



OPEN

## Molecular subtype identification of cerebral ischemic stroke based on ferroptosis-related genes

Yufeng Wang<sup>1,3</sup>✉, Xinjuan Xu<sup>1,3</sup>, Xinjun Shui<sup>1</sup>, Ruilin Ren<sup>1</sup> & Yu Liu<sup>2</sup>

Cerebral ischemic stroke (CIS) has the characteristics of a high incidence, disability, and mortality rate. Here, we aimed to explore the potential pathogenic mechanisms of ferroptosis-related genes (FRGs) in CIS. Three microarray datasets from the Gene Expression Omnibus (GEO) database were utilized to analyze differentially expressed genes (DEGs) between CIS and normal controls. FRGs were obtained from a literature report and the FerrDb database. Weighted gene co-expression network analysis (WGCNA) and protein–protein interaction (PPI) network were used to screen hub genes. The receiver operating characteristic (ROC) curve was adopted to evaluate the diagnostic value of key genes in CIS, followed by analysis of immune microenvironment, transcription factor (TF) regulatory network, drug prediction, and molecular docking. In total, 128 CIS samples were divided into 2 subgroups after clustering analysis. Compared with cluster A, 1560 DEGs were identified in cluster B. After the construction of the WGCNA and PPI network, 5 hub genes, including MAPK3, WAS, DNAJC5, PRKCD, and GRB2, were identified for CIS. Interestingly, MAPK3 was a FRG that differentially expressed between cluster A and cluster B. The expression levels of 5 hub genes were all specifically highly in cluster A subtype. It is noted that neutrophils were the most positively correlated with all 5 real hub genes. PRKCD was one of the target genes of FASUDIL. In conclusion, five real hub genes were identified as potential diagnostic markers, which can distinguish the two subtypes well.

Cerebral ischemic stroke (CIS) is a symptom of the focal or complete neurological deficit caused by cerebral blood supply dysfunction, hypoxia and ischemia. CIS has the characteristics of a high incidence and mortality rate, which endangers human physical and mental health<sup>1,2</sup>. The lack of blood supply and reperfusion lead to the interruption of energy and redox homeostasis in the cerebra, caused the metabolic disorder of lipid oxides in cells, and then leads to abnormal metabolism under the catalysis of iron-dependent phospholipid peroxidation. The aggregation of these products destroys the redox balance and the homeostasis of iron metabolism in cells, triggering regulatory cell necrosis, which is called ferroptosis<sup>3,4</sup>. Ferroptosis is significantly different from apoptosis and cell necrosis at the morphological, biological, and gene levels<sup>5</sup>. In recent years, a series of research reports demonstrated that ferroptosis is closely related to CIS<sup>6–8</sup>.

Ferroptosis is caused by a redox imbalance between the production of oxidants and antioxidants, which can be precisely regulated at multiple levels<sup>9</sup>. Increasing evidence supported ferroptosis as a mechanism of acute ischemic stroke in vivo<sup>10</sup>. During the cerebral ischemic episode, iron overload is found in the damaged area of the cerebra<sup>11,12</sup>. The imbalance of iron metabolism, including its overload and deficiency, has been shown to impact ischemic stroke outcomes<sup>13</sup>. The regulation of ferroptosis can intervene the process of stroke in animal models, suggesting that ferroptosis is a novel potential target for the treatment of stroke<sup>14</sup>. In addition, ferroptosis-related gene (FRG) signature had been reported to have the potential as ischemic stroke diagnostic markers<sup>15</sup>.

In weighted gene co-expression network analysis (WGCNA), the genes with similar expression patterns were clustered into a module, and the correlation between the module and phenotype was analyzed<sup>16,17</sup>. WGCNA can be used to find genes with key functions, help identify potential mechanisms, and explore candidate biomarkers<sup>18,19</sup>. Herein, we first clustered samples based on expression patterns of FRGs to obtain different subtypes of CIS. Based on the differential expressed genes (DEGs) between different subtypes, we used WGCNA to screen the key modules of diseases and get the hub genes related to ferroptosis. Later, we also constructed a multi-factor regulatory network for the screened key hub genes to analyze their expression in CIS and subclusters.

<sup>1</sup>Department of Neurosurgery, Shanxi Cardiovascular Hospital, No.18, Yifen Street, Taiyuan City 030024, Shanxi Province, China. <sup>2</sup>Department of Surgical, Peking University First Hospital Taiyuan, Taiyuan, China. <sup>3</sup>These authors contributed equally: Yufeng Wang and Xinjuan Xu. ✉email: wangyufeng1979@126.com

## Methods

### Datasets acquisition and preprocessing

We used the keywords "stroke" and "homo sapiens" to filter the gene expression profile data in the Gene Expression Omnibus (GEO) database. Then the corresponding dataset was screened using the following criteria: (1) sample size is  $\geq 5$ ; (2) normal controls were included in the dataset. The exclusion criteria were as follows: (1) studies performed at the cell line or animal level; (2) single cohort study; (3) repeated or overlapping studies. Three datasets of CIS, GSE16561, GSE22255, and GSE58294, were included in our study. The GSE16561 includes peripheral blood samples of 39 CIS patients and 24 healthy controls, GSE22255 includes peripheral blood samples of 20 CIS patients and 20 normal controls, and GSE58294 includes peripheral blood samples of 69 patients with CIS and 23 normal controls. After downloading the submitted gene expression matrix file, the gene expression profile were annotated using the GPL platform annotation file, and the gene probes were converted into gene symbols. The combat function in the R package "SVA"<sup>20</sup> was used to remove batch effects.

### Acquisition of FRGs

The list of FRGs were obtained from the FerrDb (<http://www.zhounan.org/ferrdb/index.html>) database, a database dedicated to ferroptosis regulators and ferroptosis-disease associations<sup>21</sup>, supplemented by a previous literature<sup>22</sup>. A total of 267 FRGs were obtained finally. Among which, 216 FRGs were detected in three CIS datasets.

### Identification of DEGs among different groups

After data preprocessing of the above three datasets, the "limma" (3.36.5) package was used to obtain the DEGs between CIS and normal controls<sup>23</sup>. The screened standard was false discovery rate (FDR)  $< 0.05$  and  $|\log_2 \text{fold change (FC)}| > 0.24$ . Differentially expressed FRGs were further screened and displayed in the heat map. Samples were grouped using the "ConsensusClusterPlus" (1.44.0) package in R<sup>25</sup>, and the parameters were set as  $\text{maxK} = 5$ ,  $\text{reps} = 1000$ ,  $\text{pItem} = 0.8$ ,  $\text{pFeature} = 1$ ,  $\text{clusterAlg} = \text{"pam"}$ . The optimal  $k$  was determined according to cumulative distribution function (CDF) and area under CDF curve. The DEGs among different CIS subtypes were screened based on  $\text{FDR} < 0.05$  and  $|\log_2 \text{FC}| > 0.3$ . The principal component analysis (PCA) using the `prcomp` function of R<sup>26</sup> was performed to evaluate the accuracy of grouping.

### Functional analysis and gene set enrichment analysis (GSEA)

To analyze the DEGs involved in biochemical processes and pathways in the development of CIS, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment<sup>27,28</sup> and Gene Ontology (GO) analysis were performed using Database for Annotation, Visualization, and Integrated Discovery (DAVID, <https://david.ncifcrf.gov/>)<sup>29,30</sup> with  $p < 0.05$ . Top 20 enrichment results were visualized by the "GOplot" of R. In addition, GSEA was conducted to find out enriched gene sets with hallmark gene sets (`h.all.v2022.1.Hs.symbols.gmt`) downloaded from Molecular Signatures Database (MSigDB). Enrichment results with  $\text{FDR} < 0.05$  were considered statistically significant.

### WGCNA

The R-package "WGCNA" (1.67)<sup>31,32</sup> was used to construct a scale-free gene co-expression network. The "hclust" function was utilized to cluster the sample data to delete outliers. The "pickSoftThreshold" function was used to calculate an appropriate soft threshold power. The height was set to 0.90. Then, the adjacency matrix was calculated and transformed into the topological overlap matrix (TOM) and the corresponding dissimilarity matrix ( $1 - \text{TOM}$ ). Genes with similar expression patterns were clustered together and divided into modules with default parameters according to the "cutreeDynamic" function. The identified modules were truncated and merged with a height of 0.25. The model characteristic gene (ME) is defined as the first principal component in each module, which can summarize the expression pattern of all genes in the module. To determine the key modules most related to CIS, the "moduleEigengenes" function was applied to calculate the ME of each module, and the correlation between ME and CIS was analyzed by the "Pearson" method. The module with the highest correlation with CIS was selected as the hub module. The candidate hub genes were selected based on the module connectivity and clinical trait relationship. The gene significance (GS) and module membership (MM) were calculated to select candidate hub genes with  $\text{GS} > 0.5$  and  $\text{MM} > 0.8$ .

### Screening hub genes

All genes in the hub module were imported into the STRING (<https://cn.string-db.org/>) database<sup>33</sup> to build the protein-protein interaction (PPI) network and visualized by Cytoscape (3.9.0)<sup>34</sup>. The CytoHubba, plug-in in Cytoscape, was used to screen the central genes in the PPI network (gene of the top 40-degree). By overlapping the candidate hub genes in the hub module and the central genes in the PPI network, the real hub genes were obtained. The Pearson correlation among the real hub genes was calculated and the expression of the real hub genes between two groups was pairwise compared. The receiver operating characteristic (ROC) curve analysis was adopted to evaluate the accuracy of the real hub genes in distinguishing different CIS subtypes, and normal controls and CIS patients.

### Immune microenvironment analysis

CIBERSORT<sup>35</sup> was used to calculate the proportion of 22 immune cells in each sample based on gene expression patterns. The difference in immune cells among different subgroups were calculated and visualized using a bar plot. The percentage of different immune cells in each sample was shown in the column accumulation chart. In

addition, Pearson correlation coefficient method was used to calculate the correlation between real hub genes and differential immune cells.

### Construction of transcription factor (TF) regulatory network, drug prediction and molecular docking

In order to explore the effect of TFs regulating gene expression on the occurrence and development of CIS, the regulation network of TF-hub gene was built based on TRRUST (<https://www.grnpedia.org/trrust/>) database<sup>36</sup>. In addition, real hub gene related drugs were identified from the DGIdb (<https://dgidb.org/>) database<sup>37</sup>, followed by molecular docking by AutoDock Vina 1.1.2 software<sup>38</sup>. The docking results were visualized through PyMol 2.5<sup>39</sup>.

### Statistical analysis

R software (v3.5.3) was used for all statistical analysis. The "limma" (3.36.5) package was used to obtain the DEGs. Consensus clustering analysis was performed using the "ConsensusClusterPlus" (1.44.0) package in R. The PCA using the `prcomp` function of R was performed to evaluate the accuracy of grouping. The R-package "WGCNA" (1.67) was used to construct a scale-free gene co-expression network and select candidate hub genes with  $GS > 0.5$  and  $MM > 0.8$ . CIBERSORT was used to perform immune infiltration analysis. Unless otherwise stated, a  $p$ -value  $< 0.05$  was deemed statistically significant.

## Results

### DEGs in the CIS samples

The combined dataset included 67 healthy controls and 128 CIS samples. A total of 13,521 genes in the three datasets were intersected after removing the batch effect. A total of 1995 DEGs were identified in the CIS group, of which 1518 genes were up-regulated, and 477 genes were down-regulated (Fig. 1A,B). Totally, 46 differentially expressed FRGs were identified, of which 41 were up-regulated and 5 were down-regulated (Fig. 1C).

### DEGs between two CIS subtypes

According to the CDF value and CDF area change,  $k = 2$  exhibited the best unsupervised clustering competence (Fig. S1). A total of 128 CIS samples were divided into 2 subgroups after cluster analysis: cluster A and B (Table 1). Compared with cluster A, 1560 DEGs were identified in cluster B (1004 up-regulated and 556 down-regulated, Fig. 2A,B). Among which, 34 differentially expressed FRGs were screened, of which 17 genes were up-regulated and 17 genes were down-regulated (Fig. 2C). Result of PCA was displayed in Fig. 2D. GO enrichment results showed that most of the DEGs were distributed in cytosol, nucleoplasm, extracellular exosome, and nucleus, and involved in regulation of transcription and immune response, protein binding, and other biological processes (Fig. 3A–C). KEGG enrichment analysis showed that DEGs were enriched in NF-kappa B signaling pathway, HIF-1 signaling pathway, TGF-beta signaling pathway, and metabolic pathways (Fig. 3D). GSEA results indicated that the pathways of MYOGENESIS, COAGULATION, APICAL JUNCTION, HEME METABOLISM, HEDGEHOG SIGNALING, KRAS SIGNALING DN, REACTIVE OXYGEN SPECIES PATHWAY, P53 PATHWAY, HYPOXIA, UV RESPONSE UP, INFLAMMATORY RESPONSE, and XENOBIOTIC METABOLISM were more active in cluster A, while MYC TARGETS V1, E2F TARGETS, OXIDATIVE PHOSPHORYLATION, G2M CHECKPOINT, UNFOLDED PROTEIN RESPONSE, MTORC1 SIGNALING, ANDROGEN RESPONSE, DNA REPAIR, PROTEIN SECRETION, and MYC TARGETS V2 were more active in cluster B. The pathways with  $FDR < 0.01$  were displayed in Fig. S2.

### Construction of WGCNA

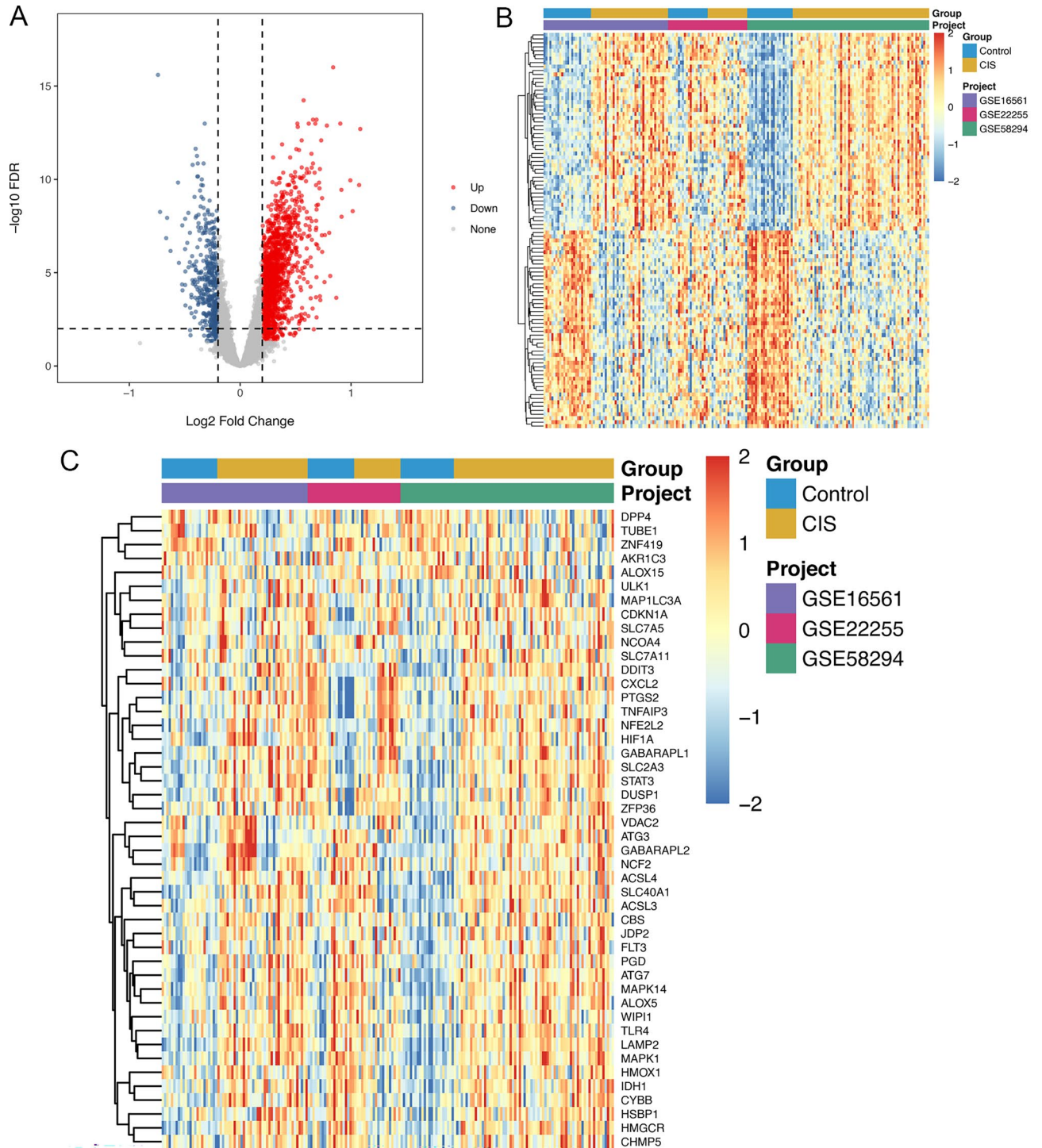
WGCNA was utilized to analyze 1560 DEGs from 128 samples to identify CIS-related genes. The power  $\beta = 10$  was selected to ensure a scale-free topology (Fig. 4A). Then, 6 modules were determined (Fig. 4B,C). Among the six modules, the black module had the highest correlation with CIS with Pearson  $r = 0.76$  ( $P = 1e-25$ , Fig. 4D). Therefore, the black module was considered the hub module, which including 508 genes.

### Identification of the real hub genes

A total of 508 genes in the black module were imported into the STRING database to construct a PPI network, which includes 429 nodes and 1674 edges (Fig. 5A), and the top 40 ranked degree was defined as the central node genes. Fifty-two genes were selected as candidate hub genes in the black module according to  $GS > 0.5$  and  $MM > 0.8$  (Fig. 5B). By overlapping the candidate hub genes and the central node genes, five intersection genes were considered as the real hub genes, including MAPK3, WAS, DNJC5, PRKCD, and GRB2. The correlation analysis showed that there were strong correlations among the five real hub genes (Fig. 5C). MAPK3 had a strong correlation with the other 4 hub genes, especially with WAS (coefficient = 0.78). Among which, MAPK3 was a FRG differentially expressed between cluster A and cluster B. These 5 real hub genes all exhibited specifically highly expressed in cluster A (Fig. 5D–H).

### ROC analysis of real hub genes

To further evaluate the accuracy of these 5 hub genes in distinguishing cluster A and cluster B in patients with CIS, ROC of the 5 hub genes were performed. The results (Fig. 6A) showed that the AUC values of all 5 genes were relatively high, showing good sensitivity and specificity. We also evaluated the diagnostic values of 5 genes in distinguishing CIS patients from normal controls (Fig. 6B). The results demonstrate that the AUC value of WAS gene was relatively high, with good sensitivity and specificity.

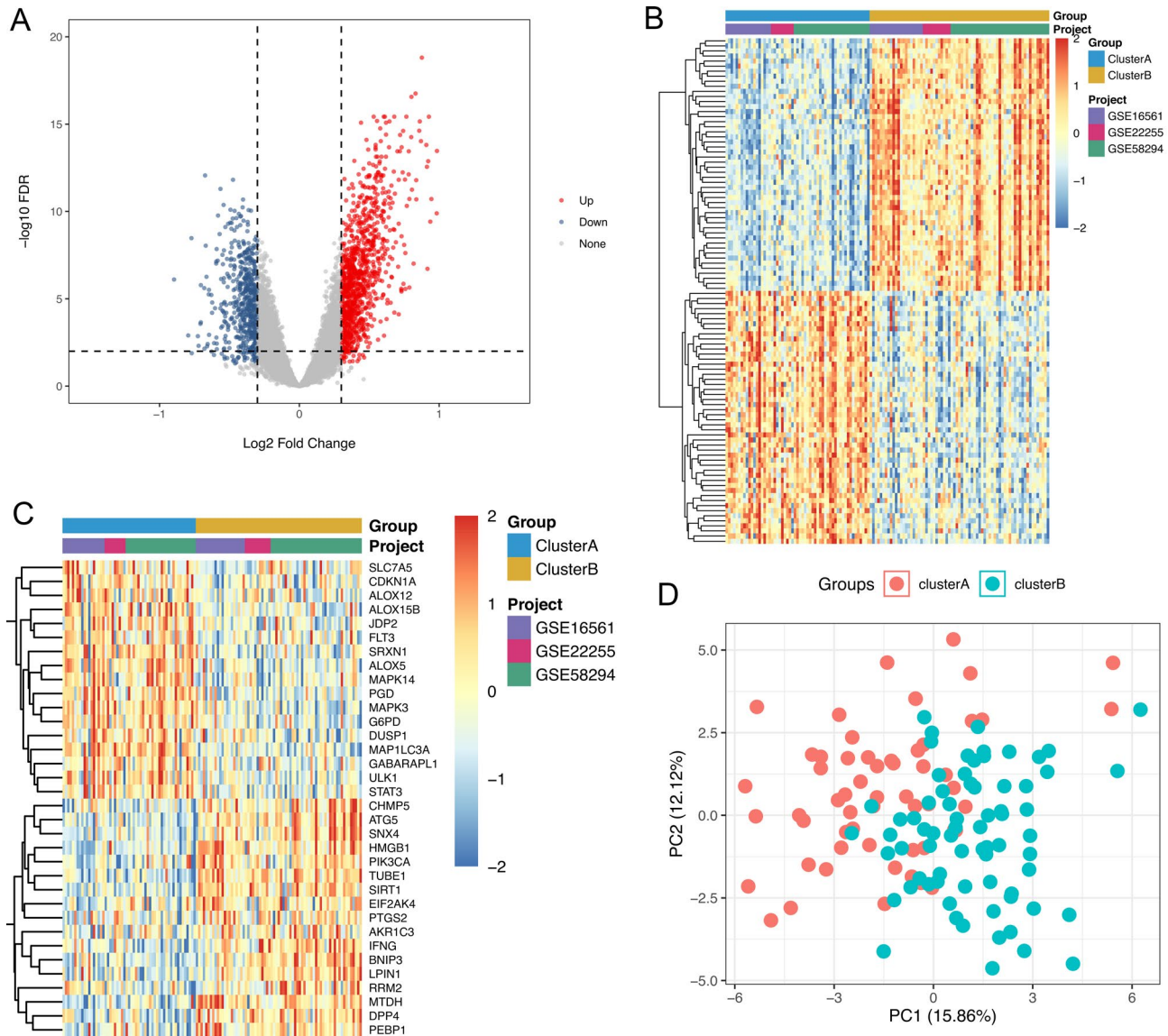


**Figure 1.** DEGs between CIS patients and healthy controls. (A) volcanic map of 1995 DEGs between CIS group and healthy controls; (B) heat map of 1995 DEGs; (C) heat map of 46 differentially expressed FRGs between CIS group and health controls.

GEO ID	Cluster A	Cluster B
GSE16561	18	21
GSE22255	9	11
GSE58294	30	39
Total	57	71

**Table 1.** The consistent clustering grouping of CIS samples.





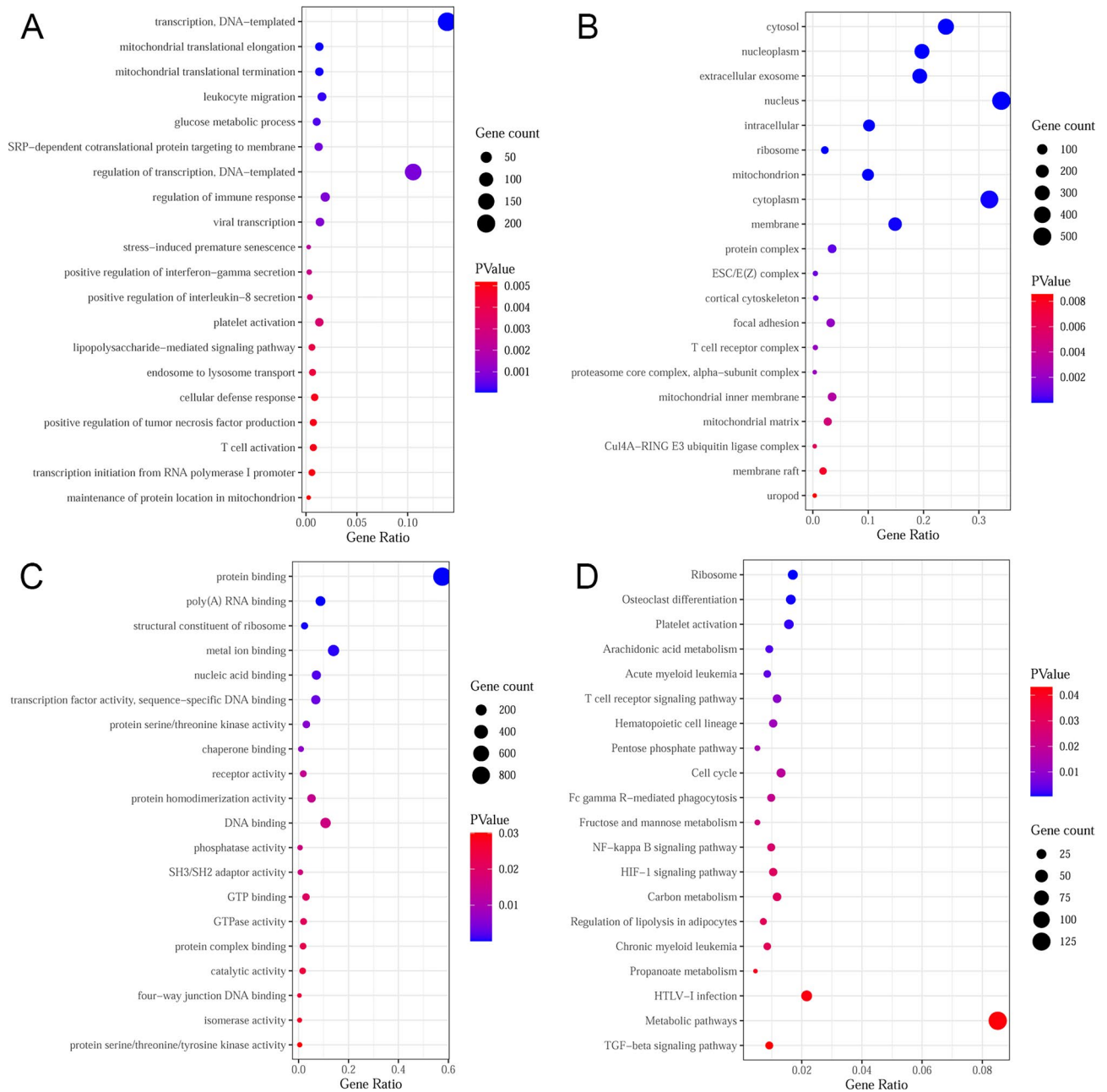
**Figure 2.** DEGs between cluster A and B in CIS patients. **(A)** volcanic map of 1560 DEGs between cluster A and cluster B in CIS patients; **(B)** heat map of 1560 DEGs between cluster A and cluster B; **(C)** heat map of 34 differentially expressed FRGs between cluster A and cluster B; **(D)** Results from PCA for cluster A and cluster B in CIS patients.

### Immune microenvironment analysis

Based on the immune microenvironment analysis, the immune infiltration levels of NK cells activated and Eosinophils were the most significantly increased, and Macrophages M2 and Neutrophils were the most significantly decreased in cluster B subtype, compared with cluster A subtype (Fig. 7A). The heat map of relative abundance of 22 immune cells is shown in Fig. 7B. It is noted that neutrophils were the most positively correlated with all 5 real hub genes (Fig. 7C).

### TF regulatory network, drug prediction and molecular docking of 5 real hub genes

Based on TRRUST database, only two hub genes, MAPK3 and WAS, related TFs were identified. Twist family bHLH transcription factor 1 (TWIST1) regulated MAPK3. ETS proto-oncogene 1, transcription factor (ETS1), MYB proto-oncogene, transcription factor (MYB), Sp1 transcription factor (SP1), and Spi-1 proto-oncogene (SPI1) regulated WAS (Fig. 8). In addition, 5 real hub genes related drugs were queried based on the DGIdb database. The result showed that only MAPK3 and PRKCD were found to be targeted by multiple drugs. For example, PRKCD was one of the target gene of FASUDIL (Fig. 9A). The binding energy of PRKCD and FASUDIL was  $-7.06$  kcal/mol, forming a hydrogen bond with ASP-36 residues (Fig. 9B).

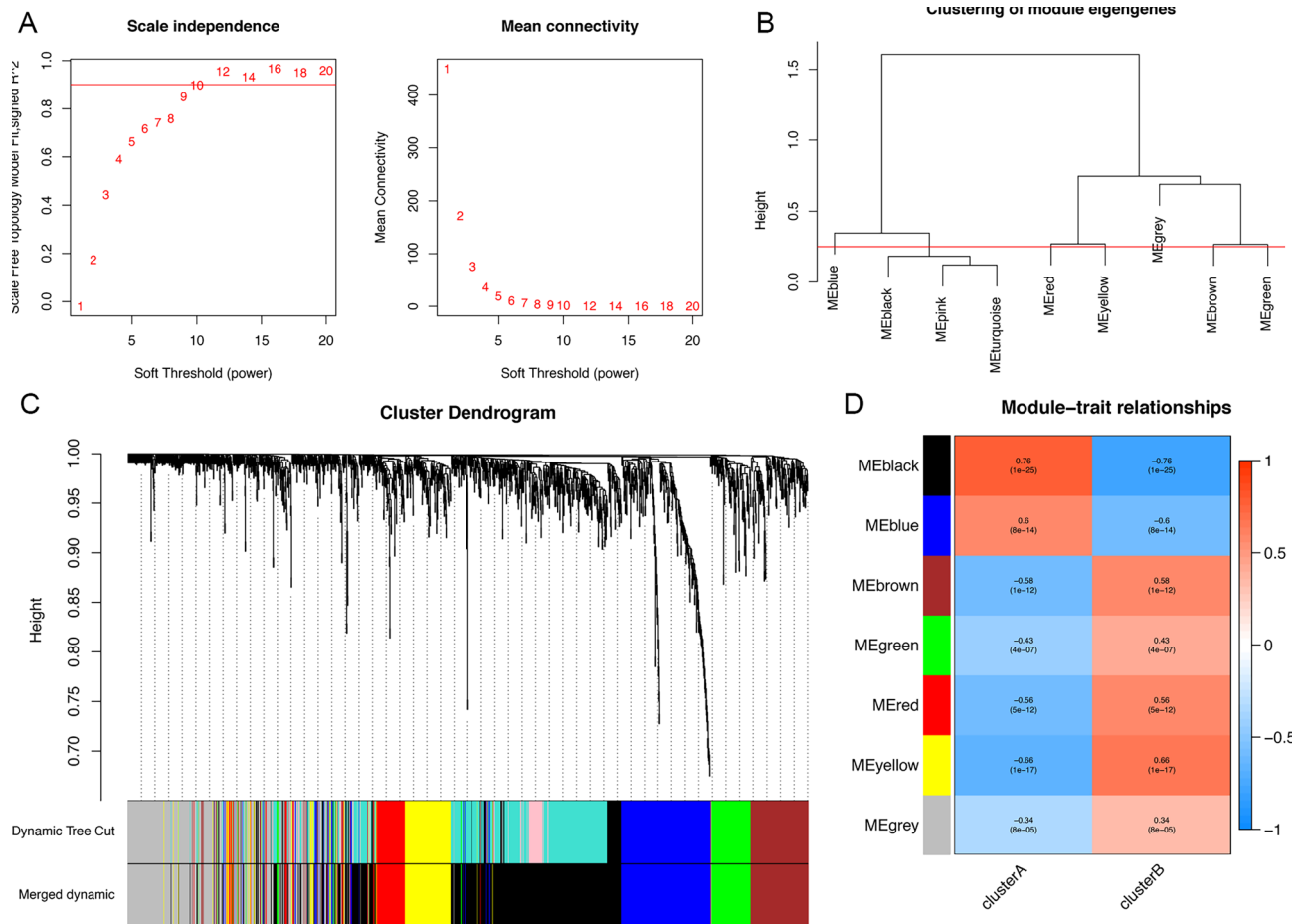


**Figure 3.** Top 20 enrichment results of DEGs between cluster A and cluster B in CIS patients. (A) BP; (B) CC; (C) MF; (D) KEGG.

### Discussion

CIS refers to the brain tissue necrosis caused by stenosis or occlusion of cerebral blood supply arteries (carotid artery and vertebral artery) and insufficient cerebral blood supply. The pathogenesis of ischemic stroke is complex. At present, reported pathological mechanisms of CIS include abnormal energy metabolism, excitotoxicity of amino acids, oxidative damage of free radicals and apoptosis<sup>40–44</sup>. At the cellular level, hypoperfusion, hypoxia, and glucose deficiency cause insufficient energy production in brain metabolism, thus leading to some toxic metabolites accumulating locally, such as excitatory toxic products, acidic metabolites, oxidative stress products, and inflammatory mediators, resulting in cerebral cell dysfunction or death<sup>45</sup>. Various pathophysiological mechanisms had been reported to be involved in stroke, including angiogenesis, oxidative stress, autophagy formation, inflammatory response, and apoptosis<sup>46</sup>.

Inflammation is an important part of the pathogenesis of CIS. Dead or damaged cells release damage-associated molecular patterns in trauma (DAMP), thus leading to activated microglia, morphological changes occurring and secreting cytokines. Followed by astrocytes secreting cytokines and chemokines (including matrix metalloproteinases (MMPs) family expression), the chemokines participate in the immune response after injury, destructed the early blood-brain barrier after stroke. Further, peripheral lymphocytes and neutrophils transfer



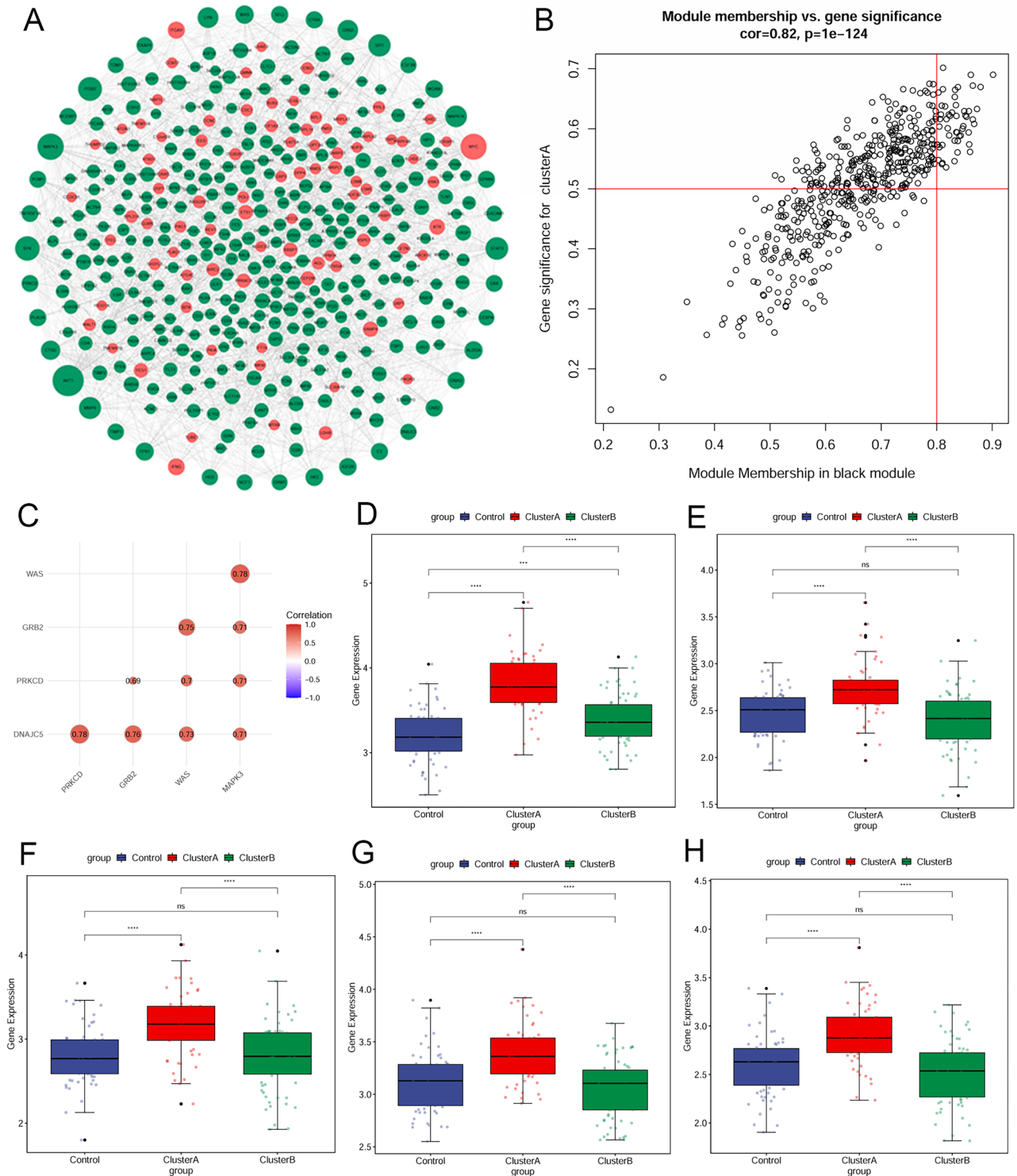
**Figure 4.** Construction of WGCNA. (A) Determination of soft thresholding power in the WGCNA; (B) the merged modules for clustering by WGCNA; (C) clustering dendrogram of genes, and different colors represent different modules; (D) the heat map of module-trait relationship in CIS, the black module has the highest correlation with CIS.

to the damaged brain area, and release proinflammatory cytokines, ROS and MMP, which aggravate brain parenchymal injury and blood-brain barrier damage<sup>11</sup>.

The pathological process of CIS is complex and involves multiple signaling pathways. Sonic hedgehog (Shh) signaling pathway exerts neuroprotective effects in ischemic stroke<sup>47</sup>. A previous review summarized the roles of p53-mediated ferroptosis signaling pathway underlying CIS<sup>48</sup>. It is reported that NF-kappa B could regulate microglial polarization after ischemic stroke<sup>49</sup>. The HIF-1 signaling pathway is activated under hypoxic conditions and plays important roles in the pathological processes of ischemic stroke<sup>50</sup>. In addition, it has been confirmed that activation of the HIF-1 signaling pathway is associated with ferroptosis<sup>51</sup>. It had been suggested that TGF- $\beta$ 1/Smad3 signaling pathway exhibited a protective role in CIS<sup>52</sup>. Functional enrichment analysis showed that DEGs between clusters A and B were enriched in NF-kappa B signaling pathway, HIF-1 signaling pathway, and TGF-beta signaling pathway. GSEA results indicated hedgehog signaling, p53 pathway, hypoxia, and inflammatory response were more active in cluster A.

As a “prototypic” or “classical” mitogen-activated protein (MAP) kinase, activation of MAPK3 protects cells from ischemic injury<sup>53</sup>. We found that MAPK3 was increased in cluster A compared to cluster B. Consistent with our study, Huang et al.<sup>54</sup> found that MAPK3 was involved in the acute myocardial infarction process as a FRG. It is also necessary to analyze the correlation between this up-regulation and clinical classification in combination with the clinical characteristics of patients, and the degree of ischemia or nerve damage. On the other hand, the expression level of MAPK3 did not show difference between CIS patients and normal controls, which need to be further verified in the study of expanding the sample size and strictly controlling the control parameters. WAS is the coding gene of wiskott aldrich syndrome protein (WASP), which is involved in relaying signals from the surface of blood cells to the actin cytoskeleton<sup>55</sup>. Salvi and Thanabalu<sup>56</sup> reported that expression of WASP was increased under hypoxic conditions. Our results also showed that WAS had the strongest correlation with MAPK3. However, the regulatory pathway and interaction of these two genes in CIS are still unclear, which needs further exploration and clarification.

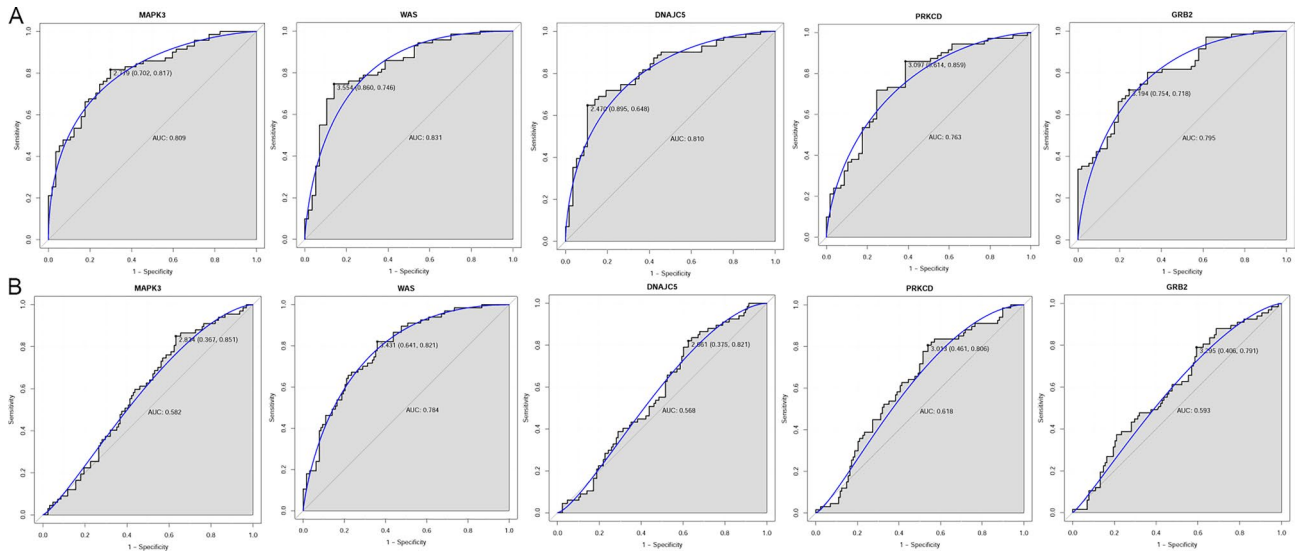
DnaJ heat shock protein family (Hsp40) member C5 (DNAJC5) is a member of DNAJ family, and primarily expressed in neurons<sup>57</sup>. Currently, no report mentioned its roles in CIS, thus its roles and functions need further exploration. Protein Kinase C Delta (PRKCD) has been implicated in mediating ischemic and reperfusion



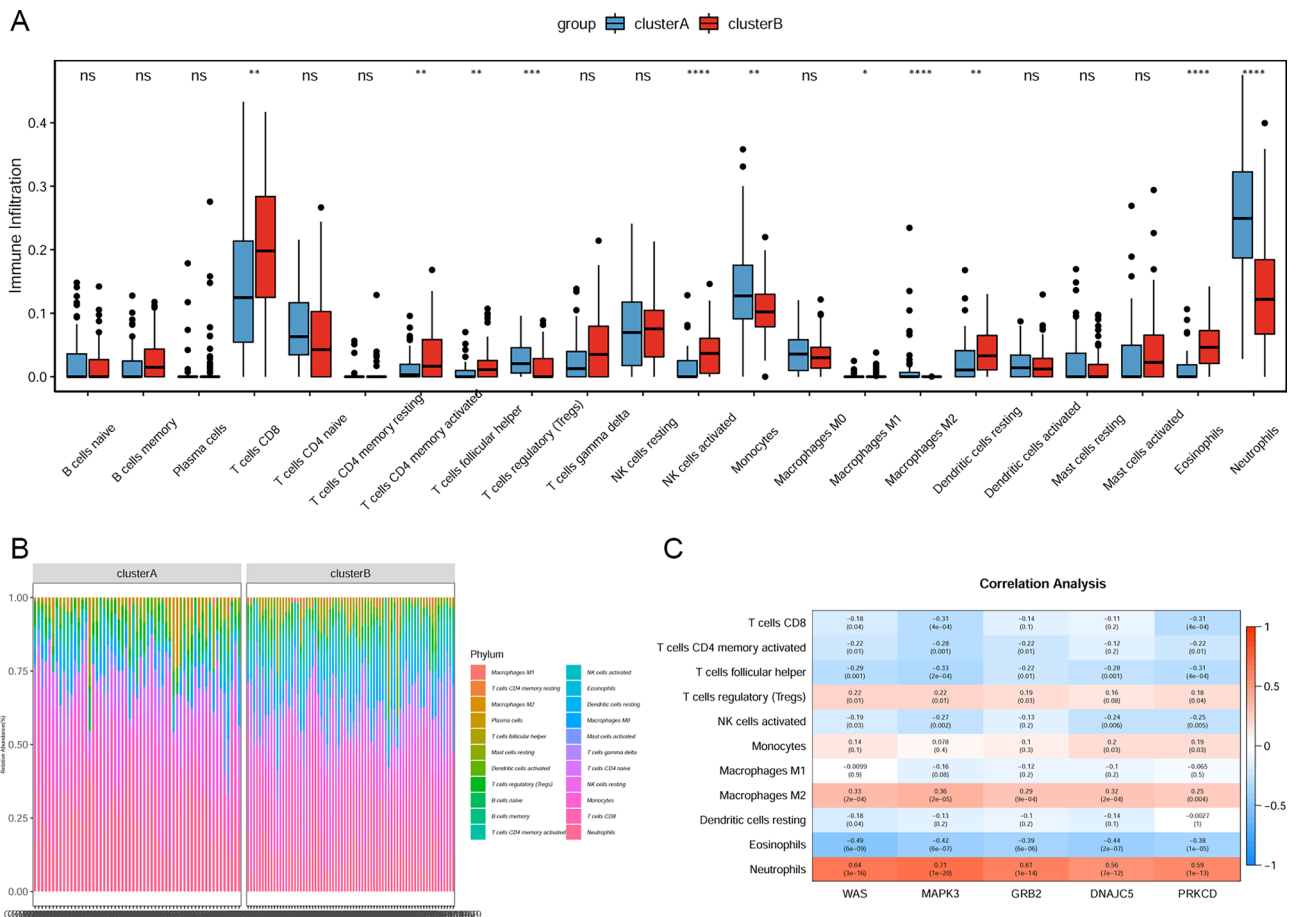
**Figure 5.** Identification of the real hub genes. (A) PPI network of 508 genes in the black module; (B) Scatterplot of genes (module membership versus gene significance) in the black module; (C) Correlation analysis among the five real hub genes; (D–H) The expression levels of five genes in normal controls, cluster A, and cluster B. (D) WAS; (E) DNAJC5; (F) PRKCD; (G) GRB2; (H) MAPK3.

damage in stroke-reperfusion<sup>58</sup>. Its upregulation is correlated with the occurrence of ischemic stroke. At present, several omics studies have shown that growth factor receptor-bound protein 2 (GRB2) is regulated by noncoding genes in ischemic stroke<sup>59,60</sup>, and its activation is critical for the activation of the Ras signaling pathway and downstream MAPK<sup>61</sup>. ROC analysis indicated that five real hub genes possessed potential diagnostic values,

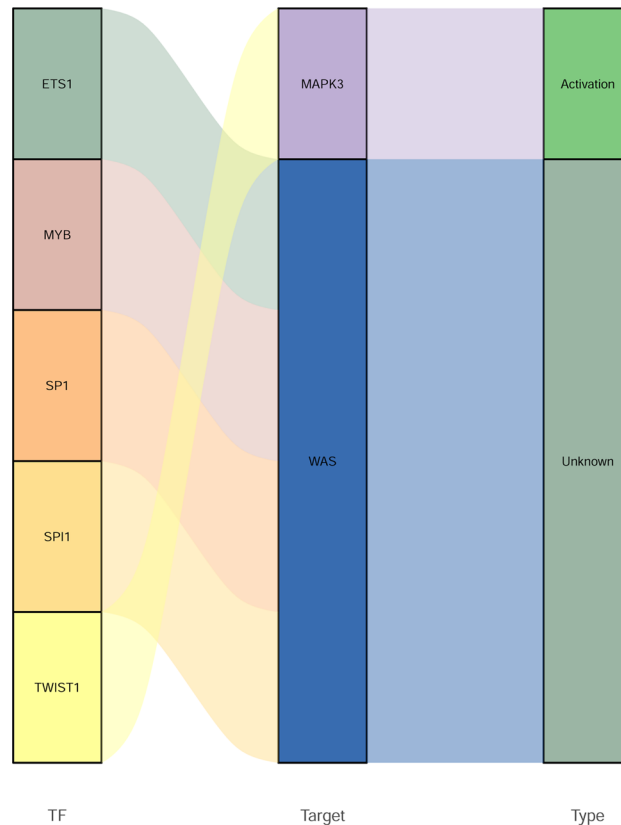




**Figure 6.** The ROC curve of key genes in CIS patients. (A) ROC analysis of MAPK3, WAS, DNAJC5, PRKCD, and GRB2 between cluster A and cluster B; (B) ROC analysis of MAPK3, WAS, DNAJC5, PRKCD, and GRB2 between CIS patients and normal control.



**Figure 7.** Immune microenvironment analysis between cluster A subtype and cluster B subtypes. (A) boxplot of immune infiltration of 22 immune cells; (B) the heat map of relative abundance of 22 immune cells; (C) pearson correlation analysis between immune cells and 5 real hub genes.

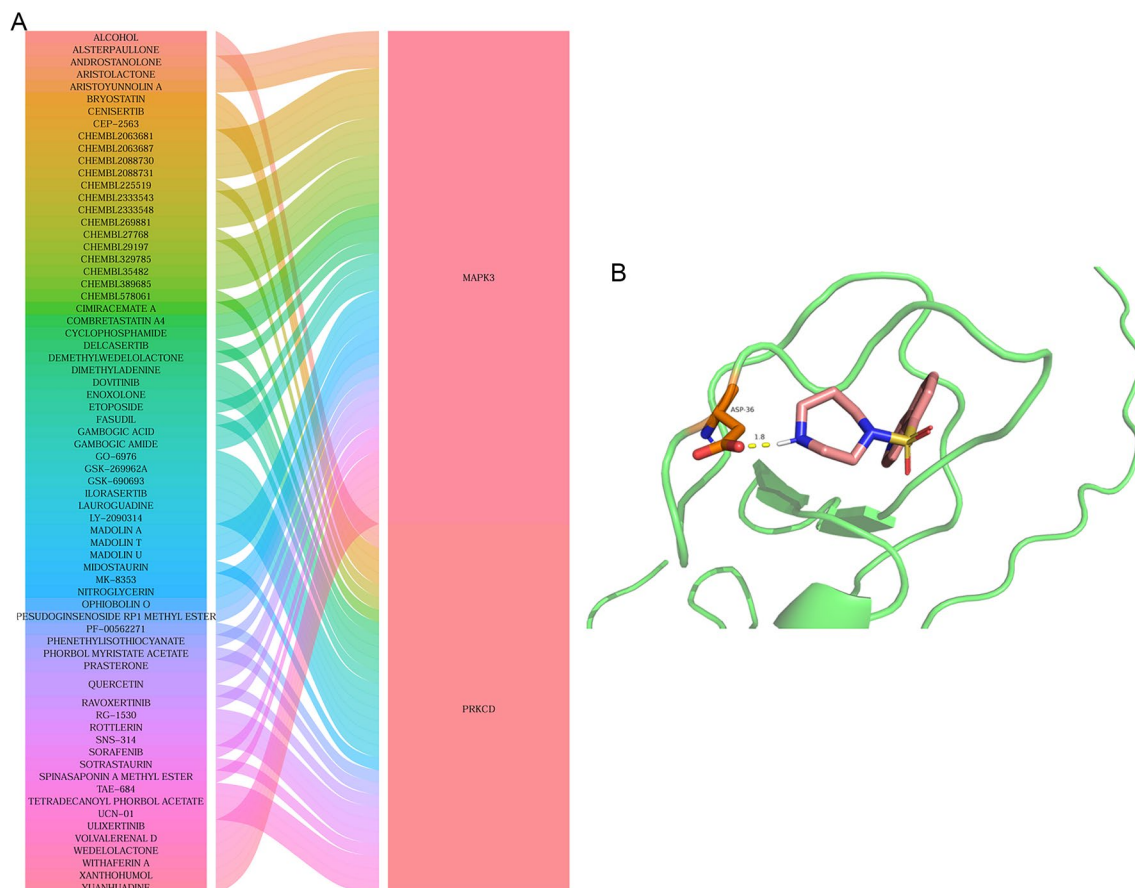


**Figure 8.** The regulatory network between TFs and 5 real hub genes.

which can distinguish the two subtypes well. The association of hub genes with the disease characters, for example, age, sex, and duration of CIS even clinical classification, was not included in the analysis, which makes the clinical characteristics of cluster A and cluster B unclear. So it is needed to further establish a more reasonable patients' cohort for biological information analysis.

Based on the immune microenvironment analysis, the infiltration levels of neutrophils were the most significantly decreased in cluster B subtype, compared with cluster A subtype. Moreover, neutrophils were the most positively correlated with MAPK3, WAS, DNJC5, PRKCD, and GRB2. Infiltrating neutrophils increase blood–brain barrier permeability, which cause microvascular disorder. It is reported that neutrophil extracellular traps are significantly increased in the plasma of ischemic stroke patients and related to stroke severity and mortality<sup>62</sup>. Thus it can be seen that neutrophils play important roles in the development of CIS. Based on TRRUST database, MAPK3 and WAS related TFs were identified. TWIST1 regulated MAPK3. ETS1, MYB, SP1, and SPI1 regulated WAS. It is found that MACC1-AS1 exerts a protective role in hypoxia-induced brain microvascular endothelial cells in CIS via regulating miR-6867-5p/TWIST1<sup>63</sup>. CASC15 promotes cerebral ischemia/reperfusion injury via miR-338-3p/ETS1 axis in acute IS<sup>64</sup>. MiR-150 and its target MYB can form a negative feedback loop to control the level of post-stroke vascular endothelial growth factor (VEGF) expression<sup>65</sup>. In IS, miR-1224 contributes to dysfunction of natural killer cell by targeting Sp1 signaling<sup>66</sup>. In stroke, FTX transcript, XIST regulator (FTX) can regulate angiogenesis through miR-342-3p/SPI1 axis<sup>67</sup>. It is indicated that MAPK3 and WAS may be involved in the process of CIS under the regulation of TFs of TWIST1, ETS1, MYB, SP1, and SPI1. In addition, related drugs associated with MAPK3, WAS, DNJC5, PRKCD, and GRB2 were queried based on the DGIdb database. The result showed PRKCD was one of the target gene of FASUDIL. Moreover, in the analysis of molecular docking, the binding score of PRKCD with FASUDIL was lower than  $-6.0$  kcal/mol, indicating a better binding affinity. It is suggested that PRKCD could be a drug target of FASUDIL in the treatment of CIS. It is noted that neutrophils was the most positively correlated with all 5 real hub genes. PRKCD was one of the target gene of FASUDIL.

In summary, our results showed that 5 key hub genes, MAPK3, WAS, DNJC5, PRKCD, and GRB2, which were highly expressed in cluster A and can distinguish cluster A from cluster B of CIS. Among the 5 genes, MAPK3 was a FRG and WAS showed potential diagnostic value in distinguishing CIS patients from normal controls. In addition, we plan to collect a larger sample size for further validation of these five real hub genes identified in our study, followed by cell- and animal-level experiments to investigate their specific roles in CIS.



**Figure 9.** Drug prediction and molecular docking. (A) drug prediction of 5 real hub genes based on the DGIdb database; (B) molecular docking of FASUDIL-PRKCD.

### Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 18 July 2023; Accepted: 31 January 2024

Published online: 23 April 2024

### References

- Emergency Medical Quality Control Center of National Health and Health Commission, Emergency Physicians Branch of Chinese Medical Association, and Emergency Committee of World Federation of Chinese Medicine Societies. Expert consensus on emergency diagnosis and treatment of acute ischemic stroke in China. *Chin. J. Crit. Care Med.* **38**, 281–287 (2018).
- Sarkar, S., Chakraborty, D., Bhowmik, A. & Ghosh, M. K. Cerebral ischemic stroke: cellular fate and therapeutic opportunities. *Front. Biosci.* **24**, 435–450 (2019).
- Liang, J., Han, R. & Zhou, B. Metabolic reprogramming: Strategy for ischemic stroke treatment by ischemic preconditioning. *Biology* **10**, 424 (2021).
- Jiang, X., Stockwell, B. R. & Conrad, M. Ferroptosis: Mechanisms, biology and role in disease. *Nat. Rev. Mol. Cell. Biol.* **22**, 266–282 (2021).
- Li, J. *et al.* Ferroptosis: Past, present and future. *Cell. Death Dis.* **11**, 88 (2020).
- Li, C. *et al.* Nuclear receptor coactivator 4-mediated ferritinophagy contributes to cerebral ischemia-induced ferroptosis in ischemic stroke. *Pharmacol. Res.* **174**, 105933 (2021).
- Millan, M. *et al.* Targeting pro-oxidant iron with deferoxamine as a treatment for ischemic stroke: Safety and optimal dose selection in a randomized clinical trial. *Antioxidants* **10**, 1270 (2021).
- Zhou, Y., Liao, J., Mei, Z., Liu, X. & Ge, J. Insight into crosstalk between ferroptosis and necroptosis: Novel therapeutics in ischemic stroke. *Oxid. Med. Cell. Longev.* **2021**, 9991001 (2021).
- Tang, D., Chen, X., Kang, R. & Kroemer, G. Ferroptosis: Molecular mechanisms and health implications. *Cell Res.* **31**, 107–125 (2021).
- Liu, Y. *et al.* Ferroptosis: An emerging therapeutic target in stroke. *J. Neurochem.* **160**, 64–73 (2022).
- Shi, K. *et al.* Global brain inflammation in stroke. *Lancet Neurol.* **18**, 1058–1066 (2019).
- Rroji, O., Kumar, A., Karuppagounder, S. S. & Ratan, R. R. Epigenetic regulators of neuronal ferroptosis identify novel therapeutics for neurological diseases: HDACs, transglutaminases, and HIF prolyl hydroxylases. *Neurobiol. Dis.* **147**, 105145 (2021).
- Guo, J., Tuo, Q. Z. & Lei, P. Iron, ferroptosis, and ischemic stroke. *J. Neurochem.* **165**, 487–520 (2023).
- Xu, Y. *et al.* Role of ferroptosis in stroke. *Cell. Mol. Neurobiol.* **43**, 205–222 (2023).

15. Chen, G., Li, L. & Tao, H. Bioinformatics identification of ferroptosis-related biomarkers and therapeutic compounds in ischemic stroke. *Front. Neurol.* **12**, 745240 (2021).
16. Tang, Y., Ke, Z. P., Peng, Y. G. & Cai, P. T. Co-expression analysis reveals key gene modules and pathway of human coronary heart disease. *J. Cell. Biochem.* **119**, 2102–2109 (2018).
17. Nomiri, S. *et al.* Exploiting systems biology to investigate the gene modules and drugs in ovarian cancer: A hypothesis based on the weighted gene co-expression network analysis. *Biomed. Pharmacother.* **146**, 112537 (2022).
18. Presson, A. P. *et al.* Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. *BMC Syst. Biol.* **2**, 95 (2008).
19. Qu, X. *et al.* Weighted gene co expression network analysis (WGCNA) with key pathways and hub-genes related to micro RNAs in ischemic stroke. *IET Syst. Biol.* **15**, 93–100 (2021).
20. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
21. Zhou, N. & Bao, J. FerrDb: A manually curated resource for regulators and markers of ferroptosis and ferroptosis-disease associations. *Database* **2020**, 1–10 (2020).
22. Liang, J. Y. *et al.* A novel ferroptosis-related gene signature for overall survival prediction in patients with hepatocellular carcinoma. *Int. J. Biol. Sci.* **16**, 2430–2441 (2020).
23. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
24. Kroon, E. E. *et al.* Neutrophil extracellular trap formation and gene programs distinguish TST/IGRA sensitization outcomes among *Mycobacterium tuberculosis* exposed persons living with HIV. *PLoS Genet.* **19**, e1010888 (2023).
25. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
26. Lakota, K. *et al.* International cohort study of 73 anti-Ku-positive patients: association of p70/p80 anti-Ku antibodies with joint/bone features and differentiation of disease populations by using principal-components analysis. *Arthritis Res. Ther.* **14**, R2 (2012).
27. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
28. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–d592 (2023).
29. Sherman, B. T. *et al.* DAVID: A web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).
30. da Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
31. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
32. Langfelder, P. & Horvath, S. Fast R functions for robust correlations and hierarchical clustering. *J. Stat. Softw.* **46**, 11 (2012).
33. Szklarczyk, D. *et al.* The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
34. Otasek, D., Morris, J. H., Boucas, J., Pico, A. R. & Demchak, B. Cytoscape automation: Empowering workflow-based network analysis. *Genome Biol.* **20**, 185 (2019).
35. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
36. Han, H. *et al.* TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* **46**, 380–386 (2018).
37. Freshour, S. L. *et al.* Integration of the drug-gene interaction database (DGIdb 40) with open crowdsourcing efforts. *Nucleic Acids Res.* **49**, 1144–1151 (2021).
38. El-Hachem, N., Haibe-Kains, B., Khalil, A., Kobeissy, F. H. & Nemer, G. AutoDock and AutoDockTools for protein-ligand docking: Beta-site amyloid precursor protein cleaving enzyme 1 (BACE1) as a case study. *Methods Mol. Biol.* **1598**, 391–403 (2017).
39. Baugh, E. H., Lyskov, S., Weitzner, B. D. & Gray, J. J. Real-time PyMOL visualization for Rosetta and PyRosetta. *PLoS ONE* **6**, e21931 (2011).
40. Villa, R. F., Gorini, A., Ferrari, F. & Hoyer, S. Energy metabolism of cerebral mitochondria during aging, ischemia and post-ischemic recovery assessed by functional proteomics of enzymes. *Neurochem. Int.* **63**, 765–781 (2013).
41. Russo, E. *et al.* Energy metabolism analysis of three different mesenchymal stem cell populations of umbilical cord under normal and pathologic conditions. *Stem Cell Rev. Rep.* **16**, 585–595 (2020).
42. Tejada, G. S. *et al.* Brain ischaemia induces shedding of a BDNF-scavenger ectodomain from TrkB receptors by excitotoxicity activation of metalloproteinases and gamma-secretases. *J. Pathol.* **238**, 627–640 (2016).
43. Amantea, D. & Bagetta, G. Excitatory and inhibitory amino acid neurotransmitters in stroke: from neurotoxicity to ischemic tolerance. *Curr. Opin. Pharmacol.* **35**, 111–119 (2017).
44. Patel, D. & Wairkar, S. Biotechnology-based therapeutics for management of cerebral stroke. *Eur. J. Pharmacol.* **913**, 174638 (2021).
45. Datta, A. *et al.* Cell death pathways in ischemic stroke and targeted pharmacotherapy. *Transl. Stroke Res.* **11**, 1185–1202 (2020).
46. Chen, B. & Jin, W. A comprehensive review of stroke-related signaling pathways and treatment in western medicine and traditional Chinese medicine. *Front. Neurosci.* **17**, 1200061 (2023).
47. Liu, L., Zhao, B., Xiong, X. & Xia, Z. The neuroprotective roles of sonic hedgehog signaling pathway in ischemic stroke. *Neurochem. Res.* **43**, 2199–2211 (2018).
48. Xu, S., Li, X. & Wang, Y. Regulation of the p53-mediated ferroptosis signaling pathway in cerebral ischemia stroke (Review). *Exp. Ther. Med.* **25**, 113 (2023).
49. Zhao, S. C. *et al.* Regulation of microglial activation in stroke. *Acta Pharmacol. Sin.* **38**, 445–458 (2017).
50. Pan, Z., Ma, G., Kong, L. & Du, G. Hypoxia-inducible factor-1: Regulatory mechanisms and drug development in stroke. *Pharmacol. Res.* **170**, 105742 (2021).
51. Li, Y. *et al.* Inhibitor of apoptosis-stimulating protein of p53 inhibits ferroptosis and alleviates intestinal ischemia/reperfusion-induced acute lung injury. *Cell Death Diff.* **27**, 2635–2650 (2020).
52. Zhu, H. *et al.* TGF- $\beta$ 1/Smad3 signaling pathway suppresses cell apoptosis in cerebral ischemic stroke rats. *Med. Sci. Monit.* **23**, 366–376 (2017).
53. Lips, D. J. *et al.* MEK1-ERK2 signaling pathway protects myocardium from ischemic injury in vivo. *Circulation* **109**, 1938–1941 (2004).
54. Huang, D. *et al.* Machine learning revealed ferroptosis features and a novel ferroptosis-based classification for diagnosis in acute myocardial infarction. *Front. Genet.* **13**, 813438 (2022).
55. Kurisu, S. & Takenawa, T. The WASP and WAVE family proteins. *Genome Biol.* **10**, 226 (2009).
56. Salvi, A. & Thanabalu, T. Expression of N-WASP is regulated by HIF1 $\alpha$  through the hypoxia response element in the N-WASP promoter. *Biochem. Biophys. Res. Commun.* **453**, 13–21 (2014).
57. Stetler, R. A. *et al.* Heat shock proteins: Cellular and molecular mechanisms in the central nervous system. *Prog. Neurobiol.* **92**, 184–211 (2010).
58. Bright, R. *et al.* Protein kinase C delta mediates cerebral reperfusion injury in vivo. *J. Neurosci.* **24**, 6880–6888 (2004).
59. Zhang, L., Liu, B., Han, J., Wang, T. & Han, L. Competing endogenous RNA network analysis for screening inflammation-related long noncoding RNAs for acute ischemic stroke. *Mol. Med. Rep.* **22**, 3081–3094 (2020).



60. Jin, K. *et al.* Microarray analysis of hippocampal gene expression in global cerebral ischemia. *Ann. Neurol.* **50**, 93–103 (2001).
61. Schlaepfer, D. D., Hanks, S. K., Hunter, T. & van der Geer, P. Integrin-mediated signal transduction linked to Ras pathway by GRB2 binding to focal adhesion kinase. *Nature* **372**, 786–791 (1994).
62. Vallés, J. *et al.* Neutrophil extracellular traps are increased in patients with acute ischemic stroke: prognostic significance. *Thromb. Haemost.* **117**, 1919–1929 (2017).
63. Yan, G., Zhao, H. & Hong, X. LncRNA MACC1-AS1 attenuates microvascular endothelial cell injury and promotes angiogenesis under hypoxic conditions via modulating miR-6867-5p/TWIST1 in human brain microvascular endothelial cells. *Ann. Transl. Med.* **8**, 876 (2020).
64. Chen, C., Wang, L., Wang, L., Liu, Q. & Wang, C. LncRNA CASC15 promotes cerebral ischemia/reperfusion injury via miR-338-3p/ETS1 axis in acute ischemic stroke. *Int. J. Gen. Med.* **14**, 6305–6313 (2021).
65. Zhang, S., Chen, A. & Chen, X. A feedback loop involving microRNA-150 and MYB regulates VEGF expression in brain microvascular endothelial cells after oxygen glucose deprivation. *Front. Physiol.* **12**, 619904 (2021).
66. Feng, Y. *et al.* miR-1224 contributes to ischemic stroke-mediated natural killer cell dysfunction by targeting Sp1 signaling. *Front. Physiol.* **18**, 133 (2021).
67. Gao, Q. & Wang, Y. LncRNA FTX regulates angiogenesis through miR-342-3p/SPI1 axis in stroke. *Neuropsychiatr. Dis. Treat.* **17**, 3617–3625 (2021).

### Author contributions

Yufeng Wang designed the project. Yufeng Wang and Xinjun Shui analyzed the data. Xinjuan Xu, Ruilin Ren, and Yu Liu interpreted the data. All authors took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published.

### Funding

The authors declare that they have no competing interests.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-53327-2>.

**Correspondence** and requests for materials should be addressed to Y.W.

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

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024