jiangsu Vocational College of Medicine, Yancheng, Jiangsu, P.R. China
3 School of Public Foundation, Jiangsu Vocational College of Medicine, Yancheng, Jiangsu, P.R. China
4 School of Basic Medicine, Jiangsu Vocational College of Medicine, Yancheng, Jiangsu, P.R. China
5 Institute of Biotechnology, Jiangsu Vocational College of Medicine, Yancheng, Jiangsu, P.R. China
6 Department of Science and Technology, Jiangsu Vocational College of Medicine, Yancheng, Jiangsu, P.R. China

* Liping Liu and Qin Wu contributed equally to this work and should be considered co-first authors

Corresponding Author: Jihu Sun, e-mail: ywjkc1@163.com

Source of support: Funding was provided by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 17KJB360003)

Background: Alzheimer disease (AD) is a common and fatal subtype of dementia that remains a challenge to diagnose and treat. This study aimed to identify potential biomarkers that influence the prognosis of AD.

Material/Methods: A total of 6 gene expression profiles from the Gene Expression Omnibus (GEO) database were assessed for their potential as AD biomarkers. We identified differentially expressed genes (DEGs) using the prediction analysis for microarray (PAM) algorithm and obtained hub genes through the analysis of the protein-protein interaction (PPI) network and module analysis.

Results: We identified 6 gene expression profiles from the GEO database and assessed their potential as AD biomarkers. Shared gene sets were extracted and integrated into large expression profile matrices. We identified 2514 DEGs including 68 upregulated- and 2446 downregulated genes through analysis of the limma package. We screened 379 significant DEGs including 68 upregulated and 307 downregulated genes for their ability to distinguish AD from control samples using PAM algorithm. Functional enrichment of the 379 target genes was produced from Database for Annotation, Visualization and Integrated Discovery.(DAVID) and included histone function, beta receptor signaling, cell growth, and angiogenesis. The downregulated genes were significantly enriched in MAPK signaling, synaptic signaling, neuronal apoptosis and AD associated pathways. Upon analysis of the PPI network, 32 hub genes including ENO2, CCT2, CALM2, ACACB, ATP5B, MDH1, and PP2CA were screened. Of these hub genes, NFKBIA and ACACB were upregulated and 29 genes were downregulated in AD patients.

Conclusions: We screened 379 significant DEGs as potential biomarkers of AD using PAM and obtained 32 hub genes through PPI network and module analysis. These findings reveal new potential AD biomarkers with prognostic and therapeutic value.

MeSH Keywords: Alzheimer Disease • Biological Markers • Microarray Analysis • Protein Array Analysis

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/919249

2776 3 9 39

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Background

Alzheimer disease (AD) is the most common subtype of dementia which is officially listed as the sixth leading cause of death worldwide. Recent estimates indicate that AD ranks third behind heart disease and cancer as the major cause of death in the elderly [1]. AD is characterized by deposits of amyloid-beta (Aβ) plaques, and intracellular neurofibrillary tangles (NFT) in the neocortical and limbic regions of the brain [2,3]. As a group of unknown primary degenerative and an irreversible progressive brain diseases, AD causes neuronal cell apoptosis and brain atrophy [4] and slowly destroys memory, cognitive ability [5,6] and the ability of the body to perform basic bodily functions such as walking and swallowing, seriously affecting quality of life [7]. The risk factors for the onset and development of AD closely relate to oxidative stress, mitochondrial dysfunction, inflammation, glutamatergic excitotoxicity, low neurotrophic factors and neurogenesis [8]. It thus appears that AD is a complex disease for which suitable therapeutic approaches are not currently recognized. Although drug development is improving, the complexity of AD makes therapeutic approaches challenging. To improve AD therapy, a deeper understanding of the molecular mechanisms causing the disease are required [9].

To further understand the mechanisms of AD pathogenesis, high-throughput gene expression data has been investigated and substantial progress has been made in reconstructing gene regulatory networks. Network-based approaches [10] including protein-protein interaction (PPI) networks have been applied and are informative and powerful for discovering disease mechanisms. PPI networks can be reconstructed from protein domains, gene expression data, and structure-based information [11] which serve to regulate protein activity, the scaffolding of multi-protein complexes, and enzyme-substrate interactions [12]. PPI networks are altered in many disease states [13–15] and their targeting offers hope for disease treatment. An array of studies has integrated gene expression and PPI data to identify protein complexes [16], small subnetworks [17], and biomarkers [18] in disease states. Transcription factor networks are also a key determinant of cell fate decisions during mammalian development and adult tissue homeostasis is disrupted in disease [19].

The purpose of this study was to identify biomarkers that influence AD prognosis. Six gene expression profiles were obtained from the Gene Expression Omnibus (GEO) database, and 2514 differentially expressed genes (DEGs) were initially screened. Of the 379 DEGs identified, 68 were upregulated and 307 were downregulated identified through prediction analysis for microarray (PAM) algorithms that discriminate AD from normal samples. Upregulated genes were significantly enriched in histone function, beta receptor signaling, cell growth, and angiogenesis. Downregulated genes were significantly enriched in MAPK signaling, synaptic signaling, neuronal apoptosis, and Alzheimer associated pathways. A total of 32 hub genes including ENO2, CCT2, CALM2, ACACB, ATP5B, MDH1, and PPP2CA were screened based on PPI networks and module analysis. Of these hub genes, NFKBIA and ACACB were upregulated, whilst 29 genes were downregulated in AD patients. These findings reveal new AD biomarkers with prognostic and therapeutic value.

Material and Methods

Data download and preprocessing

Six sets of gene expression profiles related to AD were collected from the GEO database (*https://www.ncbi.nlm.nih.gov/geo/*) including GSE1297 [20], GSE28146 [21], GSE4757 [22], GSE5281 [23], GSE48350 [24], and GSE11882 [25]. GSE1297 was based on the GPL96 platform, whilst others were based on GPL570. The probes corresponding to multiple genes were removed and multiple probes corresponding to the genes were added. Mean expression values were then assessed. The gene sets were scored with z-score standards and extracted using shared gene sets. The genes were then integrated into large gene expression profiles. Finally, due to the batch effects of the GPL96 and GPL570 platforms, the removeBatchEffect function was employed from the limma package.

Screening of significant differentially expressed genes

Brain samples were collected from Alzheimer's disease Research Center (ADRC) brain banks for subsequent gene expression profiling. All AD patients met the Alzheimer disease and Related Disorders Association criteria for the clinical diagnosis of AD. Controls were cognitively normal healthy controls with no disease history including cardiovascular risk factors. The limma package [26] in R was used to identify significant DEGs (fold change >1.2; $P<0.05$).

Establishment of the PAM model

PAM algorithms (*http://www-stat.stanford.edu/tibs/PAM/index.html*) were used to identify potential biomarkers with diagnostic value in AD. The centroid methodology [27] was employed to obtain accurate classifications and depict each gene class. The sorting approach can shrink each class centroid to an overall centroid based on the threshold values to solve data classification difficulties. Moreover, 10-fold cross-validations were performed to guarantee forecasting accuracy. For the aforementioned gene integration profiles, we used limma analysis to build the PAM model and identify potential biomarkers of AD.

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

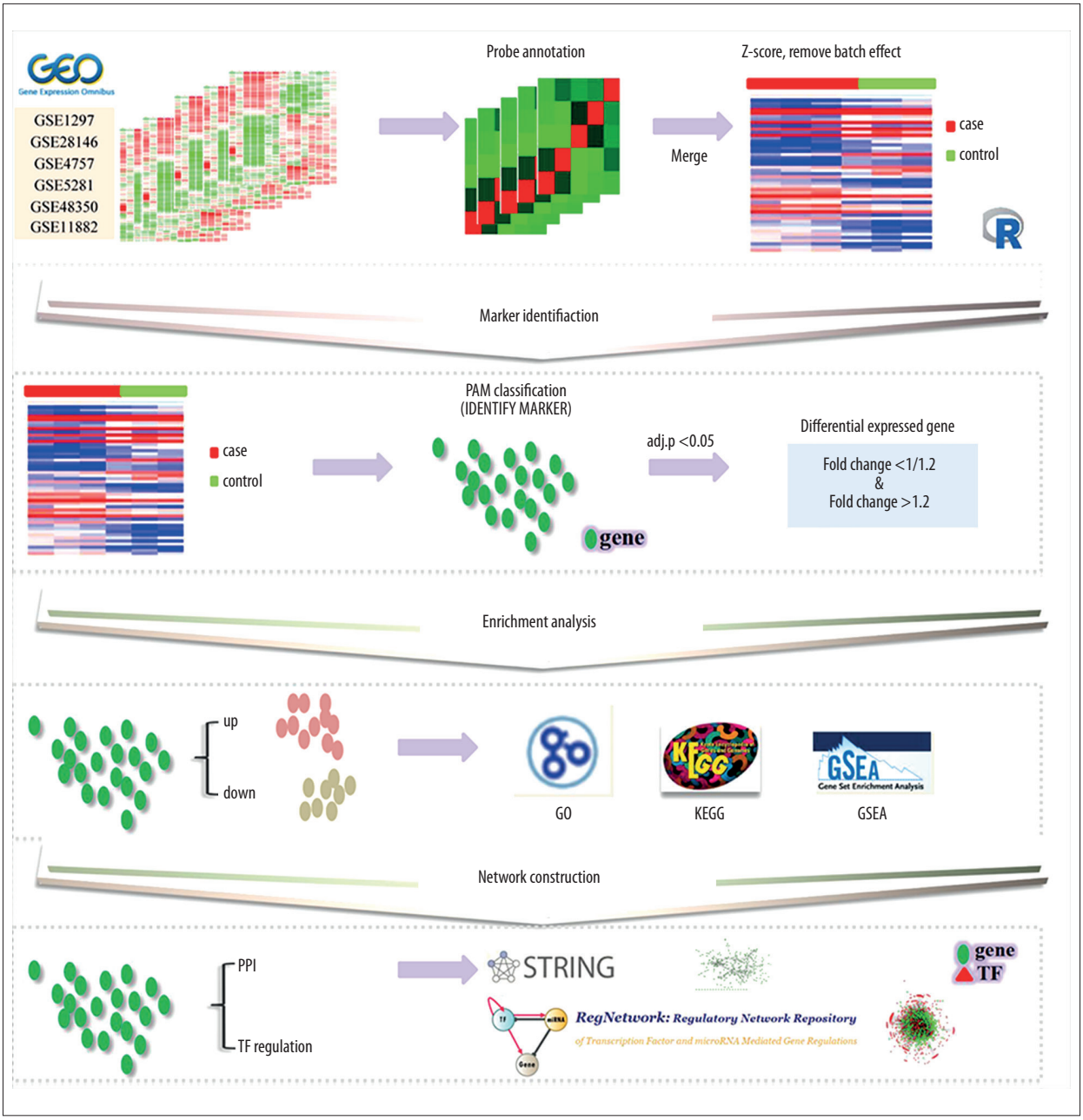


Figure 1. Identification of potential AD biomarkers. Six gene expression profiles were downloaded from the GEO database and z-score standardization was performed to assess expression profiles. The DEGs between normal and AD samples (fold change >1.2; *P*<0.05 were obtained through the limma package by PAM. GO and KEGG gene enrichment was analyzed by DAVID. PPI and transcription factor network analysis of the genes were further characterized. AD – Alzheimer disease; GEO – Gene Expression Omnibus; DEGs – differentially expressed genes; PAM – prediction analysis for microarray; GO – Gene Ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes; DAVID – Database for Annotation, Visualization and Integrated Discovery; PPI – protein-protein interaction.

Gene Set enrichment analysis

The function of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the potential biomarkers was analyzed by Database for Annotation, Visualization and Integrated Discovery.(DAVID) (*https://david.ncifcrf.gov/*, 6.7 version) with Benjamini *P*<0.05. Subsequent studies involving the significant functional enrichment of GO and KEGG analysis were used in the analysis.

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 1. Expression spectrum data statistics.

GSE	Platform	Case	Control	Total	Number of probes in the original expression profile	The number of genes corresponding to the annotation
GSE1297	GPL96	22	9	31	22283	12433
GSE28146	GPL570	22	8	30	54675	20486
GSE4757	GPL570	10	10	20	54675	20486
GSE5281	GPL570	74	87	161	54675	20486
GSE48350	GPL570	253	0	253	54675	20486
GSE11882	GPL570	0	173	173	54675	20486

GSE – Gene Set Enrichment.

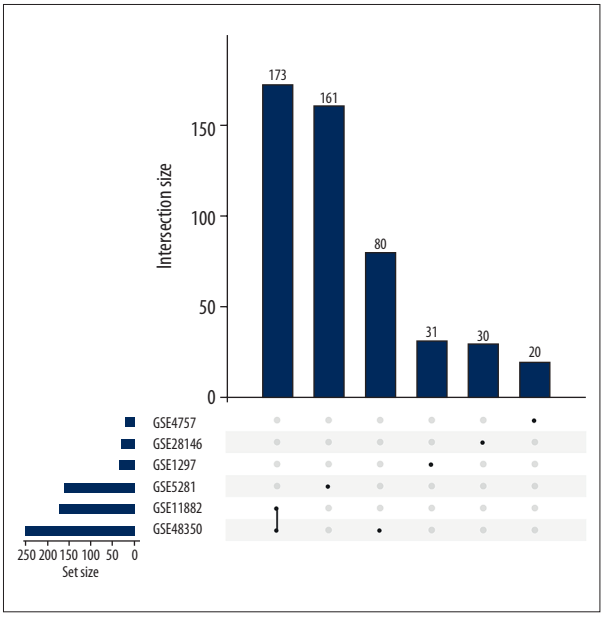


Figure 2. Sample overlap in the data sets. The bar graph corresponds to the number of samples in each data set. Black solid lines correspond to GEO datasets. Lines between the solid origin describe that the same sample exists between GSE11882 and GSE48350. GSE48350 contained 173 normal samples of GSE11882. GEO – Gene Expression Omnibus.

Gene Set Enrichment Analysis (GSEA) enrichment analysis

We used C2 curated genes as background gene sets using the MsigDB database (*http://software.broadinstitute.org/gsea/msigdb/*, version 6.2). The screening of upregulated and downregulated gene sets was performed using GSEA enrichment analysis, specific parameters were as follows:

- Numbers of permutations: 1000
- Collapsed datasets due to gene symbols: false
- Enrichment statistic: classic
- Metrics for gene ranking: Signal2Noise
- Gene list sorting model: real
- Gene list ordering mode: descending
- Max size: excluding larger sets: 1000
- Min size: excluding larger sets: 3

Establishment of the PPI network

"Multiple Protein" functional modules of the String database (*http://string-db.org/*, v10.5) were used to construct the PPI network which was further visualized using Cytoscape (version 3.6.1) software. Moreover, we analyzed the betweenness, tightness, degree and distribution of hub genes in the PPI network using the igraph R package. We searched for whole functional modules through the fastest greedy algorithm and analyzed GO and KEGG pathway enrichment analysis of the genes in these modules. Hub genes in the PPI network were identified and the proteins of these hub genes were further characterized based on degree size of the network.

Transcription factor analysis

Transcription factor gene data of the RegNetwork database (*http://regnetworkweb.org/*) was used to characterize the transcription factor regulatory networks associated with the target genes. The RegNetwork describes transcription factor to genes from the perspective of TransFac, TRED and JASPAR database resources. The RegNetwork describes a more comprehensive transcription factor regulation network. RegNetwork describes transcriptional regulation between transcription factor and target genes from the 3 database resources of TransFac, TRED, and JASPAR, and the RegNetwork's transcription factor regulation network is more comprehensive. Amongst them, the RegNetwork database contains 149 841 transcriptional regulations between transcription factor and target gene regulatory relationships in humans. Target and hub genes were mapped into the transcription factor regulatory network separately to obtain each transcription factor regulatory network which were visualized using Cytoscape. Furthermore, transcription

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 2. The demographic characteristics of the participant.

		GSE1297	GSE28146	GSE4757	GSE5281	GSE48350
Status	Case	22	22	10	87	253
	Control	9	8	10	74	0
nft	>20	12	–	–	–	–
	≤20	19	–	–	–	–
braak	>4	21	–	–	–	42
	≤4	10	–	–	–	35
age	>70	30	29	–	141	72
	≤70	1	1	–	18	101
mmse	>20	15	–	–	–	3
	≤20	16	–	–	–	50
Sex	Male	13	18	–	103	124
	Female	18	12	–	58	129
pmi	>2.5	23	–	–	–	–
	≤2.5	8	–	–	–	–
Cell Type	Layer III neurons	–	–	–	82	–
	Pyramidal neuron	–	–	–	79	–
individual	AA	–	–	–	–	33
	C	–	–	–	–	140
brain region	Entorhinal cortex	–	–	–	–	54
	Hippocampus	–	–	–	–	62
	Postcentral gyrus	–	–	–	–	68
	Superior frontal gyrus	–	–	–	–	69

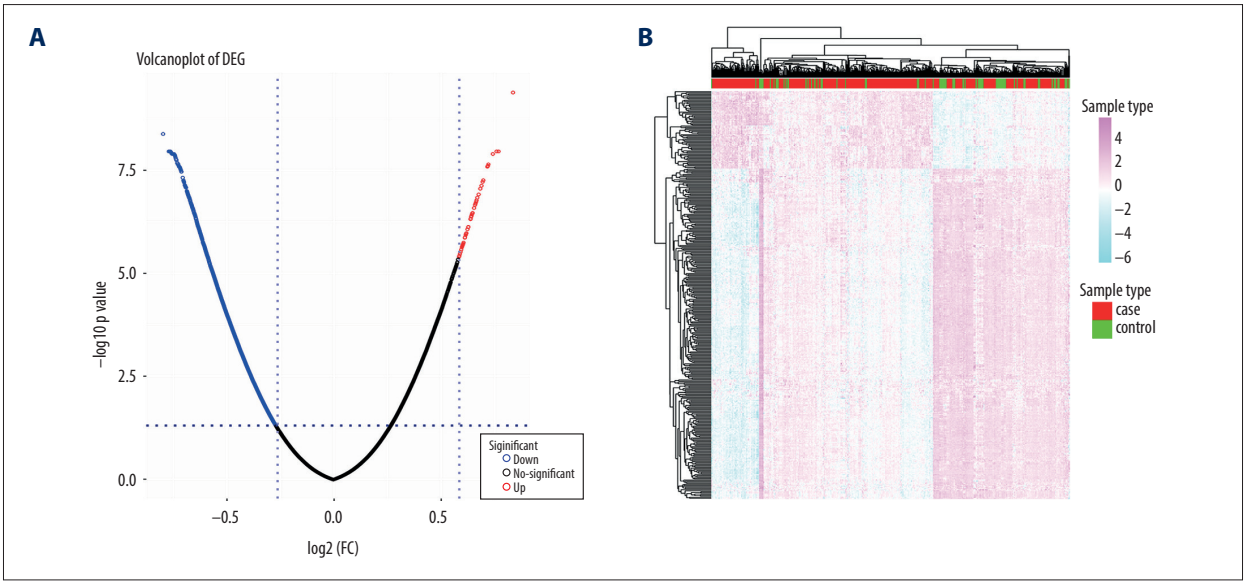


Figure 3. Visualization of significant DEGs. (**A**) Volcano plots of the DEGS. Blue lines: significant downregulated genes; red lines: upregulated genes; black lines: insignificant differences. The value of the horizontal line: -log10 (0.05); the 2 vertical lines: log2 (1/2) and log (2) from left to right. (**B**) DEGs in the sample. Columns: sample; rows: genes; red: case sample; green: normal sample. DEGs – differentially expressed genes.

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

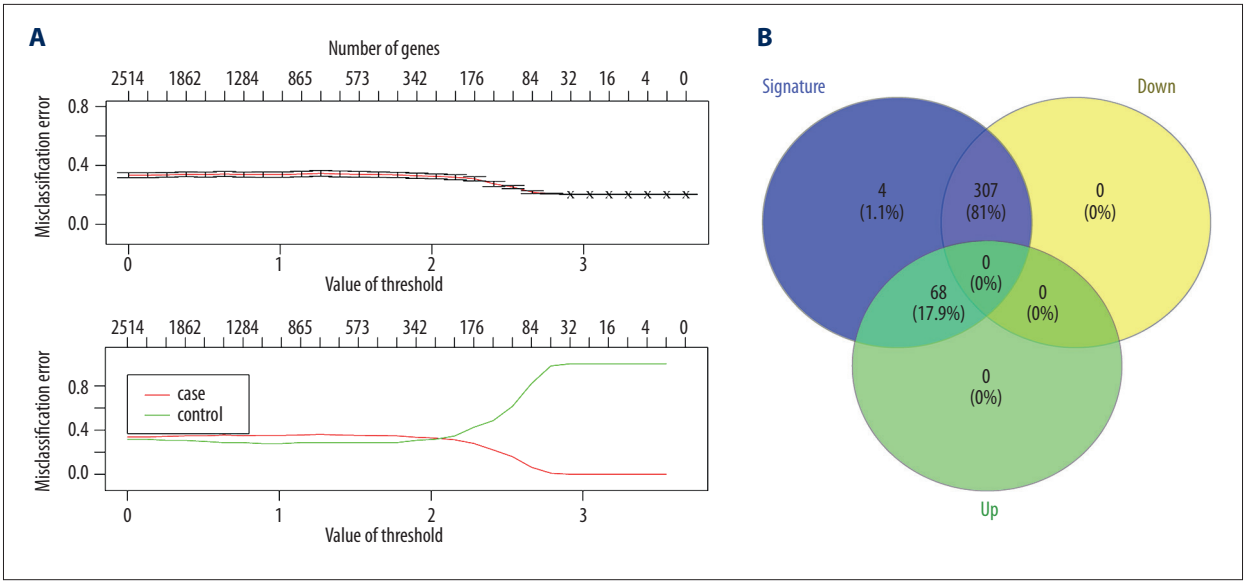


Figure 4. Biomarkers of AD using the PAM method. (**A**) Error rate distribution of PAM classifier training; (**B**) distribution diagram of upregulated and downregulated signature genes using PAM; blue: PAM signature; yellow: downregulated genes; green: upregulated genes. AD – Alzheimer disease; PAM – prediction analysis for microarray.

factors regulating target and hub genes were functionally annotated by DAVID.

Results

Data downloads and preprocessing

Six sets of AD gene expression profiles were download from the GEO database. The PAM model was built and designed and is presented in Figure 1.

Screening of significant differentially expressed genes

We downloaded the gene expression profiles of AD including GSE1297, GSE28146, GSE4757, GSE5281, GSE48350, and GSE11882 (Table 1). For each set of expression profiles, probes corresponding to the multiple genes were respectively deleted, whilst for multiple probes corresponding to the same gene, the mean expression values of the probes were taken as the expression values of the genes. Next, z-score standardization was performed for the 6 expression profiles. A total of 12 399 shared genes from 6 expression profiles were extracted, and 6 sets of expression profiles were integrated in large gene expression profiles which contained 12 399 genes and 668 samples. As the origin of the expression profiles were from GPL96 and GPL570 platforms, the platform differences were regarded as batch effects, and the data was further processed by the removeBatchEffect of the limma package. In Figure 2, the bar graph corresponds to the number of samples in each data set and each black solid origin corresponds to the corresponding GEO data sets. The lines between the solid origins explain that the same sample is present in both GSE11882 and GSE48350 datasets. GSE48350 contained 173 samples of GSE11882. Finally, we analyzed GSE1297, GSE28146, GSE4757, GSE5281, and GSE48350 datasets that included 12 399 genes, and 495 samples (394 AD and 101 normal samples). For these samples, the demographic characteristics of the patients (sex, age, stage, etc.) are shown in Table 2. A total of 2514 DEGs (fold change >1.2; $P<0.05$) were obtained through the expression profiles of the limma package. Amongst them, the DEGs contained 2446 downregulated genes and 68 upregulated genes (Figure 3A, 3B).

Establishment of the PAM model

The PAM algorithm was used to identify potential biomarkers with diagnostic value in AD. Objects were trained in the pamr package of R to ensure minimal errors between cross-validation sets and test sets using 10-fold cross-validation. As shown in Figure 4A, the threshold was selected as 2.65 and we identified a total of 379 significant DEGs. As shown in Figure 4B, there were 307 downregulated genes and 68 upregulated genes. These DEGs represented potential biomarkers of AD.

Gene Set enrichment analysis

Both upregulated and downregulated genes were subjected to GO and KEGG pathway enrichment analysis using DAVID (*https://david.ncifcrf.gov/*, version 6.7) (Benjamini $P<0.05$). Amongst them, we identified 307 downregulated genes that were significantly enriched in MAPK signaling pathway, synaptic

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

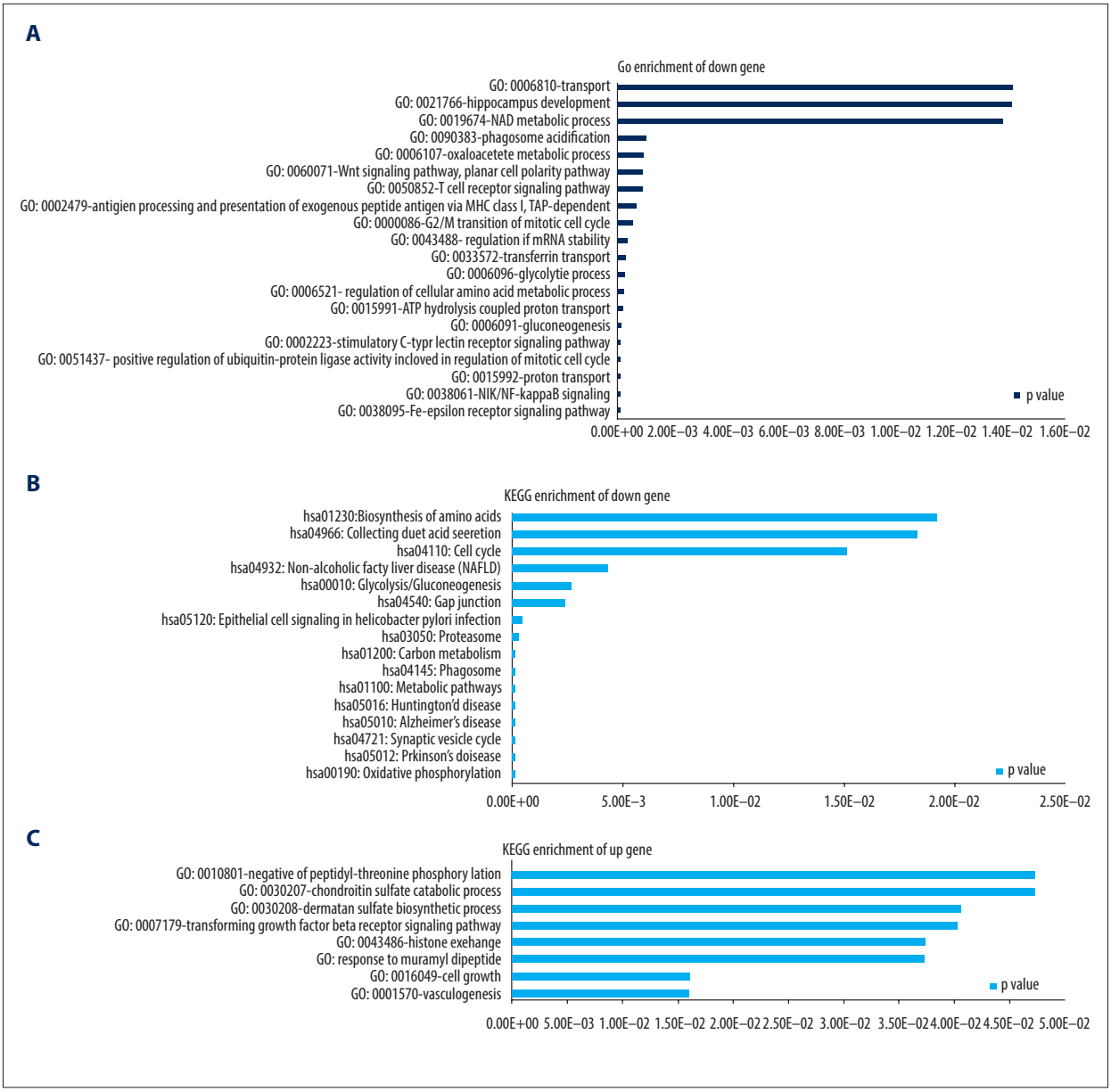


Figure 5. Functional annotation of the potential biomarkers. (**A, B**) Downregulated genes; deep blue: annotated results of GO term enrichment; light blue: KEGG pathway enrichment. (**C**) upregulated genes following KEGG pathway enrichment. GO – Gene Ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes.

signaling, neuronal projection development, neuronal apoptosis, protein folding, and Alzheimer-related pathways (Figure 5A, 5B). The 68 upregulated genes were significantly enriched in cell growth, histone modifications and angiogenesis (Figure 5C).

GSEA enrichment analysis

We selected C2 curated gene sets from the MsigDB database (*http://software.broadinstitute.org/gsea/msigdb/*, version 6.2) as background gene sets. The upregulated and downregulated genes were analyzed by GSEA enrichment using the parameters described in the materials and methods. Since the upregulated gene sets contained only 68 genes, the GSEA requires that the input gene sets are greater than 100. We performed GSEA analysis on the 307 downregulated genes and imprints in the GSEA display. At *P*-values <0.05, the downregulated genes were significantly enriched in pathways including: MARTORIATI_MDM4_TARGETS_FETAL_LIVER_UP, CHOW_RASSF1_TARGETS_DN, YANG_BREAST_CANCER_ESR1_BULK_UP, APPIERTO_RESPONSE_TO_FENRETINIDE_DN, KEGG_CYSTEINE_AND_METHIONINE_METABOLISM,

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

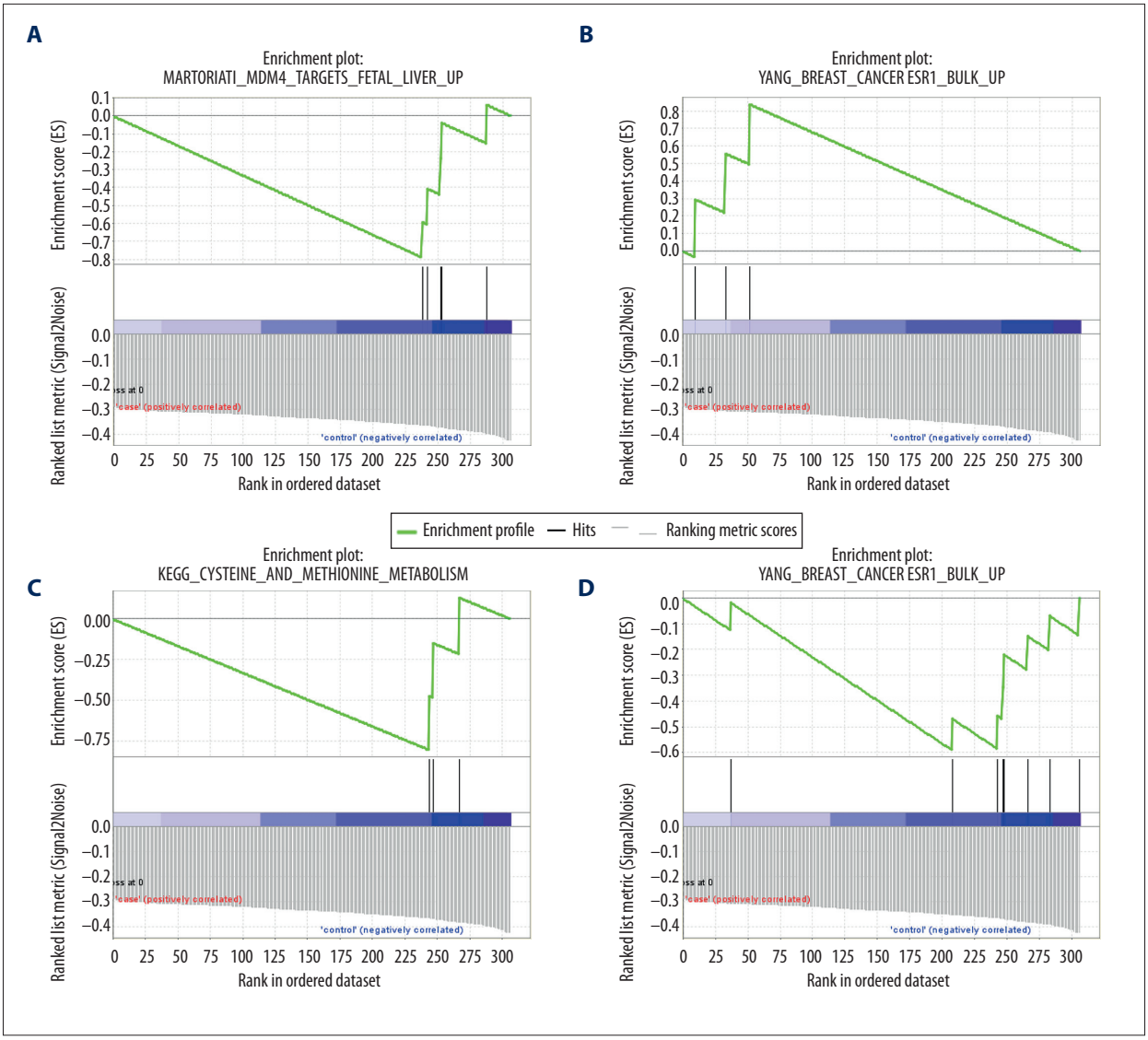


Figure 6. Downregulated genes visualized by Gene Set Enrichment Analysis (GSEA). (**A**) MARTORIATI_MDM4_TARGETS_FETAL_LIVER_UP. (**B**) YANG_BREAST_CANCER_ESR1_BULK_UP. (**C**) KEGG_CYSTEINE_AND_METHIONINE_METABOLISM. (**D**) MULLIGHAN_NPM1_MUTATED_SIGNATURE_2_UP.

MULLIGHAN_NPM1_MUTATED_SIGNATURE_2_UP and VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_DN. The GSEA display results are shown in Figure 6.

PPI network analysis

"Multiple Protein" functional modules of the String database were used to construct the PPI network of 379 significant DEGs (Figure 7A), and the network contained 281 nodes and 1177 edges. For the PPI network of the 379 significant DEGs, we used the igraph R package to calculate the topological properties of the network including the betweenness and tightness of the network, which were 270.6948 and 0.0006465125, respectively. The genes with a degree greater than or equal to 16 were regarded as hub genes (Figure 7B). The maximum degree of 36 was for ENO2 and others included CCT2, CALM2, ACACB, ATP5B, MDH1, PPP2CA, PSMD14, ATP5C1, LDHA, CCT4, COPS5, TXN, PGK1, NDUFA4, ACTR10, IMMT, ATP5F1, NDUFAB1, CUL1, PSMB7, SKP1, MTHFD1, TUBA4A, PSMA5, TUBA1B, TUBA1C, HINT1, NME1, PSMA1, TUBB4B, and NFKBIA. Of the 32 hub genes, NFKBIA and ACACB were upregulated and 29 genes were downregulated in AD patients. Figure 7C shows the degree distribution of the genes in the network. Figure 7D shows the hub scores of the 281 genes, where black points indicate hub scores of the non-hub genes and red scores indicate hub scores of the hub genes. The hub scores for the hub nodes were significantly higher. Document verification of these genes is shown in Table 3.

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

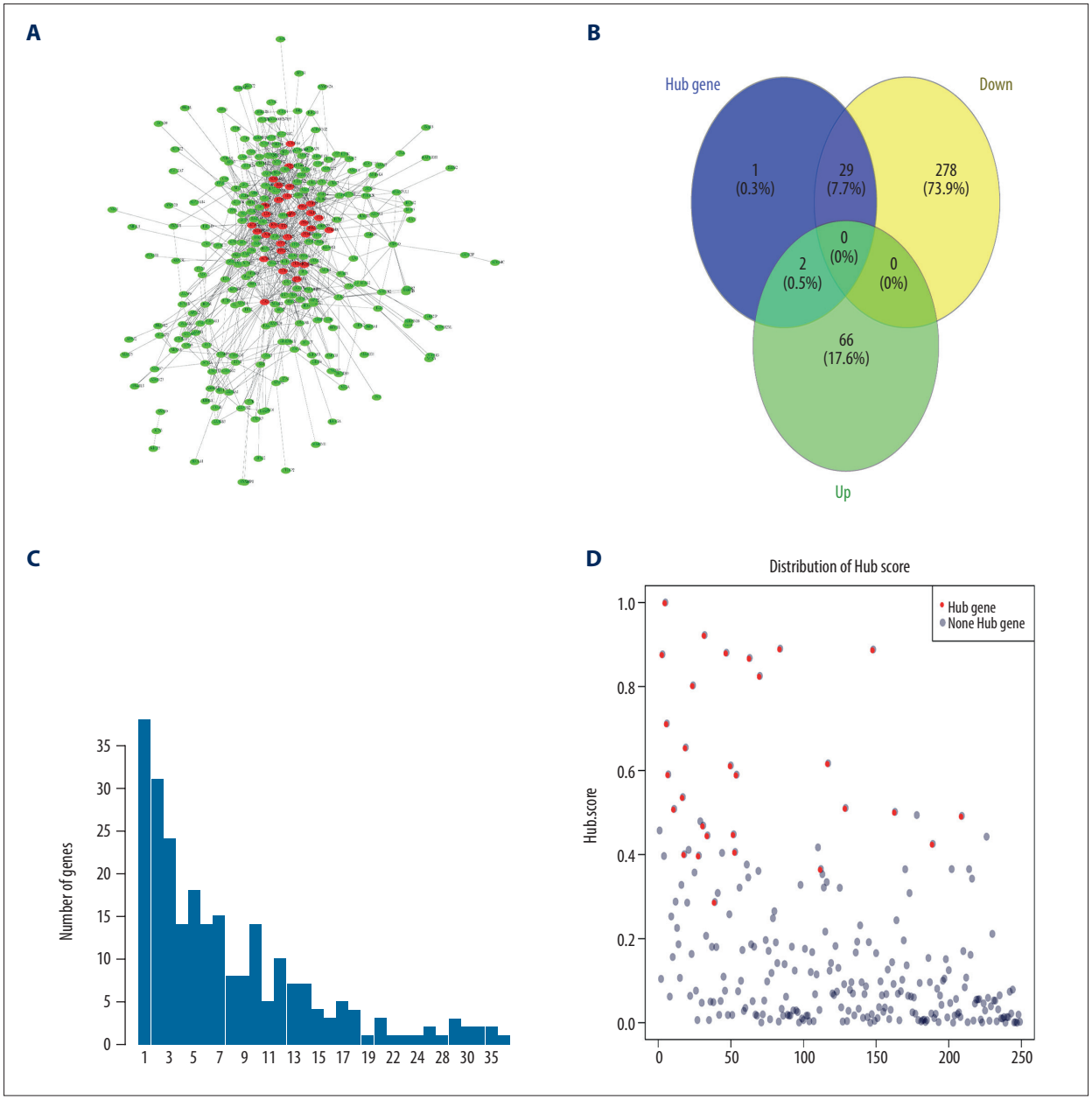
Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Using functional annotation of the 32 hub genes using DAVID, the genes were enriched in the regulation of telomerase activity and function, carbon metabolism, and TGF-beta signaling. As shown in Figure 7E and 7F, the functions of GO and KEGG of the 32 hub genes were correctly displayed.

For the PPI network, we identified 9 functional modules (Figure 8A). Amongst them, the largest number of genes focused on the first, second, and third modules which contained 55, 67, and 40 genes, respectively. The genes in the 3 modules are enriched and analyzed. Figure 8B and 8C show that the first module mainly involved protein folding, glycolysis, metabolic pathways, and amino acid synthesis. Figure 8D and 8E showed that the second module involved the regulation of neuronal apoptosis, neurotransmitter secretion, cell histamine responses and other important functions. Figure 8F and 8G show that the third module was enriched for important functions such as protein catabolism, protein repair, TGF-signaling pathway and cell cycle regulation.

Transcription factor analysis

Data were downloaded from the RegNetwork database (http://regnetworkweb.org/) to obtain regulatory data for transcriptional factors and genes. The 379 most significant DEGs were then mapped to the network. Finally, a transcription factor



This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

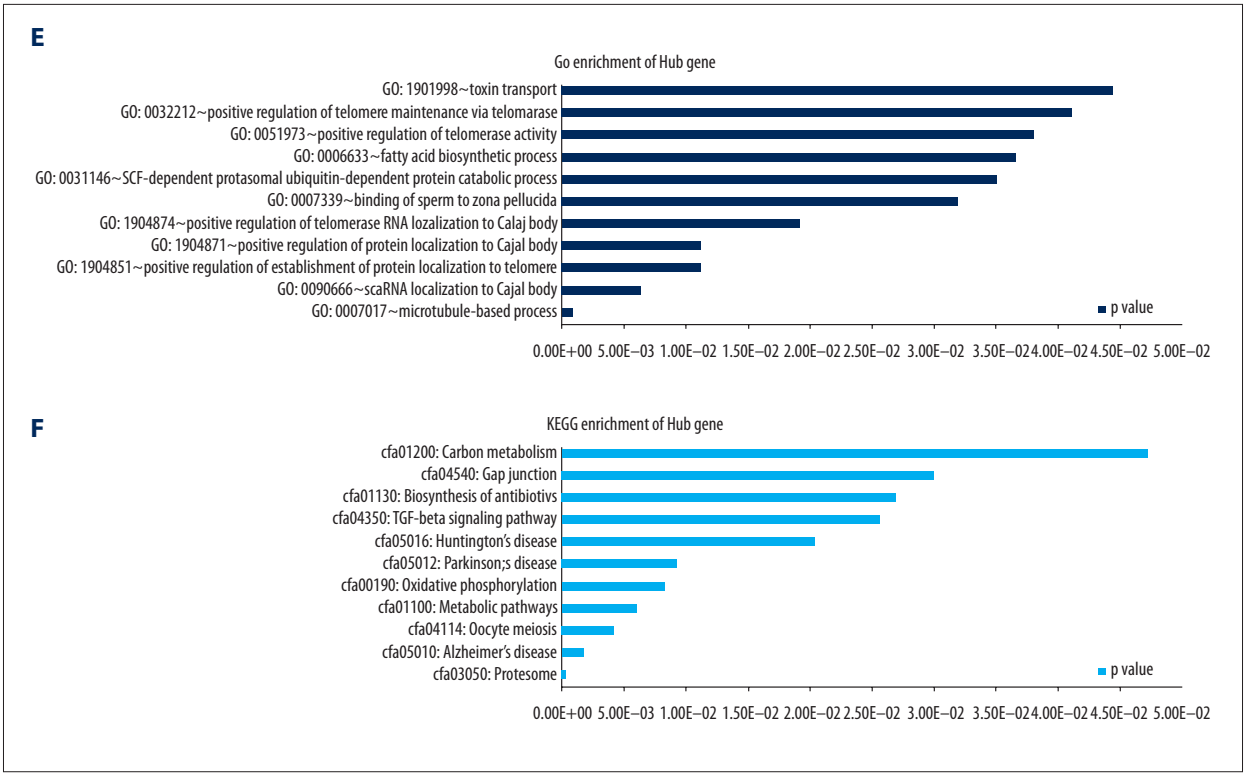


Figure 7. PPI network. (**A**) Network diagram of PPI; red dots: 32 hub genes. (**B**) Distribution of upregulated and downregulated genes. (**C**) Degree distribution diagram of genes in the interaction network. (**D**) Score distribution of hub nodes with higher degrees; red: hub nodes; gray: non-hub nodes. (**E**) GO function analysis of hub genes. (**F**) KEGG function analysis of the hub genes. PPI – protein-protein interaction; GO – Gene Ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes.

Table 3. The Literature Validation of Hub Gene in protein-protein interaction network.

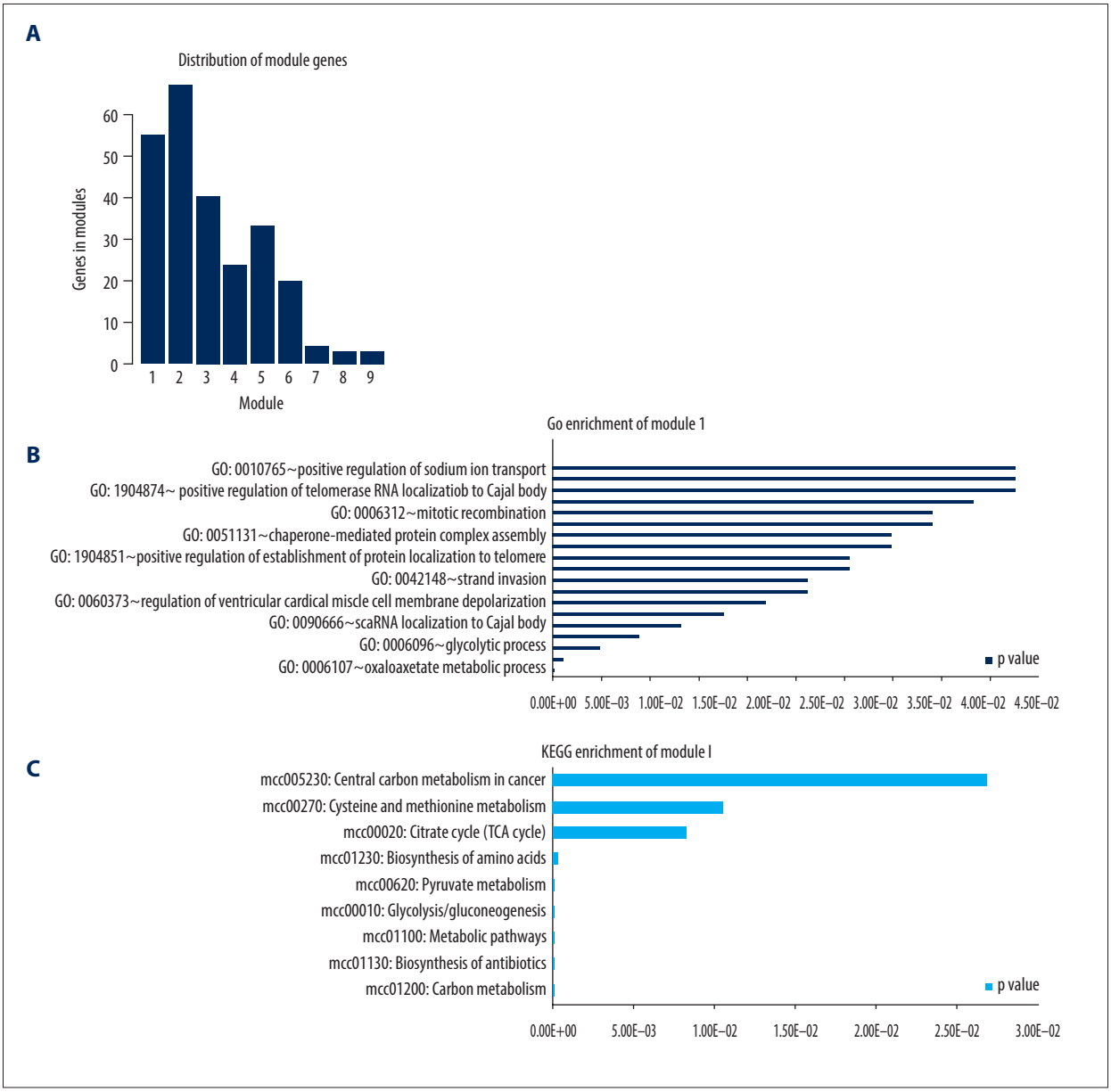
Degree	Gene	PubMed
36	*ENO2*	29458411,28885659, 28127669, 28671048
35	*CCT2, CALM2*	19332537, 25812852
34	*ACACB, ATP5B*	27911303, 24558171
30	*MDH1, PPP2CA*	28966679,26811028, 28623007
29	*PSMD14, ATP5C1, LDHA*	27050411,29848990, 21541279, 29079469
28	*CCT4*	27929117
27	*COPS5, TXN*	23926111,
24	*PGK1*	21672555, 20570715
23	*NDUFA4*	28578378, 30054583
22	*ACTR10*	
21	*IMMT, ATP5F1, NDUFAB1*	28966679, 28869321
19	*CUL1*	12354302
18	*PSMB7, SKP1, MTHFD1, TUBA4A*	12354302,17691219, 20217437
17	*PSMA5, TUBA1B, TUBA1C, HINT1, NME1*	18923405, 25311278
16	*PSMA1, TUBB4B, NFKBIA*	26202100

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

regulatory network containing 753 points and 3174 edges were obtained (Figure 9A) which contained 404 transcription factors and 359 regulated genes. In Figure 9A, the red triangles indicate transcription factors; green circles indicate genes; and arrows indicate the transcriptional regulations between transcription factor and target gene regulation. Through DAVID's functional enrichment analysis, we found 404 transcription factors that regulate neuronal apoptosis, neuronal differentiation, calcineurin-NFAT signaling cascades, protein complex assembly, cellular insulin responses, Ras protein signaling, Toll-like receptor signaling pathways, T cell signaling pathways, Huntington's disease-related pathways, and DNA replication (Figure 9B, 9C).

Considering the functions of the 32 hub genes in previous PPI networks, we characterized the post-transcriptional regulatory networks of the 32 genes and obtained regulatory relationships between transcription factor and the hub genes. As shown in Figure 9D, for these 32 hub genes, 31 (~96.875%) were regulated by 12 transcription factors. These 12 transcription factors are associated with protein transcription, drug responses, cell apoptosis, ErbB signaling, cell proliferation, and Jak-STAT signaling (Figure 9E, 9F).



This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

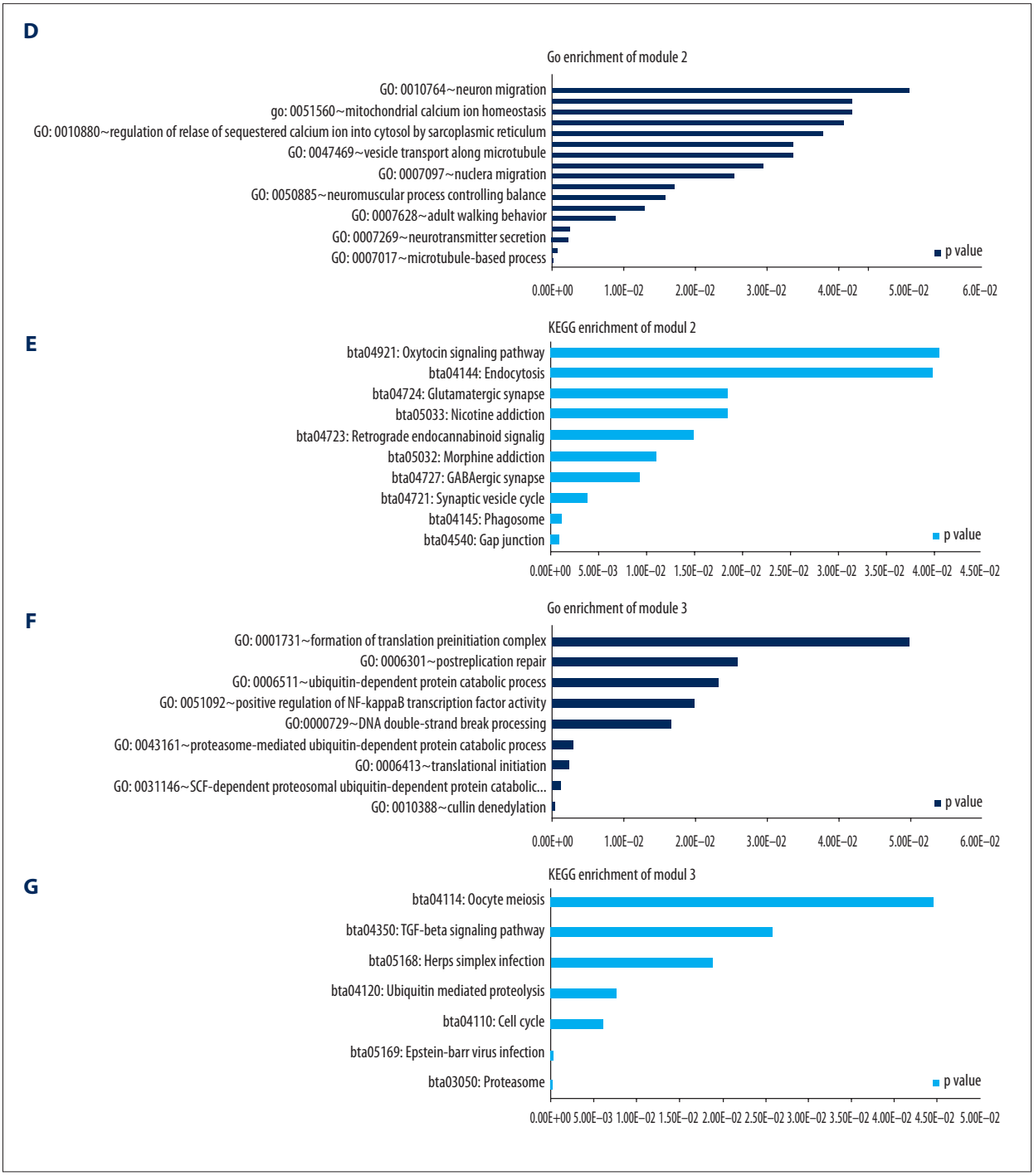


Figure 8. Module analysis in the PPI network. (**A**) Degree distribution in the interaction network. (**B, C**) GO and KEGG function analysis of module 1. (**D, E**) GO and KEGG function analysis of module 2. (**F, G**) GO and KEGG function analysis of module 3. PPI – protein-protein interaction; GO – Gene Ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes.

Discussion

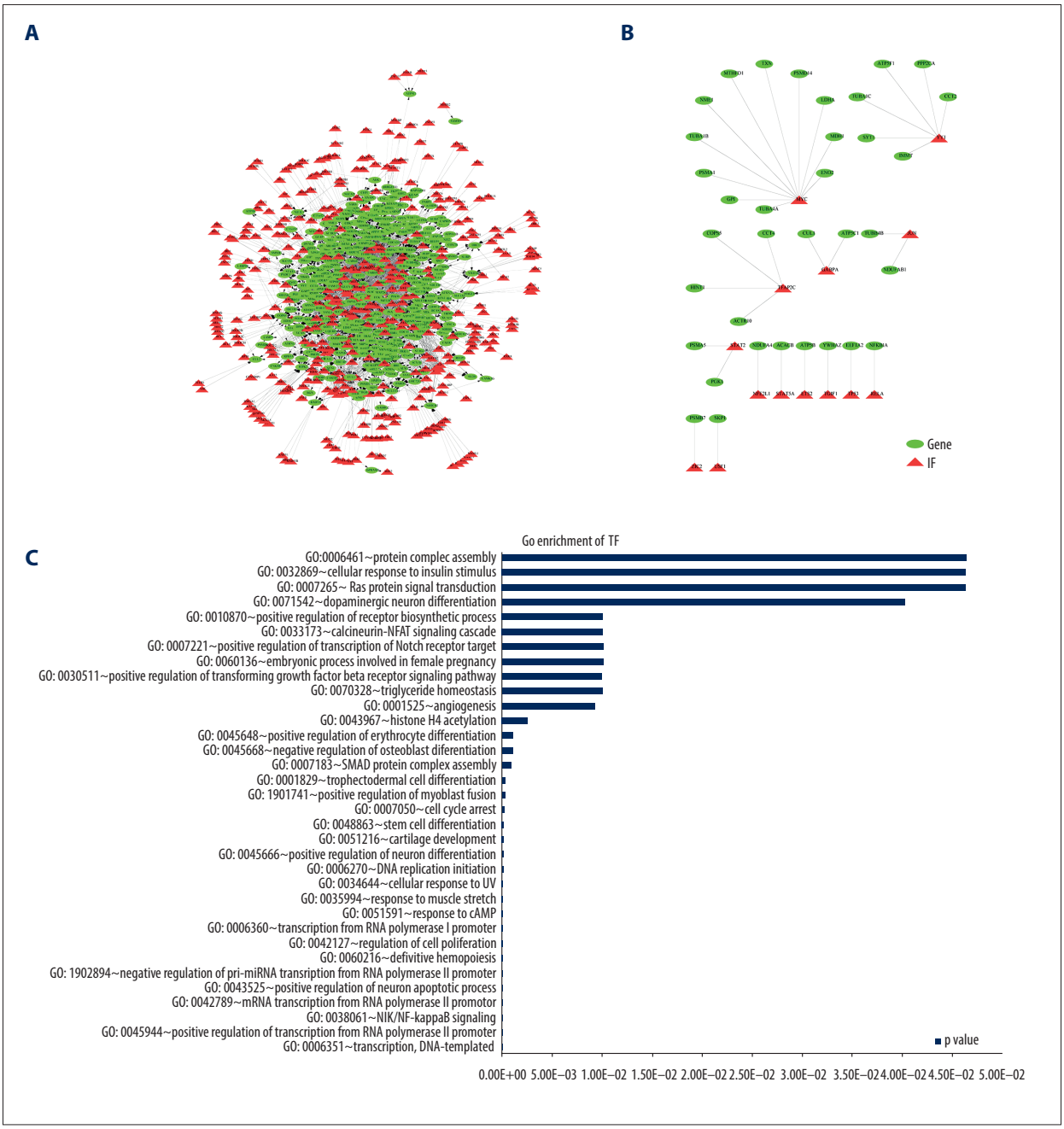
Dementia represents a broad category of brain diseases that cause the loss of cognitive functioning including thinking, remembering, reasoning, and behavioral abilities that interfere with daily life activities [27]. AD is the most common subtype of dementia that contributes to 60–70% of cases [28] and ultimately leads to death [29–31]. To systematically understand the pathogenesis of AD and provide a diagnosis we integrated 6 gene expression profiles of the GEO database of AD, including

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

2514 DEGs screened through the limma package, and the PAM algorithm to identify 379 significant DEGs genes including 68 upregulated genes and 307 downregulated genes to differ AD from normal samples. The upregulated genes were significantly enriched for histone modifications, beta receptor signaling pathways, cell growth, and angiogenesis. The downregulated genes were significantly enriched in MAPK signaling, synaptic signaling, neuronal apoptosis and AD pathways. We obtained 32 hub genes based on the PPI networks in addition to module analysis including ENO2, CCT2, CALM2, ACACB, ATP5B, MDH1, PPP2CA, PSMD14, ATP5C1, LDHA, CCT4, COPS5, TXN, PGK1, NDUFA4, ACTR10, IMMT, ATP5F1, NDUFAB1, CUL1, PSMB7, SKP1, MTHFD1, TUBA4A, PSMA5, TUBA1B, TUBA1C, HINT1, NME1, PSMA1, TUBB4B, and NFKBIA which represented potential biomarkers with diagnostic value in AD. Of these 32 hub genes, the maximum degree amongst the genes was ENO2, NFKBIA, and ACACB which were upregulated, whilst 29 genes were downregulated in AD patients.

We screened 379 significant DEGs as potential biomarkers of AD using PAM and obtained 32 hub genes through the PPI network and module analysis. Most genes with larger degrees



This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

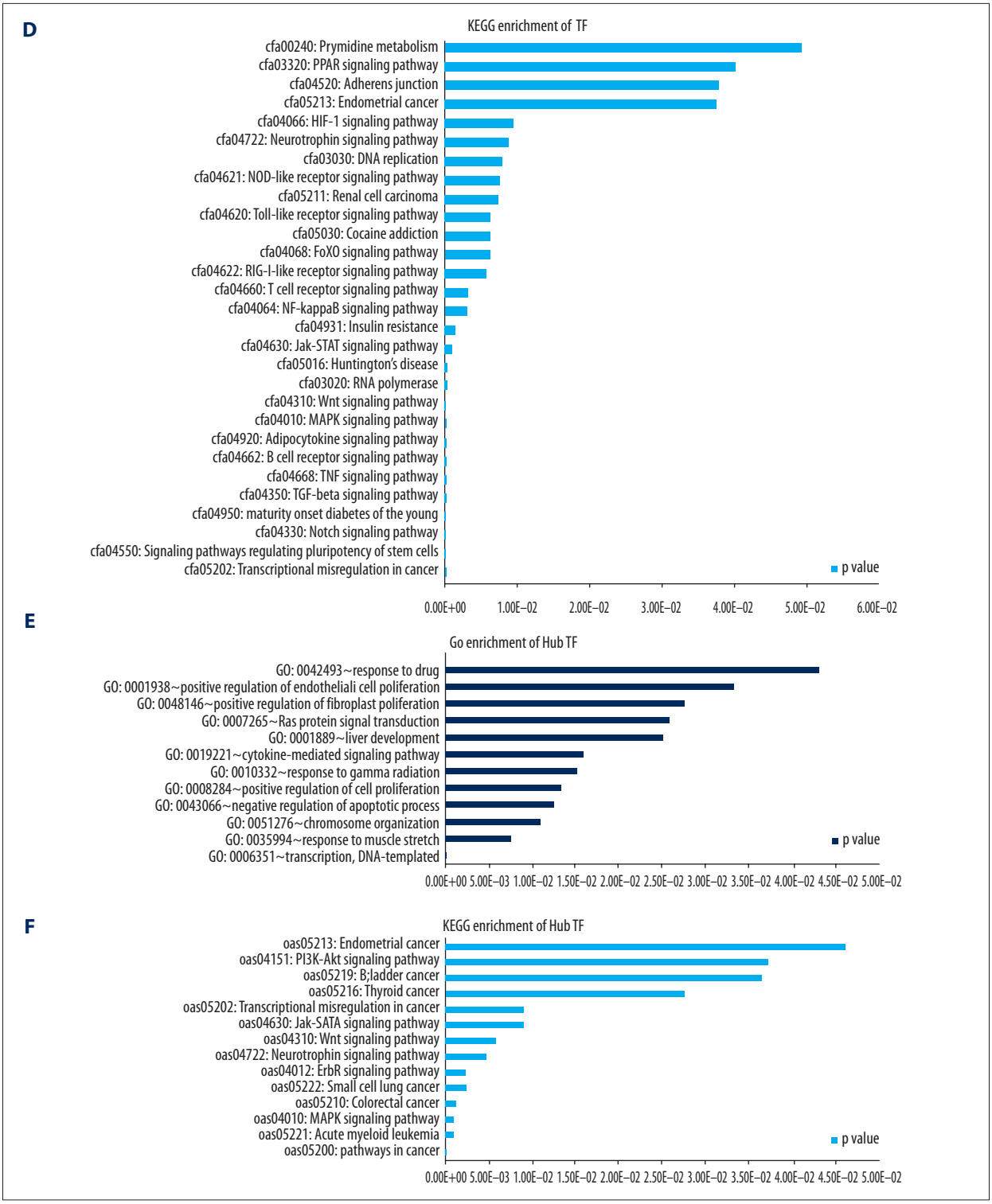


Figure 9. Transcription factor regulatory network. (**A**) Transcriptional factor regulatory network containing 753 points and 3174 edges, red triangle: transcription factors; green circle: 404 transcription factors and 359 genes. (**B**) Transcription factor regulatory network of hub genes which contain 49 points and 35 edges. There were 31 hub genes and 12 transcription factor genes. (**C, D**) GO function and KEGG pathway enrichment results of 404 transcription factors. (**E, F**) GO functions and KEGG pathway enrichment results of 12 transcription factors regulating 31 hub genes. GO – Gene Ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes.

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

were known to cause AD in the PPI network, whilst others were novel and not previously reported. ENO2 (enolase 2) was the highest scoring hub gene which is crucial for the catalysis of phosphoenolpyruvate in glial and neuronal cell populations. The enzyme is downregulated in AD [32]. The upregulated genes included ACACB and NFKBIA. ACACB (acetyl-CoA carboxylase beta) is a rate-limiting enzyme involved in mitochondrial fatty acid oxidation and plays a key role in fatty acid metabolism [33]. NFKBIA (NFKB inhibitor alpha) is involved in inflammatory responses and is upregulated in AD [34]. Of the other hub genes, CALM2 (calmodulin 2) is a member of the calmodulin gene family and regulates intracellular signaling [35]; MDH1 (malate dehydrogenase 1) is highly expressed in brain tissue [36]; PGK1 (phosphoglycerate kinase 1) is associated neuronal disease [37]; ACTR10 (actin related protein 10) is expressed in brain tissue [38]; and NME1 (NME/NM23 nucleoside diphosphate kinase 1) is downregulated in patients with AD[39]. For these 32 hub genes, 31 (~96.875%) were regulated by 12 transcription factors. These 12 transcription factors were associated with protein transcription, drug responses, cell apoptosis, ErbB signaling, cell proliferation and Jak-STAT signaling.

Conclusions

We identified potential AD biomarkers with diagnostic value using multiple sets of chip data and through the construction and analysis of PPI and transcription factor regulatory networks. These findings contribute to our understanding of AD pathogenesis and systematically provide new directions for the diagnosis and treatment of AD.

References:

1. Murphy SL, Xu JQ, Kochanek KD: Deaths: final data for 2010. National Vital Statistics Reports, 2003; 50(15): 1–118
2. Hampel H, Shen Y, Walsh DM et al: Biological markers of amyloid beta-related mechanisms in Alzheimer's disease. Exp Neurol, 2010; 223(2): 334–46
3. Verma M, Vats A, Taneja V: Toxic species in amyloid disorders: Oligomers or mature fibrils. Ann Indian Acad Neurol, 2015;18(2): 138–45
4. Jin M, Shepardson N, Yang T et al: Soluble amyloid β-protein dimers isolated from Alzheimer cortex directly induce Tau hyperphosphorylation and neuritic degeneration. Proc Natl Acad Sci, 2011; 108(14): 5819–24
5. Alinejad S, Aaseth J, Abdollahi M et al: Clinical aspects of opium adulterated with lead in Iran: A review. Basic Clin Pharmacol Toxicol, 2018; 122(1): 56–64
6. Giuffrida ML, Caraci F, Pignataro B et al: β-amyloid monomers are neuroprotective. J Neurosci, 2009; 29(34): 10582–87
7. Korolev IO: Alzheimer's disease: A clinical and basic science review. Medical Student Research Journal, 2014; 4: 24–33
8. Ganguli M, Chandra V, Kamboh MI et al: Apolipoprotein E polymorphism and Alzheimer disease: the Indo-US cross-national dementia study. Arch Neurol, 2000; 57(6): 824–30
9. Rao VS, Srinivas K, Kumar GNS, Sujin GN: Protein interaction network for Alzheimer's disease using computational approach. Bioinformation, 2013; 9(19): 968–72
10. Bhardwaj N, Lu H: Correlation between gene expression profiles and protein-protein interactions within and across genomes. Bioinformatics, 2005; 21(11): 2730–38
11. Huang Y, Sun X, Hu G: An integrated genetics approach for identifying protein signal pathways of Alzheimer's disease. Comput Methods Biomech Biomed Engin, 2011; 14(4): 371–78
12. Zhong M, Lee GM, Sijbesma E: Modulating protein–protein interaction networks in protein homeostasis. Curr Opin Chem Biol, 2019; 50: 55–65
13. Thul PJ, Åkesson L, Wiking M et al: A subcellular map of the human proteome. Science, 2017; 356(6340): 806–7
14. Huttlin EL, Bruckner RJ, Paulo JA et al: Architecture of the human interactome defines protein communities and disease networks. Nature, 2017; 545: 505–9
15. Li Z, Ivanov AA, Su R et al: The OncoPPi network of cancer-focused protein–protein interactions to inform biological insights and therapeutic strategies. Nat Commun, 2017; 8: 14356
16. Feng J, Jiang R, Jiang T: A max-flow-based approach to the identification of protein complexes using protein interaction and microarray data. IEEE/ACM Trans Comput Biol Bioinform, 2011; 8(3): 621–34
17. Huang Y, Zhang J, Huang Y: Computational identification of proteins subnetwork in Parkinson's disease study. Anti-Counterfeiting, Security and Identification (ASID) 2012 International Conference on IEEE; 2012
18. Jahid MJ, Ruan J: Identification of biomarkers in breast cancer metastasis by integrating protein-protein interaction network and gene expression data. IEEE International Workshop; 2011
19. Wilkinson AC, Nakauchi H, Göttgens B: Mammalian transcription factor networks: Recent advances in interrogating biological complexity. Cell Syst, 2017; 5(4): 319–31
20. Blalock EM, Geddes JW, Chen KC et al: Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. Proc Natl Acad Sci USA, 2004; 101(7): 2173–78
21. Blalock EM, Buechel HM, Popovic J et al: Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. J Chem Neuroanat, 2011; 42(2): 118–26
22. Dunckley T, Beach TG, Ramsey KE et al: Gene expression correlates of neurofibrillary tangles in Alzheimer's disease. Neurobiol Aging, 2006; 27(10): 1359–71
23. Liang WS, Dunckley T, Beach TG et al: Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain. Physiol Genomics, 2007; 28(3): 311–22
24. Berchtold NC, Cribbs DH, Coleman PD et al: Gene expression changes in the course of normal brain aging are sexually dimorphic. Proc Natl Acad Sci USA, 2008; 105(40): 15605–10
25. Berchtold NC, Coleman PD, Cribbs DH et al: Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. Neurobiol Aging, 2013; 34(6): 1653–61
26. Ritchie ME, Phipson B, Wu D et al: limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res, 2015; 43(7): e47
27. Bickart KC, Brickhouse M, Negreira A et al: Atrophy in distinct corticolimbic networks in frontotemporal dementia relates to social impairments measured using the Social Impairment Rating Scal. J Neurol Neurosurg Psychiatry, 2014; 85(4): 438–48
28. Ballard C, Gauthier S, Corbett A et al: Alzheimer's disease. Lancet, 2011; 377: 1019–31
29. Wu Q, Sun J, Song X et al: Blocking beta 2-adrenergic receptor inhibits dendrite ramification in a mouse model of Alzheimer's disease. Neural Regen Res, 2017; 12: 1499–506
30. Brookmeyer R, Evans DA, Hebert L et al: National estimates of the prevalence of Alzheimer's disease in the United States. Alzheimers Dement, 2011; 7(1): 61–73
31. Reitz C, Mayeux R: Alzheimer disease: Epidemiology, diagnostic criteria, risk factors and biomarkers. Biochem Pharmacol, 2014; 88(4): 640–51
32. Brooks WM, Lynch PJ, Ingle CC et al: Gene expression profiles of metabolic enzyme transcripts in Alzheimer's disease. Brain Res, 2007; 1127: 127–35

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

33. Ma L, Murea M, Snipes JA et al: An ACACB variant implicated in diabetic nephropathy associates with body mass index and gene expression in obese subjects. PLoS One, 2013; 8(2): e56193
34. Li X, Long J, He T et al: Integrated genomic approaches identify major pathways and upstream regulators in late onset Alzheimer's disease. Sci Rep, 2015; 5(1): 12393
35. Wang X, L Michaelis M, K Michaelis E: Functional genomics of brain aging and Alzheimer's disease: focus on selective neuronal vulnerability. Curr Genomics, 2010; 11(8): 618–33
36. Schmitz M, Llorens F, Pracht A et al: Regulation of human cerebrospinal fluid malate dehydrogenase 1 in sporadic Creutzfeldt-Jakob disease patients. Aging (Albany NY), 2016; 8(11): 2927
37. Li X, Zheng Y, Lu Z: PGK1 is a new member of the protein kinome. Cell Cycle, 2016; 15(14): 1803–4
38. Xu J, Patassini S, Rustogi N et al: Regional protein expression in human Alzheimer's brain correlates with disease severity. Commun Biol, 2019; 2(1): 43
39. Ansoleaga B, Jové M, Schlüter A et al: Deregulation of purine metabolism in Alzheimer's disease. Neurobiol Aging, 2015; 36(1): 68–80

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]