

# Identification of Molecular Biomarkers Associated With Stroke Progression Using WGCNA and SVM-RFE

**Si-tong Liu**

Guangxi Medical University

**You Zhang**

First Affiliated Hospital of GuangXi Medical University

**Xin-gui Wu** (✉ [wxingui200061@aliyun.com](mailto:wxingui200061@aliyun.com))

First Affiliated Hospital of GuangXi Medical University

**Chang-xing Lu**

Guangxi Medical University

**Qi-Ping Hu**

Guangxi Medical University

---

## Research Article

**Keywords:** stroke, WGCNA, machine learning, biomarkers, GSVA

**Posted Date:** January 6th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1221609/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** Stroke is the second most common cause of death worldwide and the leading cause of long-term severe disability with neurological impairment worsening within hours after stroke onset and being especially involved with motor function. So far, there are no established and reliable biomarkers to prognose stroke. Early detection of biomarkers that can prognose stroke is of great importance for clinical intervention and prevention of clinical deterioration of stroke.

**Methods:** TGSE119121 dataset was retrieved from the Gene Expression Integrated Database (Gene Expression Omnibus, GEO) and weighted gene co-expression network analysis (WGCNA) was conducted to identify the key modules that could regulate disease progression. Moreover, functional enrichment analysis was conducted to study the biological functions of the key module genes. The GSE16561 dataset was further analyzed by the Support Vector Machines coupled with Recursive Feature Elimination (SVM-RFE) algorithm to identify the top genes regulating disease progression. The hub genes revealed by WGCNA were associated with disease progression using the receiver operating characteristic curve (ROC) analysis. Subsequently, functional enrichment of the hub genes was performed by deploying gene set variation analysis (GSVA). The changes at gene level were transformed into the changes at pathway level to identify the biological function of each sample. Finally, the expression level of the hub gene in the rat infarction model of MCAO was measured using RT-qPCR for validation.

**Results:** WGCNA analysis revealed four hub genes: DEGS1, HSDL2, ST8SIA4 and STK3. The result of GSVA showed that the hub genes were involved in stroke progression by regulating the p53 signal pathway, the PI3K signal pathway, and the inflammatory response pathway. The results of RT-qPCR indicated that the expression of the four HUB genes was increased significantly in the rat model of MCAO.

**Conclusion:** Several genes, such as DEGS, HSDL2, ST8SIA4 and STK3, were identified and associated with the progression of the disease. Moreover, it was hypothesized that these genes may be involved in the progression stroke by regulating the P53 signal, the PI3K signal, and the inflammatory response pathway, respectively. These genes have potential prognostic value and may serve as biomarkers for predicting stroke progression. The early identification of the patients at risk of progression is essential to prevent clinical deterioration and provide a reference for future research.

## 1. Introduction

Stroke is a serious threat to human health. Globally, it is the second most common cause of death and the main cause of long-term severe disability (Feigin et al.,2014). Stroke can be ischemic or hemorrhagic, but ischemic stroke is the main type, accounting for about 87% of all stroke cases (Radu et al.,2017). Ischemic stroke is a severe clinical syndrome with a high burden of neurological disability and mortality (Evers et al.,2004). Current treatments for ischemic stroke injuries have proven to be inadequate, partly because of the incomplete understanding of the cellular and molecular mechanisms involved in ischemic stroke. Early identification of patients at risk of progression is becoming increasingly important because

of the high fatality rate and narrow treatment window for stroke. The identification of the relevant genes regulating the progression of stroke can promote the understanding of the underlined mechanisms of stroke, facilitate the prognosis of stroke, and even be the basis for the development of new treatments.

With the development of high-throughput sequencing and bioinformatics, the availability of sequencing data has enabled the discovery of key genes, networks, and pathways related to stroke. However, identifying these features remains challenging. Weighted gene co-expression network analysis (WGCNA) is a network-based screening method and a data exploration tool, which can divide genes with similar expression patterns into several biologically meaningful co-expression modules, analyze the relationship between gene modules and clinical features, and finally evaluate the importance of genes in feature-related modules. The latter group of significant genes constitute candidate biomarkers and therapeutic drug targets (Zhang et al.,2005). WGCNA has been widely used in a variety of medical fields, such as oncology (Xiang et al.,2019), neurological (Katrinli et al.,2019), and psychiatric disorders (Guo et al.,2019), chronic and infectious diseases (Vieira et al.,2019). Moreover, most of the conclusions drawn by WGCNA can be further confirmed by bioinformatics analysis or biological experiments, reinforcing thus the high reliability of WGCNA.

Support Vector Machines coupled with Recursive Feature Elimination (SVM-RFE) is a hybrid combination of a dimensionality reduction technique with an established classification method called Support Vector Machines (SVM). The so-called recursive feature elimination is used to initially obtain a features ranking in its iteration. In every iteration, the score of each feature is calculated according to certain rules, and the feature with the lowest score is removed, with this process being repeated until all genes have been ranked. SVM-RFE belongs to the category of wrapper dimensionality reduction techniques since it uses the weight vector of the hyperplane produced by the SVM model to select and delete features. Then, the training dataset is used to train the SVM classification model, and all the steps are repeated until the features in the candidate list are completely deleted, and the last gene to be deleted is the most significant one for the classifier (Guyon et al.,2002).

In this study, gene expression profile GSE119121 dataset was retrieved from the Gene Expression Omnibus database (GEO) and weighted gene co-expression network analysis (WGCNA) was performed to identify the key modules associated with disease progression, and to functionally annotate the key modules with genetic ontology (GO) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway terms. Then, the GSE16561 dataset was analyzed with the SVM-RFE algorithm to identify the gene with the highest score regarding disease progression. The key genes of each module revealed WGCNA were used as inputs to the SVM-RFE method to obtain the best stable results for predicting phenotypes.

## **2. Results**

### **2.1 Weighted co-expression network construction and identification of key modules**

There were totally 47 samples of transcriptional data in GSE119121, including control (n= 8), 1-hour ischemia (n = 8), 2-hours ischemia (n = 8), 3-hour ischemia (n = 8), 6-hour ischemia (n=7) and 24-hour ischemia (n= 8) groups, respectively. All samples were included in the co-expression network. The soft threshold  $\beta$  is determined by the function "sft\$PowerEstimate" to reassure that the reconstructed network possesses the scale-free topology feature. The absolute value of the correlation coefficient was greater than 0.8, and the data selection threshold was set to 10 (Figure 2) for the present manuscript's data. Then, the dynamic hierarchical tree cutting algorithm was used to detect the co-expression modules, and the minimum height of the merged modules was set to 0.25. Finally, 14 modules are obtained for the detection of stroke-related gene modules by WGCNA (Figure 3). Firstly, the samples were clustered to delete the contour samples (Figure 4). Then, WGCNA was used to identify the genes that are highly co-expressed in each group. Modules are colored randomly. A total of 8 gene modules were identified (Figure 5), showing the genetic significance of each tissue module, with the most significant being the brown ( $r=0.65$ ,  $p=5e-05$ ) and the black  $r = -0.92$ ,  $p=2e-20$  ones. Both these modules were highly correlated with stroke (Figure 6). Therefore, the brown and black modules were selected for subsequent verification.

## 2.2 Pathway analysis for specific modules

All genes in the Brown module were selected for GO and KEGG analysis. Our results (figure 6) of GO and KEGG analysis demonstrated that the brown module (figure 7) was mainly enriched in signaling mechanisms, such as innate immune system, regulation of defense response, production of cytokines, inflammatory response, cellular redox homeostasis, endoplasmic reticulum stress response and so on. Among them, the regulation of defense response, the production of cytokines, inflammatory response, and the endoplasmic reticulum stress response were all related to stroke, which is in agreement with previous findings. It is thus speculated that the impact of the brown module on stroke progression was closely related to the above pathways. The interaction diagram of each module was shown in the PPI diagram. In addition, the pathway analysis results showed that the black module (figure 8) was associated with many processes including defense response, receptor internalization, congenital immune response, negative regulation of genome replication, negative regulation of immune response, and positive regulation of DNA metabolic processes.

### Feature selection for machine learning

Specific stroke samples were extracted for analysis from the GSE16561 dataset of the GEO database, including the control group (n=24) and stroke patients (n=39) to further identify the key genes that regulate stroke progression by performing SVM-RFE feature screening and subsequent validation. The results showed that the SVM-RFE feature selection method could achieve high accuracy and low error rate (figure 8). In addition, the top 100 genes were selected according to their score and the intersection with the genes of the black module and the brown module (annex 1). The results showed that 4 genes were prioritized, and these were DEGS1, HSDL2, ST8SIA4, and STK3 (figure 9).

After further comparison of the differences of the four hub genes between normal patients and stroke patients, the results showed that the four hub genes were significantly increased in the stroke group ( $P < 0.01$ ), suggesting that these four genes can act as risk factors for the development of stroke.

## 2.3 Detection efficiency of hub genes

the "pROC" package was used to analyze the receiver operating characteristic (ROC) curve to assess the predictive ability of the four hub genes on stroke progression. The results showed that the expression of the four genes could significantly distinguish the stroke progression, and the areas under the curve of the DEGS1, HSDL2, ST8SIA4, STK3 genes were 0.884, 0.911, 0.902, and 0.984, respectively (figure 10). However, the specific molecular mechanisms involved in the four genes should be further explored.

## 2.4 Pathway analysis of hub genes

The differential feature enrichment of four hub genes was conducted using the GSEA algorithm to study the correlation between DEGS1, HSDL2, ST8SIA4 and STK3, and the functional characteristics of stroke progression. The results showed that the overexpression of the DEGS1 group was mainly enriched in the pathways ANGIOGENESIS, IL2\_STAT5\_SIGNALING, TGF\_BETA\_SIGNALING, INFLAMMATORY\_RESPONSE, APOPTOSIS and PI3K\_AKT\_MTOR\_SIGNALING. The enriched pathways for the HSDL2 group were mainly ADIPOGENESIS, NOTCH\_SIGNALING, P53\_PATHWAY, INTERFERON\_ALPHA\_RESPONSE, PI3K\_AKT\_MTOR\_SIGNALING and IL6\_JAK\_STAT3\_SIGNALING. The main enriched pathways of the ST8SIA4 group were REACTIVE\_OXYGEN\_SPECIES\_PATHWAY, WNT\_BETA\_CATENIN\_SIGNALING, PI3K\_AKT\_MTOR\_SIGNALING, HYPOXIA, INFLAMMATORY\_RESPONSE and IL6\_JAK\_STAT3\_SIGNALING. The main enriched pathways of the STK3 group were INFLAMMATORY\_RESPONSE, APOPTOSIS, PI3K\_AKT\_MTOR\_SIGNALING, INTERFERON\_ALPHA\_RESPONSE, HYPOXIA, IL6\_JAK\_STAT3\_SIGNALING, and other signaling pathways (figure 11). It is observed that the GSEA results show that most of the hub genes were related to the p53 signal pathway, the PI3K signal pathway, and the inflammatory response pathway, indicating that stroke progression is associated with these pathways.

## 2.5 Validation of hub genes

As shown in figure 12, the results of real-time photoluminescence (RT-PCR) showed that the DEGS1, HSDL2, ST8SIA4 and STK3 genes are overexpressed in the stroke group in comparison to the control group with the most pronounced difference being the one of the HSDL2. The PCR results validated the main findings of the discovery transcriptomics analysis. The four hub genes can be used as biomarker primers to influence stroke progression (See attachment 2 for primers in Annex 2).

## 3. Discussion

Ischemic stroke is one of the most common cerebrovascular diseases, and its mortality rate is increasing year by year. Many stroke survivors are being left with various neurological defects, leading to different degrees of impaired quality of life, and this leads to a great burden on patients, their families, and society (Benjamin et al.,2018). The occurrence and development of ischemic stroke are closely related to genetic changes. The identification of genes that regulate the progress of stroke can promote the understanding of the underlined mechanisms of stroke, predict the progress of stroke, and contribute to the development of new treatments.

In this study, the GSE119121 dataset was retrieved from the GEO database, and 14 modules were obtained after the co-expression module analysis with the dynamic hierarchical tree cutting algorithm. 8 gene modules were identified by WGCNA among the groups of the overexpressed genes. The most prominent of these were the brown module and the black module. GO and KEGG analysis showed that the brown module was mainly enriched in signal mechanism such as innate immune system, regulation of defense response, production of cytokines, inflammatory response, endoplasmic reticulum stress response, and so on, with all of them being highly related to stroke. This is consistent with previous reports. In recent years, the role of inflammasome-mediated inflammatory pathways in various central nervous system diseases has attracted increasing attention (Cribbs et al.,2012). Cheon, Mu et al. reported increased levels of pro-inflammatory cytokines, such as IL-1 $\beta$  or IL-18, in the brains of both animals and stroke patients (Cheon et al.,2018) (Mu et al.,2018). In addition, anti-inflammatory therapy also has a neuroprotective effect on stroke. Then, the GSE16561 dataset was analyzed using the SVM-REF algorithm to identify the most related to the disease genes, and intersect them with the genes of the black and the brown modules, ending up with a final list of four hub genes: DEGS1, HSDL2, ST8SIA4 andSTK3. The four hub genes were verified in the second data set, and all four hub genes were significantly increased in the stroke group ( $P < 0.01$ ), suggesting that these four genes are risk factors for stroke development.

DEGS1 is a protein-coding gene. Karsaig has identified DEGS1 as a pathogenic gene involved in myelin degeneration in the central nervous system and peripheral nervous system (Kraveka et al.,2007). Previous studies have shown that sphingolipid is increasingly regarded as an important cellular mediator in tumor and inflammatory hypoxia (Riechers et al.,2016). Sphingolipid is also an important part of the cell membrane, which is functionally related to basic processes, such as cell differentiation, neuronal signal transduction, and myelin formation. Sphingolipid imbalance is the chief culprit of various neurological diseases (Karsai et al.,2019). Defects in sphingolipid synthesis or degradation can lead to various nervous system diseases. The high expression of DEGS1 after cerebral ischemia promotes the synthesis of ceramides, which may affect the balance between sphingolipids and dihydrosphingolipids. DEGS1 enzyme participates in catalyzing biosynthesis of ceramide from the head and controls the process from dihydroceramide to ceramide, which is essential for maintaining the balance regulation between sphingolipid and dihydrosphingolipid (Rodriguez et al.,2014). Previous studies have shown that ceramide can be used as a predictor of ischemic stroke (Mohamud et al.,2019). Alsanafi believes that manipulating dihydroceramide levels may represent a new treatment strategy for cancer. This leads to the accumulation of dihydroceramide, which ultimately leads to the activation of endoplasmic reticulum

stress and autophagy before cancer cells die (Alsanafi et al.,2020). The polyubiquitin of DEGS1 seems to change its function from promoting apoptosis to survival, which plays a vital role in the mechanism of stroke treatment. The confirmation of this validation could suggest that DEGS1 could also be used as a target for the treatment of stroke.

HSDL2 is peroxisome located in human, mouse, and rat cells, and a potential key regulator of lipid metabolism. HSDL2 can regulate lipid metabolism and participate in cell proliferation and apoptosis. Abnormal lipid changes are the pathological basis of stroke, HSDL2 may be involved in the occurrence and development of stroke by regulating lipid metabolism (Zhang et al.,2018). Recent evidence shows that HSDL2 is highly expressed in breast cancer (Dong et al.,2019), lung adenocarcinoma (Shi et al.,2020), thyroid papillary carcinoma (Zeng et al.,2019), ovarian cancer (Zhang et al.,2019), and human neurogliocytoma (Ruokun et al.,2016). The overexpression of HSDL2 promotes apoptosis, but its mechanism in cerebral infarction should be further studied.

STK3 is a pro-apoptotic kinase in inhibiting growth. This kinase promotes chromatin condensation during apoptosis and has been proved to be the substrate of Caspase-6 in biochemistry. Caspases 3, 6, and 7 are involved in the execution of apoptosis. Casp6 has become an important marker for Huntington's disease, Alzheimer's disease, and cerebral ischemia, and it is activated in the early stages of the disease. After brain injury, CASP6 induces neuronal death by triggering an apoptotic cascade, which in turn affects the recovery of nerve function, and can be used as a potential therapeutic target for neurological diseases (Girling et al.,2018).CASP6 also plays a key role in axonal degeneration, which further pinpoints the importance of this protease in the pathways of neurodegeneration (Riechers et al.,2016).

ST8SIA4, a protein-coding gene, mainly synthesizes PolySia, which is a complex sugar that participates in many processes of nerve development, such as neuroblast migration, outward neurite growth, axon pathfinding, axonal bundle formation, and synaptic formation. PolySia is almost completely attached to the neural cell adhesion molecule (NCAM) in the brain. NCAM is an adhesive glycoprotein located on the cell surface of neurons and immune cells and is involved in many important biological functions, including cell adhesion and migration, synaptic formation, learning and memory consolidation, and social interaction. PolySia-modified NCAM is thought to be related to many specific functions of the brain (Burgess et al.,2008). In addition, polySia can be used as a reservoir of neurological (biological) active molecules, such as neurotrophic proteins (e.g. BDNF) (Kanato et al.,2008), catecholamine neurotransmitters (e.g., dopamine) (Isomura et al.,2011), growth factors (e.g., fibroblast growth factor (FGF2, bFGF) (Ono et al.,2012)).

The differential feature enrichment of the four hub genes was further performed using the GSVA algorithm to study the correlation between the hub gene and the functional characteristics of stroke progression. The results showed that most of the hub genes are involved in the pathological process of stroke through the inflammatory response pathway, the PI3K signal pathway, and the P53 signal pathway, while they were also associated with the prognosis.

The inflammatory response pathway includes processes such as inflammatory cell infiltration, apoptosis, and oxidative stress. Several studies have shown that apoptosis is triggered by stroke. Caspase is a marker protein in the process of cell apoptosis, and important factors include caspase-3, caspase-8, and caspase-9. When subjected to stress, cytochrome C released by the mitochondria binds to procaspase-9 /Apaf-1 to activate and cleaved caspase-9 (Ying et al.,2020). The lytic caspase-9 further copes with other caspase members, initiating the caspase cascade, and then triggering apoptosis (Li et al.,2019). Activated caspase-8 cooperatively cleaves and activates the caspase of downstream effector molecules caspase-1, caspase-3, caspase-6, and caspase-7 and amplifies the apoptosis signals (Buschhaus et al.,2018). Clinically, in the acute phase of cerebral infarction, certain effects, such as reducing secondary brain injury and promoting nerve regeneration, have been achieved by inhibiting the process of apoptosis.

The PI3K/Akt signaling pathway is involved in various cellular processes, and it has been proved that the activation of this pathway is related to the occurrence and development of angiogenesis. The negative regulation of angiogenesis promotes the genes of thrombosis, vascular permeability, and inflammation, protecting thus vascular function (Lugovaya et al.,2020). On the one hand, RelA, NF- $\kappa$  B/ Rel family plays an important role in the response to inflammation and immunity, which may constitute the regulatory signal of PI3K-AKT, and then regulate the pathway of apoptosis (Gao et al.,2019). On the other hand, several studies have confirmed that drugs can improve stroke symptoms by regulating the PI3K/Akt pathway (Lugovaya et al.,2020). A recent study has also reported that PI3K/Akt regulates apoptosis and that the activation of the PI3K/Akt pathway after stroke plays a protective role in neuronal apoptosis (An et al.,2017).

Under the pathological condition of stroke, p53 plays an important role in the regulation of apoptosis and the cell cycle (Guo et al.,2020). The increase of cyclin chaperone D level will affect the process of cells entering the S phase under the regulation of p53(Saito et al.,2005). The degradation of p53 hinders its role in the regulation of apoptosis (Gao et al.,2020). MDM2 is the ubiquitin ligase of p53 and plays a central role in regulating the stability of p53. Akt mediates the phosphorylation of MDM2 in Ser166 and Ser186 and increases its interaction with p300, so MDM2 mediates the ubiquitination and degradation of p53(Liu et al.,2020). Phosphorylation of MDM2 also blocks its binding to p19ARF and increases the p53 degradation (Wang et al.,2020). It can be seen that many genes are involved in stroke through their interactions.

## 4. Conclusion

In this study, several potential biomarkers and related pathways of ischemic stroke were identified, based on the comprehensive weighted gene co-expression network and machine learning analyses with WGCNA and SVM-RFE, respectively, with this being a significant contribution to improve diagnosis and treatment of ischemic stroke.

## 5. Materials And Methods

## 5.1 Data collection

The Series Matrix File data GSE119121 was downloaded from the NCBI GEO public database, and included transcriptomics data of 47 samples from rats, including control (n=8), 1-hour ischemia (n=8), 2-hour ischemia (n=8), 3-hour ischemia (n=8), 6-hour ischemia (n=7), and 24-hour ischemia (n=8) and samples. This dataset was used to reconstruct a co-expression network using WGCNA. The GSE16561 matrix file included a total of 63 samples, consisted of the control group (n=24) and stroke patients (n=39). This dataset was used for SVM-RFE feature screening and subsequent validation. The flowchart of the overall analysis is shown in Figure 1.

## 5.2 Identification of key modules by WGCNA

By, Gene coexpression modules were reconstructed using the weighted gene co-expression network (WGCNA), and the correlation between gene networks and phenotypes was explored. Then, the core genes of the network were identified detecting the hubs of the modules. The WGCNA-R package was used to construct the co-expression network of all genes in the GSE119121 dataset screening the genes with the top 5000 variances with this algorithm and applying a soft threshold of 10. The weighted adjacency matrix was transformed into a topological overlap matrix to estimate the network connectivity degree, and the hierarchical clustering method was used to construct the clustering tree structure of the TOM matrix. Different branches of the cluster tree represented different gene modules, and different colors represented different modules. Genes were then classified using the weighted correlation coefficient, according to their expression patterns. In specific, genes with similar patterns were classified into one module, and tens of thousands of genes were divided into multiple modules through gene expression patterns.

## 5.3 Functional enrichment analysis of gene modules

The Metascape web tool ([www.metascape.org](http://www.metascape.org)) was used to obtain the biological functions and signal pathways involved in the modules of interest in WGCNA (black and brown modules in this study having the highest correlation with phenotype). This analysis enabled the annotation and visualization of the genes in specific modules using gene ontology and Kyoto gene genome encyclopedia pathway terms. Min overlap  $\geq 3$  and  $p \leq 0.01$  were considered as thresholds to infer statistically significant findings.

## 5.4 SVM- RFE

SVM-RFE is a machine learning method based on support vector machines (Huang et al.,2014, Guo.,2014), which searches for the optimal features by recursively deleting features from the feature vector generated by SVM. In this study, the R package "E1071" was used to perform SVM-RFE analysis,

and the key genes regulating disease progression in the data set GSE16561 were found and used for subsequent verification.

## 5.5 Gene Set Variation Analysis

Gene Set Variation Analysis (GSVA) is a non-parametric and unsupervised method to evaluate transcriptome gene enrichment. GSVA converts gene level changes into pathway level changes by comprehensively scoring the gene sets of interest, and then annotates samples based on the biological functions. In this study, gene sets were downloaded from the Molecular signatures database (v7.0), and each gene set was scored comprehensively by the GSVA algorithm to evaluate the potential changes in the biological function of different samples.

## 5.6 Verification

Six SD rats were purchased from the Experimental Animal Center of Guangxi Medical University. The rats were randomly divided into two groups: the Middle Cerebral Artery Occlusion (MCAO) group (n = 3) and the control group (n = 3). The models of the focal brain ischemia-reperfusion rats were performed by the Middle Cerebral Artery Occlusion (MCAO) (Longa et al.,1989). The rats were anesthetized 24h later and then their brain tissue samples were collected. The total RNA was extracted by Trizol and reversely transcribed into cDNA using the YEASEN reverse transcription kit. RT-qPCR was performed with a qPCR kit. The relative expression was measured by the 2-delta CT method. Our research was approved by the Ethics Committee of Guangxi Medical University.

## Abbreviations

GEO:Gene Expression Omnibus;WGCNA:weighted gene co-expression network analysis;SVM-RFE:Support Vector Machines coupled with Recursive Feature Elimination; ROC:receiver operating characteristic curve;GO:gene set variation analysis,KEGG:genetic ontology Kyoto Encyclopedia of Genes and Genomes;RT-PCR:real-time photoluminescence;MCAO:Middle Cerebral Artery Occlusion;NCAM:neural cell adhesion molecule.

## Declarations

### Author Contributions

LST was the first author who performed most of the experiments and data analysis. ZY contributed equally to this manuscript. LST and ZY performed the animal experiments. WXG designed, performed, and supervised the study and wrote the paper. WXG, HQP and LCX provided suggestions and guidance for this study and revised the paper. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the National Nature Science Foundation of China (no. 81700759).

## Data Availability Statement

The original contributions presented in the study are included in the article/[Supplementary Material](#) and further inquiries can be directed to the corresponding authors.

## Ethics Statement

The animal study was reviewed and approved by the Animal Ethics Committee of the Beijing Institute of Basic Medical Sciences.

## Consent for publication

Not applicable.

## Informed consent

For this type of study, formal consent was not required.

## Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Author details

1. 2. Department of TCM, First Affiliated Hospital of Guangxi Medical University, Nanning 530000, China.3. 4. Guangxi Medical University, Nanning 530000, China.

## References

1. V. L. Feigin, M. H. Forouzanfar, R. Krishnamurthi et al., "Global and regional burden of stroke during 1990-2010: findings from the Global Burden of Disease Study 2010," *Lancet*, vol. 383, no. 9913, pp. 245–255, 2014
2. Radu RA, Terecoasa EO, Bajenaru OA, Tiu C (2017) Etiologic classification of ischemic stroke: where do we stand? *Clin Neurol Neurosurg* 159:93–106
3. Evers SM, Struijs JN, Ament AJ, et al. International comparison of stroke cost studies. *Stroke*. 2004; 35:1209–1215.
4. Zhang Bin., Horvath Steve.(2005). A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*, 4(undefined), Article17. doi:10.2202/1544-6115.1128

5. Xiang Z, Li J, Song S, Wang J, Cai W, Hu W, Ji J, Zhu Z, Zang L, Yan R, Yu Y. A positive feedback between IDO1 metabolite and COL12A1 via MAPK pathway to promote gastric cancer metastasis. *J Exp Clin Cancer Res.* 2019 Jul 17;38(1):314. doi: 10.1186/s13046-019-1318-5. PMID: 31315643; PMCID: PMC6637527.
6. Katrinli S, Lori A, Kilaru V, Carter S, Powers A, Gillespie CF, Wingo AP, Michopoulos V, Jovanovic T, Ressler KJ, Smith AK. Association of HLA locus alleles with posttraumatic stress disorder. *Brain Behav Immun.* 2019 Oct;81:655-658. doi: 10.1016/j.bbi.2019.07.016. Epub 2019 Jul 13. PMID: 31310798; PMCID: PMC6754776.
7. Guo Y, Ma J, Xiao L, Fang J, Li G, Zhang L, Xu L, Lai X, Pan G, Chen Z. Identification of key pathways and genes in different types of chronic kidney disease based on WGCNA. *Mol Med Rep.* 2019 Sep;20(3):2245-2257. doi: 10.3892/mmr.2019.10443. Epub 2019 Jun 28. PMID: 31257514; PMCID: PMC6691232.
8. Vieira SE, Bando SY, de Paulis M, Oliveira DBL, Thomazelli LM, Durigon EL, Martinez MB, Moreira-Filho CA. Distinct transcriptional modules in the peripheral blood mononuclear cells response to human respiratory syncytial virus or to human rhinovirus in hospitalized infants with bronchiolitis. *PLoS One.* 2019 Mar 7;14(3): e0213501. doi: 10.1371/journal.pone.0213501. PMID: 30845274; PMCID: PMC6405118.
9. Guyon I; Weston J; Barnhill S; et al. Gene Selection for Cancer Classification using Support Vector Machines[J]. *Machine Learning*, 2002, 46(1-3):389-422.
10. Huang Mei-Ling., Hung Yung-Hsiang., Lee W M., Li R K., Jiang Bo-Ru.(2014). SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier. *ScientificWorldJournal*, 2014(undefined), 795624. doi:10.1155/2014/795624
11. Guo Pi., Luo Youxi., Mai Guoqin., Zhang Ming., Wang Guoqing., Zhao Miaomiao., Gao Liming., Li Fan., Zhou Fengfeng.(2014). Gene expression profile based classification models of psoriasis. *Genomics*, 103(1), 48-55. doi:10.1016/j.ygeno.2013.11.001
12. Longa EZ, Weinstein PR, Carlson S, Cummins R. Reversible middle cerebral artery occlusion without craniectomy in rats. *Stroke.* 1989 Jan;20(1):84-91. doi: 10.1161/01.str.20.1.84. PMID: 2643202.
13. Benjamin Emelia J., Virani Salim S., Callaway Clifton W., Chamberlain Alanna M., Chang Alexander R., Cheng Susan., Chiuve Stephanie E., Cushman Mary., Delling Francesca N., Deo Rajat., de Ferranti Sarah D., Ferguson Jane F., Fornage Myriam., Gillespie Cathleen., Isasi Carmen R., Jiménez Monik C., Jordan Lori Chaffin., Judd Suzanne E., Lackland Daniel., Lichtman Judith H., Lisabeth Lynda., Liu Simin., Longenecker Chris T., Lutsey Pamela L., Mackey Jason S., Matchar David B., Matsushita Kunihiro., Mussolino Michael E., Nasir Khurram., O'Flaherty Martin., Palaniappan Latha P., Pandey Ambarish., Pandey Dilip K., Reeves Mathew J., Ritchey Matthew D., Rodriguez Carlos J., Roth Gregory A., Rosamond Wayne D., Sampson Uchechukwu K A., Satou Gary M., Shah Svati H., Spartano Nicole L., Tirschwell David L., Tsao Connie W., Voeks Jenifer H., Willey Joshua Z., Wilkins John T., Wu Jason Hy., Alger Heather M., Wong Sally S., Muntner Paul., American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee.(2018).

- Heart Disease and Stroke Statistics-2018 Update: A Report From the American Heart Association. *Circulation*, 137(12), e67-e492. doi:10.1161/CIR.0000000000000558
14. Cribbs DH, Berchtold NC, Perreau V, Coleman PD, Rogers J, Tenner AJ, Cotman CW (2012) Extensive innate immune gene activation accompanies brain aging, increasing vulnerability to cognitive decline and neurodegeneration: a microarray study. *Journal of neuroinflammation* 9:179.
  15. Cheon SY, Kim EJ, Kim SY, Kim JM, Kam EH, Park JK, Koo BN (2018) Apoptosis Signal-regulating Kinase 1 Silencing on Astroglial Inflammasomes in an Experimental Model of Ischemic Stroke. *Neuroscience* 390:218-230.
  16. Mu Yuyuan., Wang Zhichao., Zhou Jiexin., Tan Chenxi., Wang Hongjiao.(2018). Correlations of Post-stroke Depression with Inflammatory Response Factors. *Iran J Public Health*, 47(7), 988-993.
  17. Kravka JM, Li L, Szulc ZM, Bielawski J, Ogretmen B, Hannun YA, Obeid LM, Bielawska A. Involvement of dihydroceramide desaturase in cell cycle progression in human neuroblastoma cells. *J Biol Chem*. 2007 Jun 8;282(23):16718-28. doi: 10.1074/jbc.M700647200. Epub 2007 Feb 5. PMID: 17283068; PMCID: PMC2084375.
  18. Riechers SP, Butland S, Deng Y, Skotte N, Ehrnhoefer DE, Russ J, Laine J, Laroche M, Pouladi MA, Wanker EE, Hayden MR, Graham RK. Interactome network analysis identifies multiple caspase-6 interactors involved in the pathogenesis of HD. *Hum Mol Genet*. 2016 Apr 15;25(8):1600-18. doi: 10.1093/hmg/ddw036. Epub 2016 Feb 11. PMID: 26908611.
  19. Karsai G, Kraft F, Haag N, Korenke GC, Hänisch B, Othman A, Suriyanarayanan S, Steiner R, Knopp C, Mull M, Bergmann M, Schröder JM, Weis J, Elbracht M, Begemann M, Hornemann T, Kurth I. DEGS1-associated aberrant sphingolipid metabolism impairs nervous system function in humans. *J Clin Invest*. 2019 Mar 1;129(3):1229-1239. doi: 10.1172/JCI124159. Epub 2019 Feb 11. PMID: 30620338; PMCID: PMC6391115
  20. Rodriguez-Cuenca S, Barbarroja N, Vidal-Puig A. Dihydroceramide desaturase 1, the gatekeeper of ceramide induced lipotoxicity. *Biochim Biophys Acta*. 2015 Jan;1851(1):40-50. doi: 10.1016/j.bbaliip.2014.09.021. Epub 2014 Oct 2. PMID: 25283058.
  21. Mohamud Yusuf A, Hagemann N, Hermann DM. The Acid Sphingomyelinase/ Ceramide System as Target for Ischemic Stroke Therapies. *Neurosignals*. 2019;27(S1):32-43. doi: 10.33594/000000184. PMID: 31778304.
  22. Alsanafi M, Kelly SL, McNaughton M, Merrill AH Jr, Pyne NJ, Pyne S. The regulation of p53, p38 MAPK, JNK and XBP-1s by sphingosine kinases in human embryonic kidney cells. *Biochim Biophys Acta Mol Cell Biol Lipids*. 2020 Apr;1865(4):158631. doi: 10.1016/j.bbaliip.2020.158631. Epub 2020 Jan 15. PMID: 31954175.
  23. Zhang DY, Liu Z, Lu Z, Sun WL, Ma X, Zhang P, Wu BQ, Cui PY. Lentivirus-mediated overexpression of HSDL2 suppresses cell proliferation and induces apoptosis in cholangiocarcinoma. *Onco Targets Ther*. 2018 Oct 17;11: 7133-7142. doi: 10.2147/OTT.S176410. PMID: 30410369; PMCID: PMC6200078.

24. Dong B, Yang Y, Han A, Zhang S, Lin Z, Wang Y, Piao J. Ectopic expression of HSDL2 is related to cell proliferation and prognosis in breast cancer. *Cancer Manag Res.* 2019 Jul 12;11: 6531-6542. doi: 10.2147/CMAR.S205316. PMID: 31372054; PMCID: PMC6634268.
25. Shi Y, Mao Z, Huang Y, Sun Y, Cao Q, Yin X, Huang J, Zhang Q. Knockdown of HSDL2 inhibits lung adenocarcinoma progression via down-regulating AKT2 expression. *Biosci Rep.* 2020 Apr 30;40(4): BSR20200348. doi: 10.1042/BSR20200348. PMID: 32211805; PMCID: PMC7138907.
26. Zeng J, Ma X, Wang J, Liu R, Shao Y, Hou Y, Li Z, Fang Y. Down-regulated HSDL2 expression suppresses cell proliferation and promotes apoptosis in papillary thyroid carcinoma. *Biosci Rep.* 2019 Jun 4;39(6): BSR20190425. doi: 10.1042/BSR20190425. PMID: 31101684; PMCID: PMC6549096.
27. Zhang X, Wang Y. Identification of hub genes and key pathways associated with the progression of gynecological cancer. *Oncol Lett.* 2019 Dec;18(6):6516-6524. doi: 10.3892/ol.2019.11004. Epub 2019 Oct 18. PMID: 31788113; PMCID: PMC6865827.
28. Ruokun C, Yake X, Fengdong Y, Xinting W, Laijun S, Xianzhi L. Lentivirus-mediated silencing of HSDL2 suppresses cell proliferation in human gliomas. *Tumour Biol.* 2016 Nov;37(11):15065-15077. doi: 10.1007/s13277-016-5402-6. Epub 2016 Sep 22. PMID: 27658777.
29. Girling Kimberly D, Demers Marie-Josée, Laine Jean et al. Activation of caspase-6 and cleavage of caspase-6 substrates is an early event in NMDA receptor-mediated excitotoxicity.[J] .*J Neurosci Res*, 2018, 96: 391-406.
30. Riechers SP, Butland S, Deng Y, Skotte N, Ehrnhoefer DE, Russ J, Laine J, Laroche M, Pouladi MA, Wanker EE, Hayden MR, Graham RK. Interactome network analysis identifies multiple caspase-6 interactors involved in the pathogenesis of HD. *Hum Mol Genet.* 2016 Apr 15;25(8):1600-18. doi: 10.1093/hmg/ddw036. Epub 2016 Feb 11. PMID: 26908611.
31. Burgess A, Wainwright SR, Shihabuddin LS, Rutishauser U, Seki T, Aubert I. Polysialic acid regulates the clustering, migration, and neuronal differentiation of progenitor cells in the adult hippocampus. *Dev Neurobiol.* 2008 Dec;68(14):1580-90. doi: 10.1002/dneu.20681. PMID: 18844212.
32. Kanato Y, Kitajima K, Sato C. Direct binding of polysialic acid to a brain-derived neurotrophic factor depends on the degree of polymerization. *Glycobiology.* 2008 Dec;18(12):1044-53. doi: 10.1093/glycob/cwn084. Epub 2008 Sep 16. PMID: 18796648.
33. Isomura R, Kitajima K, Sato C. Structural and functional impairments of polysialic acid by a mutated polysialyltransferase found in schizophrenia. *J Biol Chem.* 2011 Jun 17;286(24):21535-45. doi: 10.1074/jbc.M111.221143. Epub 2011 Apr 4. PMID: 21464126; PMCID: PMC3122212.
34. Ono S, Hane M, Kitajima K, Sato C. Novel regulation of fibroblast growth factor 2 (FGF2)-mediated cell growth by polysialic acid. *J Biol Chem.* 2012 Feb 3;287(6):3710-22. doi: 10.1074/jbc.M111.276618. Epub 2011 Dec 12. PMID: 22158871; PMCID: PMC3281739.
35. Ying J., Wu J., Zhang Y., et al. Ligustrazine suppresses renal NMDAR1 and caspase-3 expressions in a mouse model of sepsis-associated acute kidney injury. *Molecular and Cellular Biochemistry.* 2020;464(1-2):73–81. doi: 10.1007/s11010-019-03650-4.

36. Li J., Liu Z., Wang L., Xu H., Wang Y. Thousand and one kinase 1 protects MCAO-induced cerebral ischemic stroke in rats by decreasing apoptosis and pro-inflammatory factors. *Bioscience Reports*. 2019;39doi: 10.1042/bsr20190749.
37. Buschhaus J. M., Humphries B., Luker K. E., Luker G. D. A caspase-3 reporter for fluorescence lifetime imaging of single-cell apoptosis. *Cells*. 2018;7 doi: 10.3390/cells7060057.
38. Lugovaya A. V., Emanuel V. S., Kalinina N. M., Ivanov A. M., Artemova A. V. Apoptosis and autophagy in the pathogenesis of acute ischemic stroke (review of literature) *Russian Clinical Laboratory Diagnostics*. 2020;65(7):428–434. doi: 10.18821/0869-2084-2020-65-7-428-434.
39. Gao X., Li H., Dong Z., et al. Relationship between Erk1/2 signal pathway and nerve cell apoptosis rats with ischemic stroke. *Cellular and Molecular Biology*. 2019;65(7):127–131. doi: 10.14715/cmb/2019.65.7.22.
40. An H., Duan Y., Wu D., et al. Phenothiazines enhance mild hypothermia-induced neuroprotection via PI3K/Akt regulation in experimental stroke. *Scientific Reports*. 2017;7:p. 7469. doi: 10.1038/s41598-017-06752-5.
41. Guo L., Kang J. S., Kang N. J., Choi Y. W. S-petasin induces apoptosis and inhibits cell migration through activation of p53 pathway signaling in melanoma B16F10 cells and A375 cells. *Archives of Biochemistry and Biophysics*. 2020;692:p. 108519. doi: 10.1016/j.abb.2020.108519.
42. Saito A., Hayashi T., Okuno S., Nishi T., Chan P. H. Modulation of p53 degradation via MDM2-mediated ubiquitylation and the ubiquitin-proteasome system during reperfusion after stroke: role of oxidative stress. *Journal of Cerebral Blood Flow & Metabolism*. 2005;25(2):267–280. doi: 10.1038/sj.jcbfm.9600028.
43. Gao Xinping., Li Qing., Chen Guobin., He Haipeng., Ma Ying.(2020). MAGEA3 promotes proliferation and suppresses apoptosis in cervical cancer cells by inhibiting the KAP1/p53 signaling pathway. *Am J Transl Res*, 12(7), 3596-3612.
44. Liu K., Xue B., Bai G., Zhang W. F-box protein FBXO31 modulates apoptosis and epithelial-mesenchymal transition of cervical cancer via inactivation of the PI3K/AKT-mediated MDM2/p53 axis. *Life Sciences*. 2020;259:p. 118277. doi: 10.1016/j.lfs.2020.118277.
45. Wang Y., Zhao J., Zhang C., Wang P., Huang C., Peng H. MDM2 promotes the proliferation and inhibits the apoptosis of pituitary adenomas cells by directly interacting with p53. *Endokrynologia Polska*. 2020 doi: 10.5603/EP.a2020.0053.

## Figures

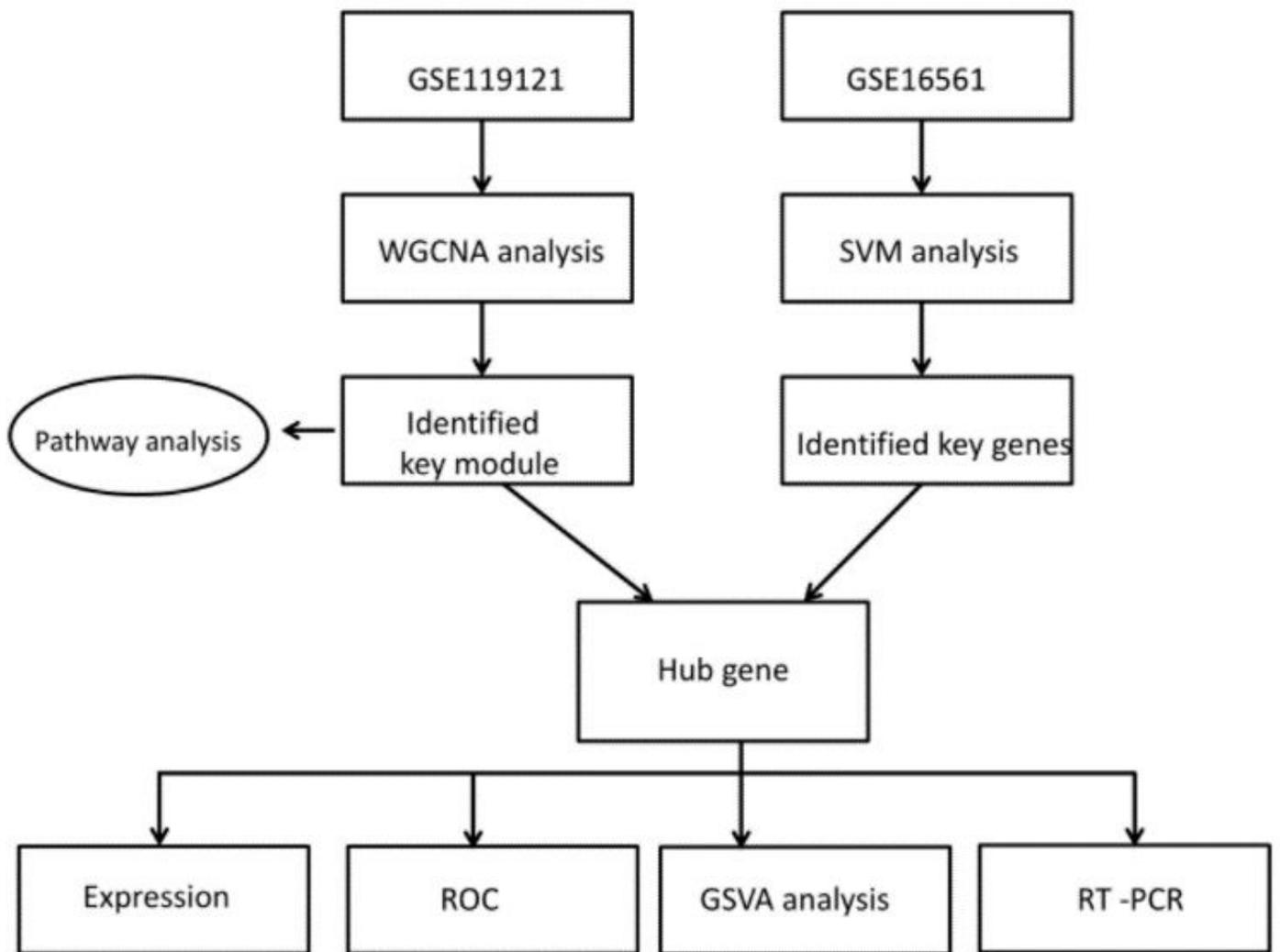
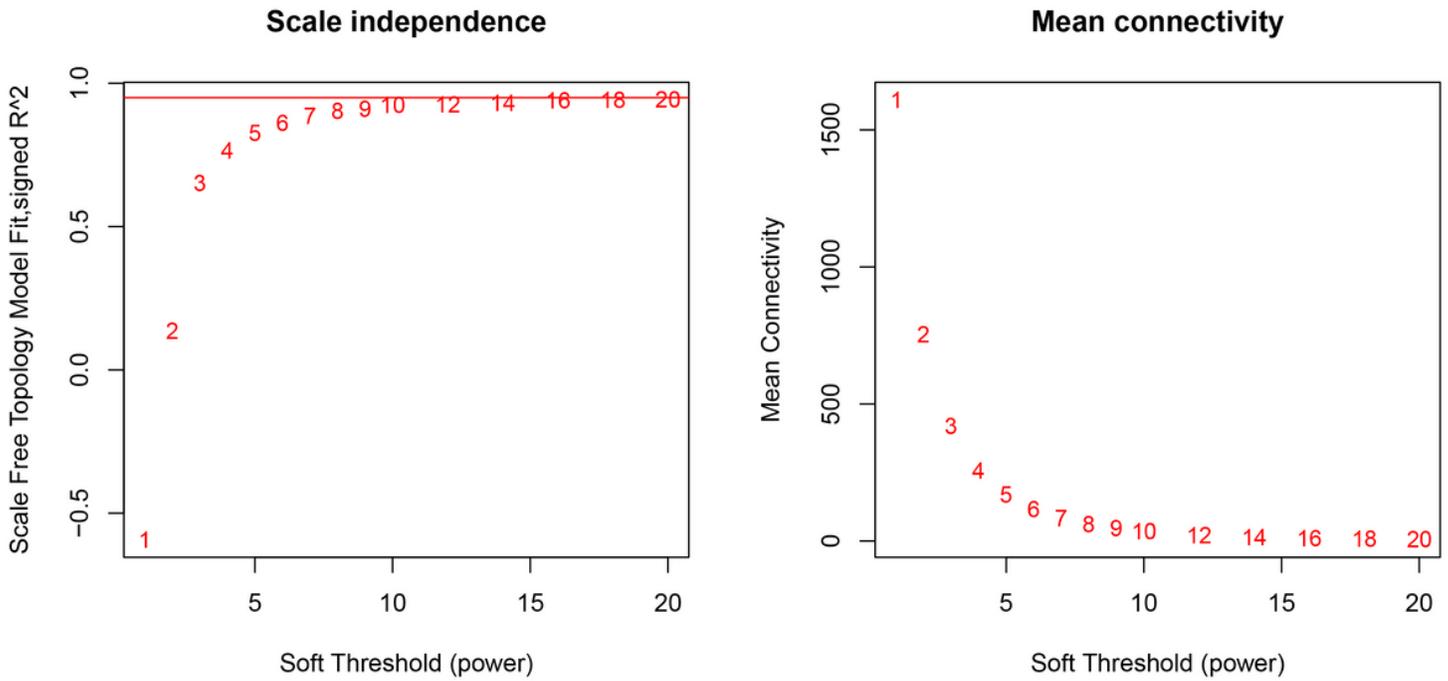


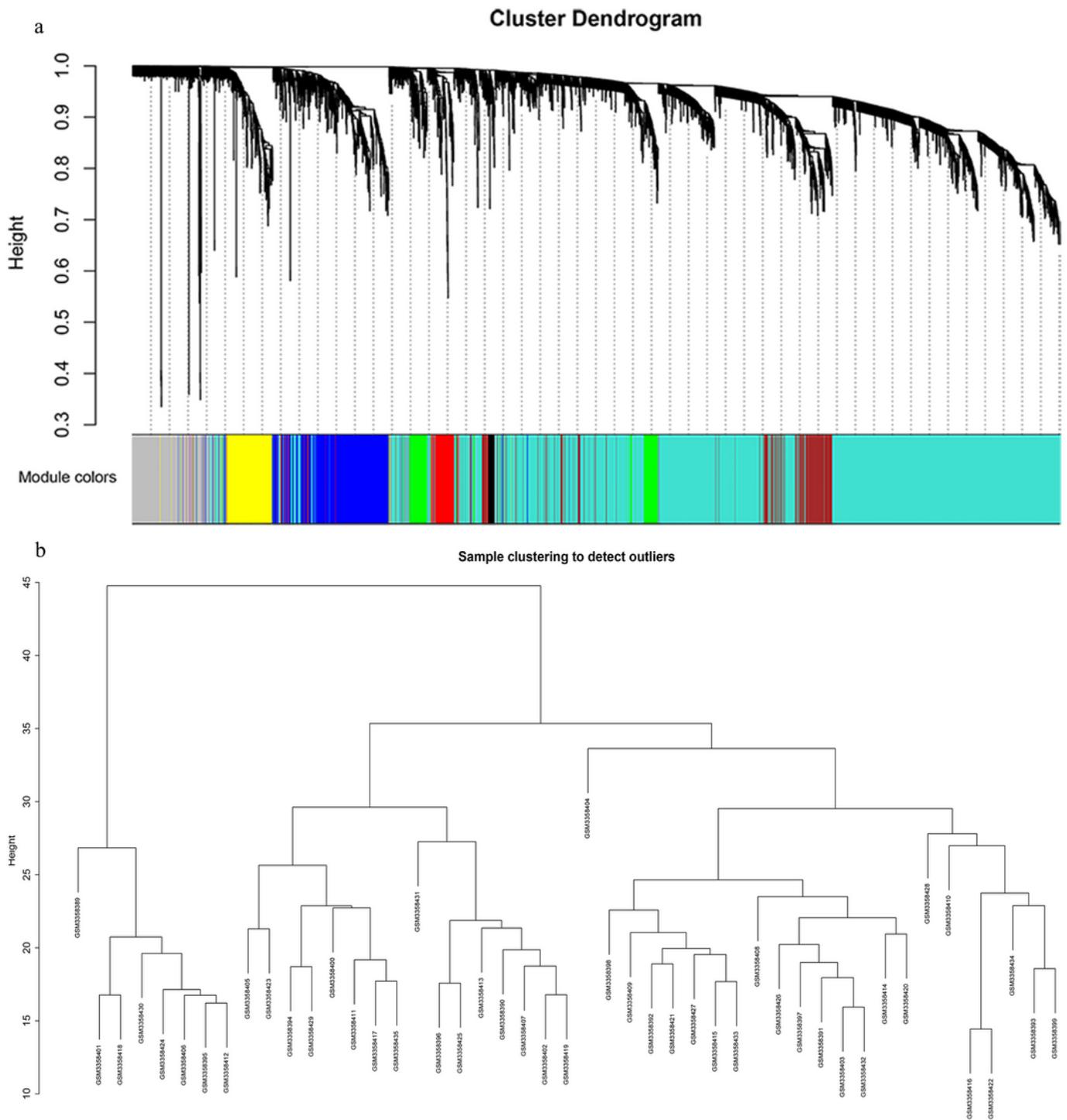
Figure 1

Flowchart illustrates the bioinformatic analysis process performed on the retrieved stroke datasets.



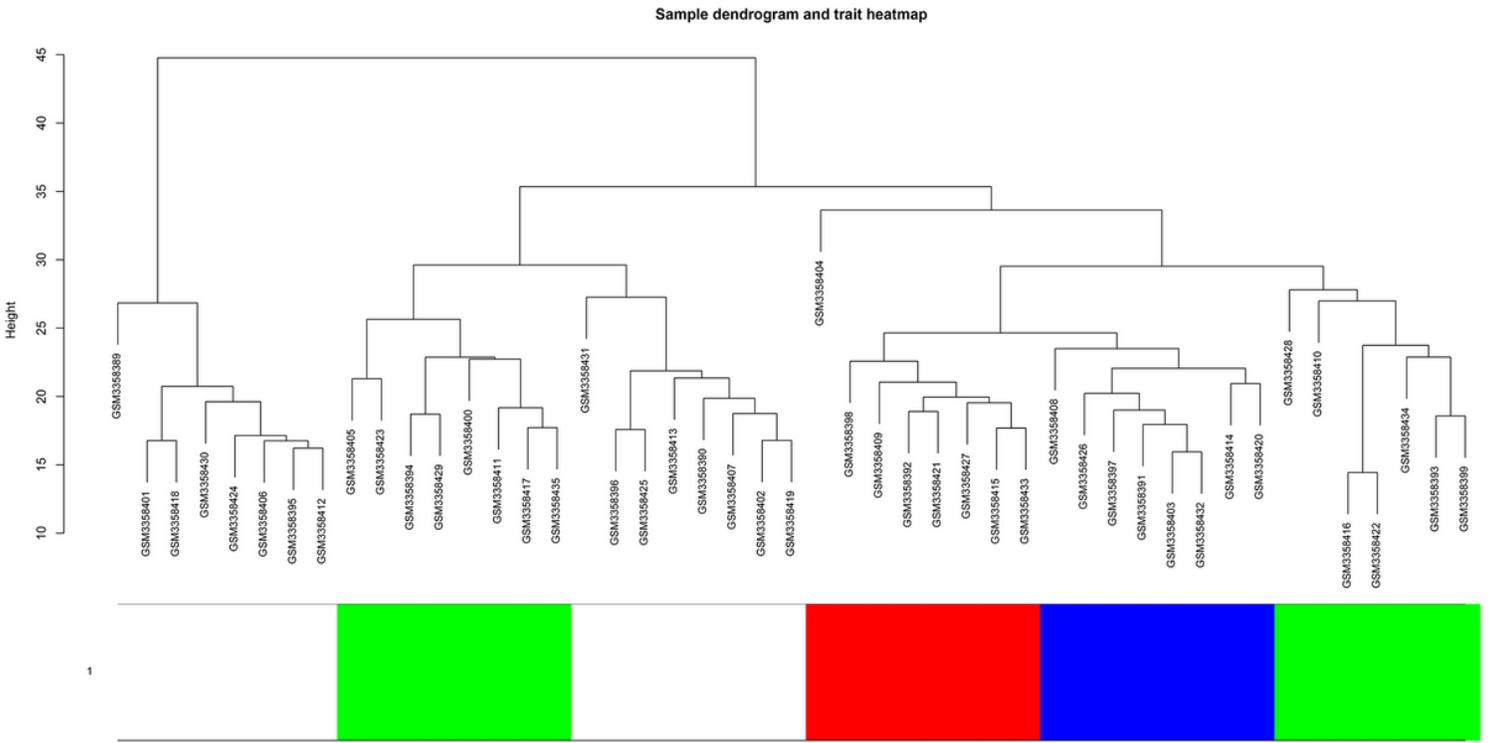
**Figure 2**

Scale-free fitting exponential analysis of all kinds of soft threshold power. Determination of appropriate soft-thresholding values ( $\beta$ ). (A) The scale-free fit index with  $\beta$  values ranging from 1 to 20. (B) The mean connectivity of the supposed  $\beta$  value was ranging from 1 to 20. The left panel shows the effect of the deployed soft threshold power on the scale-free topology fitting index, and the right panel shows its effect on average connectivity. (B): Gene clustering tree based on hierarchical clustering of dissimilarity degree based on adjacency relationship (tree diagram)



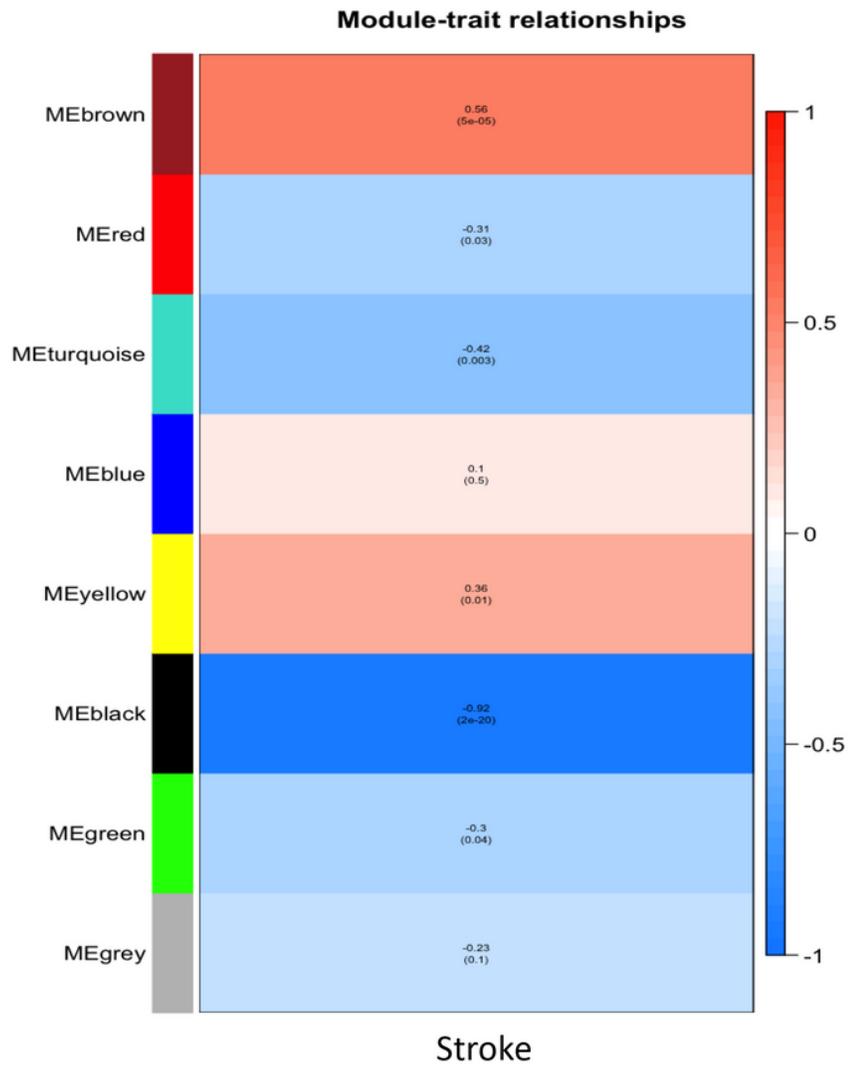
**Figure 3**

Division of co-expression gene module (A) A tree map showing how many genes are clustered using average linkage hierarchical clustering based on the different measures of TOM. (B) Clustering tree of genes. Gene clustering tree obtained by hierarchical clustering based on adjacency similarity (tree view) gene clustering tree using the dissimilarity degree hierarchical clustering of adjacency relationship (tree view)



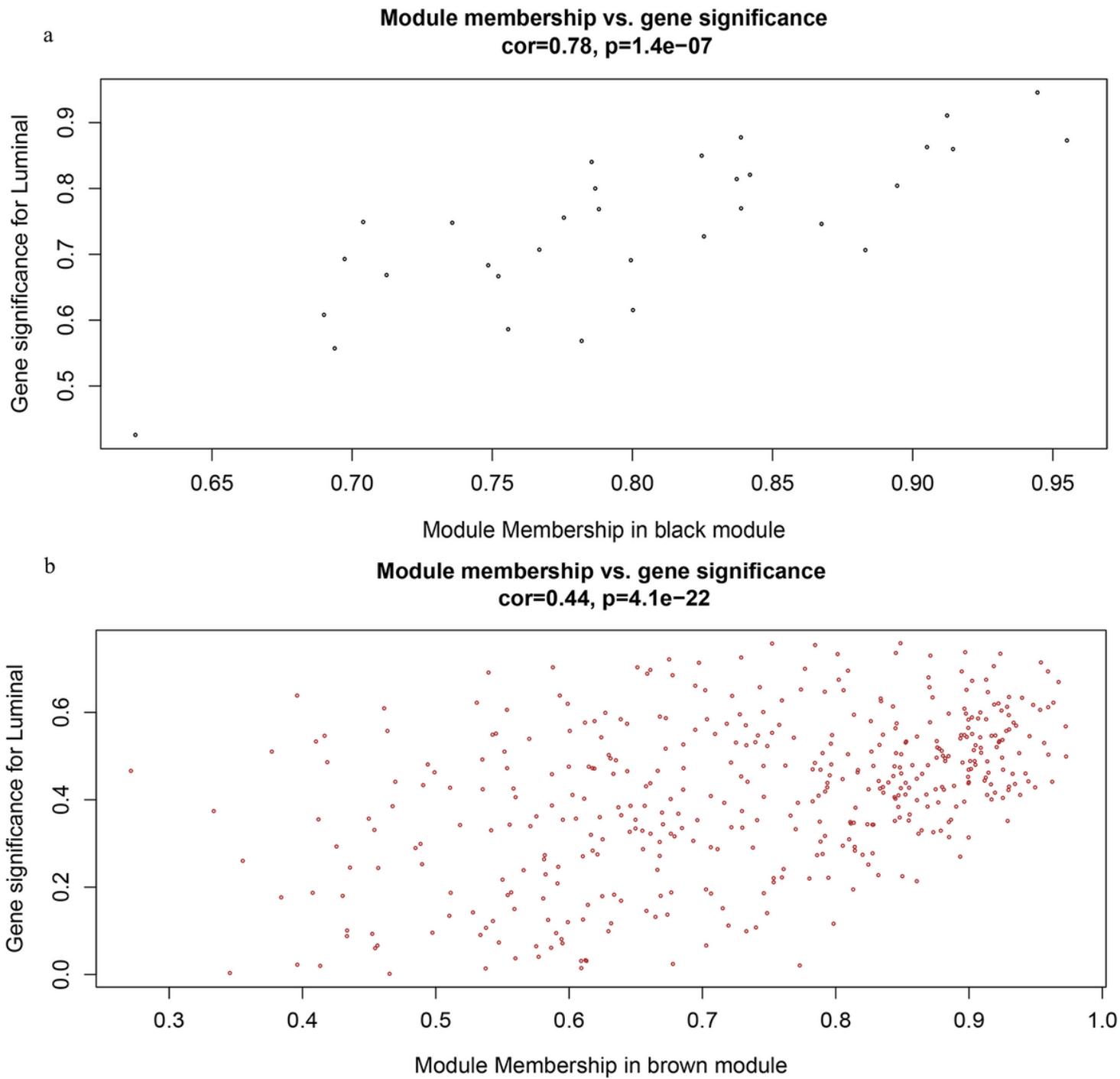
**Figure 4**

Functional correlation of modules to identify modules related to stroke progression. Each row corresponds to a module feature gene, and each column corresponds to a clinical feature. Each cell contains a corresponding correlation in the first row and a P-value in the second row. The table is color-coded by correlation according to the color legend.



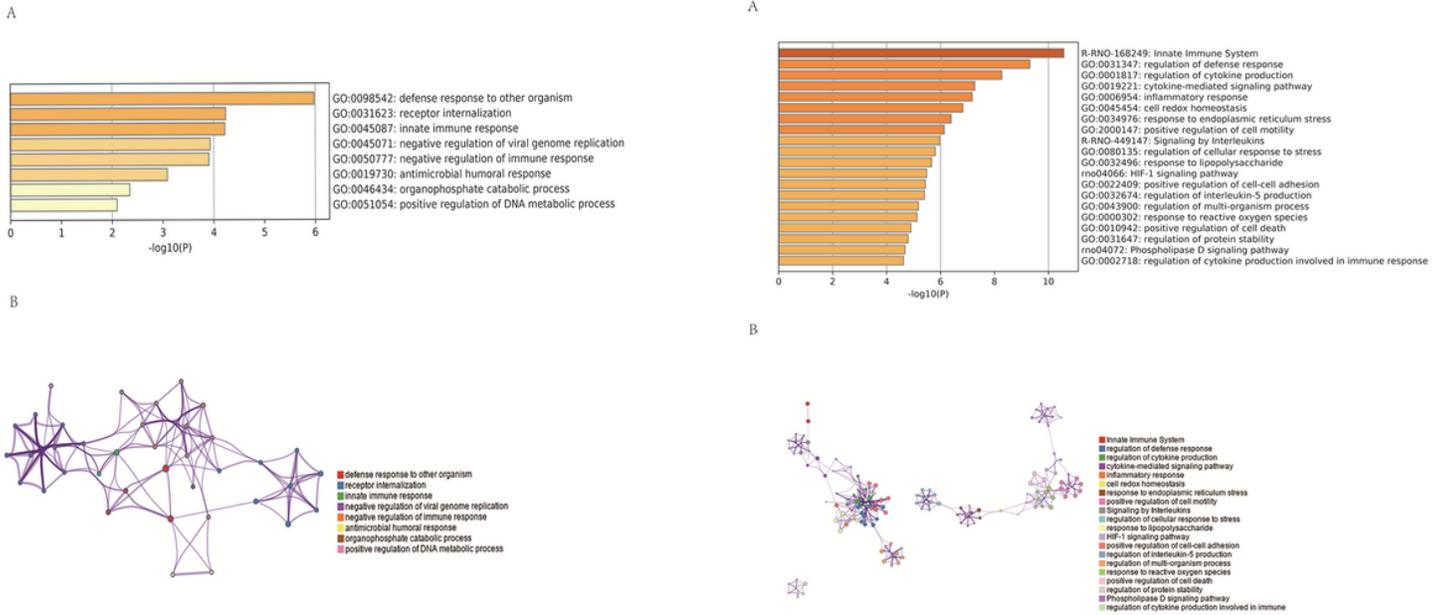
**Figure 5**

Module-character relationship. Each row corresponds to a module feature gene and the column corresponds to a feature. Each cell contains the corresponding correlation and p-value.



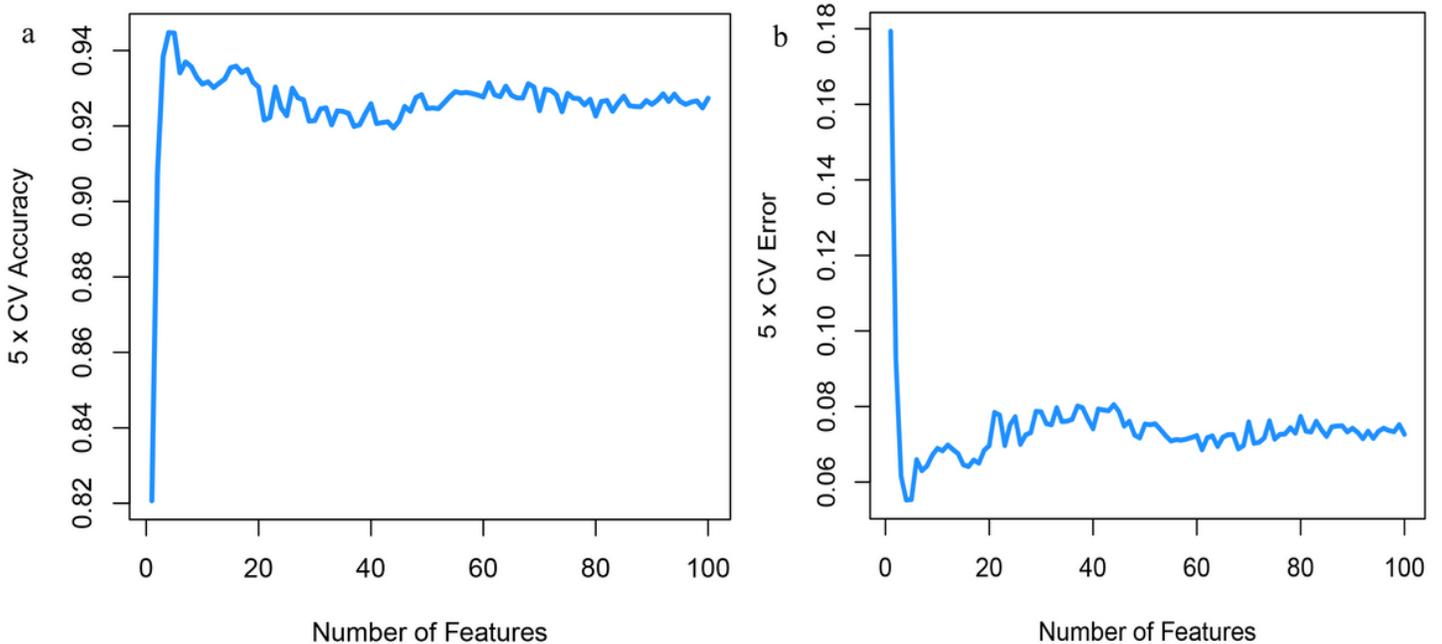
**Figure 6**

The relationship between gene meaning and module. The correlation between modular traits and the meaning of gene expression.



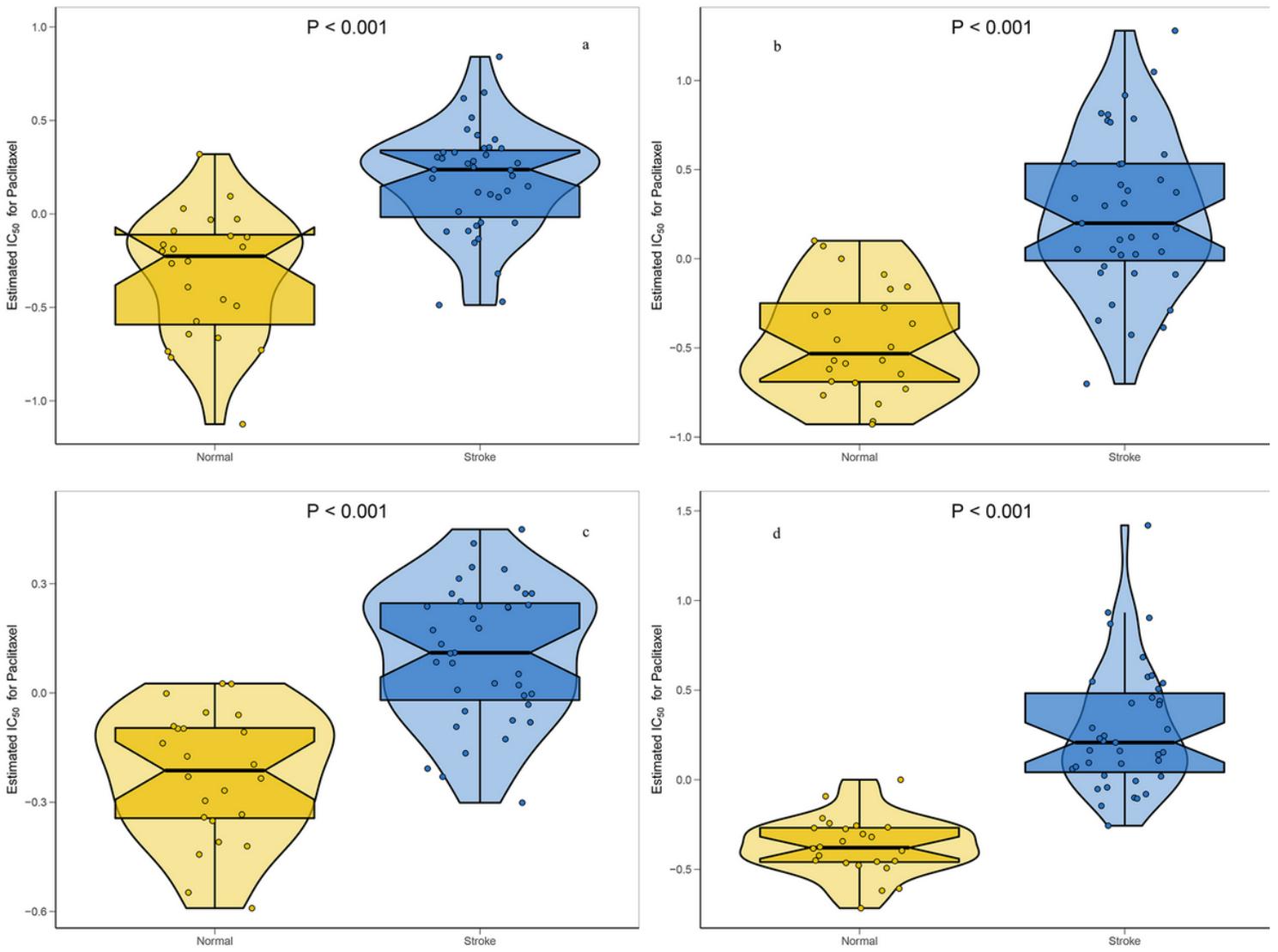
**Figure 7**

The KEGG path result diagram of the module (6) the brown module (7) the black module. Map A represents the pathway of module gene enrichment, the longer the band is, the more significant the P value is, and the map B represents the interaction between pathways.



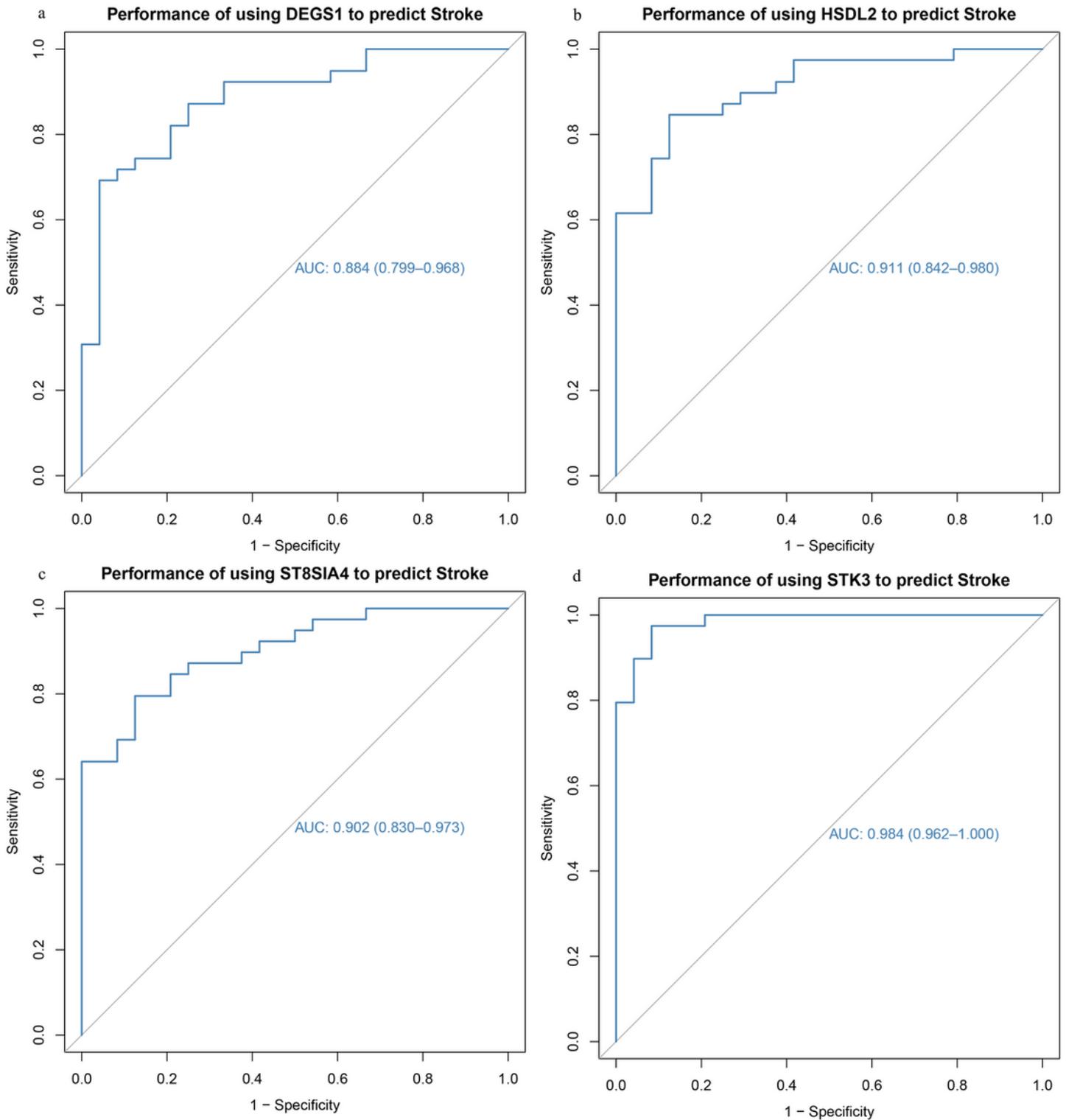
**Figure 8**

Box line diagram



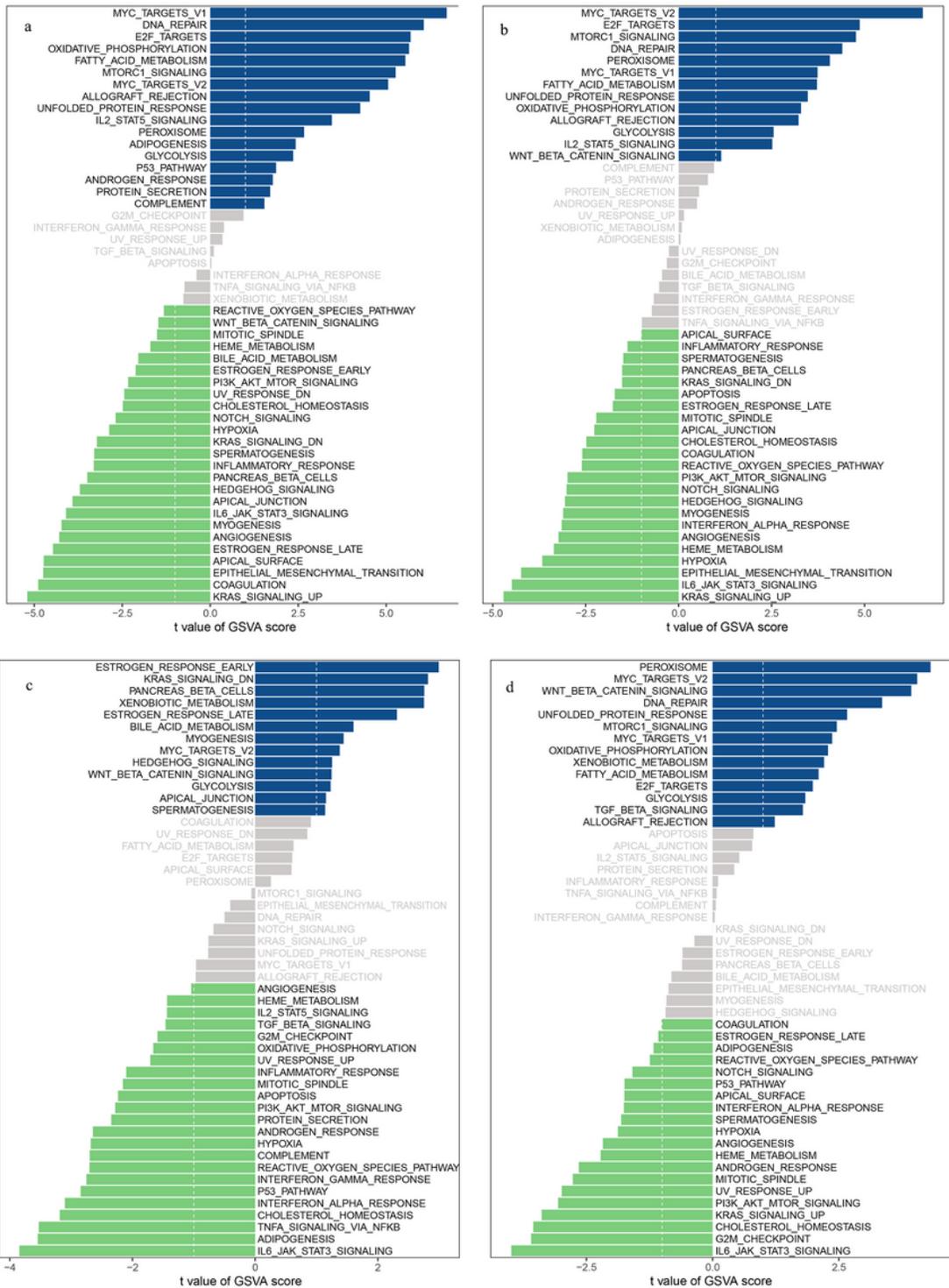
**Figure 9**

The expression of four central genes in the second data set.



**Figure 10**

The area under the curve (AUC) values obtained from the ROC curve analysis were, 0.884 (95% CI HSDL2 AUC 0.848) for the DEGS1, 0.911 (95% CIRV 0.842) for the HSDL2, 0.902 (95% CIRV 0.830) for the ST8SIA4, and 0.984 (95% CIRV 0.962) 1.000 for the STK3, respectively.



**Figure 11**

GSVA analysis revealed differences in pathways. The differences of pathway between high expression of the core gene and low expression of VS gene. Blue represents the enriched pathways of the underexpressed group, and green represents the enriched pathway of the overexpressed group.

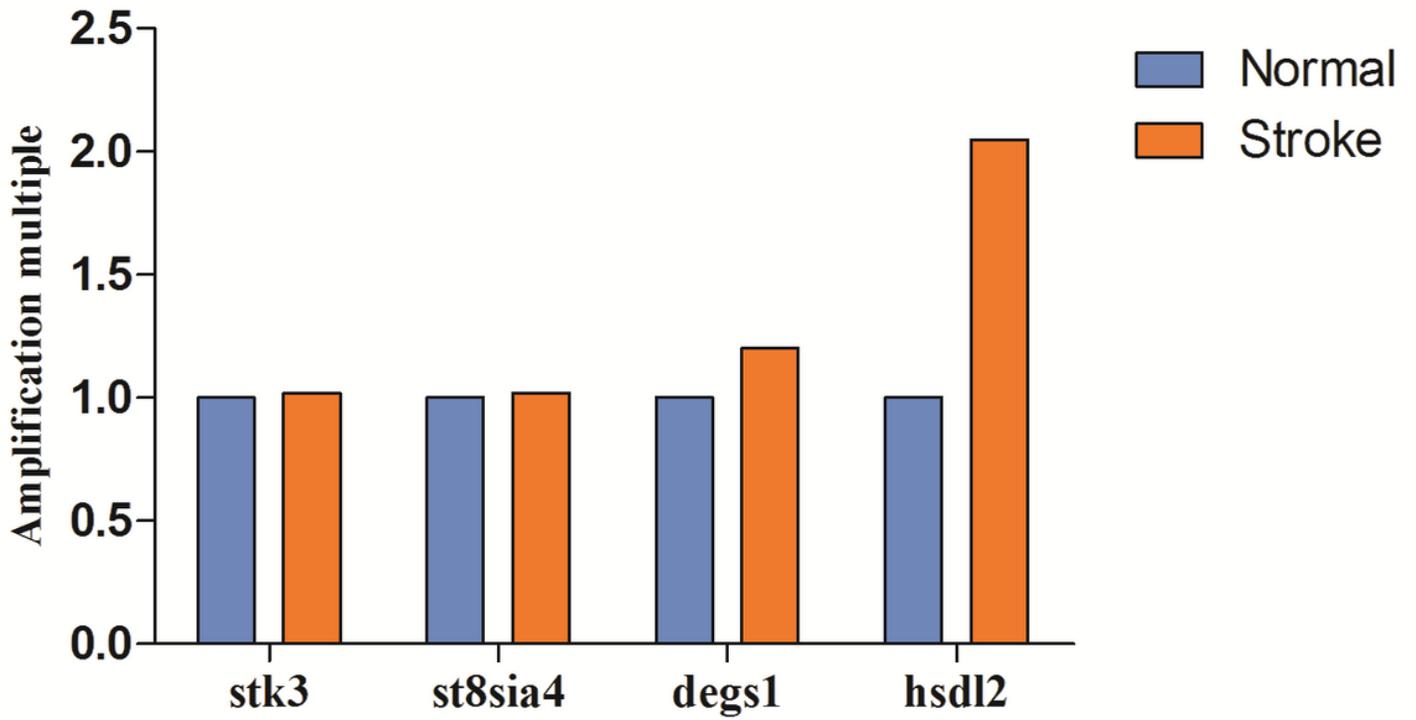


Figure 12

Validation of hub gene using qRT-PCR. The expression of four genes in brain tissue was detected by qRT-PCR and showed a significant increase of expression compared to controls. GAPDH was used as an internal control.