Identification of Dysregulated Mechanisms and Potential Biomarkers in Ischemic Stroke Onset

Objective: Ischemic stroke (IS) is a major cause of severe disability. This study aimed to identify potential biomarkers closely related to IS diagnosis and treatment.

Methods: Profiles of gene expression were obtained from datasets GSE16561, GSE22255, GSE112801 and GSE110993. Differentially expressed mRNAs between IS and controls were then subjected to weighted gene co-expression network analysis as well as multiscale embedded gene co-expression network analysis. The intersection of the two sets of module genes was subjected to analyses of functional enrichment and of microRNAs (miRNAs) regulation. Then, the area under receiver operating characteristic curves (AUC) was calculated to assess the ability of genes to discriminate IS patients from controls. IS diagnostic signatures were constructed using least absolute shrinkage and selection operator regression.

Results: A total of 234 common co-expression network genes were found to be potentially associated with IS. Enrichment analysis found that these genes were mainly associated with inflammation and immune response. The aberrantly expressed miRNAs (hsa-miR-651-5p, hsa-miR-138-5p, hsa-miR-374a-3p) in IS had regulatory effects on IS-related genes and were involved in brain-related diseases. We used the criterion AUC > 0.7 to screen out 23 hub genes from IS-related genes in the GSE16561 and GSE22255 datasets. We obtained an 8-gene signature (ADCY4, DUSP1, ATP5F1, DCTN5, EIF3G, ELAVL1, EXOSC7 and PPIE) from the training set of GSE16561 dataset, which we confirmed in the validation set of GSE16561 dataset and in the GSE22255 dataset. The genes in this signature were highly accurate for diagnosing IS. In addition, the 8-gene signature significantly correlated with infiltration by immune cells.

Conclusion: These findings provide new clues to molecular mechanisms and treatment targets in IS. The genes in the signature may be candidate markers and potential gene targets for treatments.

Keywords: ischemic stroke, bioinformatics analysis, miRNAs, immune inflammatory response, candidate markers

Introduction

Ischemic stroke (IS) occurs when a thrombus or embolus obstructs a large cerebral artery, blocking blood flow to certain parts of the brain.\(^1\) IS is a leading cause of death, accounting for over 6 million deaths annually worldwide.\(^2\) Despite increasing understanding of IS pathophysiology and improvements in laboratory and clinical research, the diagnosis and treatment of IS remain challenging.\(^3\),\(^4\) Biomarkers can serve as a valuable component of the diagnostic evaluation approach, guiding treatment and prognosis.

Patients with IS usually present multiple risk factors, so the development of new treatments depends on more thorough understanding of IS pathogenesis. IS patients usually present with a rapid interruption of blood supply, which leads to neurological failure due to hypoxia and brain tissue damage.\(^2\) Multiple studies have shown...
that neuroinflammatory response is an important pathological mechanism of IS, which disrupts the blood-brain barrier by blocking and increases patient mortality. Disruption of the blood-brain barrier results in the migration of various inflammatory cells, such as macrophages, monocytes, and T lymphocytes to the ischemic site, where they release proinflammatory factors. This creates a complex inflammatory network that indirectly affects the integrity of the blood-brain barrier. The inflammatory state of regulatory cells, against the secretion of harmful factors, may maintain the homeostasis of the blood-brain barrier caused by inflammation and promote functional recovery in patients with IS. The inflammation and reperfusion cause cerebral ischemia and increase production of reactive oxygen species and cell death. These mechanistic insights may help in the development of treatments for the prevention of secondary brain injury after stroke.

MicroRNAs (miRNAs), endogenously expressed RNA molecules, inhibit mRNA translation and play a key role in the pathophysiological process of ischemic stroke injury. The target genes regulated by miRNAs can mediate inflammatory response, oxidative stress, cell proliferation and apoptosis, making them potentially useful for diagnosis and therapy in IS. There is evidence that brain miRNAs are altered in ischemic stroke and that these changes lead to significant alterations in the expression of genes involved in neuroinflammation. MiRNAs are new diagnostic and therapeutic molecules for the treatment of ischemic stroke, the prevention of neurodegeneration and brain damage, and the promotion of long-term recovery.

In the present study, we explored molecular dysregulation and potential biomarkers in IS by using gene expression profiles from patients in public databases. By combining weighted gene co-expression network analysis (WGCNA) and multiscale embedded gene co-expression network analysis (MEGENA), we established diagnostic gene signatures for IS that, at the same time, may provide novel therapeutic targets.

**Materials and Methods**

**Data Collection**

All data in this study were obtained from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). GSE16561 contained the mRNA expression profiles of peripheral whole blood from 39 acute IS patients and 24 healthy control subjects based on the GPL6883 platform. GSE22255 contained the mRNA expression profiles of peripheral blood mononuclear cells (PBMCs) from 20 IS patients and 20 sex- and age-matched controls based on the GPL570 platform. GSE112801 contained the miRNA expression profiles of whole blood samples from 39 IS patients and 10 age-matched controls based on the GPL18573 platform. GSE110993 contained the miRNA expression profiles of the peripheral blood plasma samples from 20 IS patients and 20 healthy control subjects based on the GPL15456 platform. Sample quality were evaluated using principal component analysis (PCA).

**Differential Expression Analysis**

Differences in mRNA and miRNA expression between IS patients and controls were analyzed on data from GSE16561, GSE22255, and GSE112801 using the limma R package. The DEseq2 R package was used to analyze differences in miRNA expression between IS patients and controls in GSE110993. Differentially expressed mRNAs (DEmRs) and differentially expressed miRNAs (DEmiRs) were defined as those whose expression differences were associated with $P < 0.05$.

**Construction of Co-Expression Network**

WGCNA was performed for DEmRs in GSE16561 using the WGCNA routine in R. Gene co-expression networks were identified based on topological overlap. The soft thresholding power $\beta$ was chosen based on scale-free topology criteria. Then a topological overlap measure (TOM) matrix was transformed from the adjacency matrix. Next, modules were identified using hierarchical clustering and dynamic tree cutting. MEGENA was performed using the MEGENA routine in R. This analysis generates multiscale networks of different possible gene-gene interactions. Default parameters were used to calculate a planar filtered network (PFN) from gene expression correlations emerging from the MEGENA.

**Functional and Pathway Enrichment of DEmRs**

The clusterProfiler routine in R was used to analyze DEmRs for enrichment in Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Biological processes (BPs) were
included in the analysis of GO enrichment. $P < 0.05$ was set as the threshold of statistical significance. Enrichment results were displayed through gene set variation analysis (GSVA) using the GSVA routine in R. The clusterProfiler package was also used to perform gene set enrichment analysis (GSEA), the results of which were displayed using the fgsea routine in R.

**Prediction of DEmiR Targets**

Targetscan (http://www.targetscan.org/vert_72/) was used to predict the target genes regulated by DEmiRs.

**LASSO Regression**

Areas under receiver operating characteristic curves (AUCs) were calculated for IS-related genes using pROC routine in R in order to assess the ability of the genes to discriminate IS patients from controls. Genes with AUC > 0.7 in both the GSE16561 and GSE22255 datasets were used to build a binomial least absolute shrinkage and selection operator regression (LASSO) model using the glmnet routine in R. The training set was 75% of samples (randomly selected) in the GSE16561 dataset. When performing LASSO regression, we retained all variables with non-zero coefficients. The remaining 25% of samples in the GSE16561 dataset and all samples in the GSE25504 dataset served as the validation set.

**Immune Cell Infiltration**

The levels of infiltration by different types of immune cells in GSE16561 were evaluated using CIBERSORT (https://cibersort.stanford.edu/) and ssGSEA in the GSVA routine in R. Immune cells expressed as 0 were excluded from the analysis. A set of marker genes for immune cell types analyzed by ssGSEA were obtained from Bindea et al. Differences in infiltration of immune cells between IS patients and controls were calculated using the limma R package. We also evaluated potential correlations between the hub genes and immune cells using Pearson correlation analysis. Results associated with $P$ value $< 0.05$ were considered significant.

**Results**

**Gene Networks of DEmRs in IS Patients Based on WGCNA and MEGENA**

The flowchart of this study is shown in Figure 1. Based on PCA, IS patients were weakly distinguishable from healthy controls in the GSE16561 and GSE22255 datasets (Supplementary Figure S1). To identify IS-associated genes, we first performed differential analysis of gene expression between IS patients and controls in GSE16561. A total of 5700 DEmRs were obtained (Figure 2A). We then performed WGCNA on these DEmRs to construct a co-expression network. The power of $\beta=12$ (scale-free $R^2=0.9$) was set as the appropriate soft-thresholding value to satisfy the scale-free network criteria (Figure 2B). Twelve co-expression modules were identified (Figure 2C), and there were strong correlations among different genes within each module (Figure 2D). Correlation heatmap of module genes showed the TOM among all genes.

We also analyzed differential gene expression in the GSE22255 dataset, giving 2390 DEmRs between IS patients and controls (Figure 3A). By comparing with the DEmRs in GSE16561, we identified that 96 DEmRs were upregulated and 261 DEmRs were downregulated in IS (Figure 3B). These DEmRs were considered as common genes and may be strongly associated with IS.

Using MEGENA, we constructed a co-expression network of common genes (Figure 3C). We identified 33 modules and 346 module genes. Among them, the largest module, C1_3, consisted of 51 genes, while module C1_10 consisted of 51 genes and C1_2 consisted of 35 genes (Figure 3D). Then, 234 module genes common to the two co-expression networks of WGCNA and MEGENA were considered as IS-associated genes.

**Functional and Pathway Enrichment of Module Genes**

To identify IS-related molecular dysregulation, we analyzed the enrichment IS-associated genes. GO analysis showed that the IS-associated genes were involved mainly in activated cellular response of interleukin (IL)-1, positive regulation of IL-17 secretion, and regulation of eosinophil differentiation of biological processes (BPs), as well as the inhibition of oxidative phosphorylation, mRNA processing, and ATP metabolic processes (Figure 4A).

The IS-associated genes showed enrichment of the following KEGG pathways: the activated IL-17 signaling pathway, retrograde endocannabinoid signaling, ovarian steroidogenesis, inhibited oxidative phosphorylation, Parkinson disease, and Alzheimer’s disease (Figure 4B). GSEA results showed enrichment of focal adhesion, NOD-like receptor signaling, and chemokine signaling in IS, whereas oxidative phosphorylation, Parkinson disease, and Huntington disease were inhibited (Figure 4C). The intersection genes involved in these enrichment results
may be dysregulated in IS, and most appear to be associated with immune and inflammatory responses.

Network of miRNAs Regulating the IS-Associated Genes

To identify the regulatory mechanisms of the IS-associated genes, we first analyzed DEmiRs. In GSE112801, we detected 558 DEmiRs (Figure 5A), among which 50 were up-regulated and 108 down-regulated in IS, based on comparison with DEmiRs in GSE110993 (Figure 5B). Among the predicted target genes of these dysregulated miRNAs, we identified nine IS-associated genes: ATP5F1, CLNS1A, CNOT7, ELAC1, EXOSC1, MAT2A, PDHA1, SEH1L, and TRAP1. These genes are regulated by the following four miRNAs involved in brain disease: hsa-miR-651-5p, hsa-miR-138-5p, hsa-miR-9-3p, and hsa-miR-374a-3p (Figure 5C).

Identification of Gene Markers of IS

To identify potential biomarkers for IS, we calculated AUCs for IS-associated genes in GSE22255 and GSE16561 (Figure 6A). We obtained 23 hub genes that gave AUC > 0.7 in both datasets (Figure 6B). Then we randomly split all IS samples in the GSE16561 dataset into a training set (75%) and a validation set (25%). We employed LASSO regression to screen further the 23 hub genes in the training set. Finally, we obtained 8 candidate genes with non-zero coefficients (Figure 6C).

We performed ROC curve analysis to evaluate the diagnostic performance of the 8-gene signature in the training set (Figure 6D), obtaining an AUC of 0.998. The AUC in the validation set was 0.929 (Figure 6E). We also used GSE22255 as an external dataset to validate the gene signature, obtaining an AUC of 0.825 (Figure 6F).

In the gene signature, ADCY4 and DUSP1 were upregulated in IS, while ATP5F1, DCTN5, EIF3G, ELAVL1, EXOSC7, and PPIE were downregulated (Figure 7A–B). The expression heatmap of the 8-gene signature in GSE22255 and GSE16561 is shown in Figure 7C.

Immune Cell Infiltration in IS

Among the IS-related molecular mechanisms, we found a large number of immune inflammatory responses. When we calculated the infiltration levels of immune cells in IS, we found that monocytes and neutrophils were the most abundant (Figure 8A). IS patients showed significantly
higher levels of macrophages M0, monocytes, and neutrophils than controls, but lower levels of naive CD4 T cells, activated NK cells, memory B cells, resting memory CD4 T cells, and CD8 T cells than controls (Figure 8B). Correlation analysis revealed that 8-gene signature correlated significantly with the levels of immune cell infiltration in blood samples (Figure 8C).

**Discussion**

We performed co-expression network analysis of gene expression data in IS patients, allowing us to identify potential biomarkers and target genes associated with IS. We identified an 8-gene signature that can predict the occurrence of IS, and we found that the signature correlated with immune cell infiltration. We also identified

![Figure 2](image-url)
a network of miRNAs regulating the expression of IS-associated genes.

WGCNA and MEGENA complement each other by maximally retaining IS-specific module genes in the gene networks. The analysis may be particularly reliable when it focuses on genes common to the two co-expression networks. When we explored the GO terms and KEGG pathways enriched in these intersection genes, we found enrichment of inflammation and immune response. This is consistent with studies linking the immune system to IS-associated inflammation that can lead to cell death on one hand, but can also support stroke recovery on the other hand. In our enrichment results, a large number of interleukin-related biological functions and signaling pathways were found to be involved in IS-related dysregulation. IL-1 is one of the most important cytokines in the development of ischemic injury, and it is a potential therapeutic target for IS. In the neuroinflammatory response of IS, IL-17A signaling is associated with post-stroke neurogenesis. IL-17 has the ability to disrupt the blood-brain barrier, and it contributes to neurodegeneration by
Figure 4 Analysis of biological processes and KEGG pathways enriched in genes related to IS. (A) The main 10 functional terms of biological process (BP) enrichment analysis. Orange columns are activated terms and blue columns are inhibited terms in IS relative to controls. (B) The main 10 signaling pathways of Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. Orange columns are activated terms and blue columns are inhibited terms in IS relative to controls. (C) Significantly activated or inhibited KEGG pathways in IS, based on GSEA. Abbreviations: NES, normalized enrichment score; pval, P values; padj, adjusted P values.
increasing the oxidative stress response. Oxidative stress and inflammatory activity are early events in the cascade of cerebral ischemic injury. The relationship between eosinophils and IS is currently unclear. Several studies have shown positive correlation between the number of eosinophils and the prognosis of IS. Additionally, activation of focal adhesion kinase contributes to angiogenesis, neurogenesis and functional recovery after IS. NOD-like receptor (NLR), which drives an inflammatory signaling pathway, shows promise as a therapeutic target in IS. This suggests that the dysregulated genes we identified may contribute to IS by modulating neuroimmune and inflammatory responses.

Interestingly, GO and KEGG analysis as well as GSEA showed that oxidative phosphorylation was suppressed in IS. Oxidative phosphorylation plays an important role in the response and recovery of neurons after oxidative stress, and dysfunction of this pathway can lead to ATP depletion, cellular signaling defects, and dysregulated apoptosis. Our finding of a loss-of-function of oxidative phosphorylation in IS should be explored as a potential risk factor for IS.

Several of the DEMiRs in IS that we identified have previously been linked to brain injury. The miR-651 can protect against focal cerebral ischemic injury, and may be a regulator of genes related to Alzheimer’s disease. The target gene we identified, ATP5F1, was confirmed to be downregulated in AD. ATP5F1 drives mitochondrial ATP production, and defects in mitochondrial ATP synthesis have previously been reported in amyotrophic lateral
Figure 6 Potential candidate genes for IS. (A) Areas under the receiver operating characteristic curve (AUCs) for dysregulated genes in the GSE16561 and GSE22255 datasets. (B) Mean AUCs of hub genes. Red represents genes upregulated in IS; blue, genes downregulated in IS. (C) Distribution of LASSO coefficients for the 8-gene signature. (D–F) Receiver operating characteristic curves for the gene signature in the (D) training set of GSE16561, (E) validation set of GSE16561 and (F) GSE22255.
Figure 7 Expression of signature genes in IS patients and controls. (A) The expression of signature genes in IS patients and controls in the GSE16561 dataset. **P<0.001. (B) Expression of signature genes in IS patients and controls in the GSE22255 dataset. (C) Heatmap of signature genes in IS patients and controls in the GSE16561 and GSE22255 datasets. Black type represents upregulation; blue type, downregulation.
Figure 8 Differences between IS patients and controls in brain tissue infiltration by immune cell types. (A) Bar plot of levels of infiltration by 18 types of immune cells in the brains of IS patients. (B) Differences in immune cell infiltration between IS patients and controls. Red indicates significantly greater infiltration in IS; blue, significantly less infiltration. (C) Correlations between immune cell infiltration and expression of signature genes in IS. *P < 0.05, **P < 0.01.

Abbreviation: FC, fold change; Abs(cor), absolute (correlation).
sclerosis (ALS), Huntington disease.\(^{41-43}\) The miR-138-5p alleviates neurological impairment in IS by promoting astrocyte proliferation and inhibiting inflammatory response.\(^{44}\) The miR-9-3p is upregulated in the brain tissue of IS patients,\(^{45}\) and it may be a diagnostic marker for experimental and human mild head injury.\(^{46}\) MiR-9 upregulation may target the expression of BDNF to promote neurodegeneration of dopaminergic neurons caused by neurotoxins and cell apoptosis.\(^{47}\) Hypoxia or reoxygenation downregulate miR-374a-3p in endothelial cells,\(^{48}\) so future work should explore whether this downregulation is associated with IS.

Here, we identified eight potential gene biomarkers associated with IS. All these genes have previously been associated with various aspects of IS, suggesting the reliability of our analyses. ADCY4 is abnormally expressed in stroke and is involved in renin-angiotensin signaling.\(^{49,50}\) DUSP1 is upregulated in IS and may have diagnostic value.\(^{51,52}\) In fact, this member of the oxidative phosphorylation pathway is associated with multiple neurodegenerative diseases,\(^{53}\) though little has been published about a role in IS. DCTN5 is a bipolar disorder-associated protein that encodes a subunit of the motor complex important for retrograde dendritic transport in neurons.\(^{54}\) EIF3G is involved in apoptosis and has been associated with neuropsychiatric disorders.\(^{55,56}\) ELAVL1 is a potential target for miR-9, and it plays a critical role in the progression of inflammation.\(^{57}\) EXOSC7 is found upregulated expression in animal models of stroke.\(^{58}\) Downregulation of PPIE is associated with cognitive dysfunction in IS.\(^{51,59}\)

Macrophages that substantially infiltrate stroke lesions in IS patients may adopt a pro-inflammatory phenotype that enhances brain damage or a catabolic inflammatory phenotype that enhances brain repair.\(^{60}\) Monocytes in the blood may mediate neuroinflammatory responses that can strongly affect IS outcomes, and these cells may serve as the basis of new therapeutic approaches.\(^{1}\) Neutrophils are also important targets of strategies to treat and even prevent IS.\(^{61}\) The delayed arrival of T cells in stroke lesions means that they may be targeted to treat IS during a longer window than other therapies.\(^{62}\) In conclusion, immune responses after IS may prove to be highly useful as targets for developing stroke treatments.

Our study has some limitations. First, the data we analyzed were from public databases and the sample size was small, and lacked validation on clinical samples for key outcomes. Second, the gene signature could not be reconstructed in the validation cohort in order to validate its prognostic value in IS. Then, the expression profiles of miRNAs in this study were obtained from different sample types. In which whole blood may contain many cells that can store miRNAs and release them, which affects miRNA expression. We also did not examine further how the genes in the signature may influence neuroinflammation in IS. Follow-up studies of biochemical and animal studies should be done to verify and extend key results.

## Conclusion

Module genes of IS were identified by applying WGCNA and MEGENA, and they were found to be involved mainly in inflammatory immune responses. We also identified a signature of 8 genes (ADCY4, DUSP1, ATP5F1, DCTN5, EIF3G, ELAVL1, EXOSC7, and PPIE) that may be useful as biomarkers of IS. If these biomarkers can be verified in experimental studies, our findings may help in the early diagnosis and targeted treatment of IS. Understanding these regulatory mechanisms and developing new target genes may lead to the development of new therapeutic strategies.

## Data Sharing Statement

The raw data, analyses and codes in this study can be obtained from the corresponding author upon reasonable request.

## Acknowledgments

This study was supported by the Scientific Research Project of Guangxi Health Commission (Z20200212, Z20200146, Z20210200 and Z20210683), the High-Level Medical Expert Training Program of Guangxi “139” Plan Funding (G201903049), the Project of Guigang Scientific Research and Technology Development Plan (Guigang 190802 Shezi), and the Guangxi Medical and Health Key Discipline Construction Project (Department of Emergency Medicine and Department of Geriatrics, The First People’s Hospital of Nanning).

## Disclosure

The authors report no conflicts of interest related to this work.

## References


