PORTLAND PRESS

Research Article

Bioinformatics method combined with logistic regression analysis reveal potentially important miRNAs in ischemic stroke

Zhiqiang Wei[1,*], Xingdi Qi[2,*], Yan Chen[1,*], Xiaoshuang Xia[1], Boyu Zheng[3], Xugang Sun[3], Guangming Zhang[3], Ling Wang[3], Qi Zhang[3], Chen Xu[3], Shihe Jiang[3], Xiulian Li[3], Bingxin Xie[3], Xiaohui Liao[3] and Ai Zhu[3]

[1]Department of Neurology, The Second Hospital of Tianjin Medical University, Tianjin 300211, P.R. China; [2]Public Administration, The Second Hospital of Tianjin Medical University, Tianjin 300211, P.R. China; [3]Department of Geriatric, The Second Hospital of Tianjin Medical University, Tianjin 300211, P.R. China

Correspondence: Zhiqiang Wei (weizhiqiangdoctor@yeah.net)

Purpose: The present study aimed to investigate the comprehensive differential expression profile of microRNAs (miRNAs) by screening for miRNA expression in ischemic stroke and normal samples.
Methods: Differentially expressed miRNA (DEM) analysis was conducted using *limma* R Bioconductor package. Target genes of DEMs were identified from TargetScanHuman and miRTarBase databases. Functional enrichment analysis of the target genes was performed using *clusterProfiler* R Bioconductor package. The miRNA-based ischemic stroke diagnostic signature was constructed via logistic regression analysis.
Results: Compared with the normal cohort, a total of 14 DEMs, including 5 up-regulated miRNAs and 9 down-regulated miRNAs, were identified in ischemic stroke patients. These DEMs have 1600 regulatory targets. Using a logistic regression model, the top five miRNAs were screened for constructing an miRNA-based ischemic stroke diagnostic signature. Using the miRNA–mRNA interaction pairs, two target genes (specificity protein 1 (*SP1*) and Argonaute 1 (*AGO1*)) were speculated to be the primary genes of ischemic stroke.
Discussion and conclusion: Here, several potential miRNAs biomarkers were identified and an miRNA-based diagnostic signature for ischemic stroke was established, which can be a valuable reference for future clinical researches.

Introduction

Stroke is an acute predominant cerebrovascular disease, which causes brain tissue damage owing to a sudden rupture of blood vessels in the brain or the inability of blood to flow into the brain because of vascular occlusion [1]. In 2010, the incidence of stroke was approximately 16.9 million, which added to a pool of 33 million stroke survivors worldwide. Currently, stroke is considered to be the second leading cause of death after ischemic heart disease [2]. Typical symptoms of stroke include sudden unilateral weakness, numbness, visual loss, diplopia, altered speech, ataxia, and non-orthostatic vertigo [3,4]. Stroke can be categorized into ischemic and hemorrhagic stroke. The incidence of ischemic stroke is higher than that of hemorrhagic stroke, accounting for 60–70% of the total number of strokes [5,6]. Cardioembolic stroke accounts for approximately 25–30% of ischemic stroke cases, and 25% of such cases pass all diagnostic tests and have no known causes. Because of the acute onset, treatment difficulty, and the lack of detection methods, such cases pose a great challenge to clinicians.

MicroRNAs (miRNAs) are a class of small endogenous RNAs that regulate gene expression post-transcriptionally and play a role in gene silencing and translation inhibition by binding to target

genes. The miRNAs are a highly conserved class of tissue-specific genes that have been found in all eukaryotic cells preserved across species since their discovery in 1993 [7,8]. In general, they are short RNA molecules measuring 19–25 nucleotides in size. A single miRNA can target hundreds of mRNAs and influence the expression of many genes often involved in a functional interacting pathway [9]. Appropriate maintenance of miRNA expression is required for a balanced physiological environment because these small molecules influence almost every genetic pathway from cell cycle checkpoint and cell proliferation to apoptosis, with a wide range of target genes [10]. In recent years, miRNA regulation has been extensively studied for their role in biological processes (BPs) as well as in the development and progression of various human diseases including ischemic stroke [11,12].

In mammals, the brain exhibits a high level and activity of several miRNAs that show region-specific expression [13,14]. Studies conducted using conditional knock-out Dicer (RNA enzymes are critical for miRNAs biogenesis) have demonstrated the indispensable functional significance of miRNAs in controlling processes that include cellular differentiation, proliferation, synaptic morphogenesis, and vascular formation [15–17], indicating an overlap between ischemic stroke and miRNAs. It has been suggested that stroke alters the expression levels of many miRNAs in human blood and brain [18,19].

Despite decades of research, treatment for ischemic stroke is limited to thrombolytic therapy and symptom management. Moreover, a considerable number of patients remain asymptomatic and cannot be detected at onset. To this end, a more comprehensive approach to predict potential ischemic stroke patients based on the differential expression of specific miRNAs is required. To address these issues, we performed bioinformatics analysis combined with machine learning algorithms to identify potential candidate diagnostic miRNAs. The present study will help us screen for ischemic stroke-associated miRNA biomarkers and can be tremendously useful for ischemic stroke patients.

Materials and methods
Study materials
The materials used in the present study were obtained from the Gene Expression Omnibus (GEO, https://ncbi.nlm.nih.gov/geo) with the accession number of GSE55937, which included 24 blood samples from healthy individuals and 24 blood samples from ischemic stroke. The miRNA expression profiles of the abovementioned samples were detected based on Affymetrix Multispecies miRNA-3 Array chip platform.

Differential expression analysis
The miRNA expression profiles were normalized using robust multi-array (RMA) method via the *affy* R Bioconductor package and standardized by logarithmic transformation. Differentially expressed miRNAs (DEMs) were screened using the *limma* R Bioconductor package by employing the criteria of absolute log-transformed fold change ($|\log2FC|$) > 0.5 and P-value ≤ 0.05.

Construction of miRNA–mRNA regulatory network
The target genes of DEMs were searched from TargetScan (Release 7.2: March 2018 www.targetscan.org) and miRTarBase (Release 7.0: 15 September 2017; mirtarbase.mbc.nctu.edu.tw). Target genes that were common between the two databases were used for constructing the miRNA–mRNA regulatory network. Cytoscape software was applied for visualizing the regulatory network.

Functional enrichment analysis
Functional enrichment analysis of the target genes of DEMs was conducted using clusterProfiler Bioconductor package. Ultimately, Gene Ontology (GO, including BP, Molecular Function, and Cellular Component) and KEGG pathways that satisfied Benjamini–Hochberg (BH)-adjusted P-value of <0.05 were retained.

Construction of logistic regression model
Here, we proposed to test if DEMs could help distinguish stroke samples from normal ones. For this purpose, logistic regression analysis was performed by considering DEMs and sample groups as continuous predictor and categorical responsory, respectively, based on the *glm* basic R function. Each DEM with a P-value of <0.05 was retained for constructing the prediction model.
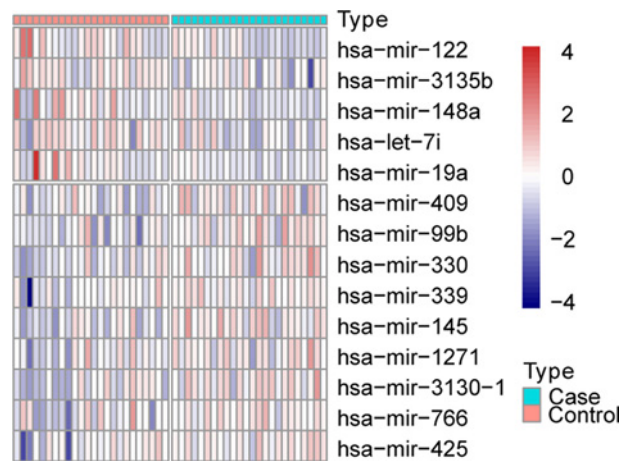
Figure 1. The overview of DEMs' expression profile between ischemic stroke and normal samples
Heat map display of the 14 DEMs in all samples. The horizontal axis is the sample and the vertical axis is miRNA. Red represents high expression and blue represents low expression.

Statistical analysis

Statistical analyses were performed using R software v3.5.2. *Affy* R Bioconductor package for the normalization of raw expression profiles. *Limma* R Bioconductor package was used for conducting differential expression analysis. A *P*-value of <0.05 was considered to be statistically significant in all of the abovementioned analyses.
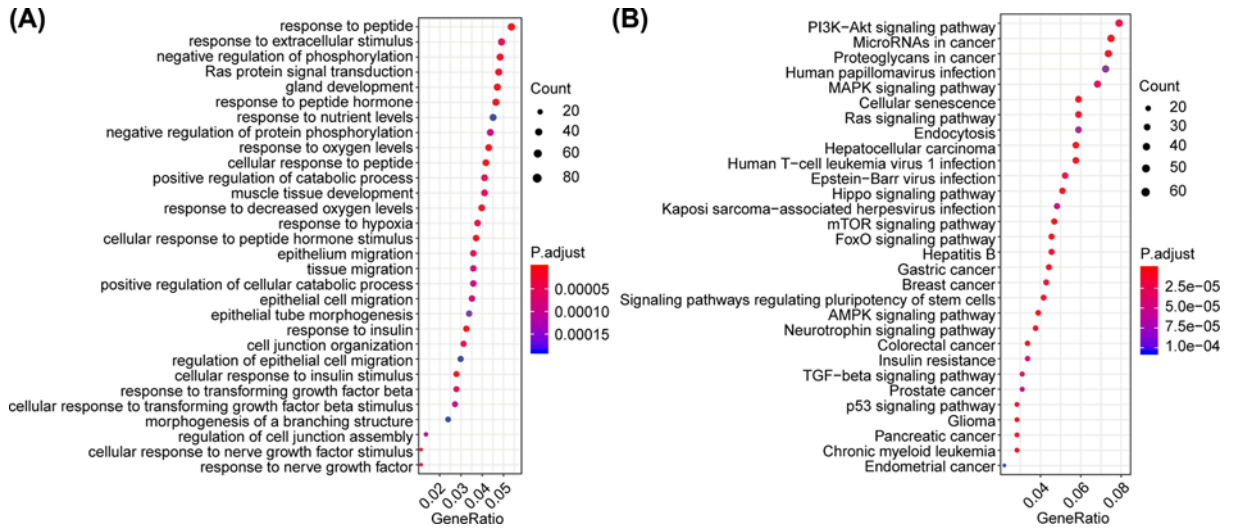
Results
DEMs

Expression of all miRNAs contained in the miRNA chip after normalization is shown in Supplementary Figure S1A, which indicates that the normalization process successfully eliminated the batch effects. After calculation, we obtained a total of 14 DEMs in ischemic stroke samples in comparison with the normal samples, including 5 up-regulated miRNAs and 9 down-regulated miRNAs, as shown in Supplementary Figure S1B. Expressions of those 14 DEMs in normal and ischemic stroke groups were illustrated as a heatmap in Figure 1.

Target genes of DEMs

A total of 1600 target genes (Supplementary Table S1) were simultaneously predicted by TargetScanHuman and miR-TarBase databases for 14 DEMs; functional enrichment analysis of these 1600 target genes led to the identification of a total of 499 and 87 significantly enriched GO terms, respectively. Figure 2A,B illustrates the top 30 most significant GO terms and KEGG pathways.

DEMs effectively characterize ischemic stroke

Correlation of expression of the 14 DEMs in ischemic stroke and normal samples is shown in Figure 3A, indicating that there was no particularly strong collinear relationship among them. Thus, all these 14 DEMs were used for the construction of logistic regression model. Receiver operating characteristic (ROC) analysis was used for evaluating the performance of the model. Consequently, the area under curve (AUC) value of the model was 0.8645, as shown in Figure 3B, which proved that the logistic model could robustly determine the sample type. More importantly, we found that the *P*-values of the five miRNAs, hsa_mir_122, hsa_mir_99b, hsa_mir_339, hsa_mir_145, and hsa_mir_3130_1, were less than 0.05, indicating that those five miRNAs had a greater contribution to the model than the remaining nine miRNAs. Hence, we reconstructed the logistic model using those five miRNAs, and it was found that the AUC value of the logistic model could reach 0.8589 (Figure 3C). Additionally, we conducted five-fold cross-validation basing on the dataset, and result illustrated high AUC value (Figure 3D). The abovementioned results proved that the model constructed based on these five miRNAs could effectively predict the sample type and should be more cost-effective for the diagnosis of ischemic stroke.

Figure 2. The GO and KEGG enrichment analysis of the 1600 target genes

GO and KEGG enrichment results are shown in (**A**) and (**B**), respectively. The horizontal axis in the figure represents the ratio of genes enriched, and the vertical axis represents the name of each BP or pathway.
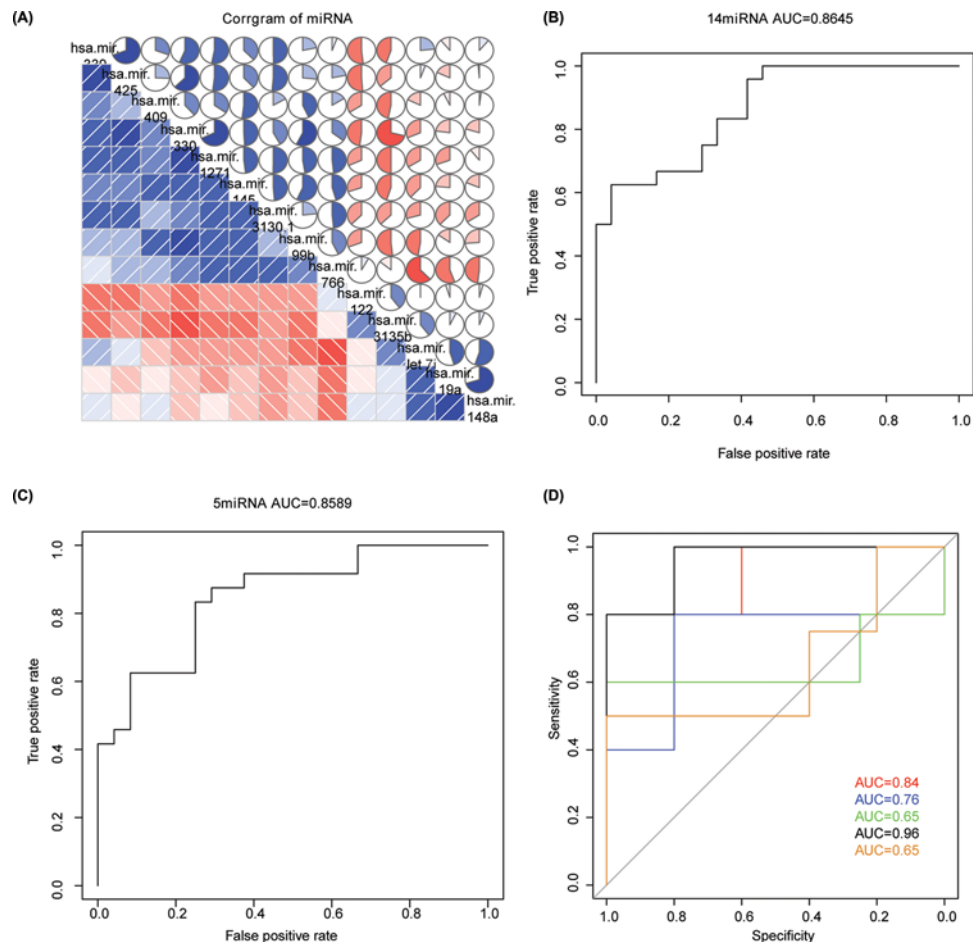
Figure 3. The collinear analysis and ROC curve of the DEMs

(**A**) The collinear analysis of the 14 DEMs. The darker the blue or red color, and the larger the area of blue or red color in the Pie, the greater the collinearity between them. (**B**) The ROC curve of the model. The AUC value of the Logistic model was found to be 0.8645. (**C**) The top five miRNAs were used to reconstruct the logistic model with the AUC value of the logistic model reached 0.8589. (**D**) The ROC curve of the model basing five-fold cross-validation.

Specificity protein 1 and Argonaute 1 are highly connected in miRNA–mRNA regulatory network

We constructed the miRNA–mRNA network for all 14 DEMs and the five DEMs that had a *P*-value of <0.05 in logistic regression analysis, as shown in Supplementary Figures S2A,B, respectively. Nodes in the network were colored according to their connectivity, i.e., number of their direct neighbors; furthermore, of the 1600 target genes, specificity protein 1 (*SP1*) and Argonaute 1 (*AGO1*) were the two genes that were regulated by at least two miRNAs in both regulatory networks. Therefore, they might be pivotal biomarkers in the development of ischemic stroke.

Discussion

Ischemic stroke is a leading cause of death and disability, resulting in over 6 million deaths per year worldwide [20]. In addition to the high mortality rate, the timely monitoring of undiagnosed stroke patients is also a very critical issue. All these issues highlight the dire need for effective forecasting targets or biomarkers. Because miRNAs are implicated in a wide variety of diseases and have been shown to be essential for diverse proper physiological functions in the human brain, it is beneficial to develop a comprehensive specific expression profile of miRNAs in ischemic stroke patients for identifying potential miRNAs candidates as well as for targeting mRNA. Here, compared with the normal samples, a total of 14 DEMs in ischemic stroke patients were screened, including 5 up-regulated and 9 down-regulated miRNAs. Moreover, 1600 target genes have been further identified. We also studied the BPs related to these genes via GO and KEGG enrichment analysis. We found that these 1600 target genes were significantly enriched in ischemic stroke-related BPs, such as response to oxygen levels and response to decreased oxygen levels. The results of this biomarker selection study demonstrate not only the rationality of our method (many related BPs can be selected) but also the importance of two processes (response to oxygen levels and response to decreased oxygen levels) in ischemic stroke. The responses to oxygen levels and to decreased oxygen levels have been researched with regard to transport, oxygen homeostasis, translation, nitrogen fixation, and angiogenesis, which are involved in hypoxia, retinal neoplasms, tumor angiogenesis, retinoblastoma, and neoplasms [21,22]. A laboratory study conducted by Perez-Alvarez using an *in vivo* model revealed that mTORC1 (mammalian target of rapamycin complex-1), a protein complex downstream of PI3K-Akt pathway, was dysregulated after ischemic stroke and oxygen–glucose deprivation [23]. This evidence highlights the importance of understanding the relation between response to oxygen levels and ischemic stroke. Few other topics concerning response to oxygen levels and response to degraded oxygen levels in ischemic stroke may provide scope for more novel studies in the future.

We further reconstructed the logistic model and identified top five miRNAs (hsa_mir_122, hsa_mir_99b, hsa_mir_339, hsa_mir_145, and hsa_mir_3130_1). Previously, based on the screening of miRNA functional synergistic network, miR-145, miR-122, and miR-99b have been implicated to be associated with ischemic stroke by participating in the processes of post-ischemic neuronal damage and thrombosis, respectively, [24]. Based on our search, we found that the hsa_mir_339 and hsa_mir_3130_1 have not been well studied in ischemic stroke, which provide a very valuable starting point for future biomarker selection studies. The miR-339 and miR-3130 were mainly reported to be tumor suppressors in previous studies [25–27] owing to their biological roles in the suppression of cell proliferation. Additionally, Martinez et al. [28] also illustrated that miR-339 in the cerebellum and plasma of rats could be perturbed by *in vitro* stimulation with agents ethanol and caffeine; this indicates the potential of miR-339 as a novel biomarker for ischemic stroke. From the miRNA–mRNA regulatory network, two target genes (*SP1* and *AGO1*) are speculated to be the primary genes of ischemic stroke. AGO1 plays critical roles in RNA interference among the many regulators participating in miRNA formation. According to the findings reported by Shi et al., the expression of miR-103 is modulated by hypoxia-inducible factor 1α, which can target AGO1 to promote tumor vessel formation. Meanwhile, miR-103 can substantially affect angiogenesis and vascular density after ischemic stroke by targeting vascular endothelial growth factor (VEGF) [29]. SP1 is a member of a family of transcription factors that include SP2, SP3, and SP4; these factors are implicated in various essential BPs and have been established to play important roles in cell growth, differentiation, apoptosis, and carcinogenesis. SP1 reportedly interacts with zinc finger protein 179 (Znf179), which is a neuroprotective factor for the accumulation of reactive oxygen species (ROS). Znf179 autoregulation through Sp1-dependent mechanism plays an important role in neuroprotection, and NGF-induced Sp1 signaling may help attenuate more extensive (ROS-induced) damage following brain injury [30]. Both genes are related to ischemic stroke on some levels, and both those genes deserve to be investigated further in detail.

In conclusion, in light of the fact that no gold standard treatment is currently available and that disease-specific prediction for ischemic stroke remains unreliable despite the presence of several standard criteria, we summarized the differential expression profile of miRNAs in ischemic stroke. The manner of expression of five miRNAs as well as of the two specific target genes (*SP1* and *AGO1*) may provide new insights into the discovery of therapeutic biomarkers.

Because many pathological states are known to alter miRNA profiles and functions, understanding those changes and developing new target genes to rectify them might lead to the formulation of novel therapeutic strategies.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

Author Contribution

Z.W. made substantial contributions to the study conception and design. Z.W., Y.C. and X.Q. provided valuable instructions and suggestions for this paper, written this article and analyzed the data, and helped for revising the manuscript. X.X., B.Z., X.S., G.Z., L.W., helped for revising the manuscript. Q.Z., C.X., S.J., X.L., B.X., X.L., Z.A. participated in the collection and analysis of the data. All authors read and approved the final manuscript.

Abbreviations

AGO1, Argonaute 1; BP, biological process; DEM, differentially expressed miRNA; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; miRNA, microRNA; NGF, neuronal growth factor; ROS, reactive oxygen species; SP1, specificity protein 1; Znf179, zinc finger protein 179.

References

1 Hankey, G.J. (2017) Stroke. *Lancet* **389**, 641–654, https://doi.org/10.1016/S0140-6736(16)30962-X

2 Katan, M. and Luft, A. (2018) Global burden of stroke. *Semin. Neurol.* **38**, 208–211, https://doi.org/10.1055/s-0038-1649503

3 Kamel, H. and Healey, J.S. (2017) Cardioembolic stroke. *Circ. Res.* **120**, 514–526, https://doi.org/10.1161/CIRCRESAHA.116.308407

4 Rabinstein, A.A. (2017) Treatment of acute ischemic stroke. *Continuum* **23**, 62–81

5 Meschia, J.F. and Brott, T. (2018) Ischaemic stroke. *Eur. J. Neurol.* **25**, 35–40, https://doi.org/10.1111/ene.13409

6 Kim, J.S. and Caplan, L.R. (2016) Clinical stroke syndromes. *Front. Neurol. Neurosci.* **40**, 72–92, https://doi.org/10.1159/000448303

7 Vishnoi, A. and Rani, S. (2017) MiRNA biogenesis and regulation of diseases: an overview. *Methods Mol. Biol.* **1509**, 1–10, https://doi.org/10.1007/978-1-4939-6524-3'1

8 Tiwari, A., Mukherjee, B. and Dixit, M. (2018) MicroRNA key to angiogenesis regulation: miRNA biology and therapy. *Curr. Cancer Drug Targets* **18**, 266–277, https://doi.org/10.2174/1568009617666170630142725

9 Verjans, R., van Bilsen, M. and Schroen, B. (2017) MiRNA deregulation in cardiac aging and associated disorders. *Int. Rev. Cell Mol. Biol.* **334**, 207–263, https://doi.org/10.1016/bs.ircmb.2017.03.004

10 Backes, C., Meese, E. and Keller, A. (2016) Specific miRNA Disease biomarkers in blood, serum and plasma: challenges and prospects. *Mol. Diagn. Ther.* **20**, 509–518

11 Heydari, E., Alishahi, M., Ghaedrahmati, F., Winlow, W., Khoshnam, S.E. and Anbiyaiee, A. (2020) The role of non-coding RNAs in neuroprotection and angiogenesis following ischemic stroke. *Metab. Brain Dis.* **35**, 31–43, https://doi.org/10.1007/s11011-019-00485-2

12 Lu, T.X. and Rothenberg, M.E. (2018) MicroRNA. *J. Allergy Clin. Immunol.* **141**, 1202–1207, https://doi.org/10.1016/j.jaci.2017.08.034

13 Olejniczak, M., Kotowska-Zimmer, A. and Krzyzosiak, W. (2018) Stress-induced changes in miRNA biogenesis and functioning. *Cell. Mol. Life Sci.* **75**, 177–191, https://doi.org/10.1007/s00018-017-2591-0

14 Hrovatin, K. and Kunej, T. (2018) Classification of miRNA-related sequence variations. *Epigenomics* **10**, 463–481, https://doi.org/10.2217/epi-2017-0126

15 Le, T.D., Zhang, J., Liu, L. and Li, J. (2017) Computational methods for identifying miRNA sponge interactions. *Brief. Bioinform.* **18**, 577–590

16 Wang, D., Xin, L., Lin, J.H., Liao, Z., Ji, J.T. et al. (2017) Identifying miRNA-mRNA regulation network of chronic pancreatitis based on the significant functional expression. *Medicine (Baltimore)* **96**, e666

17 Li, G., Morris-Blanco, K.C., Lopez, M.S., Yang, T., Zhao, H. et al. (2018) Impact of microRNAs on ischemic stroke: from pre- to post-disease. *Prog. Neurobiol.* **163–164**, 59–78, https://doi.org/10.1016/j.pneurobio.2017.08.002

18 Sorensen, S.S., Nygaard, A.B., Nielsen, M.Y., Jensen, K., Christensen, T. et al. (2014) miRNA expression profiles in cerebrospinal fluid and blood of patients with acute ischemic stroke. *Transl. Stroke Res.* **5**, 711–718, https://doi.org/10.1007/s12975-014-0364-8

19 Li, P., Teng, F., Gao, F., Zhang, M., Wu, J. and Zhang, C. (2015) Identification of circulating microRNAs as potential biomarkers for detecting acute ischemic stroke. *Cell. Mol. Neurobiol.* **35**, 433–447, https://doi.org/10.1007/s10571-014-0139-5

20 Rodgers, H. (2013) Stroke. *Handb. Clin. Neurol.* **110**, 427–433, https://doi.org/10.1016/B978-0-444-52901-5.00036-8

21 Choudhry, H. and Harris, A.L. (2018) Advances in hypoxia-inducible factor biology. *Cell Metab.* **27**, 281–298, https://doi.org/10.1016/j.cmet.2017.10.005

22 Chen, R., Lai, U.H., Zhu, L., Singh, A., Ahmed, M. and Forsyth, N.R. (2018) Reactive oxygen species formation in the brain at different oxygen levels: the role of hypoxia inducible factors. *Front. Cell Dev. Biol.* **6**, 132, https://doi.org/10.3389/fcell.2018.00132

23 Perez-Alvarez, M.J., Villa Gonzalez, M., Benito-Cuesta, I. and Wandosell, F.G. (2018) Role of mTORC1 controlling proteostasis after brain ischemia. *Front. Neurosci.* **12**, 60, https://doi.org/10.3389/fnins.2018.00060

24 He, W., Chen, S., Chen, X., Li, S. and Chen, W. (2016) Bioinformatic analysis of potential microRNAs in ischemic stroke. *J. Stroke Cerebrovasc. Dis.* **25**, 1753–1759, https://doi.org/10.1016/j.jstrokecerebrovasdis.2016.03.023

25 Banerjee, H., Krauss, C., Worthington, M., Banerjee, N., Walker, R.S. et al. (2019) Differential expression of efferocytosis and phagocytosis associated genes in tumor associated macrophages exposed to African American patient derived prostate cancer microenvironment. *J. Solid Tumors* **9**, 22–27, https://doi.org/10.5430/jst.v9n2p22

26 Luo, A., Zhou, X., Shi, X., Zhao, Y., Men, Y. et al. (2019) Exosome-derived miR-339-5p mediates radiosensitivity by targeting Cdc25A in locally advanced esophageal squamous cell carcinoma. *Oncogene* **38**, 4990–5006, https://doi.org/10.1038/s41388-019-0771-0

27 Weber, C.E., Luo, C., Hotz-Wagenblatt, A., Gardyan, A., Kordass, T. et al. (2016) miR-339-3p is a tumor suppressor in melanoma. *Cancer Res.* **76**, 3562–3571, https://doi.org/10.1158/0008-5472.CAN-15-2932

28 Rossetto, I.M.U., Cagnon, V.H.A., Lizarte, F.S.N., Tirapelli, L.F., Tirapelli, D.P.C. et al. (2019) Ethanol and caffeine consumption modulates the expression of miRNAs in the cerebellum and plasma of UChB rats. *Life Sci.* **229**, 180–186, https://doi.org/10.1016/j.lfs.2019.05.016

29 Shi, F.P., Wang, X.H., Zhang, H.X., Shang, M.M., Liu, X.X. et al. (2018) MiR-103 regulates the angiogenesis of ischemic stroke rats by targeting vascular endothelial growth factor (VEGF). *Iranian J. Basic Med. Sci.* **21**, 318–324

30 Chuang, J.Y., Kao, T.J., Lin, S.H., Wu, A.C., Lee, P.T. et al. (2017) Specificity protein 1-zinc finger protein 179 pathway is involved in the attenuation of oxidative stress following brain injury. *Redox. Biol.* **11**, 135–143, https://doi.org/10.1016/j.redox.2016.11.012