BMC Neuroscience

RESEARCH

Open Access

Identification of key genes associated with oxidative stress in ischemic stroke via bioinformatics integrated analysis

Gaiyan Li[1], Yu Cheng[2], Shanshan Ding[1], Qianyun Zheng[1], Lanqiong Kuang[2], Ying Zhang[1*] and Ying Zhou[3*]

Abstract

Background Ischemic stroke (IS) is a common cerebrovascular disease. Although the formation of atherosclerosis, which is closely related to oxidative stress (OS), is associated with stroke-related deaths. However, the role of OS in IS is unknown.

Methods OS-related key genes were obtianed by overlapping the differentially expressed genes (DEGs) between IS and normal control (NC) specimens, IS-related genes, and OS-related genes. Then, we investigated the mechanism of action of key genes. Subsequently, protein–protein interaction (PPI) network and machine learning algorithms were utilized to excavate feature genes. In addition, the network between feature genes and microRNAs (miRNAs) was established to investigate the regulatory mechanism of feature genes. Finally, quantitative PCR (qPCR) was utilized to validate the expression of feature genes with blood specimens.

Results A total of 42 key genes related to OS were acquired. Enrichment analysis indicated that the key genes were associated with oxidative stress, reactive oxygen species, lipid and atherosclerosis, and cell migration-related pathways. Then, 6 feature genes (*HSPA8*, *NCF2*, *FOS*, *KLF4*, *THBS1*, and *HSPA1A*) related to OS were identified for IS. Besides, 6 feature genes and 255 miRNAs were utilized to establish a feature genes-miRNA network which contained 261 nodes and 277 edges. At last, qPCR results revealed that there was a trend for higher expression of *FOS*, *KLF4*, and *HSPA1A* in IS specimens than in NC specimens. Additionally, *HSPA8* expression was significantly decreased in the IS specimens, which was consistent with the findings of the GEO database analysis.

Conclusion In conclusion, 6 feature genes (*HSPA8*, *NCF2*, *FOS*, *KLF4*, *THBS1*, and *HSPA1A*) related to OS were mined by bioinformatics analysis, which might provide a new insights into the evaluation and treatment of IS.

Clinical trial number: Not applicable.

Keywords Ischemic stroke, Oxidative stress, Machine learning algorithm, WGCNA

*Correspondence:
Ying Zhang
zhangying032317@163.com
Ying Zhou
yingzhouwu@hotmail.com
Full list of author information is available at the end of the article

Introduction

Ischemic stroke (IS) is characterized by high incidence, disability, high mortality and recurrence. Besides, IS is a major cause of disability and cognitive impairment, accounting for about 5.2% of all deaths worldwide [1]. Brain tissue necrosis and focal neuron defect, induced by thrombosis or embolism, are the basic pathology of IS. IS progresses rapidly, necessitating timely detection. Early detection and diagnosis of IS are very important for the treatment and rehabilitation of IS.

The standard treatment of ischemic stroke (IS) includes intravenous thrombolysis, arterial thrombectomy, and supportive therapy. The purpose of these therapies designed to quickly recover brain blood flow, reduce brain damage [2]. Thrombolytic therapy, particularly with recombinant tissue plasminogen activator (rtPA), requires rapid diagnosis within a narrow therapeutic time window and immediate initiation to dissolve the thrombus and reinfuse the affected brain area [3]. However, thrombolytic therapy is not suitable for all IS patients. Bleeding or recent history of surgery are contraindications to intravenous thrombolysis. Arterial thrombectomy is an interventional therapy, that shows significant efficacy even within 24 h after symptom onset. However, its success is highly dependent on the timeliness of treatment and the specific condition of the patient [4]. The effect of this therapy strategy is significant affected by the intervention time and patient selection criteria, emphasizing the importance of precision medical treatment in IS. Exploring new treatments is essential to improve the prognosis of stroke and reduce its social and economic burden.

Oxidative stress (OS) is one of the important mechanisms of ischemic cerebral (IS). Studies [5] have shown that OS is related to neurological function damage in patients, which can accelerate neuronal damage. As a result, targeted inhibition of OS has been as a neuroprotective strategy for IS treatment. The overproduction of reactive oxygen species (ROS) is central to oxidative stress (OS), resulting in an imbalance of oxidation and antioxidation functions, thus causing a series of OS cascades [6]. In clinical, some drugs can directly inhibit the production of ROS, thereby inhibiting oxidative stress. For instance, antioxidants, such as vitamin C and vitamin E, can inhibit the activity of nicotinamide adenine dinucleotide phosphate oxidase (NADPH) and reduce the production of ROS [7]. However, the available antioxidant strategies are limited to preclinical studies, and more clinical trials are needed to verify the clinical curative effect [8]. Notably, only a few studies have assessed the genes related to oxidative stress-induced IS. Genes such as Nrf2, SOD1, SOD2, and GPX1 have been identified to play a role in OS [9–12], while the complete gene network of OS has not been fully characterized. Therefore, Further genetic studies to identify all the genes involved in OS response is very necessary.

In recent years, Genetic studies provide a new insight for revealing the disease susceptibility of ischemic stroke. In this process, bioinformatics plays an indispensable role. Through the comprehensive application of genome-wide association studies (GWAS), expression quantitative trait loci (eQTLs) and transcriptome analysis, bioinformatics technology can mine IS-related genetic from large genetic data sets. With the continuous accumulation of genetic data and the continuous innovation of bioinformatics tools, it provides us a new opportunities and strategies for understanding the genetic research of IS.

In our study, bioinformatics integrated analysis was used to further analyze the oxidative stress-related genes involved in IS, and the targeted miRNAs of key genes were predicted. This provides a new vision for us to identify potential biomarkers and explore therapeutic strategies for IS.

Materials and methods

Data source

The Gene Expression Omnibus (GEO) database is the primary database for storing, retrieving and analysing high-throughput gene expression data. From these, we obtained datasets (GSE16561 and GSE180470) that were directly related to the IS, with a consistent sample source, method of acquisition, and a small difference in the number of samples and controls. GSE16561 (peripheral whole blood, GPL6883, 39 IS specimens and 24 normal control (NC) specimens) and GSE180470 (peripheral whole blood, GPL20301, 3 IS specimens and 3 NC specimens) were collected from GEO. Then, 1399 OS-related genes were retrieved from the published literature [13]. The analysis flow was shown in Fig. 1.

Differential expression analysis of GSE16561 dataset

The expression matrix of the GSE16561 dataset was normalized via the limma package (v3.50,2022) [14]. Principal component analysis (PCA) was performed on the data before and after data normalization. Empirical Bayesian smoothing of standard errors using trend = TRUE via eBayes was performed with multiple test correction, and moderated t-statistic and log-odds were calculated for each gene in each comparison. Subsequently, the differentially expressed genes (DEGs) between IS specimens and NC specimens were extracted via limma package ($p < 0.05$ and $|log2FC| > 0.5$). This improved the precision and stability of the estimates. This smoothing is particularly useful when the sample size is small, as it compensates for the instability of individual gene estimates by drawing on information from the entire dataset.

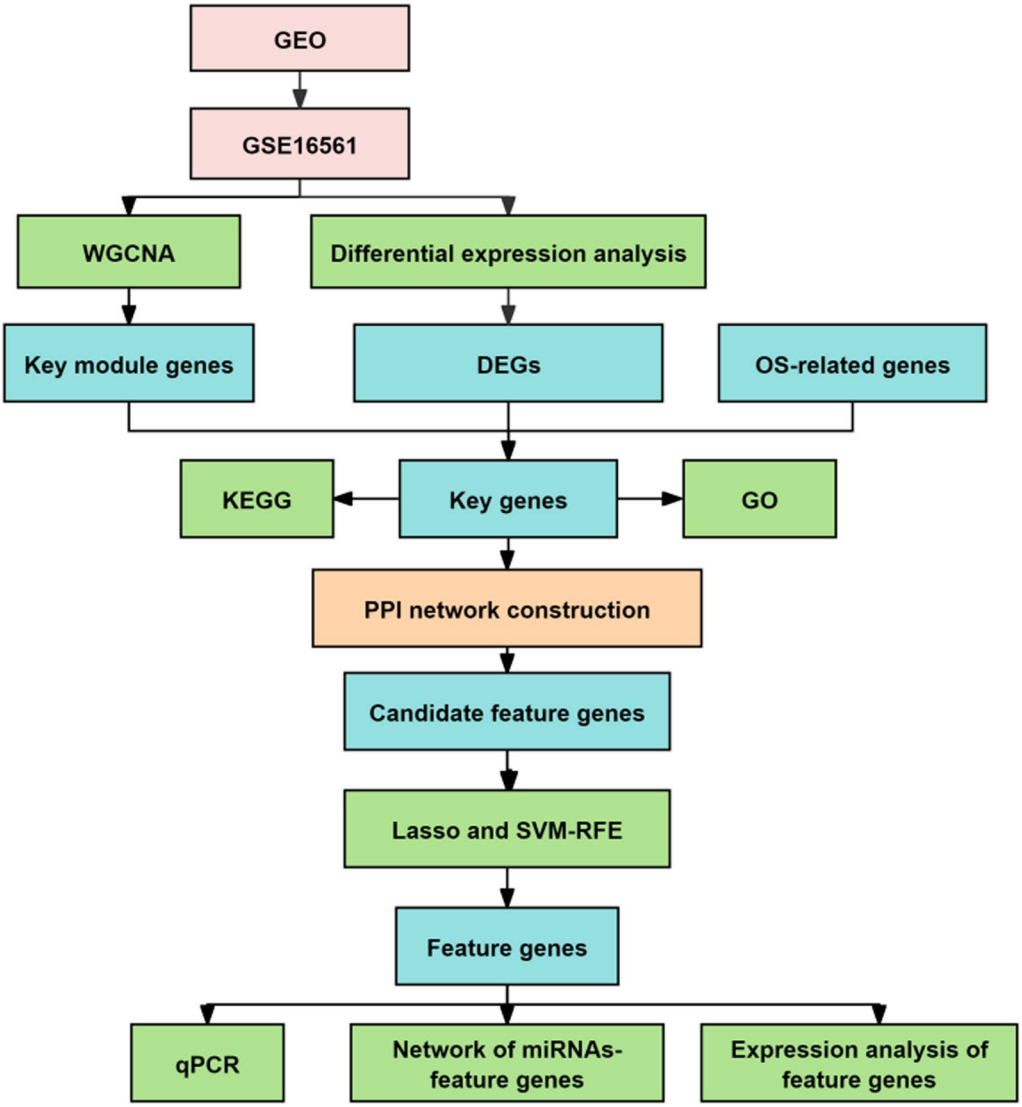

Fig. 1 Flowchart describing the schematic overview of the study design

Weighted gene co-expression network analysis (WGCNA)

The IS-related genes were determined by WGCNA package (v1.70, 2022) in the stroke dataset GSE16561 [15]. Firstly, the specimens were clustered to remove the outliers and ensure the accuracy of the subsequent analysis. Then, analysed using the WGCNA in-package (v1.70, 2022) [15] algorithm, the optimal soft threshold (β) was determined to ensure that the network approached scale-free distribution. Similarity was calculated between genes according to the adjacency, and the phylogenetic tree between genes was obtained, the higher the square of the correlation coefficient (R^2), the closer the network approximates the distribution at the no-network scale. The core of constructing the co-expression matrix is to categorize the tens of thousands of genes of the input expression matrix into dozens of modules. The modules were segmented via a dynamic tree cutting algorithm. Then, based the results of the correlation analysis, the modules with the strongest positive and negative correlation are selected as key modules, respectively. Finally, key genes were obtained by overlapping DEGs between IS and NC samples, IS-related module genes, and OS-related genes.

Functional annotation of key genes

To obtain an understanding of the metabolic pathways and gene functions of key genes, we were subjected to gene ontology (GO) and kyoto encyclopedia of genes and

genomes (KEGG) enrichment analyses. GO and KEGG enrichment analyses were implemented via clusterProfiler package (v4.2.1,2022) ($p < 0.05$ and FDR < 0.5) [16]. Multiplied check correction with "fdr". This correction ensures that the percentage of false discoveries is kept within acceptable levels while maintaining a high discovery rate.

Protein–protein interaction (PPI) network and machine learning algorithms

PPI is network with proteins as nodes involved in the same metabolic pathway, biological process, structural complex, functional association, or physical contact between proteins as edges. The PPI network was established via STRING network (https://string-db.org) (Confidence = 0.15). The genes with top 20 degree values were identified using cytoscape software (v3.7.2, 2022) for further analysis [17]. The candidate feature genes were validated via machine learning algorithms. The logistic regression (LASSO) and support vector machine-recursive feature elimination (SVM-RFE) algorithm were implemented via glmnet package (v4.1, 2022) [18] and e1071 package (v1.7, 2022), respectively. Next, the feature genes were obtained by overlapping the candidate feature genes.

Expression analysis and miRNA prediction of feature genes

The expression heat map of the feature genes was plotted via pheatmap package (v0.7.7, 2022) [19]. The microRNAs (miRNAs) interacting with feature genes were mined by NetworkAnalyst (v3.0, 2022) (https://www.networkanalyst.ca/).

Sample collection for Quantitative PCR validation

The sample was conducted in Shanghai Xuhui Central Hospital and received approval from the Ethics Committee of Shanghai Xuhui Central Hospital (reference no. 2023–030), participants and their relatives signed written informed consent for the study. A total of 20 patients came from the Department of Rehabilitation and physical center of Shanghai Xuhui Central Hospital. The data were collected from August 2023 to October 2023.

Recruitment criteria required: age control in 35–80 years old, no infection in the last 3months, no surgery in the past 3months, no cancer and no 2 Diabetes Mellitus.

Normal patients were required to have no underlying disease.

Quantitative PCR (qPCR) analysis

Firstly, 10 blood specimens were taken from IS and NC groups, and RNA was extracted from specimens with TRIzol. Second, RNA concentration was captured by NanoPhotometer N50, followed by reverse transcription of total RNA useing via SureScript-First-strand-cDNA-synthesis-kit. The qPCR reaction system was made up of 3ul of cDNA, 5ul of 2xUniversal Blue SYBR Green qPCR Master Mix, 1ul of upstream primers, and 1 ul downstream primers (Table 1). Finally, the reactions were performed on a CFX96 real-time quantitative fluorescence PCR instrument. The amplification reactions were performed as follows: pre-denaturation at 95 °C for 1 min, followed by 40 cycles, each cycle at 95 °C for 20 s, 55 °C for 20 s, and 72 °C for 30 s. Relative gene expression was measured by the $2^{-\Delta\Delta Ct}$, with GAPDH, as the internal reference gene. The primer sequences are listed in Table 2.

Table 1 The qPCR reaction system system

Component	Volume
5 × Reaction Buffer	4ul
Primer	1ul
SweScript RT I Enzyme Mix	1ul
Total RNA	0.1 ng-5ug
Nnclease-Free Water	Add to 20ul

Table 2 Related primer sequences

Primer	Sequences
KLF4 F	GGACACACGGGATGATGCTC
KLF4 R	TTCTCACCTGTGTGGGTTCG
THBS1 F	AACCTCTACTCCGGACGCAC
THBS1 R	CAGCAGGGATCCTGTGTGTA
HSPA1A F	TGTAACCCCATCATCAGCGG
HSPA1A R	AGGAAATGCAAAGTCTTGAAGC
NCF2 F	CCAGAAGCATTAACCGAGACAA
NCF2 R	CCTCGAAGCTGAATCAAGGC
FOS F	CCTAACCGCCACGATGATGT
FOS R	TCTGCGGGTGAGTGGTAGTA
HSPA8 F	TATTGGAGCCAGGCCTACAC
HSPA8 R	GTGTTGGTGGGGTTCATTGC
Endogenous -GAPDH F	CGAAGGTGGAGTCAACGGATTT
Endogenous -GAPDH R	ATGGGTGGAATCATATTGGAAC

Statistical analysis

All open databases and R software were utilized for analysis and visualization in this study. The Venn diagram was plotted via Tbtools software (v1.09, 2022) [20]. $P < 0.05$ was considered a significant difference.

Results

Identification of DEGs in GSE16561 dataset

Firstly, PCA was utilized to measure the degree of similarity between IS and NC specimens before and after

standardization. The expression patterns of DEGs were significantly different between IS and NC specimens (Fig. 2A, B). In addition, 568 DEGs were detected between IS and NC specimens, including 258 down-regulated and 310 up-regulated genes in IS specimens (Fig. 2C).

Identification of IS-related key modules by WGCNA

IS-related genes were identified via WGCNA analysis. A total of 62 specimens were utilized for subsequent analysis after excluding one outlier (GSM416554), as shown in Fig. 3A. When the soft threshold power was 3, R^2 was close to the threshold of 0.85, and mean connectivity tended to 0 (Fig. 3B). Then, 9 co-expression modules were obtained in Cluster Dendrogram (Fig. 3C). Subsequently, 6625 IS-related genes were detected in MElightyellow and MEblack depending on the module-trait relationships (Fig. 3D, Table 3). Besides, 42 key genes were obtained by overlapping 568 DEGs between IS and NC specimens, 6625 IS-related genes, and 1399 OS-related genes (Fig. 3E).

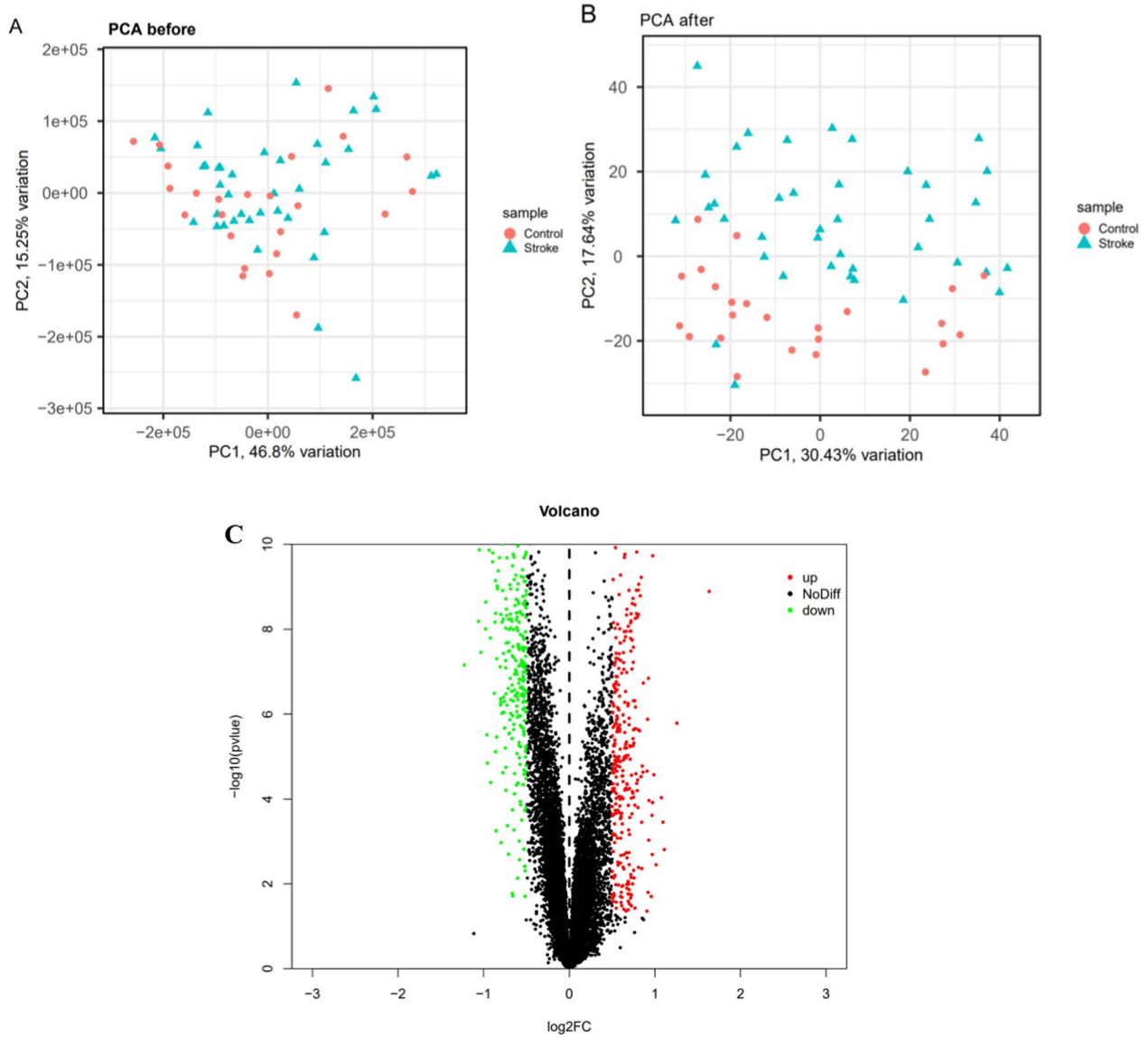

Fig. 2 Identification of DEGs inGSE16561dataset. **A**, **B** PCA was utilized to measure the degree of similarity between IS and NC specimens before and after standardization; **C** 568 DEGs were detected between IS and NC specimens,including258 down-regulated and 310 up-regulated genes in IS specimens

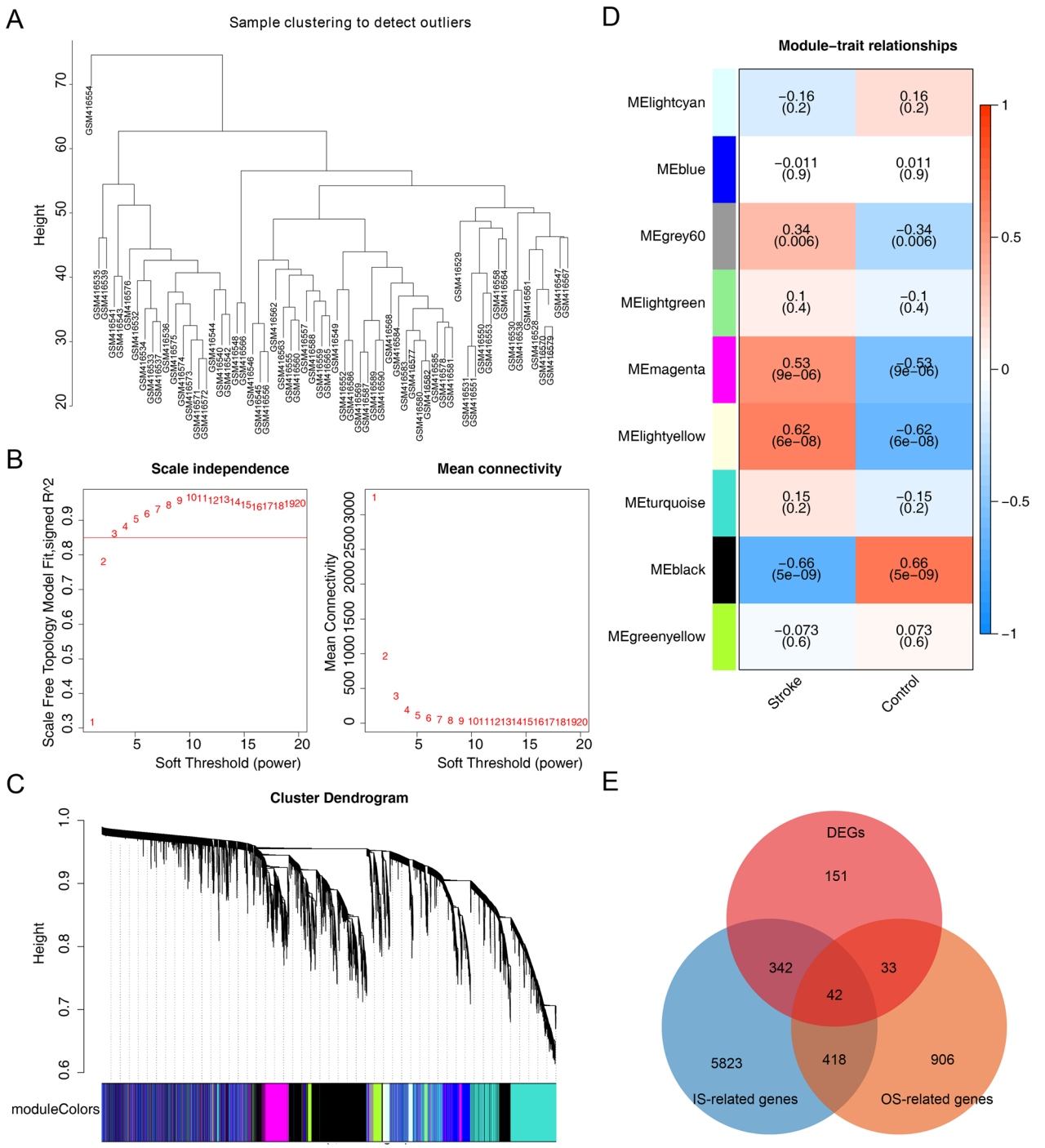

Fig. 3 Identification of IS-related key modulesby WGCNA. **A** 62 specimens were utilized for subsequent analysis after excluding one outlier (GSM416554); **B** When the soft threshold power was 3, R^2 was close to the threshold of 0.85, and mean connectivity tended to 0; **C** 9 co-expression modules were obtained in Cluster Dendrogram; **D** 6625 IS-related genes were detected in ME lightyellow and ME black depending on the module-trait relationships; **E** 42 key genes were obtained by overlapping 568 DEGs between IS and NC specimens, 6625 IS-related genes, and 1399 OS-related genes

Probing the molecular mechanism of IS based on key genes

GO and KEGG analyses were conducted to investigate the function of OS-related key genes. According to GO results, 42 key genes were associated with response to oxidative stress, cellular response to chemical stress,

Table 3 6625 IS-related genes were contained in ME lightyellow and ME black depending on the module-trait relationships

Type	Model color	Cor	p	Gene number
Stroke	Lightyellow	0.62	6.00E-08	102
	Black	−0.66	5.00E-09	6523

and reactive oxygen species-related pathways (Fig. 4A). Moreover, KEGG showed that key genes were correlated with lipid and atherosclerosis, neutrophil extracellular trap formation, and reactive oxygen species and cell migration-related pathways (Fig. 4B).

Identification of OS-related feature genes

To delve whether there were interactions between 42 key genes, we created a PPI network. The top20 proteins in the PPI network were selected, of which 5 proteins (HSPA8, LTA, STAT4, CCR7, and XBP1) were lowly expressed and 15 proteins (PTGS2, ITGAM, LCN2, KLF4, ROCK1, NCF2, FOS, CDC42, MPO, G6PD, HP, HIF1A, THBS1, HSPA1A, and LRRK2) were highly expressed in the IS specimens (Fig. 5A). Subsequently, 8 candidate feature genes (*FOS*, *ITGAM*, *HSPA1A*, *HSPA8*, *NCF2*, *THBS1*, *KLF4*, and *ROCK1*) were recognized via LASSO algorithm (Fig. 5B). Moreover, 12 candidate feature genes (*NCF2*, *HSPA8*, *CDC42*, *KLF4*, *LRRK2*, *HSPA1A*, *FOS*, *HIF1A*, *G6PD*, *THBS1*, *PTGS2*, and *XBP1*) were recognized via SVM-RFE algorithm (Fig. 5C). Finally, 6 feature genes (*HSPA8*, *NCF2*, *FOS*, *KLF4*, *THBS1*, and *HSPA1A*) were acquired by overlapping the candidate feature genes obtained through two machine learning algorithms (Fig. 5D).

The network of feature gene-miRNA

To investigate the regulatory mechanism of feature genes, we established the network between feature genes and miRNAs. The network (261 nodes and 277 edges) contained 6 feature genes and 255 miRNAs. Results revealed that hsa-mir-92a-3p regulated *HSPA8*, *HSPA1A*, and *THBS1* in Fig. 6A. Subsequently,

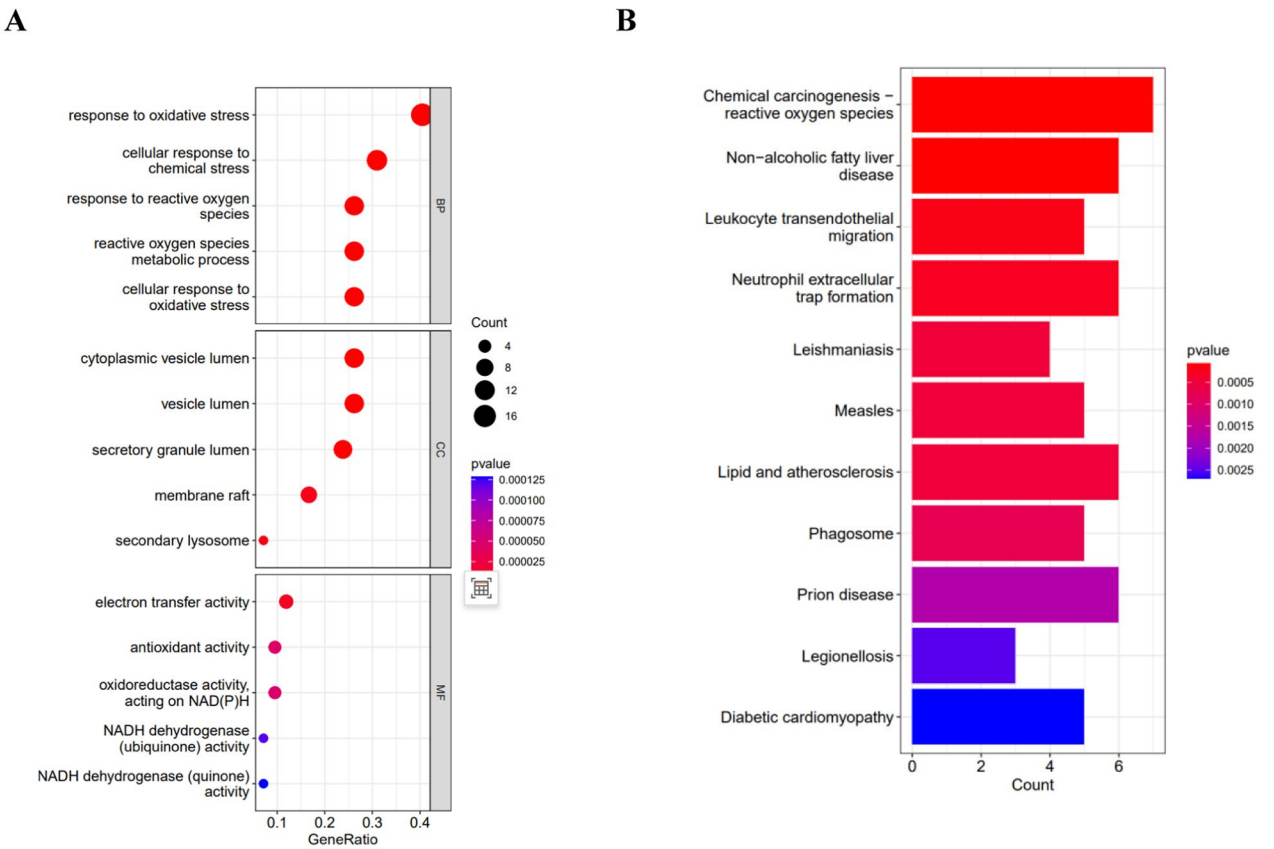

Fig. 4 GO and KEGG analyses were conducted to investigate the function of OS-related key genes. **A** According to GO results, 42 key genes were associated with response to oxidative stress, cellular response to chemical stress, and reactive oxygen species-related pathways; **B** Moreover, KEGG showed thatkey genes were correlated with lipid and atherosclerosis,neutrophil extracellular trap formation, and reactive oxygen species and cell migration-related pathways

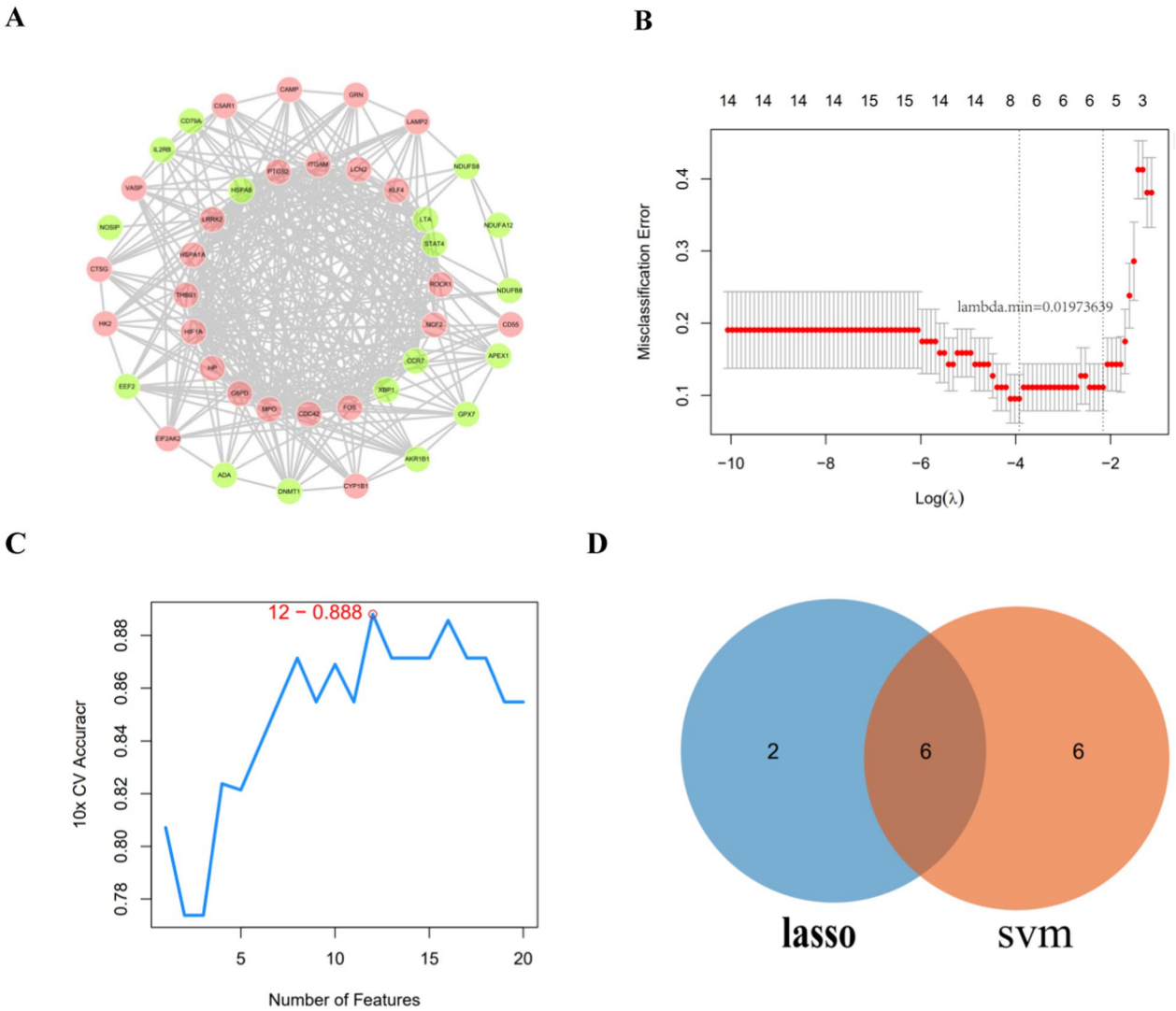

Fig. 5 PPI network and modules screening. **A** The top20 proteins in the PPI network were selected, of which 5 proteins were lowly expressed and 15 proteins were highly expressed in the IS specimens; **B** 8 candidate feature genes were recognized via LASSO algorithm; **C** candidate feature genes were recognized via SVM-RFE algorithm; **D** 6 feature genes were acquired by overlapping the candidate feature genes obtained through two machine learning algorithms

we delved the feature genes expression in GSE180470. The results demonstrated the expression trends of *HSPA8*, *KLF4*, *HSPA1A*, and *FOS* were consistent with those of GSE16561 dataset (Fig. 6B).

Participant inclusion of blood samples for qPCR

A total 10 blood specimens were taken from IS and NC groups met the inclusion criteria. IS group: age 66–79 years (73 + 5.12), 5 males (50%) and females(50%); NC group: age 52–79 years (73.8 + 8.75), 4 males(40%) and females(60%).

The qPCR validation of feature gene expression in IS and NC specimens

qPCR validation was conducted using blood specimens to assess the differences in 6 feature genes (*HSPA8*, *NCF2*, *FOS*, *KLF4*, *THBS1*, and *HSPA1A*) expression between IS and NC samples. The results showed that the expression trends of *HSPA8*, *FOS*, *KLF4*, and *HSPA1A* between IS and NC specimens were consistent with the expression in the GEO database, except for *HSPA8* expression with significant difference (Fig. 7). However, the expression trends of *THBS1* and *NCF2*

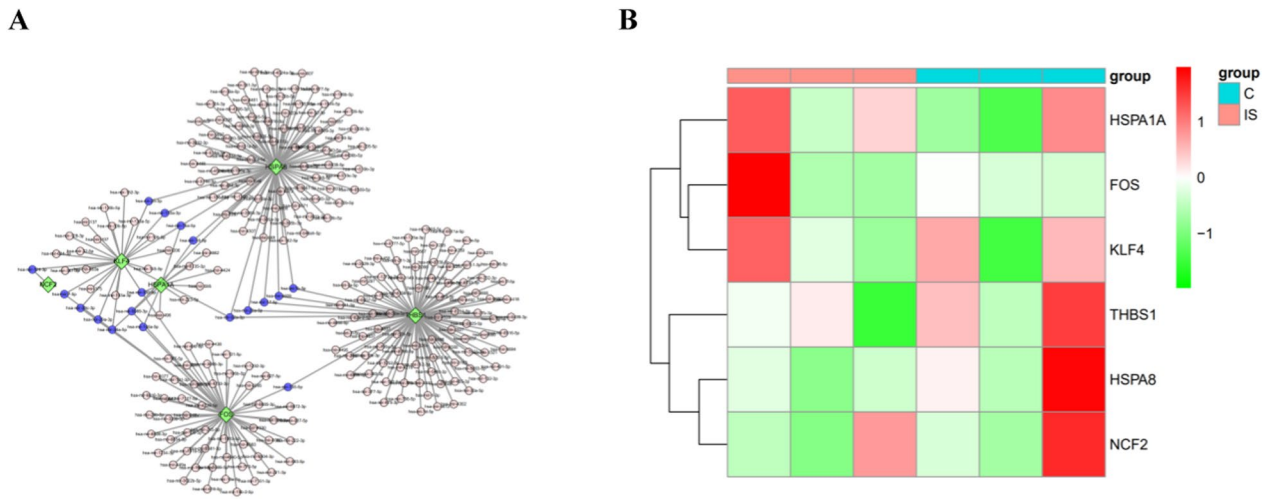

Fig. 6 The network of feature gene-miRNA. **A** hsa-mir-92a-3p regulated HSPA8,HSPA1A, andTHBS1; **B** The results of the GSE180470 dataset demonstrated the expression trends of genes were consistent with those of GSE16561 dataset

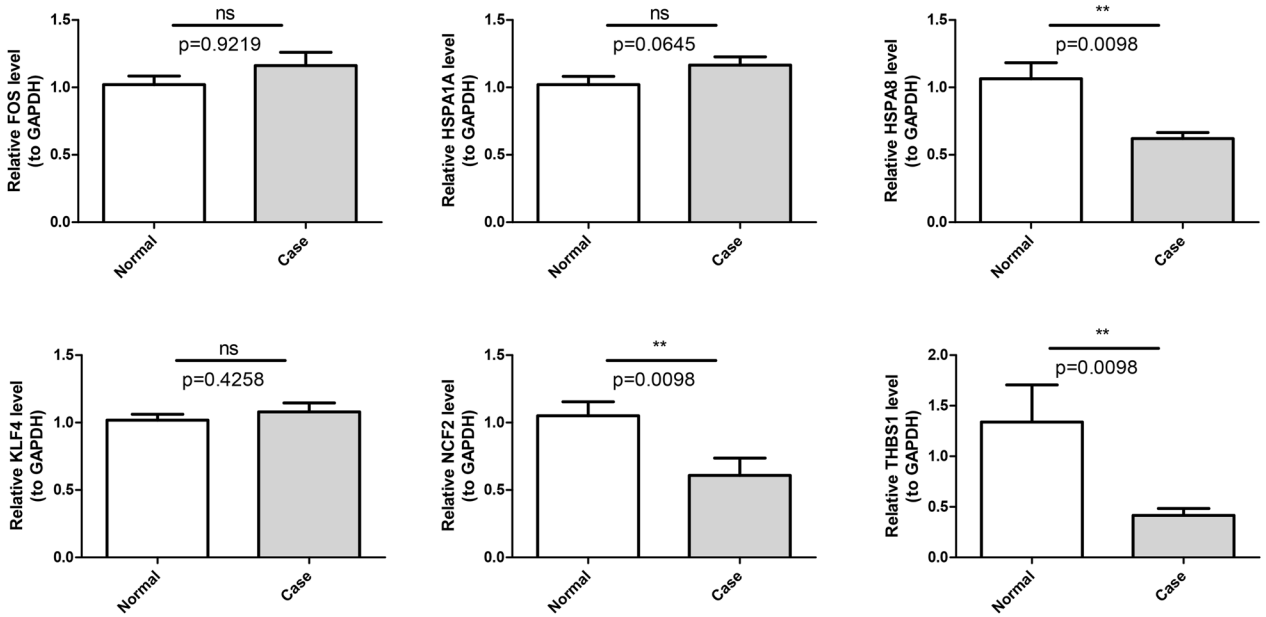

Fig. 7 qPCR validation offeaturegene expression in IS and NC specimens

were not consistent with the GEO database, which might be influenced by tissue heterogeneity (Fig. 7).

Discussion

Oxidative stress (OS) not only directly causes cell death by triggering apoptosis, necrosis and autophagy [21], but also aggravates IS injury through indirect pathways such as inflammatory response, thrombosis promotion and vascular endothelial function damage [22]. OS, which affects the pathophysiological process of IS through multiple dimensions, is the core mechanism of the disease progression of IS. In this study, 6 feature genes *(HSPA8, NCF2, FOS, KLF4, THBS1, and HSPA1A)* related to OS were identified by bioinformatics analysis, which might provide a new insight into the evaluation and treatment of IS.

Herein, GO showed that 42 genes were associated with response to oxidative stress (OS), cellular response to chemical stress, and reactive oxygen species-related pathways. Moreover, KEGG showed that the key genes

were correlated with lipid, atherosclerosis, and reactive oxygen species-related pathways. Studies have shown that IS, including atherosclerosis, cardiogenic embolism, and arteriolar occlusion, is associated with many risk factors [23]. Atherosclerosis is the most common IS subtype with many treatment options. Atherosclerosis can lead to many types of stroke, such as ischemic and hemorrhagic. About 70% of the existing cerebrovascular patients in China have IS. Notably, atherosclerotic thrombosis is one of the key causes of IS, indicating that the population with atherosclerosis is the high-risk group of IS [24].

In this study, 6 feature genes (*HSPA8*, *NCF2*, *FOS*, *KLF4*, *THBS1*, and *HSPA1A*) were identified by overlapping the candidate feature genes obtained through two machine learning algorithms. The *FOS* (FBJ osteosarcoma oncogene) is involved in the regulation of lipid metabolism and was related to intracellular lipid accumulation. Furthermore, FOS can regulate the formation of VSMC-derived foam cells in Vascular smooth muscle cells (VSMC) after activation by mtROS, which confirms the important role of *FOS* gene in oxidative stress response and lipid metabolism disorder [25]. Notably, lipid metabolism disorder is one of the risk factors of IS [26]. Studies have shown that high FOS expression may be used as a reference index for the time of brain injury [27]. In this study, *FOS* expression was high in IS patients. Therefore, future studies should explore if *FOS* gene can predict the neural recovery stage of IS.

Krüppel-like factor 4 (*KLF4)* gene can alleviate neuronal damage in IS patients by promoting lncRNA-ZFAS1 expression to inhibit Drp1m6A modification [28]. And *KLF4* can reduce cerebrovascular damage by improving vascular endothelial inflammation after IS [29]. However, the anti-inflammatory and pro-inflammatory effects of *KLF4* are environment-dependent in pathophysiology. Moreover, the promotion or inhibition of *KLF4* in atherosclerosis depends on the target genes and target cells [30]. The regulatory role of *KLF4* in diseases is difficult to clarify since it has dual regulatory functions. The serum *KLF4* level in patients with acute ischemic stroke is negatively correlated with infarct size, which indicates that *KLF4* level can reflect the severity of ischemic stroke and play a protective role in the pathogenesis. *KLF4* may also be used as a potential biomarker for predicting the prognosis of acute ischemic stroke. In conclusion, *KLF4* plays a neuroprotective role in ischemic stroke by reducing infarct size, inhibiting oxidative stress, restoring blood–brain barrier function and promoting long-term nerve recovery [31], consistent with our findings. We can further explore the related molecular mechanisms of *KLF4* in IS in future studies.

HSPA1A (heat shock 70 kDa protein 1A) and *HSPA8* (heat shock 70 kDa protein 8) belong to the heat shock 70 kDa protein family *(HSP70*) [26]. HSP70 is a chaperone protein induced by various stresses on cells and is involved in the development of various diseases [31]. *Hsp70* can improve the activity of antioxidant enzymes, such as super oxide dismutase (SOD) and accelerate the removal of oxygen free radicals, indicating that *Hsp70* can mitigate the damage caused by excessive ROS [32]. Studies have shown that the expression of Hsp70mRNA and *Hsp70* in brain tissue was extremely low in the absence of stimulant stimulation. Denaturetic protein in the damaged cells could rapidly induce the expression of *Hsp70* when focal or global brain deficiency is present. Studies have shown that OS can induce HSPA1A, leading to its release to the extracellular [33], consistent with our results.

Thrombospondin-1 (*THBS1)* is a multifunctional glycoprotein released from platelets, macrophages, and fat cells. And it is a potent regulator of angiogenesis that functions to concurrently inhibit endothelial cell migration and the release of vascular endothelial growth factor from the extracellular matrix [29]. *THBS1* was shown to be elevated in Acute ischemic stroke (AIS) and had positive predictive value at 3-months prognosis [25]. And as AIS patients with relatively higher *IGF2* and *LYVE1 l*evels and lower *THBS*1 levels were more likely to have good outcomes [29]. The *NCF2* gene is a potential diagnostic and prognostic biomarker for unstable atherosclerotic plaque (UAP) -associated stroke, and smoking may up-regulate *NCF2,* thereby accelerating plaque instability and UAP-associated stroke [34]. In this study, the expression of *THBS1* and *NCF2* genes in qPCR validation was higher in normal subjects, which may be due to differences in model species and the complex role of it's derivant in the microenvironment.

miRNAs are a group of endogenous non-coding small RNAs that regulate gene expression at the transcriptional level. Many studies have shown that miRNAs were highly expressed in atherosclerosis and played an important role in the occurrence and development of atherosclerosis. miRNA-15a is up-regulated in anterior circulation infarction and small atherosclerotic stroke. miR-15a is associated with IL-6, IGF-1 and acute cerebral ischemia in blood, indicating that it is a potential diagnostic biomarker and therapeutic target for stroke [35]. Studies have verified that miR-15a-5p could reduce the inflammatory response and artery damage in diabetic AS rats by targeting FASN, and the increased expression of miR-15a-5p may protective the treatment and treatment of AS patients [36]. Yang showed that inhibition of microRNA-15a-5p expression could reduce OS and inflammation in IS patients, indicating the importance of microRNA-15a-5p in IS [33].

In this study, results showed that *HSPA8, HSPA1A*, and *THBS1* are regulated by hsa-mir-92a-3p. *m*iR-92a belonging to microRNAs family, participates in pathophysiological process of cardiovascular diseases such as atherosclerosis, acute coronary syndrome, and myocardial infarction. And miR-92a was revealed to be highly expressed in normal vascular endothelial cells, it affected the degradation and translation of target genes by interacting with target genes with specific binding sites, which is essential for the stability of normal vascular homeostasis. Previous studies have detected the expression of *KLF*4 and miR-92a in rat myocardial tissue, confirming the target relationship between miR-92a and *KLF4,* and suggesting that the deletion of miR-92a can promote the upregulation of *KLF4,* thereby improving cardiac function and reducing myocardial cell apoptosis [37]. Meanwhile, Wang proposed that lncRNA X-Inaction-specific transcript (XIST) could affect angiogenesis and alleviate cerebrovascular injury after IS (CIS) by mediating miR-92a to regulate anti-inflammatory factor *KLF4* [38]. In addition, miR-92a-3p was significantly increased in CAD patients compared to non-CAD patients. Knockout of miR-92a-3p in EMV could eliminate EMVS-mediated effects on ECs migration and proliferation in stromal plugs, and blocked vascular network formation. And the inhibition of THBS1 gene and protein expression was eliminated [39].

Nonetheless, this study has some limitations. First, we used the GSE180470 dataset as the validation set. The sample size of this validation set was only three pairs, which limited our ability to fully explore the diversity of gene expression in the target population. Second, the mRNA expression levels of key genes were verified by qPCR experiments in our study. Although the qPCR validation results were significant, the sample size of this study was relatively small, limiting the generalizability and statistical power of the results. A large sample size is needed for further verification of results. In qPCR, GAPDH was selected as the reference gene because it showed relatively stable expression levels in our study system. We are optimistic that the discovery of these genes will provide clinicians with new diagnostic markers and prognostic indicators, which will significantly improve the accuracy of disease assessment and personalized treatment planning of IS patients. However, we recognize that the selection of the optimal reference gene may vary for different study systems and conditions. In future studies, we could further explore other potential reference genes to assess their impact on the stability of the results. At the same time, the samples will be collected and compared in different stages of IS to clarify the expression of each gene in different periods, and we will increase the related research and analysis at the protein level.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12868-024-00921-9.

Supplementary material 1.

Supplementary material 2.

Acknowledgements
Trial participants are thanked for providing blood samples.

Author contributions
GY Li and YC sorted out the data and drafted the manuscript. Y Zhang and Y Zhou made substantial contributions to study conception and design. SS Ding and QY Zheng contributed to the analysis data. LQ Kuang contributed to the charts. All authors contributed equally to the article. All authors read and agreed to the final manuscript.

Funding
This study was funded by the Shanghai Sailing Program (PROJECT NO: 20YF1444700), 2022 Shanghai Science and Technology Commission (PROJECT NO:22Y31900200) and 2023 Xuhui District Health System Peak Discipline Construction Funding Project.

Availability of data and materials
All data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
This study was implemented according to the Declaration of Helsinki, Good Clinical practice guidelines and the Consolidated Standards of Reporting Trials. This study was approved by the Ethics Committee of Shanghai Xuhui Central Hospital, Shanghai (reference: 2023-30), participants and their relatives signed written informed consent for the study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Rehabilitation, Shanghai Xuhui Central Hospital, Shanghai, China. [2]Department of Image, Shanghai Xuhui Central Hospital, Shanghai, China. [3]Department of General Practice, Shanghai Xuhui Central Hospital, Shanghai, China.

Received: 12 January 2024 Accepted: 20 December 2024
Published online: 13 January 2025

References

1. Zhao Y, Zhang X, Chen X, Wei Y. Neuronal injuries in cerebral infarction and ischemic stroke: From mechanisms to treatment (Review). Int J Mol Med. 2022;49(2):15.
2. Grotta JC. Intravenous thrombolysis for acute ischemic stroke. Continuum (Minneap Minn). 2023;29(2):425–42.
3. Sakamoto Y, Nito C, Nishiyama Y, Suda S, et al. Safety of antithrombotic therapy within 24 hours after recombinant tissue-plasminogen activator treatment for large-artery atherosclerosis stroke: insights from emergent PTA/CAS cases. J Nippon Med Sch. 2024;91(3):307–15.

4. Mitchell PJ, Yan B, Churilov L, DIRECT-SAFE Investigators, et al. Endovascular thrombectomy versus standard bridging thrombolytic with endovascular thrombectomy within 4–5 h of stroke onset: an open-label, blinded-endpoint, randomised non-inferiority trial. Lancet. 2022;400:116–25.
5. Tian X, He W, Rong Y, et al. Dl-3-n-butylphthalide protects the heart against ischemic injury and H9c2 cardiomyoblasts against oxidative stress: involvement of mitochondrial function and biogenesis. J Biomed Sci. 2017;24(1):38–40.
6. Kim S, Lee W, Jo H, Sonn SK, et al. The antioxidant enzyme Peroxiredoxin-1 controls stroke-associated microglia against acute ischemic stroke. Redox Biol. 2022;54:102347.
7. Cheng YC, Sheen JM, Hu WL, et al. Polyphenols and oxidative stress in atherosclerosis-related ischemic heart disease and stroke. Oxid Med Cell Longev. 2017;26:8526438.
8. Narne P, Pandey V, Phanithi PB. Interplay between mitochondrial metabolism and oxidative stress in ischemic stroke: An epigenetic connection. Mol Cell Neurosci. 2017;82:176–94.
9. Tonelli C, Chio IIC, Tuveson DA. Transcriptional regulation by Nrf2. Antioxid Redox Signal. 2018;29(17):1727–45.
10. Hayashi Y, Homma K, Ichijo H. SOD1 in neurotoxicity and its controversial roles in SOD1 mutation-negative ALS. Adv Biol Regul. 2016;60:95–104.
11. Pi H, Xu S, Reiter RJ, Guo P, et al. SIRT3-SOD2-mROS-dependent autophagy in cadmium-induced hepatotoxicity and salvage by melatonin. Autophagy. 2015;11(7):1037–51.
12. Lei FJ, Chiang JY, Chang HJ, et al. Cellular and exosomal GPx1 are essential for controlling hydrogen peroxide balance and alleviating oxidative stress in hypoxic glioblastoma. Redox Biol. 2023;65:102831.
13. Wu Z, Wang L, Wen Z, Yao J. Integrated analysis identifies oxidative stress genes associated with progression and prognosis in gastric cancer. Sci Rep. 2021;11(1):3292.
14. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
15. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
16. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.
17. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
18. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22.
19. Hu K. Become competent in generating RNA-Seq heat maps in one day for novices without prior R experience. Methods Mol Biol. 2021;2239:269–303.
20. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13(8):1194–202.
21. Sies H. Oxidative stress: a concept in redox biology and medicine. Redox Biol. 2015;4:180–3.
22. Filomeni G, De Zio D, Cecconi F. Oxidative stress and autophagy: the clash between damage and metabolic needs. Cell Death Differ. 2015;22(3):377–88.
23. Su XT, Wang L, Ma SM, Cao Y, Yang NN, Lin LL, Fisher M, Yang JW, Liu CZ. Mechanisms of acupuncture in the regulation of oxidative stress in treating ischemic stroke. Oxid Med Cell Longev. 2020;2020:7875396.
24. Kahles T, Brandes RP. NADPH oxidases as therapeutic targets in ischemic stroke. Cell Mol Life Sci. 2012;69(14):2345–63.
25. Miao G, Zhao X, Chan SL, Zhang L, Li Y, Zhang Y, Zhang L, Wang B. Vascular smooth muscle cell c-Fos is critical for foam cell formation and atherosclerosis. Metabolism. 2022;132:155213.
26. Gu X, Li Y, Chen S, Yang X, Liu F, Li Y, Li J, Cao J, Liu X, Chen J, Shen C, Yu L, Huang J, Lam TH, Fang X, He Y, Zhang X, Lu X, Wu S, Gu D. Association of lipids with ischemic and hemorrhagic stroke: a prospective cohort study among 267 500 Chinese. Stroke. 2019;50(12):3376–84.
27. Zhang H, Li W, et al. FOS protein expression in brain contusion cortex and hippocampus research. Chin J Histochem Cytochem. 2004;13(1):110–3.
28. Wang QS, Xiao RJ, Peng J, Yu ZT, Fu JQ, Xia Y. Bone marrow mesenchymal stem cell-derived exosomal KLF4 alleviated ischemic stroke through inhibiting N6-methyladenosine modification level of Drp1 by targeting lncRNA-ZFAS1. Mol Neurobiol. 2023;60(7):3945–62.
29. Zhang X, Wang L, Han Z, Dong J, Pang D, Fu Y, Li L. KLF4 alleviates cerebral vascular injury by ameliorating vascular endothelial inflammation and regulating tight junction protein expression following ischemic stroke. J Neuroinflammation. 2020;17(1):107.
30. Yang C, Xiao X, Huang L, et al. Role of Kruppel-like factor 4 in atherosclerosis. Clin Chim Acta. 2021;512:135–41.
31. Nagai M, Kaji H. Thermal effect on heat shock protein 70 family to prevent atherosclerotic cardiovascular disease. Biomolecules. 2023;13(5):867.
32. Choi S, Park KA, Lee HJ, Park MS, Lee JH, Park KC, Kim M, Lee SH, Seo JS, Yoon BW. Expression of Cu/Zn SOD protein is suppressed in hsp 70.1 knockout mice. J Biochem Mol Biol. 2005;38(1):111–4.
33. Suzuki H, Kosuge Y, Kobayashi K, et al. Heat-shock protein 72 promotes platelet aggregation induced by various platelet activators in rats. Biomed Res. 2017;38(3):175–82.
34. Krause M, Heck TG, Bittencourt A, Scomazzon SP, Newsholme P, Curi R, Homem de Bittencourt PI. The chaperone balance hypothesis: the importance of the extracellular to intracellular HSP70 ratio to inflammation-driven type 2 diabetes, the effect of exercise, and the implications for clinical management. Mediators Inflamm. 2015;2015:249205.
35. Lu WJ, Zeng LL, Wang Y, Zhang Y, Liang HB, Tu XQ, He JR, Yang GY. Blood microRNA-15a Correlates with IL-6, IGF-1 and Acute Cerebral Ischemia. Curr Neurovasc Res. 2018;15(1):63–71.
36. Liu Y, Liu LY, Jia Y, Sun YY, Ma FZ. Role of microRNA-15a-5p in the atherosclerotic inflammatory response and arterial injury improvement of diabetic by targeting FASN. Biosci Rep. 2019;39(7):BSR20181852.
37. Wu Q, Wang H, He F, Zheng J, Zhang H, Cheng C, Hu P, Lu R, Yan G. Depletion of microRNA-92a enhances the role of sevoflurane treatment in reducing myocardial ischemia-reperfusion injury by upregulatingKLF4. Cardiovasc Drugs Ther. 2022;37(6):1053-1064.
38. Wang C, Dong J, Sun J, Huang S, Wu F, Zhang X, Pang D, Fu Y, Li L. Silencing of lncRNA XIST impairs angiogenesis and exacerbates cerebral vascular injury after ischemic stroke. Mol Ther Nucleic Acids. 2021;26:148–60.
39. Liu Y, Li Q, Hosen MR, Zietzer A, Flender A, Levermann P, Schmitz T, Frühwald D, Goody P, Nickenig G, Werner N, Jansen F. Atherosclerotic conditions promote the packaging of functional microRNA-92a-3p into endothelial microvesicles. Circ Res. 2019;124(4):575–87.

Publisher's Note