Identifying and Ranking Potential Driver Genes of Alzheimer's Disease Using Weighted Co-Expression Network Analysis

Liang-Yong Xia
Xi'an Jiaotong University

Lihong Tang
Xi'an Jiaotong University

Hui Huang
Shanghai Jiao Tong University

Jie Luo (✉ jieluo@sjtu.edu.cn)
Shanghai Jiao Tong University

Research

RESEARCH

Identifying and ranking potential driver genes of Alzheimer's disease using weighted co-expression network analysis

Liang-Yong Xia[1,2], Lihong Tang[1], Hui Huang[1] and Jie Luo[1*]

*Correspondence:
jieluo@sjtu.edu.cn
[1]School of Biomedical
Engineering,Shanghai Jiao Tong
University, 800 Dongchuan Road,
200030 Shanghai, China
Full list of author information is
available at the end of the article

Abstract

Background: Alzheimer's disease (AD) is one of the most common neurodegenerative diseases. Identification of AD-related genes from transcriptomics provided new direction to the mechanism for finding potential targets for drug therapy.

Methods: We mined gene co-expression network modules from differentially expressed genes (DEGs) of AD and normal samples in multiple datasets by weighted gene co-expression network analysis (WGCNA). A convergent functional genomic (CFG) method was used to prioritize potential driver genes.

Results: The 7567 DEGs were enriched significantly with 61 KEGG pathway and 242 GO terms. Then, the genes in 5 AD-specific modules obtained significantly from DEGs were interconnected with well-known AD risk genes in common PPI network. Remarkably, compared to the number of Tau production-related genes, $A\beta$ play a more critical role. Lastly, the 23 potential driver genes was prioritized by CFG method from 5 AD-specific modules.

Conclusions: Identification of AD-related genes could be useful for understanding pathophysiology of AD and looking for candidates drug targets.

Keywords: Alzheimer's disease; Genetic variation; Transcriptomic; Differentially expressed genes; Weighted gene co-expression network analysis

Background

Alzheimer's disease (AD) is one of the most common neurodegenerative diseases, accounting for the majority of dementia patients[1]. AD is estimated to affect in 13.8 million individuals in the United States (US), with 7.0 million being aged 85 years or older by 2050[2, 3]. Currently, genetic factor are believed to be partially responsible for AD[4]. Genome-wide association studies (GWAS) have also revealed that some single nucleotide polymorphisms (SNPs) contribute to AD disease onset[5, 6]. These include common variants such as amyloid protein precursor (APP), presenilin-1 (PSEN1), presenilin-2 (PSEN2) and apolipoprotein E (APOE). PSEN1, PSEN2 and APP genes are clear pathogenic genes of early-onset AD[7]. APOE, as the only identified risk gene for late-onset AD, can increase the rate of cognitive decline[8]. Different microRNAs (miRNAs) are also involved in the pathophysiology of AD[9]. For example, miRNA-377 promotes cell proliferation and inhibits cell apoptosis by regulating the expression level of cadherin 13 (CDH13), thus participating in the occurrence and development of AD[10]. Long non-coding RNAs (lncRNAs) have

been widely reported to be associated with a variety of physiological and pathological processes, such as AD. Brain cytoplasmic RNA is a kind of lncRNA, and the overexpression of brain cytoplasmic may lead to synaptic/dendritic degeneration in AD[11]. Despite the fact that remarkable advances have been made in the understanding of the genetic basis of AD, there is no pharmaceutical or other intervention that can treat AD. Identification of AD-related genes from transcriptomics becomes an attractive strategy for finding potential targets for drug therapy.

Gene expression profiling of transcriptomic datasets of AD and normal brain samples has identified potential genes and contributed to the search for potential targets[12]. Recently, most studies on AD transcriptome analysis pay attention to the identification of DEGs from different brain regionsand there are only few work studied highly correlated pairs of genes to obtain gain or loss of co-expression in AD [13]. Generally, describing the degree of association between two genes can be obtained by calculating correlation coefficients such as Pearson and Spearman between expression values. But the disadvantage of a fixed threshold approach is that the threshold is artificially defined and it ignores a lot of potential correlations. At the same time, this approach can also lose information about the trend of genetic changes, and it will be difficult to describe the strength of correlation in the network. In order to solve these problems, the idea of weighting is proposed. WGCNA is widely used to describe the association pattern between gene expression in microarrays or RNA-seq, and can be divided the gene co-expression network of complex biological processes into several highly related feature modules. The feature modules represent several groups of highly coordinated gene sets, associate the modules with specific clinical features to find genes that perform key functions, and help identify potential mechanisms involved in specific biological processes and explore candidate biomarkers[14, 15].

In this paper, we aimed to identify potential driver genes for AD from DEGs using WGCNA and CFG methods based on multiple transcriptomics dataset. Firstly, the datasets was download from AlzData and ADNI database. Then genes with $|logFC| > 0.1$ and $P < 0.05$ were regarded as DEGs. We also identified AD-specific modules using WGCNA method. Lastly, we prioritized potential driver genes by the CFG method in **Figure 1**.

Materials and Methods

AD expression data collection and preprocessing

All the original microarray data regarding AD were downloaded from Alz-Data(http://www.alzdata.org/)and Alzheimer's Disease Neuroimaging Initiative(http://adni.loni.usc.edu/). We have 467 controls and 309 AD from five dataset for subsequent analyses, including EC (39 vs 39), HP(67 vs 74), FC(128 vs 104), TC(39 vs 52) and ADNI(194 vs 40). Only ADNI had $A\beta$ and Tau related data. Detailed information of each dataset is shown in Table-1. Of note, to solve data imbalanced in ADNI dataset, we randomly chose 40 samples from the control in 10 times.

AD data detection of DEGs and enrichment of biological process

In order to identify potential genes of AD, differential expression analysis was conducted by R package edgeR and the Benjamini-Hochberg's method was used to

correct for multiple comparisons[16]. For each dataset, t-test was used to identify DEGs with a cut-off of statistical significance *P<0.05* and foldchange > 0.1. Functional enrichment of the DEGs was produced from Database for DAVID, which now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For obtained list of DEGs, DAVID is able to identify enriched biological themes, particularly KEGG pathway and GO terms.

Weighted gene co-expression network analysis

We use WGCNA approach, which is good for distinguishing the strong and weak relationships, to divide the gene co-expression network of complex biological processes into several highly related feature modules. For genes i and j, the correlation coefficient is r_{ij}, we define the correlation intensity :$a_{ij} = r_{ij}^{\beta}$, which depends on the choice of power β. The β value is determined by the lowest value close to the scale-free network.

According to this idea, the correlation strength of expression levels among all genes was calculated, and the adjacency matrix was obtained. Further, the topological overlap matrix (TOM) was calculated. In simple terms, if genes i and j have a lot of the same adjacent genes, then $TOM_{i,j}$ will be high, which means that there are similar expression patterns between the two genes. Therefore, several AD-specific modules from DEGs of AD and normal samples could be obtained by WGCNA.

Protein-protein interaction(PPI) network

Mentioned AD-specific modules before, to identify the genes in these AD-specific modules and the functional relationship between well-known AD risk genes(APP, PSEN1, PSEN2, APOE or MAPT), a protein-protein interaction(PPI) network was constructed by Gephi(https://gephi.org/).

Convergent functional genomics

The potential driver genes was prioritized from AD-specific modules by CFG method, which integrated various levels of AD-related evidence [4, 17]. There were five AD-related evidence: (1) regulation of target gene expression by AD genetic variants;(2) association at least one locus significantly ;(3) physical interaction between target genes and APP, PSEN1, PSEN2, APOE or MAPT;(4) differential expression of target gene before AD pathology emergence;(5) correlation of target gene expression with AD pathology in $A\beta$ and Tau. The range of CFG score was from 0 to 5, with 5 indicating highest priority.

Results

DEG detection

A total of 776 samples and 108,302 genes from multiple transcriptomic datasets were compiled for DEGs detection(*P<0.05* and foldchange > 0.1). Besides, for ADNI dataset, we randomly chose 40 samples from the control in 10 times and selected gene with frequency greater than or equal to 3. Each red node represented DEG for five datasets**(Figure 2)**. We identified 7567 DEG(2166 EC, 1952 HP, 949 FC, 3075

TC and 3204 ADNI) for subsequent analyses. About 6% ~ 19% of the total genes could be identified as DEGs. Among the DEG list in all five datasets, the expression patterns of well-known AD risk genes, such as APP, PSEN1, PSEN2, APOE and MAPT were only slightly altered or unchanged in AD patients. In addition, 19 genes had a consistently differential expression from EC, HP, FC, TC and ADNI(**Figure 3**).

Functional annotation of AD-related DEGs

We investigated functional enrichment of the AD-related DEGs. The 7567 target genes in the network were enriched in 324 KEGG pathway and 1381 GO terms in **Figure 4**. We identified 61 KEGG pathway and 324 GO terms (*P<0.005*), respectively. As shown in Table-2, we also found several pathways have been reported to be associated with AD, including Alzheimer disease pathway, MAPK signaling pathway and AMPK signaling pathway. Top 20 significantly KEGG pathway selected was exhibited for each dataset in **Figure 5**. Besides, these GO terms are divided into ontologies based on a hierarchical relations. Specifically, DEGs related to the biological processes for synaptic-related functions were significant enriched in Table-3, such as chemical synaptic transmission, regulation of postsynaptic membrane potential, synaptic vesicle exocytosis, synaptic transmission, GABAergic, regulation of synaptic transmission, glutamatergic, synaptic vesicle endocytosis and long-term synaptic potentiation. In addition, they were associated with neuron-related processes, including neurotransmitter secretion, neuron projection morphogenesis, negative regulation of neuron apoptotic process and negative regulation of neuron projection development. These results of KEGG pathway and GO terms could contribute to understand the neuropathological process of AD.

DEG-enriched modules

We used WGCNA to divide the DEGs into several highly related gene modules. As shown in **Figure 6**, a very significant positive correlation was observed between five modules and AD for five dataset. A modular size was ranged from 96 to 142 genes that might reflect the different layers and complexity of gene regulation in the AD brain. These five DEG-enriched modules were used for identifying potential driver genes for AD etiology and pathology. We obtained a PPI network from five AD-specific modules, which contained 5 seed genes and 209 neighbors. Due to the lack of connectivity, several genes were not included in the PPI network, such as S100A6, PSAT1 and MXI1. Notably, only a few genes were linked to all five well-known AD risk genes (APP, PSEN1, PSEN2, APOE and MAPT) in **Figure 7**.

Pathology correlation $A\beta$ and Tau

We also investigated the relationship between the genes in these five AD-specific modules and AD pathology in $A\beta$ and Tau. Table-4 shows that there were significant differences in AD pathology among different datasets. Compared with the number of Tau production-related genes, AD pathology in $A\beta$ play a vital role. These results could contributed to understanding the pathology process of AD.

Potential driver genes of AD

To search potential driver genes of AD, we identified 23 related-AD genes based on CFG score($>=$4) and P-value($<$0.05) from the genes in five AD-specific module. We also list previously published gene-based studies for the identification of patients with AD in Table-5. For example, GJA1, also known as connexin 43, shows upregulated mRNA and protein levels in AD[18].

Discussion

In this study, we aimed to identify potential driver genes for AD from multiple transcriptomics dataset DEGs by WGCNA and CFG methods.

Pathway enrichment analysis was performed to interpret the function of these DEGs. KEGG pathway analysis for the 7567 DEGs were significantly enriched in one KEGG pathway "MAPK signaling pathway", which is composed of ERK, P38, and JNK. In the adult nervous system, ERK activation is necessary for synaptic plasticity and memory formation[19]. In the brains of AD patients, P38 is highly expressed. $A\beta$-induced P38 activation increases tau phosphorylation and promotes the amyloidogenic processing of APP[20, 21]. In a mouse model of AD, the JNK signaling pathway is overactivated in the spine before cognitive decline[22]. These studies indicate that the overactivation of MAPK signaling pathway could cause the occurrence of AD. Therefore, preventing MAPK overactivation is effective strategy in order to reduce $A\beta$ deposition, Tau hyperphosphorylation, neuronal apoptosis, and memory impairment. MAPKs could be potential targets for novel and effective therapeutics of AD[23, 24].

GO term analysis indicated that the 7567 DEGs were mainly involved in chemical synaptic transmission, regulation of postsynaptic membrane potential, synaptic vesicle exocytosis, synaptic transmission, GABAergic synapses, regulation of synaptic transmission, glutamatergic, synaptic vesicle endocytosis, long-term synaptic potentiation, neurotransmitter secretion, neuron projection morphogenesis, negative regulation of neuron apoptotic process and negative regulation of neuron projection development. Damage to neuronal and synaptic function has always been considered an important pathological feature of neurodegenerative diseases, and decreased synaptic activity is also considered to be the most relevant pathological feature of AD cognitive impairment[25]. For example, the downregulation of GABAergic synapses is closely related to the loss of GABAergic inhibition[26]. Studies have found that GABAergic neurotransmission is closely related to various aspects of AD pathology, including $A\beta$ toxicity and Tau hyperphosphorylation[27]. The level of GABA inhibitory neurotransmitter in AD patients was significantly reduced, suggesting that AD has insufficient synaptic function and neuronal transmission[28]. In addition, In a mouse model of AD indicate that the impairment of hippocampal neurogenesis may be mediated by GABAergic signal dysfunction or the imbalance between excitatory and inhibitory synapses[29]. Therefore, GABAergic synapses not only plays an important role in the function of the hippocampus, but also in the pathogenesis of AD.

To explore the protein interactions, a PPI network was consisted of 5 well-known AD risk genes(APP, PSEN1, PSEN2, APOE and MAPT) and 209 genes. The 14 genes with greater than four degrees (including CXCL12,BAX, IRF2, IRF5, SMAR-CC1, SP1, CASP6, DCN, FZD2, GNA13, HNF4A, ITGAV, PELP1 and SDC2)

were selected. Especially, CXCL12 plays a major role in neuroinflammation because it mediates the local immune response and attracts leukocytes, which are thought to migrate across the blood-brain barrier along the concentration gradient of chemokines to their targets. This occurs for example in AD in the vicinity of the amyloid plaques that attract and/or activate local glial cells[30].

Limitations

There are some limitations in this study. First, although we identified 23 potential driver genes of AD by the WGCNA and CFG method, these approachs could be used to prioritize genes rather than to identify true causal genes. Therefore, further biological validation of the identified genes are necessary in future studies. Second, 4 of 5 datasets were downloaded from AlzData, which only retained the common genes from different studies during the cross-platform normalization. Third, the sample size of EC, HP and TC available for analyze was still limited, and the larger sample size of FC and ADNI might have a greater influence on the results. Fourth, the rapid development of various omics provide new opportunities for understanding of AD. However, we only used transcriptomics dataset to identify potential driver genes of AD. Finally, more potential genes of AD were not considered. Deep learning has capacity to dig out more hidden gene in data and is a machine learning algorithm based on artificial neural network, which is a computational model inspired by the structure of human brain. The main difference between deep learning and traditional artificial neural network lies in the scale and complexity of network structure. The networks of deep learning have a larger number of hidden layers, while traditional artificial neural networks usually have only one hidden layer. This is due to the lack of big data and GPU hardware technical support in the last century. Due to the emergence of more powerful CPU and GPU hardware, deep learning with more hidden layers is proposed on the basis of artificial neural network, and more nodes can be used in each hidden layer[31, 32].

Conclusions

In this study, we identified potential driver genes from AD-specific modules using multiple transcriptomics datasets and observed that DEGs were enriched with several pathways significantly, which are consistent with observations from previous studies. In summary, Our results contribute to understanding pathophysiology of AD and looking for candidates drug targets.

List of abbreviations

AD: Alzheimer's disease ; **DANI**: Alzheimer's Disease Neuroimaging Initiative; **EC**: Entorhinal Cortex; **HP**: Hippocampus; **TC**: Temporal Cortex; **FC**: Frontal Cortex; **DEGs**: Differentially expressed genes; **WGCNA**: Weighted gene co-expression network analysis; **DAVID**: Database for Annotation, Visualization and Integrated Discovery; **KEGG**: Kyoto Encyclopedia of Genes and Genomes; **GO**: Gene Ontology; $A\beta$:β-amyloid; **GWAS**: Genome-wide association studies; **SNP**: Single nucleotide polymorphisms; **GCN**: Gene co-expression network ; **CFG**: Convergent functional genomic

Declarations

Author details
[1]School of Biomedical Engineering,Shanghai Jiao Tong University, 800 Dongchuan Road, 200030 Shanghai, China.
[2]Faculty of Information Technology, Macau University of Science and Technology., Avenida Wai Long, Taipa, 999078 Macau, China.

References
 1. Darby, R.R., Joutsa, J., Fox, M.D.: Network localization of heterogeneous neuroimaging findings. Brain **142**(1), 70–79 (2019)
 2. Association, A., *et al.*: 2018 alzheimer's disease facts and figures. Alzheimer's & Dementia **14**(3), 367–429 (2018)
 3. Cummings, J., Lee, G., Ritter, A., Sabbagh, M., Zhong, K.: Alzheimer's disease drug development pipeline: 2019. Alzheimer's & Dementia: Translational Research & Clinical Interventions **5**, 272–293 (2019)
 4. Xu, M., Zhang, D.-F., Luo, R., Wu, Y., Zhou, H., Kong, L.-L., Bi, R., Yao, Y.-G.: A systematic integrated analysis of brain expression profiles reveals yap1 and other prioritized hub genes as important upstream regulators in alzheimer's disease. Alzheimer's & Dementia **14**(2), 215–229 (2018)
 5. Hao, S., Wang, R., Zhang, Y., Zhan, H.: Prediction of alzheimer's disease-associated genes by integration of gwas summary data and expression data. Frontiers in genetics **9**, 653 (2019)
 6. Andrews, S.J., Fulton-Howard, B., Goate, A.: Interpretation of risk loci from genome-wide association studies of alzheimer's disease. The Lancet Neurology **19**(4), 326–335 (2020)
 7. Lanoiselée, H.-M., Nicolas, G., Wallon, D., Rovelet-Lecrux, A., Lacour, M., Rousseau, S., Richard, A.-C., Pasquier, F., Rollin-Sillaire, A., Martinaud, O., *et al.*: App, psen1, and psen2 mutations in early-onset alzheimer disease: A genetic screening study of familial and sporadic cases. PLoS medicine **14**(3), 1002270 (2017)
 8. Wijsman, E.M., Pankratz, N.D., Choi, Y., Rothstein, J.H., Faber, K.M., Cheng, R., Lee, J.H., Bird, T.D., Bennett, D.A., Diaz-Arrastia, R., *et al.*: Genome-wide association of familial late-onset alzheimer's disease replicates bin1 and clu and nominates cugbp2 in interaction with apoe. PLoS Genet **7**(2), 1001308 (2011)
 9. Femminella, G.D., Ferrara, N., Rengo, G.: The emerging role of micrornas in alzheimer's disease. Frontiers in physiology **6**, 40 (2015)
 10. Liu, F., Zhang, Z., Chen, W., Gu, H., Yan, Q.: Regulatory mechanism of microrna-377 on cdh13 expression in the cell model of alzheimer's disease. Eur. Rev. Med. Pharmacol. Sci **22**, 2801–2808 (2018)
 11. Doxtater, K., Tripathi, M.K., Khan, M.M.: Recent advances on the role of long non-coding rnas in alzheimer's disease. Neural Regeneration Research **15**(12), 2253 (2020)
 12. Patel, H., Dobson, R.J., Newhouse, S.J.: A meta-analysis of alzheimer's disease brain transcriptomic data. Journal of Alzheimer's Disease **68**(4), 1635–1656 (2019)
 13. Xiang, S., Huang, Z., Wang, T., Han, Z., Christina, Y.Y., Ni, D., Huang, K., Zhang, J.: Condition-specific gene co-expression network mining identifies key pathways and regulators in the brain tissue of alzheimer's disease patients. BMC medical genomics **11**(6), 39–51 (2018)
 14. Nangraj, A.S., Selvaraj, G., Kaliamurthi, S., Kaushik, A.C., Cho, W.C., Wei, D.Q.: Integrated ppi-and wgcna-retrieval of hub gene signatures shared between barrett's esophagus and esophageal adenocarcinoma. Frontiers in pharmacology **11**, 881 (2020)
 15. Ren, Z.-H., Shang, G.-P., Wu, K., Hu, C.-Y., Ji, T.: Wgcna co-expression network analysis reveals ilf3-as1 functions as a cerna to regulate ptbp1 expression by sponging mir-29a in gastric cancer. Frontiers in genetics **11**, 39 (2020)

16. Hackstadt, A.J., Hess, A.M.: Filtering for increased power for microarray data analysis. BMC bioinformatics **10**(1), 1–12 (2009)

17. Ayalew, M., Le-Niculescu, H., Levey, D., Jain, N., Changala, B., Patel, S., Winiger, E., Breier, A., Shekhar, A., Amdur, R., *et al.*: Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. Molecular psychiatry **17**(9), 887–905 (2012)

18. Ren, R., Zhang, L., Wang, M.: Specific deletion connexin43 in astrocyte ameliorates cognitive dysfunction in app/ps1 mice. Life sciences **208**, 175–191 (2018)

19. Du, Y., Du, Y., Zhang, Y., Huang, Z., Fu, M., Li, J., Pang, Y., Lei, P., Wang, Y.T., Song, W., *et al.*: Mkp-1 reduces aβ generation and alleviates cognitive impairments in alzheimer's disease models. Signal transduction and targeted therapy **4**(1), 1–12 (2019)

20. Gourmaud, S., Paquet, C., Dumurgier, J., Pace, C., Bouras, C., Gray, F., Laplanche, J.-L., Meurs, E.F., Mouton-Liger, F., Hugon, J.: Increased levels of cerebrospinal fluid jnk3 associated with amyloid pathology: links to cognitive decline. Journal of psychiatry & neuroscience: JPN **40**(3), 151 (2015)

21. Giraldo, E., Lloret, A., Fuchsberger, T., Viña, J.: Aβ and tau toxicities in alzheimer's are linked via oxidative stress-induced p38 activation: protective role of vitamin e. Redox biology **2**, 873–877 (2014)

22. Sclip, A., Tozzi, A., Abaza, A., Cardinetti, D., Colombo, I., Calabresi, P., Salmona, M., Welker, E., Borsello, T.: c-jun n-terminal kinase has a key role in alzheimer disease synaptic dysfunction in vivo. Cell death & disease **5**(1), 1019–1019 (2014)

23. Feld, M., Krawczyk, M.C., Sol Fustinana, M., Blake, M.G., Baratti, C.M., Romano, A., Boccia, M.M.: Decrease of erk/mapk overactivation in prefrontal cortex reverses early memory deficit in a mouse model of alzheimer's disease. Journal of Alzheimer's Disease **40**(1), 69–82 (2014)

24. Yenki, P., Khodagholi, F., Shaerzadeh, F.: Inhibition of phosphorylation of jnk suppresses aβ-induced er stress and upregulates prosurvival mitochondrial proteins in rat hippocampus. Journal of Molecular Neuroscience **49**(2), 262–269 (2013)

25. Wu, M., Fang, K., Wang, W., Lin, W., Guo, L., Wang, J.: Identification of key genes and pathways for alzheimer's disease via combined analysis of genome-wide expression profiling in the hippocampus. Biophysics Reports **5**(2), 98–109 (2019)

26. Kim, S., Kim, H., Park, D., Kim, J., Hong, J., Kim, J.S., Jung, H., Kim, D., Cheong, E., Ko, J., *et al.*: Loss of iqsec3 disrupts gabaergic synapse maintenance and decreases somatostatin expression in the hippocampus. Cell reports **30**(6), 1995–2005 (2020)

27. Kadoyama, K., Matsuura, K., Takano, M., Otani, M., Tomiyama, T., Mori, H., Matsuyama, S.: Proteomic analysis involved with synaptic plasticity improvement by gabaa receptor blockade in hippocampus of a mouse model of alzheimer's disease. Neuroscience research **165**, 61–68 (2021)

28. Schmitz, T.W., Correia, M.M., Ferreira, C.S., Prescot, A.P., Anderson, M.C.: Hippocampal gaba enables inhibitory control over unwanted thoughts. Nature communications **8**(1), 1–12 (2017)

29. Sun, B., Halabisky, B., Zhou, Y., Palop, J.J., Yu, G., Mucke, L., Gan, L.: Imbalance between gabaergic and glutamatergic transmission impairs adult neurogenesis in an animal model of alzheimer's disease. Cell stem cell **5**(6), 624–633 (2009)

30. Guyon, A.: Cxcl12 chemokine and its receptors as major players in the interactions between immune and nervous systems. Frontiers in cellular neuroscience **8**, 65 (2014)

31. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., Dean, J.: A guide to deep learning in healthcare. Nature medicine **25**(1), 24–29 (2019)

32. Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., Telenti, A.: A primer on deep learning in genomics. Nature genetics **51**(1), 12–18 (2019)

33. Tan, M.G., Lee, C., Lee, J.H., Francis, P.T., Williams, R.J., Ramírez, M.J., Chen, C.P., Wong, P.T.-H., Lai, M.K.: Decreased rabphilin 3a immunoreactivity in alzheimer's disease is associated with aβ burden. Neurochemistry international **64**, 29–36 (2014)

34. Flores, J., Noël, A., Foveau, B., Lynham, J., Lecrux, C., LeBlanc, A.C.: Caspase-1 inhibition alleviates cognitive impairment and neuropathology in an alzheimer's disease mouse model. Nature communications **9**(1), 1–14 (2018)

35. Emilsson, L., Saetre, P., Jazin, E.: Low mrna levels of rgs4 splice variants in alzheimer's disease: association between a rare haplotype and decreased mrna expression. Synapse **59**(3), 173–176 (2006)

36. Xiao, M.-F., Xu, D., Craig, M.T., Pelkey, K.A., Chien, C.-C., Shi, Y., Zhang, J., Resnick, S., Pletnikova, O., Salmon, D., *et al.*: Nptx2 and cognitive dysfunction in alzheimer's disease. Elife **6**, 23798 (2017)

37. Feldheim, J., Kessler, A.F., Schmitt, D., Salvador, E., Monoranu, C.M., Feldheim, J., Ernestus, R.-I., Löhr, M., Hagemann, C.: Ribosomal protein s27/metallopanstimulin-1 (rps27) in glioma—a new disease biomarker? Cancers **12**(5), 1085 (2020)

38. Fujita, Y., Maeda, T., Sato, C., Sato, M., Hatakeyama, H., Ota, Y., Iwabuchi, N., Tatesawa, K., Nomura, A., Zou, K., *et al.*: Engulfment of toxic amyloid β-protein in neurons and astrocytes mediated by megf10. Neuroscience **443**, 1–7 (2020)

39. Lan, Y.-L., Zhao, J., Ma, T., Li, S.: The potential roles of aquaporin 4 in alzheimer's disease. Molecular neurobiology **53**(8), 5300–5309 (2016)

40. Conti, E., Andreoni, S., Tomaselli, D., Storti, B., Brovelli, F., Acampora, R., Da Re, F., Appollonio, I., Ferrarese, C., Tremolizzo, L.: Serum dbi and biomarkers of neuroinflammation in alzheimer's disease and delirium. Neurological Sciences, 1–5 (2020)

41. Li, T., Zhu, J.: Entanglement of ccr5 and alzheimer's disease. Frontiers in aging neuroscience **11**, 209 (2019)

42. Salter, M.W., Stevens, B.: Microglia emerge as central players in brain disease. Nature medicine **23**(9), 1018 (2017)

43. Wu, Q., Kong, W., Wang, S.: Peripheral blood biomarkers cxcl12 and tnfrsf13c associate with cerebrospinal fluid biomarkers and infiltrating immune cells in alzheimer disease. Journal of Molecular Neuroscience, 1–10 (2021)

44. Kaczmarczyk, L., Reichenbach, N., Blank, N., Jonson, M., Dittrich, L., Petzold, G., Jackson, W.S.: Insights

into calcium signaling and gene expression in astrocytes uncovered with 129s4 slc1a3-2a-creert2 knock-in mice. bioRxiv (2020)

45. Gertje, E.C., Janelidze, S., van Westen, D., Palmqvist, S., Hansson, O., Mattsson, N.: White matter lesions are associated with csf biomarkers of neuroinflammation in prodromal alzheimer's disease: Biomarkers (non-neuroimaging)/novel biomarkers. Alzheimer's & Dementia **16**, 041795 (2020)

Figures

Figure 1 A flowchart of the whole study. (1) Data collection from AlzData and ADNI; (2) Data preprocessing; (3)DEGs regarded with $|logFC| > 0.1$ and $P < 0.05$; (4)Enrichment of biological process analyzed by DAVID; (5) Use WGCNA to find AD-specific module; (6) Generate a complete list from AD-specific module; (7) CFG based on AD-related evidence to prioritize driver genes of AD.

Figure 2 Enhanced Volcano for illustrating DEGs in all datasets. The red node represents DEGs with $|logFC| > 0.1$ and $P < 0.05$. (A) EC,(B) HP,(C) FC,(D) TC and (E) ADNI. Note:in ADNI dataset, DEGs by counting the frequency of 3 or above out of 10 occurrences.

Figure 3 Venn diagram for DEGs. Venn diagram is used to represent relationships between EC (blue),HP (red), FC (green), TC (yellow) and ADNI (brown).

Figure 4 Venn diagram for KEGG pathway and GO term. venn diagram is used to represent relationships between multiple datasets. (A) KEGG pathway and (B) GO term

Figure 5 Top 20 pathway of KEGG for five datasets (P-value<0.005). (A) EC, (B) HP, (C) FC, (D) TC and (E) ADNI

Figure 6 Module-trait relationships for five datasets. Each row represents different gene co-expression modules, and each column represents different clinical phenotypes. Number represent correlation coefficients and P-values are in parenthesis. Correlation strength is represented by continuous color, with red being positive, blue being negative. (A) EC, (B) HP, (C) FC, (D) TC and (E) ADNI

Figure 7 PPI network. The PPI network contained 5 well-known AD risk genes (green) and 209 (yellow) neighbors, which are from five AD-specific modules.

Tables

Table 1 Brief Descriptions for five datasets. These datasets come from AlzData and ADNI, respectively. Each dataset has multiple features. SDs are given in parentheses.

Dataset	AlzData				Alzheimer's Disease Neuroimaging Initiative
	Entorhinal Cortex	Hippocampus	Frontal Cortex	Temporal Cortex	
Abbreviation	EC	HP	FC	TC	ADNI
No.of.gene	15361	16313	11779	15462	49387
Sample size(Control/AD)	78(39/39)	141(67/74)	232(128/104)	91(39/52)	234(194/40)
Age	80(29.6)	81.7(9.6)	83(9.4)	81(8.7)	74.3(6.5)
Male/Female/Unknown	35/43/0	68/73/0	99/111/22	32/41/18	116/118/0
$A\beta$	NA	NA	NA	NA	1142.9(494.9)
Tau	NA	NA	NA	NA	25.4(11.6)

Table 2 Significant KEGG pathways obtained from DAVID (P-value<0.005). The first column is KEGG pathway ID; the second column is the name of KEGG pathway.

ID	Description	ID	Description
hsa00020	Citrate cycle (TCA cycle)	hsa04966	Collecting duct acid secretion
hsa00190	Oxidative phosphorylation	hsa05010	Alzheimer disease
hsa00260	Glycine, serine and threonine metabolism	hsa05012	Parkinson disease
hsa00620	Pyruvate metabolism	hsa05014	Amyotrophic lateral sclerosis
hsa01200	Carbon metabolism	hsa05016	Huntington disease
hsa01210	2-Oxocarboxylic acid metabolism	hsa05017	Spinocerebellar ataxia
hsa01230	Biosynthesis of amino acids	hsa05020	Prion disease
hsa01522	Endocrine resistance	hsa05022	Pathways of neurodegeneration - multiple diseases
hsa03050	Proteasome	hsa05032	Morphine addiction
hsa04010	MAPK signaling pathway	hsa05033	Nicotine addiction
hsa04070	Phosphatidylinositol signaling system	hsa05110	Vibrio cholerae infection
hsa04071	Sphingolipid signaling pathway	hsa05120	Epithelial cell signaling in Helicobacter pylori infection
hsa04110	Cell cycle	hsa05131	Shigellosis
hsa04120	Ubiquitin mediated proteolysis	hsa05132	Salmonella infection
hsa04137	Mitophagy - animal	hsa05140	Leishmaniasis
hsa04140	Autophagy - animal	hsa05145	Toxoplasmosis
hsa04144	Endocytosis	hsa05152	Tuberculosis
hsa04145	Phagosome	hsa05163	Human cytomegalovirus infection
hsa04152	AMPK signaling pathway	hsa05167	Kaposi sarcoma-associated herpesvirus infection
hsa04211	Longevity regulating pathway	hsa05169	Epstein-Barr virus infection
hsa04218	Cellular senescence	hsa05202	Transcriptional misregulation in cancer
hsa04260	Cardiac muscle contraction	hsa05205	Proteoglycans in cancer
hsa04360	Axon guidance	hsa05212	Pancreatic cancer
hsa04625	C-type lectin receptor signaling pathway	hsa05214	Glioma
hsa04666	Fc gamma R-mediated phagocytosis	hsa05215	Prostate cancer
hsa04721	Synaptic vesicle cycle	hsa05219	Bladder cancer
hsa04722	Neurotrophin signaling pathway	hsa05220	Chronic myeloid leukemia
hsa04723	Retrograde endocannabinoid signaling	hsa05223	Non-small cell lung cancer
hsa04920	Adipocytokine signaling pathway	hsa05225	Hepatocellular carcinoma
hsa04932	Non-alcoholic fatty liver disease	hsa05235	PD-L1 expression and PD-1 checkpoint pathway in cancer
hsa04961	Endocrine and other factor-regulated calcium reabsorption		

Table 3 Significant GO terms obtained from DAVID (P-value<0.005). The first column is GO terms ID; the second column is the name of GO terms.

| ID	Term
GO:0002223	stimulatory C-type lectin receptor signaling pathway
GO:0006888	ER to Golgi vesicle-mediated transport
GO:0048015	phosphatidylinositol-mediated signaling
GO:0038128	ERBB2 signaling pathway
GO:0007249	I-kappaB kinase/NF-kappaB signaling
GO:0006672	ceramide metabolic process
GO:0000165	MAPK cascade
GO:0045944	positive regulation of transcription from RNA polymerase II promoter
GO:0007269	neurotransmitter secretion
GO:0035329	hippo signaling
GO:0006120	mitochondrial electron transport, NADH to ubiquinone
GO:0042776	mitochondrial ATP synthesis coupled proton transport
GO:0070125	mitochondrial translational elongation
GO:0032981	mitochondrial respiratory chain complex I assembly
GO:0007409	axonogenesis
GO:0048812	neuron projection morphogenesis
GO:0043524	negative regulation of neuron apoptotic process
GO:0007268	chemical synaptic transmission
GO:0060078	regulation of postsynaptic membrane potential
GO:0016079	synaptic vesicle exocytosis
GO:0048813	dendrite morphogenesis
GO:0090263	positive regulation of canonical Wnt signaling pathway
GO:0009967	positive regulation of signal transduction
GO:0051932	synaptic transmission, GABAergic
GO:0046034	ATP metabolic process
GO:0070933	histone H4 deacetylation
GO:0007420	brain development
GO:0007417	central nervous system development
GO:0035357	peroxisome proliferator activated receptor signaling pathway
GO:0015986	ATP synthesis coupled proton transport
GO:0040029	regulation of gene expression, epigenetic
GO:0007399	nervous system development
GO:0051966	regulation of synaptic transmission, glutamatergic
GO:0048488	synaptic vesicle endocytosis
GO:0010977	negative regulation of neuron projection development
GO:0060071	Wnt signaling pathway, planar cell polarity pathway
GO:0006521	regulation of cellular amino acid metabolic process
GO:2000310	regulation of N-methyl-D-aspartate selective glutamate receptor activity
GO:0038061	NIK/NF-kappaB signaling
GO:0035418	protein localization to synapse
GO:0060291	long-term synaptic potentiation

Table 4 The number of gene of AD pathology in $A\beta$ and Tau are significant for all datasets.

Dataset	$A\beta$	Tau	$A\beta$ and Tau	In total
EC	21	12	7	40
HP	8	7	6	21
FC	24	10	25	59
TC	50	4	22	76
ADNI	20	14	14	48
In total	123	47	74	244

Table 5 The brief description of 23 potential driver genes of AD. The potential 23 driver genes are prioritized by the CFG method, which is integrated various levels of AD-related data including expression of target gene is regulated by eQLT, GWAS, PPI, Early_DEG, Pathology correlation $A\beta$ and Tau. The last column is list previously published gene-based studies for the identification of patients with AD. (*: P-value<0.05; **: P-value <0.01; ***: P-value<0.001)

Gene	AD-related evidence						CFG	Reference
	eQTL	GWAS	PPI	Early_DEG	Pathology			
					$A\beta$	Tau		
GJA1	2	2	PSEN1,MAPT,APOE	yes	0.388,**	0.131,ns	5	[18]
RPH3A	5	2	-	yes	-0.199,ns	-0.738,**	4	[33]
CASP6	5	0	APP,PSEN1,PSEN2,MAPT	yes	0.482,***	0.738,**	4	[34]
RGS4	1	32	-	yes	-0.419,**	-0.579,*	4	[35]
NPTX2	1	1	-	yes	-0.688,***	-0.783,***	4	[36]
RPS27	1	0	PSEN2	yes	0.503,***	0.662,**	4	[37]
MEGF10	3	8	-	yes	0.559,***	0.120,ns	4	[38]
AQP4	7	4	-	yes	0.800,***	0.275,ns	4	[39]
MYT1L	3	12	-	yes	-0.488,***	-0.583,*	4	NA
CITED2	1	0	APP,PSEN1,APOE	yes	-0.433,**	-0.772,***	4	NA
CDH7	0	1	PSEN1	yes	-0.345,*	-0.691,**	4	NA
DBI	1	1	-	yes	0.780,***	0.718,**	4	[40]
PELI2	2	0	PSEN2	yes	0.591,***	-0.107,ns	4	NA
AVEN	1	1	-	yes	0.525,***	0.008,ns	4	NA
SLA	1	0	PSEN1,MAPT	yes	0.114,ns	0.662,**	4	NA
DCN	8	0	APP,PSEN1,MAPT,APOE	yes	-0.416,**	0.546,*	4	NA
CCR5	1	0	APP	yes	0.769,***	0.616,*	4	[41]
IRF5	1	0	APP,PSEN1,PSEN2,MAPT,APOE	yes	0.879,***	0.839,***	4	[42]
CXCL12	1	0	APP,PSEN2,MAPT,APOE	yes	0.432,**	-0.069,ns	4	[43]
CREM	1	0	PSEN1,MAPT,APOE	yes	-0.439,**	-0.396,ns	4	NA
EHHADH	14	0	MAPT,APOE	yes	0.438,**	-0.022,ns	4	NA
SLC1A3	7	1	-	yes	0.651,***	0.494,ns	4	[44]
IL15	2	18	-	yes	0.623,***	0.685,**	4	[45]

Figures
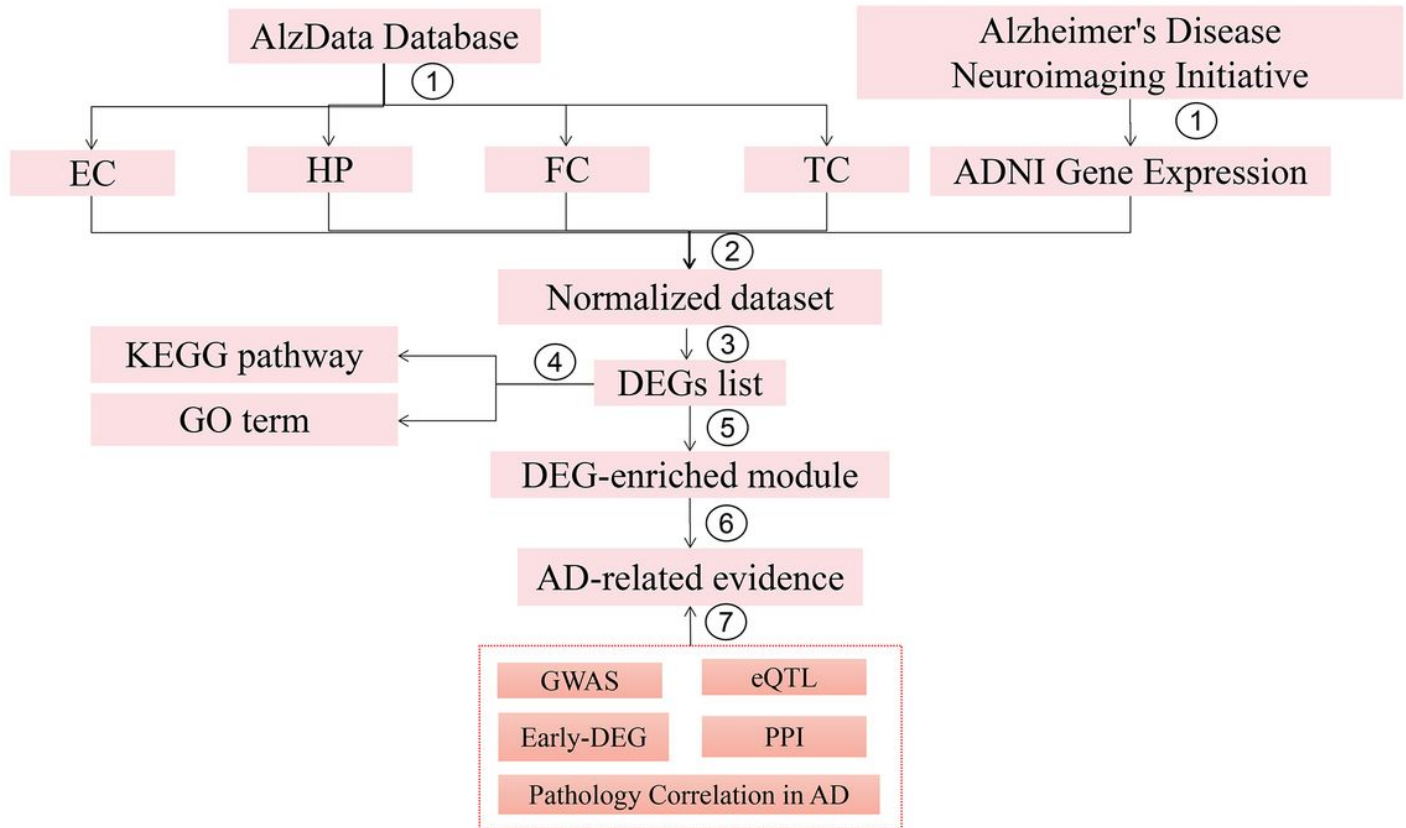
Figure 1

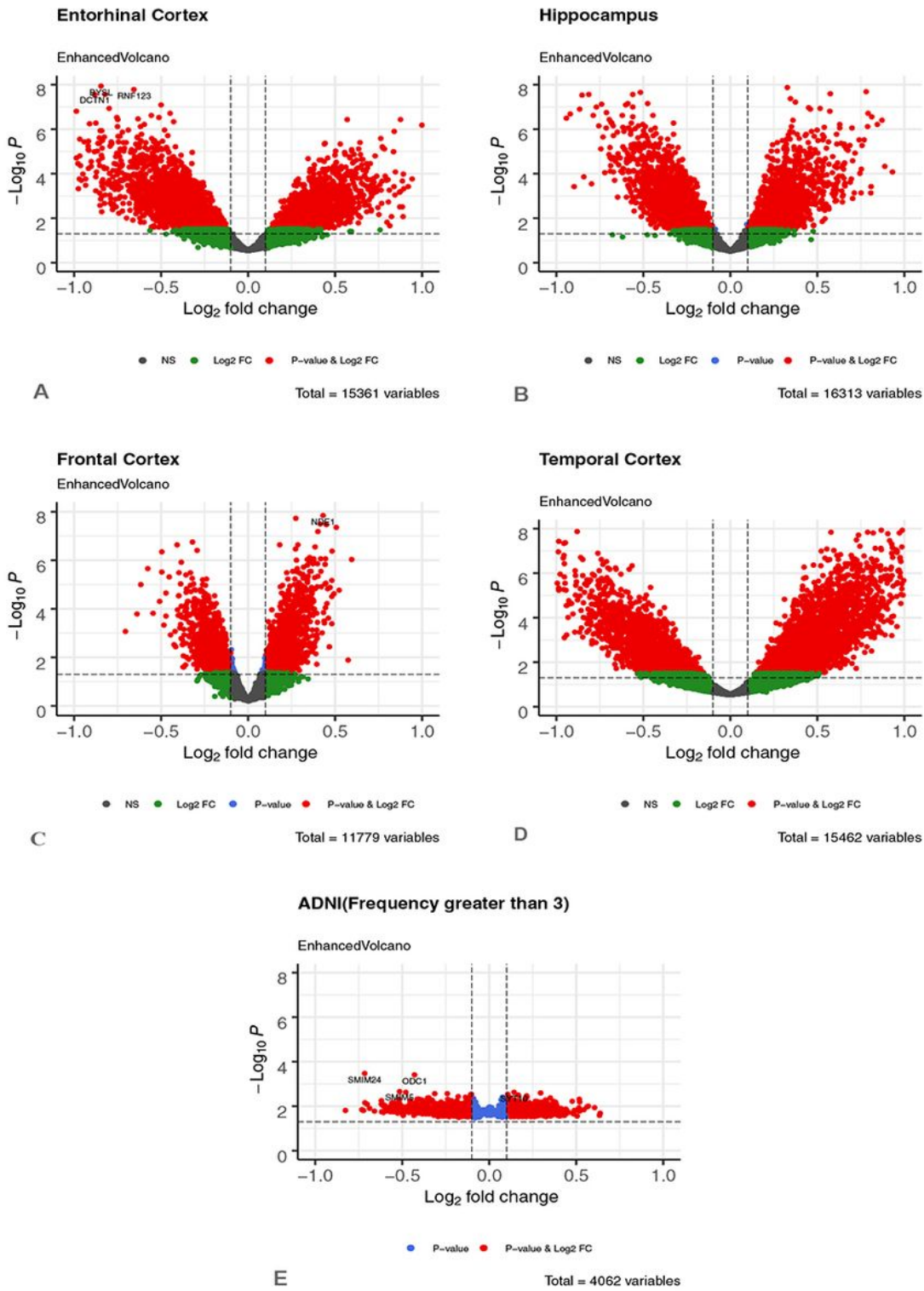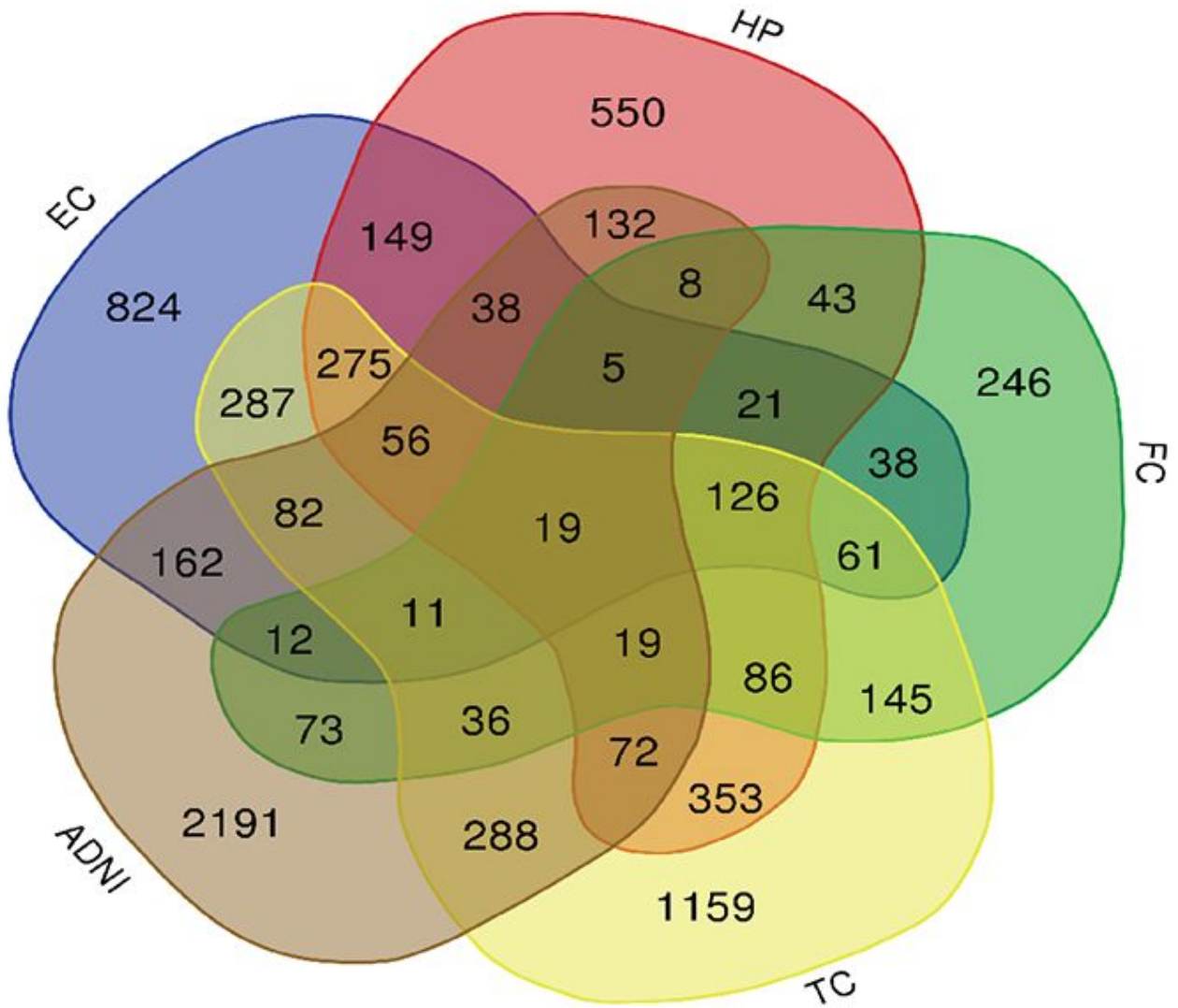"Please see the Manuscript PDF file for the complete figure caption".

Figure 2

"Please see the Manuscript PDF file for the complete figure caption".

Figure 3

Venn diagram for DEGs. Venn diagram is used to represent relationships between EC (blue),HP (red), FC (green), TC (yellow) and ADNI (brown).
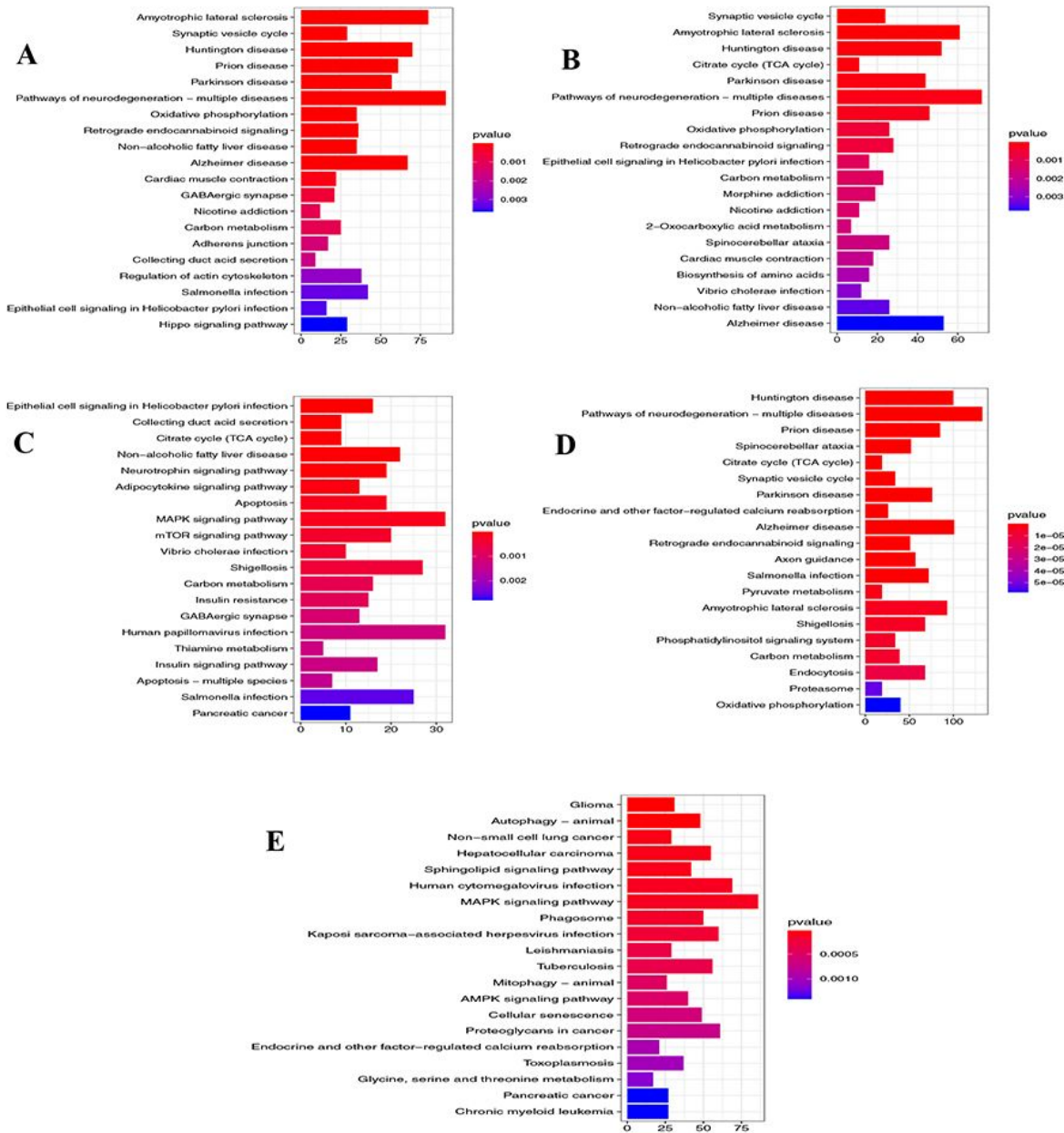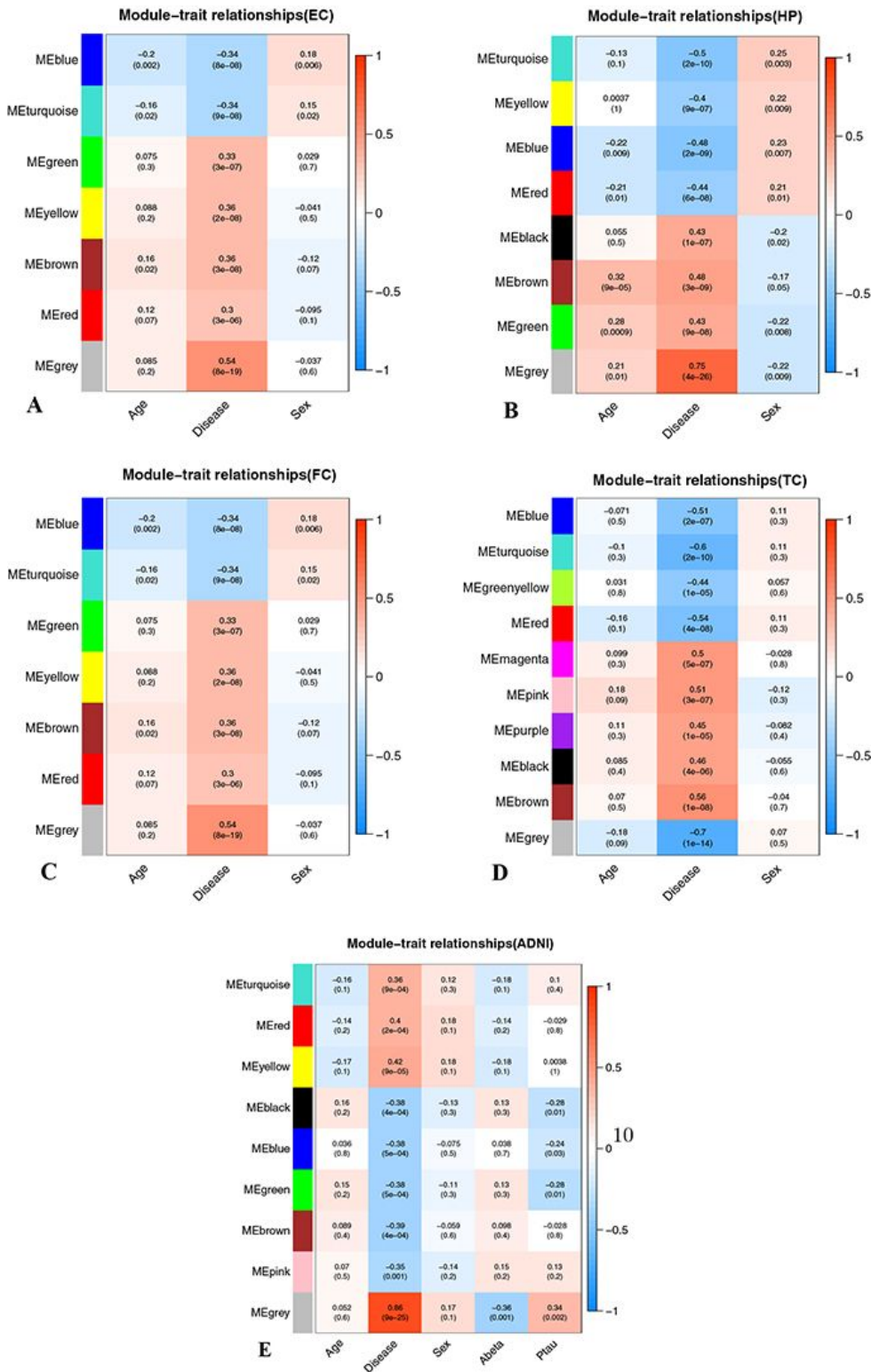
Figure 4

Venn diagram for KEGG pathway and GO term. venn diagram is used to represent relationships between multiple datasets. (A) KEGG pathway and (B) GO term
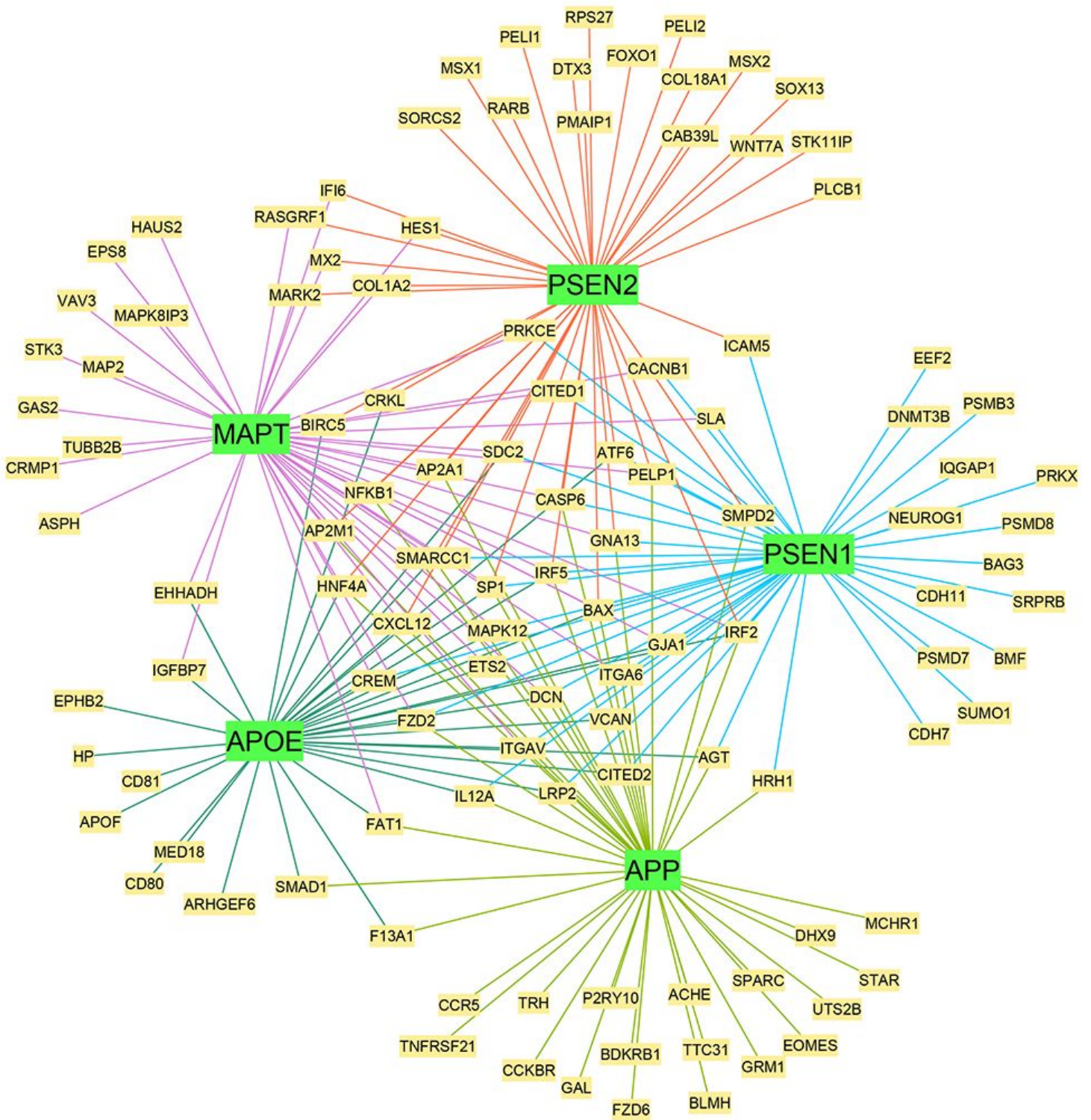
Figure 5

Top 20 pathway of KEGG for five datasets (P-value<0.005). (A) EC, (B) HP, (C) FC, (D) TC and (E) ADNI

Figure 6

Module-trait relationships for five datasets.Each row represents different gene co-expression modules, and each column represents different clinical phenotypes. Number represent correlation coeficients and P-values are in parenthesis. Correlation strength is represented by continuous color, with red being positive, blue being negative. (A) EC, (B) HP, (C) FC, (D) TC and (E) ADNI

Figure 7

PPI network. The PPI network contained 5 well-known AD risk genes (green) and 209 (yellow) neighbors, which are from five AD-specific modules.