

# ARHGDIB Plays a Novel Role in the Braak Stages of Alzheimer's Diseases via the Immune Response Mediated by Microglia

Keping Chai (✉ [ckpzjyy@126.com](mailto:ckpzjyy@126.com))

Zhejiang Hospital <https://orcid.org/0000-0002-6307-9102>

Jiwei Liang

Huazhong University of Science and Technology

Xiaolin Zhang

Huazhong University of Science and Technology

Huaqian Gu

Zhejiang Hospital

Panlong Cao

Zhejiang Hospital

Weiping Ye

Zhejiang Hospital

Huitao Tang

Zhejiang Hospital

Jingjing Liu

Zhejiang Hospital

Shufang Chen

Zhejiang Hospital

Feng Wan

Huazhong University of Science and Technology

Gang Logan Liu

Huazhong University of Science and Technology

Daojiang Shen

Zhejiang Hospital

---

## Research

**Keywords:** Braak stages, Random Forest, WGCNA, ssGSEA, Neurodegeneration

**Posted Date:** May 6th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-474315/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background:

Aberrant neurofibrillary tangles (NFT) deposits is a mainly character of Alzheimer's diseases (AD). A lot of evidence shows that at different stages of AD, the distribution region of NFT in the brain is also different. For example, the entorhinal cortex (EC) is the area where NFT deposits occurs firstly in AD. Here, we use machine learning and weighted gene co-expression network (WGCNA) to explore the relationship between cortex regions of NFT deposits and Braak stages, and reveal potential biomarkers and therapeutic targets of AD.

**Objective:** To explore the differences in gene expression patterns in multiple brain regions with distinct Braak stages and find the hub genes by using WGCNA and machine learning.

**Methods:** The transcriptional profiling data of the human entorhinal cortex, temporal cortex (TC) and frontal cortex (FC) derived from individuals ranging from Braak stages 0 to VI was obtained from the GEO database (GSE131617) in NCBI. For WGCNA analysis, first, we detected consensus modules of different brain regions and illustrated the relationships of these modules. The second, we conducted single sample gene set enrichment analysis (ssGSEA) to obtain the enrichment scores of samples and screened the enrichment genes through the best subset regression and the random forest algorithm. Third, we analyzed the relationships between consensus modules (EC-FC, EC-TC, and FC-TC) and ssGSEA enrichment scores. Next, the overlapping genes between differentially expressed genes (DEG, between Braak stage 0 and Braak stage I-VI) and genes of interest in the module were discovered. Metascape analysis were conducted to determine the function of overlapping genes, and Random Forest classifier was preformed to obtain the most significant genes from the overlapping genes. The disclosed significant genes were finally identified through network analysis.

**Results:** Preservation ((PreservEC, FC) is 0.91, (PreservEC, TC) is 0.95, (PreservTC, FC) is 0.9) of consensus modules and connectivity of different brain regions illustrated very high correlation preservation of all pairs of eigengenes across the two networks. We found that the oxidative damage pathway plays a vital role in classifying the Braak stages via ssGSEA, random forest and best subset algorithm, the imp is 0.57. Through WGCNA analysis, the black module is found highly positively correlated with oxidative damage, which is involved in immune response mainly. Through step by step filtering of the module genes by overlapping with DEGs and Random Forest classifier analysis, we found that *LYN*, *CD68*, *LAPTM5*, *IFI30*, *PI3KAP1*, *HCK* and *ARHGDIB* were co-expressed and highly correlated with oxidative damage and immune response.

**Conclusion:** The co-expression network has strong similarities in different brain regions undergoing AD. Molecules such as *IFI30*, *LYN*, *CD68*, *PTPRC*, *HCK*, *LAPTM5*, *FCERIG*, and *ARHGDIB* play an important role in the early stages of AD through the inflammatory response mediated by microglia.

## Background

Aberrant deposits of neurofibrillary tangles (NFT), the main character of Alzheimer's disease (AD)[1], which is highly related to cognitive impairment. A lot of evidence shows that at different stages of AD, the distribution region of NFT in the brain is also different. For example, the entorhinal cortex (EC) is the area where NFT deposits occurs firstly in AD[2]. However, the pathological mechanism of its formation is still unclear. Several hypotheses, such as oxidative damage, oxidative stress, insulin resist, apoE, neuroinflammation and other theories were established[3, 4]. Exploring the gene expression patterns of different brain regions, especially EC, may better help understand the mechanism of NFT formation.

Weighted gene co-expression network analysis (WGCNA) is a biology algorithm used to describe the correlation of clinical characters and gene expression based on the microarray data[5]. WGCNA can be used for clustering gene with highly correlated expression, for relating the modules to phenotypes to get the most phenotypic trait-related module, and for summarizing these co-expressed gene clusters by identification of the module eigengene or hub genes. Random forest (RF) is a more advanced machine learning algorithm based on decision tree[6]. Like other decision trees, random forests can be used for both regression and classification.

In this study, we performed ssGSEA, machine learning and WGCNA analysis on publicly accessible transcriptome data obtained from human cortex of individuals at different Braak stages. Evaluating the ssGSEA results, we found that the oxidative damage pathway play a vital role in classifying the Braak stages via random forest and best subset algorithm, the imp is 0.57. Through calculating the correlation coefficients between the modules and oxidative damage pathway, we obtained a module of interest. Next, we disclosed the overlapping genes between differentially expressed genes (DEG, between Braak stage 0 and Braak stage I- VI) and genes of interest in the module. Using these overlapping genes, we conducted metascape analysis and further identified the central players within the module through network analysis. We concluded that *C1QA*, *C1QB*, *LYN*, *CD68*, *LAPTM5*, *IFI30*, *PI3KAP1*, *HCK* and *ARHGDIB* genes are significantly associated with oxidative damage and immune response, which may be novel biomarkers involved in AD.

## Methods And Materials

### Data acquisition and preprocessing

The data used in this paper was obtained from the GEO database in NCBI (Gene Expression Omnibus, <http://www.ncbi.nlm.gov/geo>), and the data entry number is GSE131617[7]. The platform is Affymetrix Human Exon 1.0 ST Array [transcript (gene) version, HuEx-1\_0-st]. Gene expression in the cortex of Braak stages 0 (N = 13), I-II (N = 20), III-IV (N = 19) and V-VI (N = 19) was detected. The normalized data was downloaded and the expression matrix was obtained, and data filtering was performed before WGCNA analysis. For data filtering, first, neuropathological diagnosis between minimal senile change and AD was performed. The second, the gene type of *APOE* is 3\*3. 47 samples in the dataset were kept. Probe without corresponding annotation information were removed. There were about 13629 genes in the dataset.

### Single sample gene set enrichment analysis (ssGSEA)

ssGSEA is an implementation method proposed for single sample GSEA[8,9]. The difference between GSEA and ssGSEA is that ssGSEA no needs to prepare expression matrix file. The functions of gene set were acquired from Molecular Signatures Database (MSigDB) according to the describes in the review, including aging, insulin receptor pathways, oxidative stress, oxidative damage, NFT, Nicotine activity on dopaminergic neuron, etc. Performances of pathway in the gene set were quantified by ssGSEA algorithm (R package 'gsva') based on transcriptome profiling data and pathway gene sets.

### **Application of best subset regression to find the best subset of ssGSEA pathway**

The entorhinal cortical samples were grouped into individuals Braak stage 0 and Braak stages I- VI. Inputting the ssGSEA scores into best subset regression model via leaps package to predict which group the samples belong to, the best number of features as the input for next analysis.

### **Application of Random Forest algorithm to find the most important pathway and genes related to Braak stages**

The entorhinal cortical samples were grouped into individuals of Braak stage 0 or individuals of Braak stages I- VI. Inputting the overlapping genes counts and ssGSEA enrich scores into random forest classifier model via Boruta package to predict which group the samples belong to and the most important overlapping genes and to identify ssGSEA pathway for the most accurate model for grouping.

### **Construction of weighted gene co-expression network and identification of significant modules**

Data was processed using R 3.4.2 software. To ensure that the results of network construction are reliable, abnormal samples were removed. Then, the weighted gene co-expression network was constructed by WGCNA package based on R 3.4.2. First, the Pearson correlation coefficient was calculated to assess the similarity of the gene expression profiles. The second, the correlation coefficients between genes were weighted by a power function to obtain a scale-free network. A gene module is a cluster of densely interconnected genes in terms of co-expression. Then, hierarchical cluster was used to identify gene modules and different modules were represented by different colors. Dynamic treecut method was used to identify different modules, the adjacency matrix was converted to a topology overlay matrix (TOM) and modules were detected by cluster analysis during module selection.

### **Correlation analysis of gene modules with clinical phenotype**

To detect the associations of modules and clinical phenotype (ssGSEA scores), firstly, the clinical phenotype data and gene expression data were correlated using the match function. Secondly, the associations of the module eigengene (ME) and the clinicals phenotype were calculated by Pearson's correlation analysis. Modules showing significant association to oxidative damage pathway were obtained. At last, to further confirm the modules with significant correlation to oxidative damage, the correlation coefficient between the module membership (gene expression level) with gene significance (GS, for assessing the association of genes with phenotypes) was calculated using the `labelHeatmap` function, and the *p* values were obtained.

## **Finding the overlapping genes between the differentially expressed genes (DEG, between Braak stage 0 and Braak stages I- VI) and genes of interest in the module verified by WGCNA**

The entorhinal cortical samples were grouped into individuals at Braak stages 0 and individuals at Braak stages I- VI and Limma packages were performed to find the DEG[10,11]. Next, the overlapping genes between downregulated DEG and genes of interest in the module were discovered by using online Venn tools (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

## **Metascape analyses, identification of Hub genes and protein-protein interaction (PPI) analysis**

For the obtained overlapping genes, functional enrichment of Gene Ontology (GO) and KEGG pathways analyses were performed using Metascape (<https://metascape.org>)[12]. Log *P* between -23 and -8 was considered to be significant enrichment. These enrichment results were also analyzed using Cytoscape for identification of important pathway[13]. The identified hub genes were further confirmed and analyzed using String network constructed by the online database String(<http://string-db.org>)[14].

## **Exploring the cellular distribution of the identified genes**

By using the Cell marker database (<http://biocc.hrbmu.edu.cn/CellMarker/search.jsp>) , the cellular distribution of the identified important genes was further explored.

# **Results**

## **The identification consensus modules across different cortical regions**

Before WGCNA, the genes detected in GSE131617 were filtered according to the filtering procedure described in Method, and 13629 genes were obtained. Then the microarray data of 47 samples in each cortical region were read by R for Hierarchical clustering (**Supplementary Fig. 1a**). The consensus network of scale independence and mean connectivity analysis showed that when the weighted value equals to 14, the average degree of connectivity was close to 0, and scale independence was greater than 0.9, so the weighted value was set to 14 (**Supplementary Fig. 1b**). WGCNA was performed to identify consensus modules. Comparison between EC set-specific modules and EC-FC consensus modules of the global co-expression network indicates that most EC modules are preserved in FC (**Figure.1.a**). The strong overlap of the corresponding gene modules showed the similarity of cluster pattern in EC and FC region.

**Figure.1.b-g, Figure.S1c, Figure.S2 and Figure.S3** showed that the overall preservation of the three networks is positive correlation. The mean density of the three networks exceeds 0.9 in all 3 cortical regions, demonstrating that the overall structures of the co-expression networks are similar for the three cortical regions. These results indicated that the differences of these cortical regions may exist in the particular genes within the consensus network.

**ssGSEA functional enrichment analyses and key pathway identification and validation help to find the module of interest verified in WGCNA analysis in EC.**

In above results, we found overall structures of the co-expression networks are similar for the three cortical regions. In addition, An abundance of studies have shown that in the Braak stages I–II, aberrant deposits of NFT first appears in the entorhinal cortex, which is significant for finding the potential biomarkers and therapeutic targets of AD.

To explore the signaling pathways most related to Alzheimer’s disease, first, the ssGSEA analysis was performed (**Figure.2.a**). The genset of pathway related to Alzheimer’s disease can be seen in **TableS1**. The second, best subset regression was conducted to identify the representative subset (**Figure.2.b, c**). From the results, we can see that the feature number of best subsets is 8, and GO-NFT, HP-NFT, oxidative damage and axon degeneration pathway are saved in the best subset. Next, we performed the random forest algorithm based on the sklearn and boruta packages to analyze the best subset of data to find the most important features, as shown in **Figure.2.d** and **TableS2**, oxidative damage pathway was found to be the most important feature.

To identify modules which is most significantly associated with oxidative damage pathway in EC, the Pearson’s correlation coefficient between the module and oxidative damage was calculated. The highest positive association in the module trait relationship was found between black module and oxidative damage score ( $\text{cor} = 0.88, p < 0.001$ , **Figure.2e**), we also found that the black module has a high correlation with the aging and clearance pathway ( $\text{cor} = 0.69, 0.56, p < 0.001$ , **Figure.2e**). Thus, the black module was selected as module of interest in subsequent analysis.

### Identifying hub genes in black module

First, to find the DEG between Braak stage 0 and Braak stages I- VI, the EC samples were grouped into individuals at Braak stage 0 and Braak stages I- VI and Limma packages were performed. About 10% of the genes were significantly changed ( $p < 0.05$ , **Figure. 3a**). Next, we performed overlap analysis between DEGs and Top30 genes in black module by online veen tool, we found 26 genes which were in DEGs also in black module (**Figure. 3b-d**). These genes highly related to oxidative damage, suggesting that they might play an important role in oxidative damage -AD.

### Identifying hub genes functional annotation

The above identified overlapping genes were subjected to GO functional and KEGG pathway enrichment analysis. Biological processes of overlapping genes were found to focus on regulation of inflammatory response and leukocyte degranulation. Molecular functions of overlapping genes were found to focus on IgE binding, non-membrance spanning protein tyrosine kinase activity and phosphotransferase activity(**Figure. 4, Figure.S4** ).

### Identification of the most significant genes and network construction

To identify the most important genes related to oxidative damage, the overlapping genes were further filtered by RF classification. Gene counts were input into RF classifier model, and the unimportant genes, such as *C1QA*, *C1QB*, *CTSC*, *SLC2A5*, *UCP2* and others were removed (**Figure. 5a,TableS3**). To ascertain

the significance of genes and analyze the network in the corresponding modules, the PPI maps were constructed via String (**Figure. 5b**). Hub genes in network, including *PTPRC*, *LYN*, *LAPYM5*, *HCK*, *IFI30*, *ARHGDIB* and *PIK3AP1* was constructed. In the cell marker database, we found that the distribution of above genes in brain cells is very similar, mainly in microglia cell(**Figure. 5c**).

## Discussion

NFT is the major pathological character of neurodegenerative diseases, such as PD (Parkinson's disease)/AD[15]. Exploring the mechanism of NFT formation is extremely important for discovering the therapeutic targets in these diseases. In this study, we performed WGCNA, ssGSEA and machine learning analysis on the dataset GSE131617 which includes 71 samples from individuals at Braak stages between 0 and I- VI; such data from multiple samples based on the different brain regions (EC,FC,TC) is a good candidate for WGCNA analysis. First, consensus modules between different brain regions were constructed, 7 consensus modules were identified between EC and FC. **Figure.1.b-g, FigureS1c, FigureS2 and FigureS3** showed that the overall preservation of the three networks is positive correlation. The mean density of the three networks exceeds 0.9 in all 3 consensus modules, demonstrating that the overall structures of the co-expression networks are similar for the 3 cortical regions. However, the purple, pink, greenyellow and magenta module of EC were not recognized in the consensus module (EC, FC), indicating that the difference between the two regions is related to these modules. Furthermore, the black and red modules in EC that are most related to oxidative damage and clearance pathway have not been recognized in the consensus module identified by EC and TC (**FigureS2, Figure.2.e**). These showed that TC is quite different from EC in the signal pathway of oxidative damage and clearance.

Many studies have shown that the EC is the region where NFT deposits occurs firstly[16]. Therefore, studying the changes in the transcription level of the EC in the neurodegenerative diseases is extremely important for discovering early biomarkers and therapeutic targets of these diseases. In this study, when we performed ssGSEA and random forest analysis on the dataset of EC samples, we found that the unexpected oxidative damage signaling pathway is most important when distinguishing between Braak stage 0 and Braak stages I- VI rather than the signaling pathway related to NFT(**Figure.2.a-d**). This indicates that among the important basis of Braak stages, the formation of NFT is more likely due to changes in the expression level of genes related to oxidative stress pathway, rather than the NFT signaling pathway. When we analyzed the overlapping genes in the black module which is most related to oxidative damage and the DEG, we found that these genes are not only related to oxidative damage, but also related to immune response and microglia-mediated inflammation (**Figure.2e, Figure.3, Figure.4**). In order to identify genes that are most intensively related to Braak stages, we further used one of the machine learning algorithms, Random Forest, and inputted the expression matrix of the overlapping 26 genes as feature values into the model for training, and finally screened out 9 key genes(**Figure.5.a, TableS3**). When analyzing these 9 molecules, we found that most of them are expressed in microglia (**Figure.5.c**), which further indicated that microglia plays an important role in the Braak stages (0 VS I- VI).

It has been reported that activated microglia can induce the formation of NFT[17], several hypotheses can explain how the activated microglia mediates the formation of NFT, such as complement pathway, IL-CDK5 pathway, exosome secretion, etc[18–21]. However, it needs further discussion that how these molecules such as *LYN*, *HCK* and *PTPRC*, which distributed in the microglia, promote the formation of NFT. *LYN* and *HCK* as Non-receptor tyrosine-protein kinase can combine with NLRP3 which involved in the phosphorylation of tau and the formation of NFT to promote the release of IL1B from microglia[22–25]. In the co-expression network, *PTPRC* and *LAPTM5* were identified as hub genes too. *PTPRC* is not only an important regulator of T cell and B cell antigen receptor signal transduction, but also an enzyme that dephosphorylates *LYN*. It has been reported that *LAPTM5* can not only regulate the production of pro-inflammatory cytokines in macrophages, but also regulate the antigen receptor signal transduction of T cell and B cell[26]. And there is a lot of data showing that *LAPTM5* and *PTPRC* are not only co-expressed in AD/PD (**Figure.5.b**), but also in systemic lupus erythematosus, lung cancer and other diseases[27–29]. This indicated that *LAPTM5*, *PTPRC* may play a similar role in the phosphorylation of *LYN*. Moreover, in this study, we found that the decrease of the expression of these co-expressed genes at Braak stage I- VI which is negatively correlated with the degree of NFT needs further discussion. It has been reported that the expression of *LYN* in activated microglia is less than that of homeostasis microglia[30]. This indicate that *LYN* plays a role in activated microglia, and the decrease of *PTPRC* and *LAPTM5* may lead to an increase of phosphorylated *LYN*, so that it can promote the release of inflammatory factors.

In this study, we also found that *IFI30* and *FCER1G* in the co-expression network are also distributed in microglia (**Figure.5.b,c**). It has been reported that both of them are highly expressed in microglia around , which may imply that the two of them are involved in the function of clearance (**Figure.2.e**). However, in this study we found that their expression in Braak stage I- VI decreased. How their reduction in microglia promotes the formation of NFT still needs further study.

To our surprise, *ARHGDI1* was found mainly co-expressed with *LAPTM5* and *PTPRC* in the co-expression network. Its related pathways are involved in GPCR signaling pathway, apoptosis and survival Caspase cascade[32]. Through the network analysis (**Figure.5.b,c**), we speculated that it may has similar functions with *LAPTM5* and *PTPRC*. The decrease of the expression of *ARHGDI1* may also play a role in the formation of NFT. Further studies are still needed to reveal the function of *ARHGDI1* in microglia.

## Conclusions

Last but not least, through machine learning and WGCNA on microarray data from EC, FC and TC, we not only revealed the similarities and differences in the co-expression network between the above 3 brain regions, but also uncovered some novel molecules, such as *HCK*, *LAPTM5*, *IFI30* and *ARHGDI1*, which mediate the formation of NFT in the microglia of EC.

## Abbreviations

AD : Alzheimer's diseases; NFT : Neurofibrillary tangles; EC : Entorhinal cortex; TC : Temporal cortex; FC : Frontal cortex; ssGSEA : Single sample gene set- enrichment analysis; WGCNA : Weighted gene co-expression network; RF : Random forest; DEG : Differentially expressed genes; GO : Gene Ontology; PPI : Protein-protein interaction;

## Declarations

### Ethics approval and consent to participate

Not applicable

### Competing interests

The authors declare that they have no competing interests.

### Consent for publication

Not applicable.

### Acknowledgements

We acknowledge GEO database for providing their platforms and contributors for uploading their meaningful datasets.

### Authors' contributions

KPC contributed to the study design, performed the experiments, and contributed to the writing of the manuscript. JWJ contributed to the study design and to the writing of the manuscript. XLZ, HQG, PLC, WPY, HTT, JJL and SFC conducted the experiments. FW, GLL, DJS provided critical devices and contributed to the study design. The author(s) read and approved the final manuscript.

### Funding

Not applicable

### Availability of data and materials

The datasets generated and/or analysed during the current study are available in the [GEO] repository, [<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GSE131617>]

## References

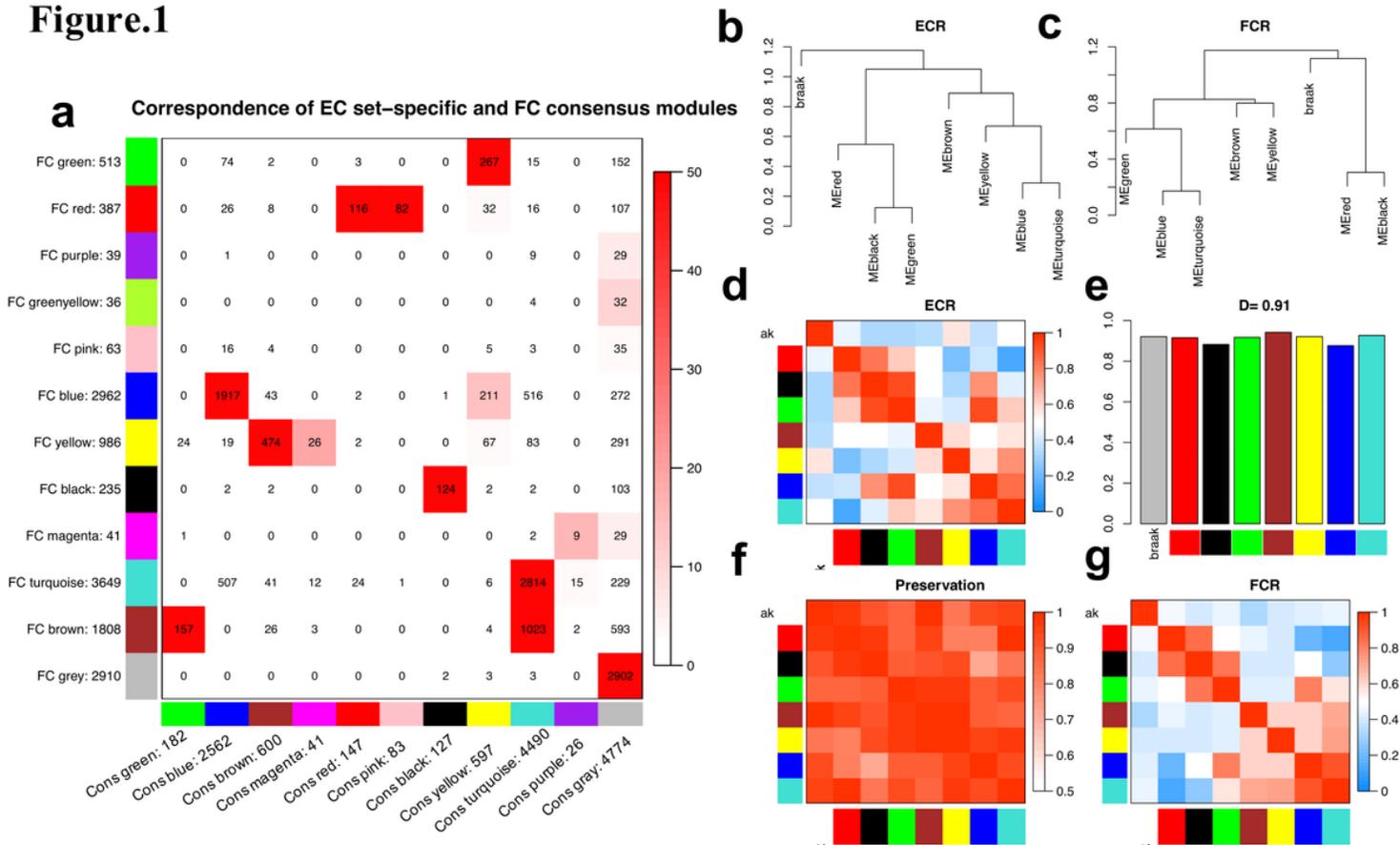
1. Dickson DW. Neurodegenerative diseases with cytoskeletal pathology: a biochemical classification. *Ann Neurol.* 1997;42:541–4.

2. Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 1991;82:239–59.
3. Solomon A, Mangialasche F, Richard E, Andrieu S, Bennett DA, Breteler M, et al. Advances in the prevention of Alzheimer's disease and dementia. *J Intern Med.* 2014;275:229–50.
4. Nakamura A, Kaneko N, Villemagne VL, Kato T, Doecke J, Doré V, et al. High performance plasma amyloid- $\beta$  biomarkers for Alzheimer's disease. *Nature.* 2018;554:249–54.
5. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
6. Sarica A, Cerasa A, Quattrone A. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. *Front Aging Neurosci.* 2017;9:329.
7. Kikuchi M, Sekiya M, Hara N, Miyashita A, Kuwano R, Ikeuchi T, et al. Disruption of a RAC1-centred network is associated with Alzheimer's disease pathology and causes age-dependent neurodegeneration. *Hum Mol Genet.* 2020;29:817–33.
8. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature.* 2009;462:108–12.
9. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102:15545–50.
10. Diboun I, Wernisch L, Orengo CA, Koltzenburg M. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. *BMC Genom.* 2006;7:252.
11. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47.
12. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019;10:1523.
13. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010;38:W214–20.
14. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 2017;45:D362–8.
15. Wang Y, Mandelkow E. Tau in physiology and pathology. *Nat Rev Neurosci.* 2016;17:5–21.
16. Cui S, Sun H, Gu X, Lv E, Zhang Y, Dong P, et al. Gene expression profiling analysis of locus coeruleus in idiopathic Parkinson's disease by bioinformatics. *Neurol Sci.* 2015;6.
17. Fan Z, Brooks DJ, Okello A, Edison P. An early and late peak in microglial activation in Alzheimer's disease trajectory. *Brain.* 2017;140:792–803.
18. Vogels T, Murgoci A-N, Hromádka T. Intersection of pathological tau and microglia at the synapse. *Acta Neuropathol Commun.* 2019;7:109.

19. Asai H, Ikezu S, Tsunoda S, Medalla M, Luebke J, Haydar T, et al. Depletion of microglia and inhibition of exosome synthesis halt tau propagation. *Nat Neurosci.* 2015;18:1584–93.
20. Quintanilla RA, Orellana DI, González-Billault C, Maccioni RB. Interleukin-6 induces Alzheimer-type phosphorylation of tau protein by deregulating the cdk5/p35 pathway. *Exp Cell Res.* 2004;295:245–57.
21. Saha P, Sen N. Tauopathy. A common mechanism for neurodegeneration and brain aging. *Mech Ageing Dev.* 2019;178:72–9.
22. Kong X, Liao Y, Zhou L, Zhang Y, Cheng J, Yuan Z, et al. Hematopoietic Cell Kinase (HCK) Is Essential for NLRP3 Inflammasome Activation and Lipopolysaccharide-Induced Inflammatory Response In Vivo. *Front Pharmacol.* 2020;11:581011.
23. Fitzer-Attas CJ, Lowry M, Crowley MT, Finn AJ, Meng F, DeFranco AL, et al. Fcγ receptor-mediated phagocytosis in macrophages lacking the Src family tyrosine kinases Hck, Fgr, and Lyn. *J Exp Med.* 2000;191:669–82.
24. Gwon Y, Kim S-H, Kim HT, Kam T-I, Park J, Lim B, et al. Amelioration of amyloid β-FcγRIIb neurotoxicity and tau pathologies by targeting LYN. *FASEB J.* 2019;33:4300–13.
25. Jevtic S, Sengar AS, Salter MW, McLaurin J. The role of the immune system in Alzheimer disease: Etiology and treatment. *Ageing Res Rev.* 2017;40:84–94.
26. Zouali M. Transcriptional and metabolic pre-B cell receptor-mediated checkpoints: implications for autoimmune diseases. *Mol Immunol.* 2014;62:315–20.
27. Salih DA, Bayram S, Guelfi S, Reynolds RH, Shoai M, Ryten M, et al. Genetic variability in response to amyloid beta deposition influences Alzheimer’s disease risk. *Brain Commun.* 2019;1:fcz022.
28. Zhang L, Zhang M, Chen X, He Y, Chen R, Zhang J, et al. Identification of the tubulointerstitial infiltrating immune cell landscape and immune marker related molecular patterns in lupus nephritis using bioinformatics analysis. *Annals of Translational Medicine.* 2020;8:18.
29. Zhang T, Yang H, Sun B, Yao F. Four hub genes regulate tumor infiltration by immune cells, antitumor immunity in the tumor microenvironment, and survival outcomes in lung squamous cell carcinoma patients. *Aging.* 2021;13:3819–42.
30. Sierksma A, Lu A, Mancuso R, Fattorelli N, Thrupp N, Salta E, et al. Novel Alzheimer risk genes determine the microglia response to amyloid-β but not to TAU pathology. *EMBO Molecular Medicine.* 2020;18.
31. Satoh J-I, Kino Y, Yanaizu M, Ishida T, Saito Y. Microglia express gamma-interferon-inducible lysosomal thiol reductase in the brains of Alzheimer’s disease and Nasu-Hakola disease. *Intractable Rare Dis Res.* 2018;7:251–7.
32. Kardol-Hoefnagel T, van Logtestijn SALM, Otten HG. A Review on the Function and Regulation of ARHGDI2/RhoGDI2 Expression Including the Hypothetical Role of ARHGDI2/RhoGDI2 Autoantibodies in Kidney Transplantation. *Transplant Direct.* 2020;6:e548.

## Figures

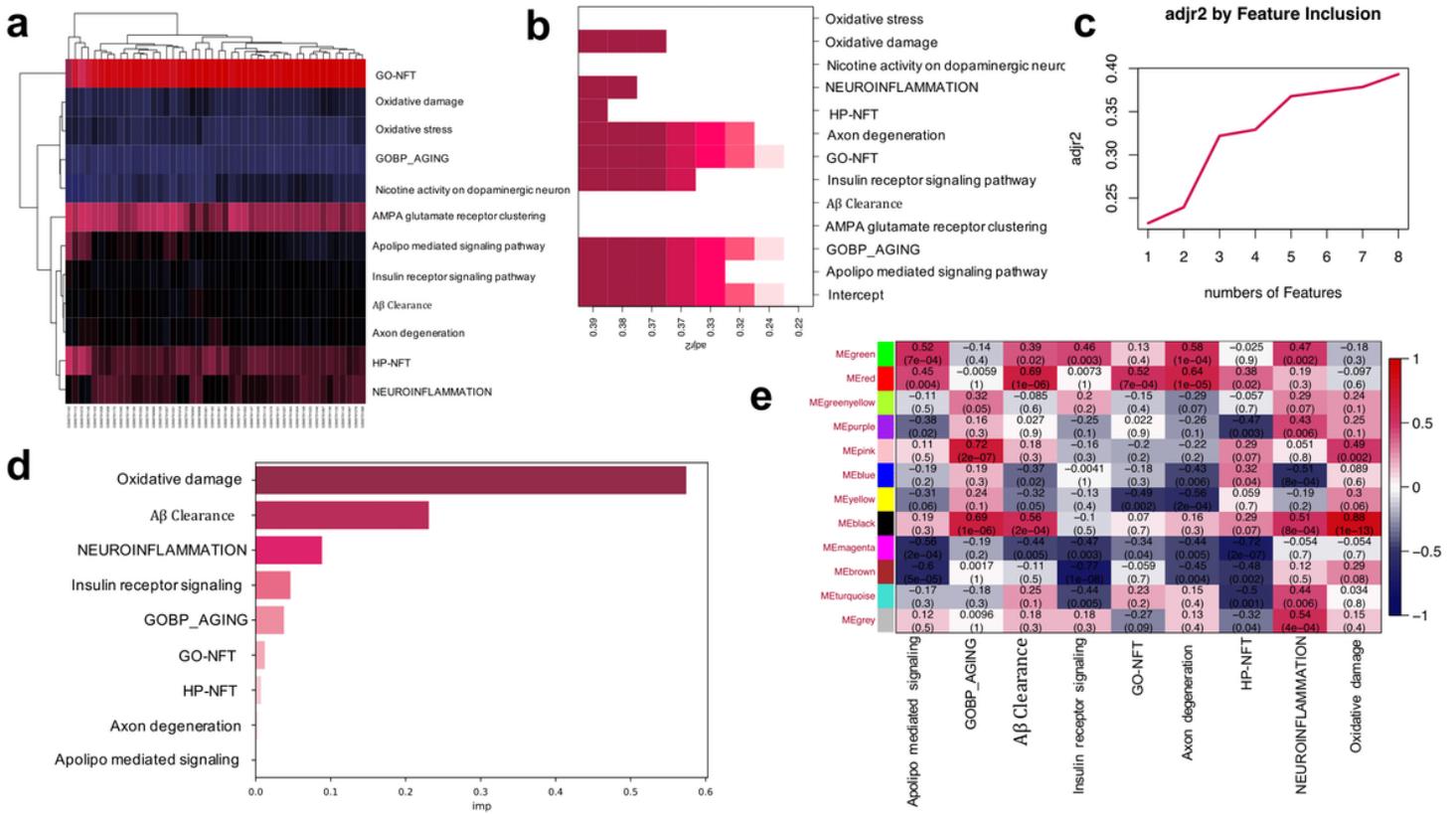
**Figure.1**



**Figure 1**

a) Comparison between EC set-specific modules and EC-FC consensus modules of the global co-expression network. The numbers in the table represent genes that shared between EC modules and consensus modules. The color code of the table is  $-\log(p)$ , where  $p$  is the  $p$  value of Fisher's exact test of the overlap of the two modules. The darker the red, the more pronounced the overlap. b-c) Clustering dendrograms of consensus module eigengenes for identifying meta-modules show presence of similar major branching pattern in EC and FC eigengene network. d, g) The heatmap shows the eigengene adjacencies in EC and FC eigengene networks. Each row and column correspond to an eigengene tagged by consensus module color. Within each heatmap, red represents high adjacency (positive correlation) and blue represents low adjacency (negative correlation) as represented by the color legend. e) Bar plot shows the preservation degree of each consensus eigengene as the height of the bar (y-axis) where each colored bar corresponds to the eigengene of the associated consensus module. The high-density value  $D$  (Preserve EC, FC) = 0.91 indicates the high overall preservation between the EC and FC networks. f) Adjacency heatmap of the preservation network between EC and FC consensus eigengene networks. The saturation of the red color indicates correlation preservation of EC and FC module eigengenes.

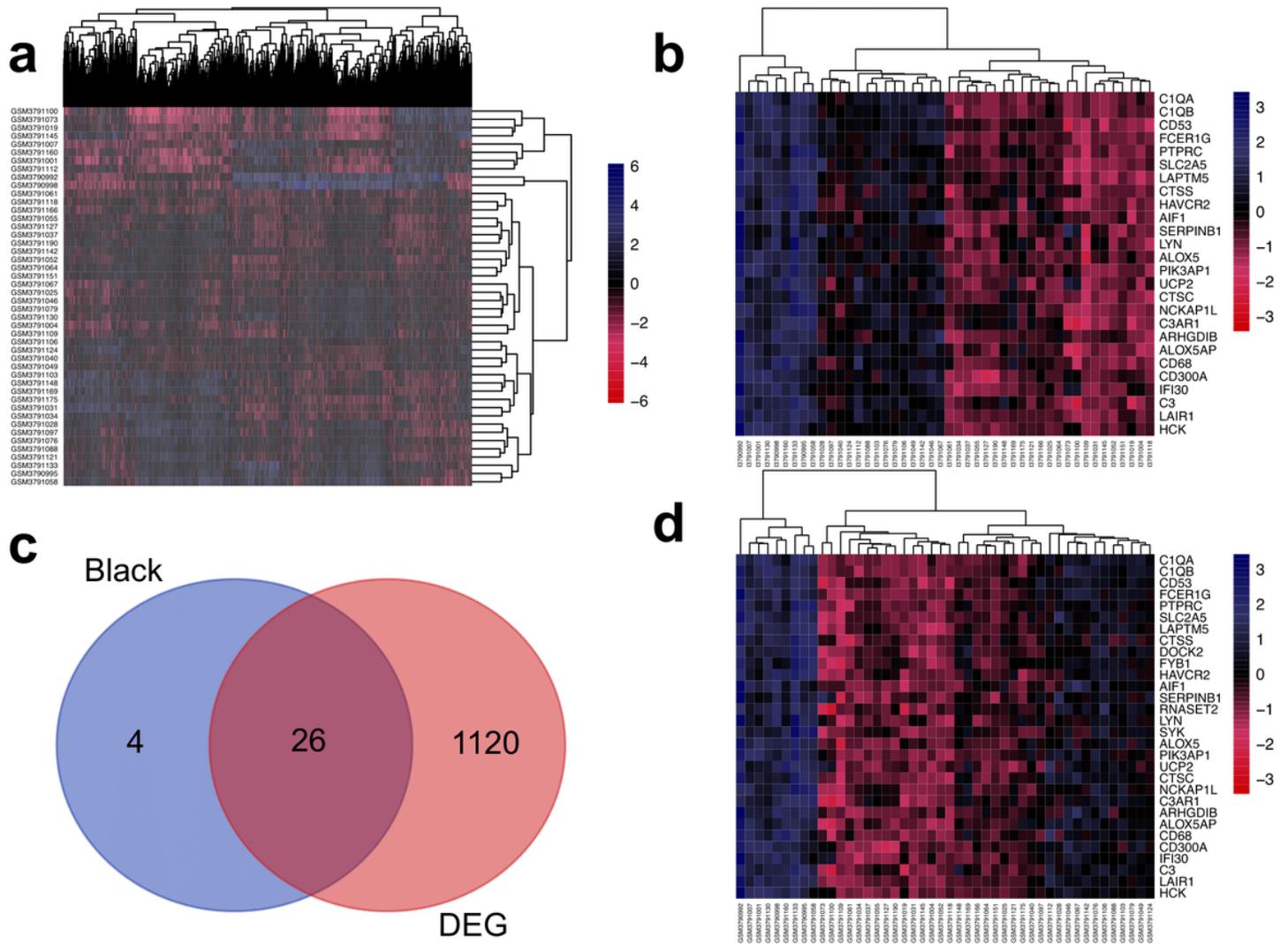
**Figure.2**



**Figure 2**

ssGSEA and WGCNA analysis of the EC data. a) Heatmap shows the ssGSEA scores of different gene set in correspond samples. b) The heatmap shows the adj R2 of Best subset regression result of each ssGSEA pathway. c) The plot shows the adj R2 of the number of features in Best subset regression model. d) The bar plot shows the importance of each ssGSEA pathway by RF model. e) Pearson correlation coefficient between the pathway and module eigengenes, numbers in brackets indicate the corresponding p values.

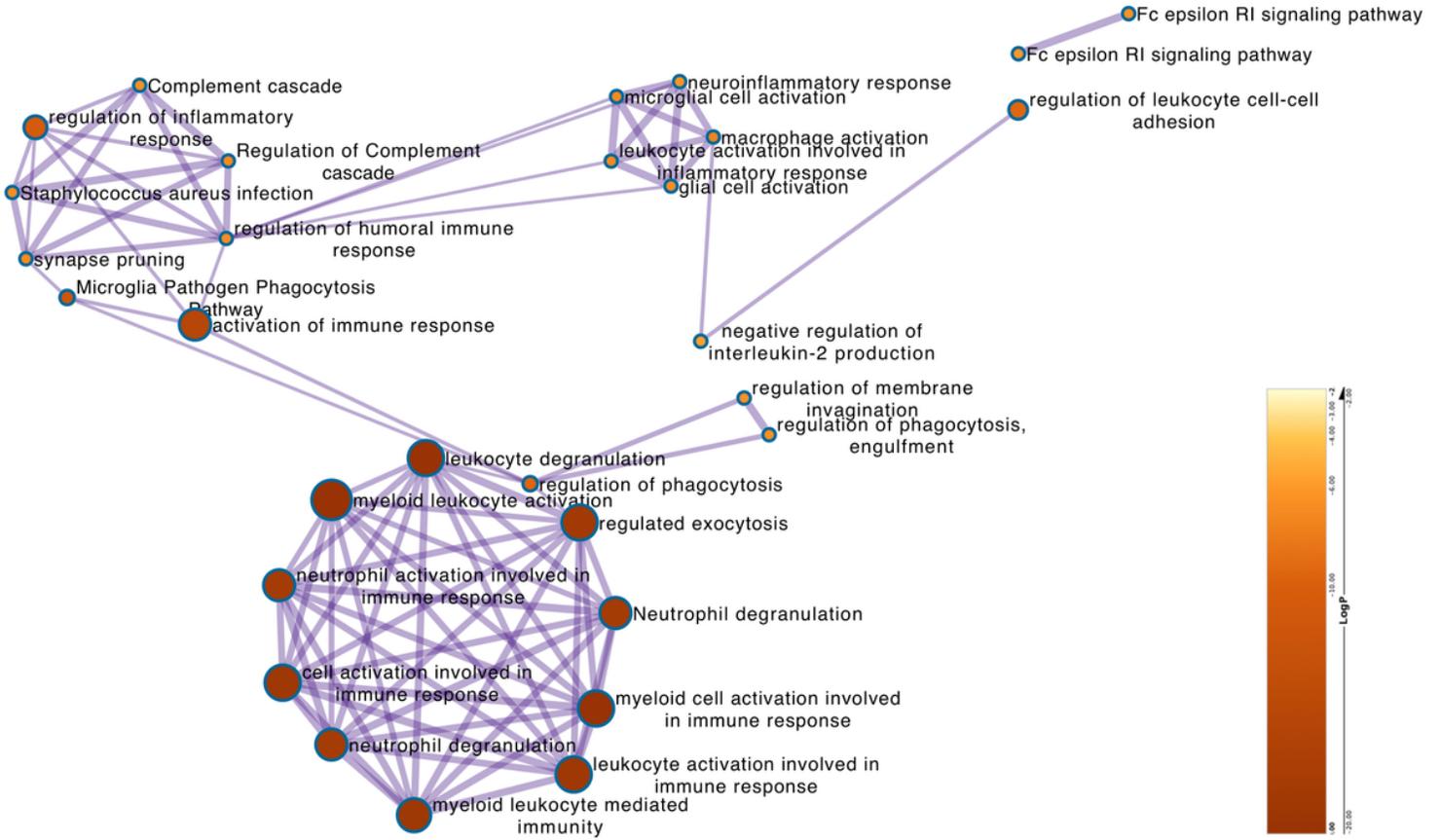
**Figure.3**



**Figure 3**

Identifying the overlapping genes between downregulated DEGs in aged group and genes in black module. a) Heatmap of the expression of DEGs. b) Heatmap of the Top30 gene expression in black module . c) Using veen tools to find the overlap genes between downregulated genes in DEGs and genes in black module. d) Heatmap showing the expression of the overlapping genes in different samples.

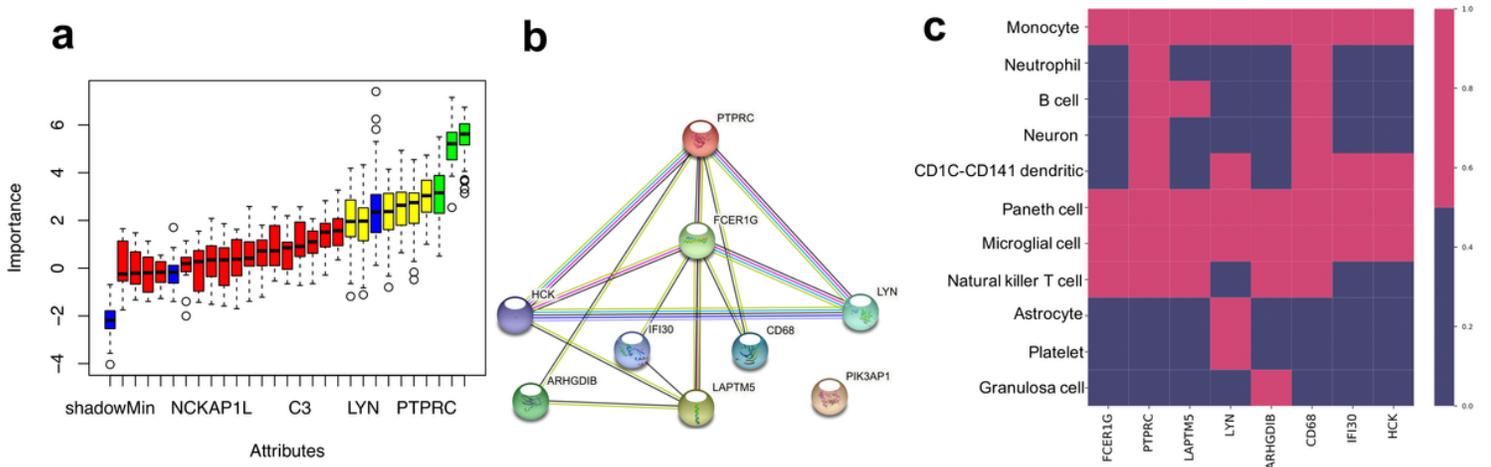
**Figure.4**



**Figure 4**

The Metascape of the overlapping genes. The network show the GO terms that the log P (-23 - -8) correlates with the significance of the enrichment.

**Figure.5**



**Figure 5**

Identifying the most important genes via RF and the cellular distribution of the important genes in the brain. a) Random Forest algorithm result. The blue box plot corresponds to the minimum, average, and maximum Z scores of a color attribute. The red, yellow, and green boxes represent the Z scores of rejected, tentative, and confirmed genes, respectively. b) The PPI network of important genes via String. c) The heatmap shows the selected genes distribution in different cell types.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FS1.pdf](#)
- [FS2.pdf](#)
- [FS3.pdf](#)
- [FigSuppliemnet.docx](#)
- [FigureS4.pdf](#)
- [TableS1.csv](#)
- [TableS2.csv](#)
- [TableS3.csv](#)