New insight in understanding and exploring optimal biomarkers in stroke: Key diagnostic markers for clinical treatment

Peng Zhang (pdcdc_hiv@126.com)
Shanghai University of Medicine and Health Sciences https://orcid.org/0000-0002-5157-3103
Jiafeng Wang
Stem Cell Research and Cellular Therapy Center, Affiliated Hospital of Guangdong Medical University
Aiqin Fan
Pudong New Area Ling qiao Community Health Service Center
Xiaolan Wang
Shanghai University of Medicine and Health Sciences
Liying Jiang
Shanghai University of Medicine and Health Sciences
Lianying Jiang
Pudong New Area People's Hospital Affiliated to Shanghai University of Medicine and Health Sciences
Changlian Lu
Shanghai University of Medicine and Health Sciences
Xuefeng Gu
Shanghai University of Medicine and Health Sciences

Research article

Keywords: Ischemic stroke; biomarkers; key regulators; microarray analysis; gene-disease

Posted Date: October 7th, 2019

DOI: https://doi.org/10.21203/rs.2.15758/v1

Abstract

Background A stroke is a complex neurological phenomenon where patients may suddenly experience paralysis, loss of vision or impaired speech because of interruption of blood flow. Stroke is the second leading cause of death and a significant reason for long-term disability worldwide. Optimal biomarkers latent in stroke remain relatively unknown, and efforts are ongoing to effectively define the molecular fingerprints of a stroke. In this study, we utilized microarray gene expression data for the in silico discovery of novel key diagnostic markers which can be utilized for clinical treatment of ischemic stroke.Methods and results We performed differential expression analysis and obtained 20 genes which have at least a two-fold change in the expression level. We also performed gene-disease association studies, functional categories enrichment and GO term enrichment analysis, tissue expression analysis, protein-protein interaction (PPI) analysis, interaction study of miRNAs that target identified biomarkers. The identified diagnostic markers were validated with the PubMed literature using publication enrichment analysis.Conclusions It was observed that most of the identified genes, including AKAP7, ARG1, S100A2, MMP9, TIMP, and CCR7 are markers for early or post-ischemic stroke.

Background

A stroke, brain stroke, or a brain attack is a complex neurological phenomenon, where patients may suddenly experience paralysis, loss of vision or impaired speech because of interruption of blood flow. It is the second leading cause of death and a significant reason of long-term disability worldwide [1, 2]. Within the United States, a stroke is the third major cause of death, and on average, one person every 40 seconds will experience a stroke, and every 4 minutes will die from a stroke, which is likely to increase due to population aging [1, 3]. In 2016, there were 5.5 million deaths and 116.4 million disability-adjusted life-years caused by stroke worldwide [4]. The global cost burden of stroke is very high which is projected to be 1.52 trillion dollars in 2050 [3].

In fact, a stroke is a medical emergency which damages the brain where no supply of blood is found, leading to neurological dysfunctions. It is caused by vascular injury to the brain, which prohibits the proper supply of blood to the brain [5]. Some of the common symptoms are headache, paralysis or numbness of the face, arms or legs, blurred or blackened vision, and difficulty in speaking and walking [6, 7]. Some of the common types of strokes are Ischemic stroke, Hemorrhagic stroke, and Transient Ischemic Attack (TIA). According to the American Stroke Association, an ischemic stroke (clots) is the most common type of stroke which accounts for 87% of all strokes. It occurs when blood supply to the brain is obstructed due to blood clots in the cerebral artery [8]. Atherosclerosis, fatty deposits lining the vessel walls, is the root cause of ischemic stroke [9]. Hemorrhagic stroke (bleeds) occurs when a weakened blood vessel in the brain ruptures and bleeds into the surrounding tissue, which further accumulates and compresses the surrounding tissues. This type of stroke can occur in response to hypertension, blood thinner use and an aneurysm [10], and accounts for 13% of total stroke cases. TIA, also called a mini-stroke, has temporary symptoms which usually last for 5 minutes. These symptoms generally do not result in permanent tissue damage, however, if a TIA attack continues for more than an hour, there will be a chance of permanent tissue damage. It primarily occurs when debris blocks the blood flow of the nervous system temporarily [11–13].

Strokes are usually treated using anti-platelet drugs, anticoagulants, thrombolytic agents, and angioplasty. The commonly used anti-platelet drug is Aspirin which is the least expensive treatment with low side effects [7, 14]. These medications help in un-sticking the platelets [15]. Anticoagulants include heparin and warfarin, which directly affect the clotting system protein. Heparin is used for a short time as compared to Aspirin. Thrombolytic agents, also known as recombinant tissue plasminogen activators (r-tPA), help in dissolving the clots by removing the obstruction and restoring blood flow to the brain [16]. In fact, r-tPA is the only FDA-approved medication for ischemic stroke which must be administered within three-hours from the onset of symptoms. Angioplasty is a procedure which involves use of a balloon-like device to open a clogged artery, such as the carotid, and to keep it open, a small wire tube is placed into the artery [17].

Literature Review

The effective treatment of critical stroke patients depends on a quick assessment of the type of stroke and general clinical status of the patients. For instance, thrombolytic therapy with r-tPA is the only FDA-approved treatment for ischemic stroke, whereas there is no effective treatment available for hemorrhagic stroke [18]. Hence, quick and objective assessment of the type of stroke during admission would improve treatment outcome. In order to diagnose acute stroke correctly, measurement of molecular characteristics from biomarkers is used. Biomarkers help to develop targets for neuroprotective therapies and monitor treatment response [19].

Several kinds of biomarkers are studied in stroke to recognize therapeutic targets, including physical markers, neuroimaging markers, electro-physiological markers, as well as histological, serum, neuronal and genetic markers [20]. Neuroimaging is a commonly used method to diagnose stroke, to estimate the severity and to predict the risk of recurrence. Histological markers are used to evaluate uncommon causes, including vasculitis and collagen vascular disease. Serum biomarkers help in the diagnosis of intracerebral hemorrhage and cerebral ischemia and testing for metabolic risk factors but interpreting them can be complicated by the effects of the blood-brain barrier, and individual heterogeneity limit its usage for early diagnosis and lesion size estimation. Further, genetic markers can be used to predict heritable cerebrovascular conditions associated with a stroke which is quite applicable in personalizing treatments [21].

Literature suggests that brain-specific protein biomarkers of glial, glial fibrillary acidic protein (GFAP) and neuronal cell injury can be detected in the cerebrospinal fluid and may also be detected within the peripheral blood supply for prompt and better decision making [18]. GFAP is released into the blood which provides important clinical information for stroke prognosis at an early stage of development as it helps in differentiating the types of strokes [20].

Despite extensive research of biomarkers in stroke, the current studies have not achieved significant evidence to draw a final conclusion. The availability of molecular and genetic data such as more than 250,000 proteins, 20,000 coding genes, and a growing number of non-coding genes, metabolites and other

molecules, all need to be utilized for the *in silico* detection and validation of genetic biomarkers of strokes [22]. Hence, there is a need to further explore high throughput genetic data (such as Microarray or RNA-seq) for the *in silico* discovery of novel key biomarkers which can be utilized for the clinical diagnosis and immediate treatment of ischemic stroke.

Methods

The gene expression profiles from peripheral whole blood of acute ischemic stroke patients, who were diagnosed with MRI and whose age were greater than 18 years, were collected in Paxgene Blood RNA tubes. The total samples of ischemic stroke patients were 39, while 24 neurologically healthy (non-stroke) controls were also collected. The complete gene expression profile is available at NCBI-GEO with Accession No. GSE16561 [23] (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16561).

Differential expression analysis

Gene expression profiling was performed on NCBI-GEO GSE16561 dataset to elucidate gene expression in chronic ischemic stroke patients and to recognize the crucial marker genes involved in ischemic stroke. The data obtained were pre-processed and normalized using a GEO2R web-based tool available at NCBI-GEO. Further, the fold change method was applied to identify differentially expressed genes (DEGs). Fold change is one of the simplest and widely used methods for DEGs and describes how the expression level of a gene changes over the two different conditions or groups (control versus disease). It is computed as a ratio of averages from control and disease samples. The levels of fold change are observed, and genes under or above a threshold are selected [24]. It is suggested in the literature that a fold change below 0.5 is considered down-regulated, whereas a fold change above 2.0 is considered up-regulated [25]. The results obtained were forwarded to functional enrichment and disease association studies, and subsequent downstream analysis.

Functional enrichment and disease association studies

Database for Annotation, Visualization and Integrated Discovery (DAVID 6.8) [26, 27] (URL: https://david.ncifcrf.gov/) was recognized as over-represented biological features for DEG sets and network modules. The terms Gene Ontology (GO) and genome encyclopedia were accepted as functional terms 35. In the DEG sets, considerably overrepresented biological features were identified using the corrected P-value < 0.05 from Benjamini-Hochberg. In order to view the outcomes, different GO term features for DEG sets were produced from the most overrepresented biological features.

Protein-Protein Interaction Studies

Protein-protein interactions (PPIs) play a pivotal role in both predicting the functions of the target protein and therapeutic potential of the targeted molecules. Most of the genes and proteins function as a result of a set of interactions [28], and understanding the physical, as well as the functional interactions, are important for the biologists. Some of the experimental methods for PPI detection are affinity purification, yeast 2 hybrid (Y2H) and tandem affinity purification (TAP). However, computational methods are also available for this purpose. In this research, we performed the PPI studies using a STRING database (https://string-db.org/), which contains experimentally validated as well as predicted PPIs, including direct (physical) and indirect (functional) interactions/associations.

MicroRNAs and gene biomarkers interaction studies

MicroRNAs are small, ~22 nucleotides long, non-coding RNAs that control gene expression at the post-transcriptional level by complementary base-pairing with the target messenger RNA (mRNA), which degrades mRNA and blocks the translation process [29]. Dysfunction of microRNAs can lead to the development and progression of several human diseases, including neurological disorders and strokes [30, 31]. Hence, the study of microRNAs targeting detected gene biomarkers in stroke is necessary. In this paper, we have utilized the miRTargetLink for Human (https://ccb-web.cs.uni-saarland.de/mirtargetlink/) tool for the interaction studies of microRNAs and gene biomarkers.

Results And Discussion

The considered gene expression datasets consist of a total of 63 samples, of which 39 belong to ischemic stroke patients and 24 are from controls who were neurologically healthy (non-stroke. To appreciate the distribution of gene expression data within these two groups, a boxplot is depicted in figure 1A. It is observed from the boxplot (figure 1A) that log2 values of gene expression lie between 1.5 to -2.0, while their 2^{nd} quartile (Mean) fluctuates between -0.5 to 0.5, suggesting the gene expression data are uniformly distributed.

Figure 1B depicts the boxplot of the gene expression profile grouped into four samples – male stroke patients, female stroke patients, male non-stroke and female non-stroke. It is observed that the female stroke group appears to be predominantly down regulated when compared to the female control group.

In order to find key diagnostic biomarkers, i.e. differentially expression genes (DEGs), we computed the fold change statistics between the two groups of samples for all the genes. We filtered DEGs with a significance level of 5% (p-value <=0.05) and have at least a two-fold change in their expression. In this way, we obtained 20 DEGs which had at least a two-fold change in the expression level between the two groups. Out of 20 DEGs, only one gene had a three-fold change in the expression level. The list of identified DEGs along with their statistics such as adjusted p-value, moderated t-statistics, B-statistics, log fold change and fold change is given in Table 1.

The scatter plot of fold change scores and log fold change scores are shown in figure 2, respectively. It is observed from both Table 1 and figure 2A that, with the exception of gene C-C motif chemokine receptor 7 (*CCR7*), all the identified DEGs are down-regulated. The profile graph of the top five highly differentially expressed genes such as ARG1, MMP9, S100A12, ORM1, FCGR3B are shown in figure 3.

Gene-Disease Association Studies

We performed the DAVID analysis of identified DEGs. The gene-disease association studies of these DEGs, along with their different scores are presented in Table 2A. It can be observed from the disease association study using Genetic Association Database (GAD) [32] that four genes with GeneBank IDs NM_004994, NM_001995, NM_006418, and NM_000570 are associated with disease term “Stroke”, while few other genes are associated with the disease terms “brain hemorrhage”, “Guillain-Barre syndrome”, or “Multiple Sclerosis”.

Functional categories and GO analysis

Functional categories analysis helps group the related genes based on their protein domain families as most co-functioning genes belong to the same protein families. Functional categories are usually derived from Gene Ontology (GO) and Pfam databases. In GO database, every GO term is represented as a node in a directed acyclic graph, and functional categories are defined as genes annotated either directly to a node or to any descendant node in the ontology. Results of the functional categories enrichment and GO term enrichment analysis of all the identified DEGs are presented in Table 2B and Table 2C, respectively. The identified DEGs associated with stroke such as MMP9 (NM_004994), ACSL1 (NM_001995), OLFM4 (NM_006418) and FCGR3B (NM_000570) are enriched with different GO terms. For instance, MMP9, OLFM4 and FCGR3B are more than four-fold enriched with the term “Secreted”, more than two-fold enriched with the term “Signal”, and three-fold enriched with the term “Disulfide bond” (Table 2B & Table 2C). Hence, these genes are responsible for the controlled release of substance by cells or tissues, transmission of information in the biological system, and catalysis of the rearrangement intrachain and interchain disulfide bonds in proteins. Any perturbation to these genes may lead the mentioned biological dysfunctions.

Tissue expression analysis

The DAVID tool integrates world-class tissue expression data including GNF-Affy, CGAP-SAGE, CGAP-EST, and Unigene-EST, where we can quickly find the most enriched gene expression patterns across thousands of normal and disease tissues for any given gene lists. Tissue expression analysis allows the identification of biomarkers and gene expression pattern discovery. Tissue expression analysis is performed by the DAVID tool with a threshold of $p < 0.05$ (Table 2D). The term “Whole Brain_3rd” is enriched with a count of 18 and 90% similarity with the identified DEGs, which states that gene expresses higher than a 3rd quartile of its expression across all the tissues. Out of 20 identified DEGs in this study, 18 genes are significantly enriched for genes expressed in the brain tissue, including four genes involved in stroke (Table 2D). These enriched brain-expression genes illustrate that gene expression profiles in peripheral blood may be relevant for quantitative metabolic phenotypes in stroke.

Protein-Protein Interaction Studies

The PPI studies were performed using the STRING database (https://string-db.org/). We performed PPI studies of only those DEGs which have been found to be associated with “stroke”, as per Genetic Association Database (Ref. Table 2). Stroke associated genes are MMP9 (GeneBank ID: NM_004994), ACSL1(GeneBank ID: NM_001995), OLFM4 (GeneBank ID: NM_006418), and FCGR3B (GeneBank ID: NM_000570). We considered three interaction sources namely, Experiments, Co-expression, and Textmining, with an interaction score of 0.90 (highest confidence) from the STRING database. The networks of these genes are shown in figure 4. These PPI networks also present KEGG and Reactome pathway analysis. For instance, MMP9 interacts with IL6, LCN2, IL1B, TNF, and CXCL8 which are involved in hsa04657 (IL-17 signaling pathway), hsa168256 (immune system), and hsa04060 (cytokine-cytokine receptor interaction) (figure 4a). Similarly, ACSL1 interacts with several other genes, which are involved in hsa03320 (PPAR signaling pathway), hsa00071 (Fatty acid degradation), and hsa01212 (Fatty acid metabolism) (figure 4b). The results suggest that these genes are involved in important biological processes that can be interrupted due to its differential expression.

MicroRNAs target studies

We performed an interaction study of miRNAs that targets identified biomarker genes using miRTargetLink for Human (https://ccb-web.cs.uni-saarland.de/mirtargetlink/). miRNA-gene interaction studies help to understand the underlying role of microRNAs in the pathway. The miRTargetLink contains both experimentally known interactions from miRTarBase and predicted interactions. The microRNAs that targets MMP9, ACSL1, OLFM4, and FCGR3B are shown in figure 5, where only experimentally validated interactions (both weak and strong) were considered. Among these miRNA targets, few have been validated within the literature to be involved in ischemic stroke pathogenesis including hsa-miR-93-3p (Upregulating SOD enzymes) targeting ACSL1 [33, 34], hsa-miR-491-5p (Inhibit cellular invasion) targeting MMP9 [35], hsa-miR-29 (Induction of Fas receptors) targeting MMP9 [36], and hsa- miR-34a-5p (NPC regulation) targeting ACSL1 [37].

Validation of identified DEGs with the PubMed Literature

The identified DEGs were validated with the PubMed literature database as publication enrichment analysis using the STRING database. The important results of the publication enrichment analysis are shown in Table 3. We observed that most of the identified DEGs are markers for early or post-ischemic stroke. For instance, peripheral blood AKAP7 expression has been detected as an early marker for lymphocyte-mediated post-stroke blood-brain barrier disruption [38]. Similarly, several immune-related genes, including ARG1, are identified in post-stroke immunosuppression and ischemic stroke severity [39]. MMP9 and different isoforms of S100 (e.g. S100A12) at the protein level have been implicated as stroke predictors. Further, baseline serum MMP9 is reported to help predict the occurrence of blood-brain barrier disruption and S100 serum protein is associated with worse clinical outcomes. Thus, MMP9 and S100 can be used as prognostic markers in ischemic stroke [23]. By neutralizing the effect of MMPs, tissue inhibitors of matrix metalloproteinases (TIMPs) are responsible for maintaining tissue proteolysis in balance [40]. Thus, TIMPs define the mechanism of regulation by neuroinflammatory stimuli. Chemokine receptor 7 (CCR7) is reported to be increased in peripheral blood leukocytes in mild to moderate ischemic stroke [41]. The orosomucoid 1 (ORM1) is a glycoprotein that suppresses lymphocyte response to lipopolysaccharides, decreases platelet aggregation, and enhances cytokine secretion (refer Table 3). ACSL1 and FCGR3B,

including MMP9, reported stroke-specific differential regulation in peripheral whole blood [38]. OLFM4, a gene associated with apoptosis, is also found to be differentially expressed in stroke by Fernandez-Cadenas et al. [42].

Conclusions

A stroke is a complex neurological phenomenon which damages the brain due to interruption of blood flow, leading to neurological dysfunctions. The global cost burden of stroke is very high and estimated to be $1.52 trillion by 2050. Ischemic stroke is the most common type of stroke, which accounts for 87% of all strokes. The commonly used antiplatelet drug is Aspirin which is the least expensive treatment with low side effects. The effective treatment of critical stroke patients depends on a quick assessment of the type of stroke and general clinical status of the patients. Hence, quick and objective assessment of the type of stroke during admission would improve treatment outcome. In order to diagnose acute stroke correctly, measurement of molecular characteristics from biomarkers are used, which help develop targets for neuroprotective therapies and monitor the response to treatment. In this study, we utilized Microarray gene expression data for the *in silico* discovery of novel key diagnostic biomarkers which can be further utilized for clinical treatment of ischemic stroke.

We obtained 20 differentially expression genes using fold change statistics and performed gene-disease association which confirmed that four genes, namely MMP9, ACSL1, OLFM4, and FCGR3B, are associated with the disease term “stroke”, while few others are associated with the disease term “brain hemorrhage”, “Guillain-Barre syndrome”, and “multiple sclerosis”. In future work, we can utilize multi-omic data integration to identify key diagnostic markers.

Declarations

Ethical approval and consent to participate: The data was retrieved from publicly available resource and hence no ethical approval is required for that the work involves purely academic dissemination.

Consent for publication: NA

Acknowledgments: NA

Availablibility of data and materials: The datasets used and/or analyzed during the current study are available from the corresponding author on ressonable request.

Competing interests: The authors declare that they have no competing interests.

Funding: This study was supported by the Seed Fund Program of Shanghai University of Medicine & Health Sciences (Grant No. SFP-18-21-01-002), The Special Program for Collaborative Innovation in Shanghai University of Medicine & Health Sciences (grant No. E1-0200-19-201205), Shanghai Municipal Education Commission (Class II Plateau Disciplinary Construction Program for Medical Technology of SUMHS, 2018-2020) and the Natural Science Foundation of Guangdong Province (2016A030313680).

Author contribution statement: PZ, JW and AF conceived and designed the study. XW, LiyingJ, LianyingJ, CL and XG provided study materials or patients and was responsible for the collection and assembly of data, data analysis and interpretation. PZ, JW and AF were involved in writing the manuscript. All authors read and approved the final manuscript.

References

1. Feigin VL, Norrving B, Mensah GA. Global Burden of Stroke. Circ Res. 2017;120:439–48.
2. Dąbrowski J, Czajka A, Zielińska-Turek J, Jaroszyński J, Furtak-Niczyporuk M, Mela A, et al. Brain Functional Reserve in the Context of Neuroplasticity after Stroke. Neural Plast. 2019;2019:9708905.
3. Lloyd-Jones D, Adams R, Carnethon M, De Simone G, Ferguson TB, Flegal K, et al. Heart disease and stroke statistics–2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation. 2009;119:480–6.
4. Feigin VL, Roth GA, Naghavi M, Parmar P, Krishnamurthi R, Chugh S, et al. Global burden of stroke and risk factors in 188 countries, during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet Neurol. 2016;15:913–24.
5. Stroke : practical management / C. Warlow, J. van Gijn, M. Dennis, J. Wardlaw, J Bamford, G. Hankey, P. Sandercock, G. Rinkel, P. Langhorne, C. Sudlow, P. Rothwell - Details - Trove. https://trove.nla.gov.au/work/35606132. Accessed 29 Jul 2019.

6. Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, et al. Guidelines for the early management of adults with ischemic stroke: a guideline from the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups: The American Academy of Neurology affirms the value of this guideline as an educational tool for neurologists. Circulation. 2007;115:e478-534.
7. Moskowitz MA, Lo EH, Iadecola C. The science of stroke: mechanisms in search of treatments. Neuron. 2010;67:181–98.
8. Stoll G, Kleinschnitz C, Nieswandt B. Molecular mechanisms of thrombus formation in ischemic stroke: novel insights and targets for treatment. Blood. 2008;112:3555–62.
9. Amarenco P, Cohen A, Tzourio C, Bertrand B, Hommel M, Besson G, et al. Atherosclerotic disease of the aortic arch and the risk of ischemic stroke. N Engl J Med. 1994;331:1474–9.
10. Grysiewicz RA, Thomas K, Pandey DK. Epidemiology of ischemic and hemorrhagic stroke: incidence, prevalence, mortality, and risk factors. Neurol Clin. 2008;26:871–95, vii.
11. Albers GW, Caplan LR, Easton JD, Fayad PB, Mohr JP, Saver JL, et al. Transient Ischemic Attack — Proposal for a New Definition. N Engl J Med. 2002;347:1713–6.
12. Johnston SC. Clinical practice. Transient ischemic attack. N Engl J Med. 2002;347:1687–92.
13. Lovett JK, Dennis MS, Sandercock P a. G, Bamford J, Warlow CP, Rothwell PM. Very early risk of stroke after a first transient ischemic attack. Stroke. 2003;34:e138-140.
14. Fisher Marc, Dávalos Antoni, Rogalewski Andreas, Schneider Armin, Ringelstein E. Bernd, Schäbitz Wolf-Rüdiger. Toward a Multimodal Neuroprotective Treatment of Stroke. Stroke. 2006;37:1129–36.
15. Amarenco P, Bogousslavsky J, Callahan A, Goldstein LB, Hennerici M, Rudolph AE, et al. High-dose atorvastatin after stroke or transient ischemic attack. N Engl J Med. 2006;355:549–59.
16. Marler JR, Tilley BC, Lu M, Brott TG, Lyden PC, Grotta JC, et al. Early stroke treatment associated with better outcome: the NINDS rt-PA stroke study. Neurology. 2000;55:1649–55.
17. Ringer AJ, Qureshi AI, Fessler RD, Guterman LR, Hopkins LN. Angioplasty of intracranial occlusion resistant to thrombolysis in acute ischemic stroke. Neurosurgery. 2001;48:1282–8; discussion 1288-1290.
18. Glushakova OY, Glushakov AV, Miller ER, Valadka AB, Hayes RL. Biomarkers for acute diagnosis and management of stroke in neurointensive care units. Brain Circ. 2016;2:28–47.
19. Simats A, García-Berrocoso T, Montaner J. Neuroinflammatory biomarkers: From stroke diagnosis and prognosis to therapy. Biochim Biophys Acta. 2016;1862:411–24.
20. Maas MB, Furie KL. Molecular biomarkers in stroke diagnosis and prognosis. Biomark Med. 2009;3:363–83.
21. Bustamante A, López-Cancio E, Pich S, Penalba A, Giralt D, García-Berrocoso T, et al. Blood Biomarkers for the Early Diagnosis of Stroke: The Stroke-Chip Study. Stroke. 2017;48:2419–25.
22. The potential role of blood biomarkers in patients with ischemic stroke: An expert opinion - Mira Katan, Mitchell SV Elkind, 2018. https://journals.sagepub.com/doi/full/10.1177/2514183X18768050. Accessed 29 Jul 2019.
23. Barr TL, Conley Y, Ding J, Dillman A, Warach S, Singleton A, et al. Genomic biomarkers and cellular pathways of ischemic stroke by RNA gene expression profiling. Neurology. 2010;75:1009–14.
24. Raza K. Analysis of Microarray Data using Artificial Intelligence Based Techniques. 2016. p. 216–39.
25. Raza* K. Reconstruction, Topological and Gene Ontology Enrichment Analysis of Cancerous Gene Regulatory Network Modules. Current Bioinformatics. 2016. http://www.eurekaselect.com/138531/article. Accessed 29 Jul 2019.
26. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57.
27. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1–13.
28. Raza K. Protein features identification for machine learning-based prediction of protein-protein interactions. bioRxiv. 2017;:137257.
29. Faiza MH, Tanveer K, Fatihi S, Wang Y, Raza K. Comprehensive overview and assessment of miRNA target prediction tools in human and drosophila melanogaster. 2017.
30. Hébert SS, Horré K, Nicolaï L, Bergmans B, Papadopoulou AS, Delacourte A, et al. MicroRNA regulation of Alzheimer's Amyloid precursor protein expression. Neurobiol Dis. 2009;33:422–8.
31. Beveridge NJ, Gardiner E, Carroll AP, Tooney PA, Cairns MJ. Schizophrenia is associated with an increase in cortical microRNA biogenesis. Mol Psychiatry. 2010;15:1176–89.
32. Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. Nat Genet. 2004;36:431–2.
33. Wang P, Liang X, Lu Y, Zhao X, Liang J. MicroRNA-93 Downregulation Ameliorates Cerebral Ischemic Injury Through the Nrf2/HO-1 Defense Pathway. Neurochem Res. 2016;41:2627–35.
34. Khoshnam SE, Winlow W, Farbood Y, Moghaddam HF, Farzaneh M. Emerging Roles of microRNAs in Ischemic Stroke: As Possible Therapeutic Agents. J Stroke. 2017;19:166–87.

35. Yan W, Zhang W, Sun L, Liu Y, You G, Wang Y, et al. Identification of MMP-9 specific microRNA expression profile as potential targets of anti-invasion therapy in glioblastoma multiforme. Brain Res. 2011;1411:108–15.
36. Schickel R, Park S-M, Murmann AE, Peter ME. miR-200c regulates induction of apoptosis through CD95 by targeting FAP-1. Mol Cell. 2010;38:908–15.
37. Liu FJ, Lim KY, Kaur P, Sepramaniam S, Armugam A, Wong PTH, et al. microRNAs Involved in Regulating Spontaneous Recovery in Embolic Stroke Model. PloS One. 2013;8:e66393.
38. O'Connell GC, Treadway MB, Petrone AB, Tennant CS, Lucke-Wold N, Chantler PD, et al. Peripheral blood AKAP7 expression as an early marker for lymphocyte-mediated post-stroke blood brain barrier disruption. Sci Rep. 2017;7:1172.
39. Petrone AB, O'Connell GC, Regier MD, Chantler PD, Simpkins JW, Barr TL. The Role of Arginase 1 in Post-Stroke Immunosuppression and Ischemic Stroke Severity. Transl Stroke Res. 2016;7:103–10.
40. Welser-Alves JV, Crocker SJ, Milner R. A dual role for microglia in promoting tissue inhibitor of metalloproteinase (TIMP) expression in glial cells in response to neuroinflammatory stimuli. J Neuroinflammation. 2011;8:61.
41. Yan J, Greer JM, Etherington K, Cadigan GP, Cavanagh H, Henderson RD, et al. Immune activation in the peripheral blood of patients with acute ischemic stroke. J Neuroimmunol. 2009;206:112–7.
42. Fernandez-Cadenas I, Del Rio-Espinola A, Domingues-Montanari S, Mendioroz M, Fernandez-Morales J, Penalba A, et al. Genes involved in hemorrhagic transformations that follow recombinant t-PA treatment in stroke patients. Pharmacogenomics. 2013;14:495–504.

Tables

Table 1. List of differentially expressed genes and their scores in brain stroke patients

ID	Adj. P. Val	P. Value	logFC	FC	Gene Symbol	Gene Title	GenBank Accession
ILMN_1812281	0.0000	0.0000	-1.6940	-3.24	ARG1	arginase 1	NM_000045
ILMN_1796316	0.0000	0.0000	-1.4304	-2.70	MMP9	matrix metallopeptidase 9	NM_004994
ILMN_1748915	0.0002	0.0000	-1.2761	-2.42	S100A12	S100 calcium binding protein A12	NM_005621
ILMN_1696584	0.0022	0.0001	-1.1837	-2.27	ORM1	orosomucoid 1	NM_000607
ILMN_2134453	0.0074	0.0006	-1.1742	-2.26	FCGR3B	Fc fragment of IgG receptor IIIb	NM_000570
ILMN_1680192	0.0097	0.0009	-1.1644	-2.24	APOBEC3A	apolipoprotein B mRNA editing enzyme catalytic subunit 3A	NM_145699
ILMN_1703049	0.0005	0.0000	-1.1421	-2.21	AKAP7	A-kinase anchoring protein 7	NM_016377
ILMN_1724533	0.0006	0.0000	-1.1193	-2.17	LY96	lymphocyte antigen 96	NM_015364
ILMN_1695157	0.0001	0.0000	-1.0993	-2.14	CA4	carbonic anhydrase 4	NM_000717
ILMN_1730454	0.0005	0.0000	-1.0948	-2.14	FOLR3	folate receptor 3	NM_000804
ILMN_1684585	0.0003	0.0000	-1.0946	-2.14	ACSL1	acyl-CoA synthetase long-chain family member 1	NM_001995
ILMN_1715131	0.0000	0.0000	1.0838	2.12	CCR7	C-C motif chemokine receptor 7	NM_001838
ILMN_1687301	0.0000	0.0000	-1.0757	-2.11	VCAN	Versican	NM_004385
ILMN_2173835	0.0354	0.0046	-1.0737	-2.10	FTH1P3	ferritin heavy chain 1 pseudogene 3	NR_002201
ILMN_1790689	0.0000	0.0000	-1.0689	-2.10	CRISPLD2	cysteine rich secretory protein LCCL domain containing 2	NM_031476
ILMN_1803819	0.0000	0.0000	-1.0330	-2.05	IQGAP1	IQ motif containing GTPase activating protein 1	NM_003870
ILMN_1807529	0.0000	0.0000	-1.0232	-2.03	PADI4	peptidyl arginine deiminase 4	NM_012387
ILMN_2197365	0.0000	0.0000	-1.0106	-2.01	RGS2	regulator of G-protein signaling 2	NM_002923
ILMN_1684982	0.0000	0.0000	-1.0078	-2.01	PDK4	pyruvate dehydrogenase kinase 4	NM_002612
ILMN_2116877	0.0123	0.0012	-1.0004	-2.00	OLFM4	olfactomedin 4	NM_006418

Table 2. Results of (A) disease association study using Genetic Association Database (GAD), (B) Functional categories enrichment analysis, (C) GO term enrichment analysis, and (D) Tissue expression enrichment analysis using DAVID tool (P-value < 0.05)

(A) Gene Association Study

Category	Term	Count	%	P-Value	Genes	Fold Enrichment	Bonferroni	Benjamini
GAD_DISEASE	Stroke	4	20	0.0318	NM_004994, NM_001995, NM_006418, NM_000570	5.3181	0.9999	0.8083
GAD_DISEASE	brain hemorrhage	2	10	0.0284	NM_004994, NM_004385	65.5101	0.9998	0.9474
GAD_DISEASE	Guillain-Barre syndrome	2	10	0.0233	NM_004994, NM_000570	80.0679	0.9992	0.9992
GAD_DISEASE	Multiple Sclerosis	4	20	0.0380	NM_004994, NM_012387, NM_000570, NM_002923	4.9611	0.9999	0.8165
GAD_DISEASE	rheumatoid arthritis	3	15	0.0251	NM_004994, NM_012387, NM_000570	11.3780	0.9995	0.9796
GAD_DISEASE	Carcinoma, Basal Cell|Carcinoma, Squamous Cell|Melanoma|Skin Neoplasms	2	10	0.0310	NM_004994, NM_012387	60.0509	0.9999	0.8545
GAD_DISEASE	Coronary Disease	2	10	0.0310	NM_004994, NM_000045	60.0509	0.9999	0.8545
GAD_DISEASE	plasma HDL cholesterol (HDL-C) levels	3	15	0.0413	NM_004994, NM_002612, NM_001995	8.6820	0.9999	0.8007
GAD_DISEASE	abdominal aortic aneurysm	2	10	0.0449	NM_004994, NM_004385	41.1777	0.9999	0.7904

(B) Functional categories enrichment analysis

Category	Term	Count	%	P-Value	Genes	Fold Enrichment	Bonferroni	Benjamini
UP_KEYWORDS	Secreted	9	45	0.0001	NM_004994, NM_000804, NM_004385, NM_005621, NM_031476, NM_000607, NM_006418, NM_000570, NM_015364	4.9612	0.0108	0.0108
UP_KEYWORDS	Signal	11	55	0.0010	NM_004994, NM_000804, NM_004385, NM_001838, NM_031476, NM_000607, NM_006418, NM_000570, NM_000717, NM_015364, NM_003870	2.8642	0.0833	0.0425
UP_KEYWORDS	Disulfide bond	10	50	0.0011	NM_004994, NM_000804, NM_004385, NM_001838, NM_031476, NM_000607, NM_006418, NM_000570, NM_000717, NM_015364	3.1543	0.0950	0.0327
UP_KEYWORDS	Innate immunity	4	20	0.0014	NM_012387, NM_145699, NM_005621, NM_015364	16.6009	0.1182	0.0309
UP_KEYWORDS	Glycoprotein	11	55	0.0020	NM_004994, NM_000804, NM_004385, NM_001838, NM_031476, NM_000607, NM_001995, NM_006418, NM_000570, NM_000717, NM_015364	2.6181	0.1655	0.0355
UP_KEYWORDS	Immunity	4	20	0.0088	NM_012387, NM_145699, NM_005621, NM_015364	8.6656	0.5431	0.1224
UP_KEYWORDS	Lipoprotein	4	20	0.0362	NM_145699, NM_016377, NM_000570, NM_000717	5.0854	0.9612	0.3714
UP_KEYWORDS	Calcium	4	20	0.0390	NM_004994, NM_012387, NM_004385, NM_005621	4.9405	0.9698	0.3545
UP_KEYWORDS	Cell membrane	7	35	0.0471	NM_005621, NM_001838,	2.3881	0.9857	0.3765

					NM_016377, NM_000570, NM_000717, NM_002923, NM_003870			

(C) GO term enrichment analysis

Category	Term	Count	%	P-Value	Genes	Fold Enrichment	Bonferroni	Benjamini
GOTERM_CC_DIRECT	GO:0005615~extracellular space	6	30	0.0083	NM_004994, NM_004385, NM_000607, NM_000045, NM_006418, NM_015364	4.2724	0.4427	0.4427
GOTERM_CC_DIRECT	GO:0005886~plasma membrane	10	50	0.0099	NM_005621, NM_001838, NM_016377, NM_001995, NM_006418, NM_000570, NM_000717, NM_002923, NM_015364, NM_003870	2.3274	0.5022	0.2944
GOTERM_CC_DIRECT	GO:0070062~extracellular exosome	8	40	0.0136	NM_004994, NM_031476, NM_000607, NM_000045, NM_006418, NM_000570, NM_000717, NM_003870	2.7297	0.6188	0.2749
GOTERM_CC_DIRECT	GO:0005576~extracellular region	6	30	0.0172	NM_004994, NM_000804, NM_004385, NM_005621, NM_031476, NM_000607	3.5745	0.7045	0.2627
GOTERM_CC_DIRECT	GO:0043005~neuron projection	3	15	0.0224	NM_000045, NM_002923, NM_003870	12.1412	0.7962	0.2724
GOTERM_CC_DIRECT	GO:0005578~proteinaceous extracellular matrix	3	15	0.0282	NM_004994, NM_004385, NM_031476	10.7368	0.8652	0.2840

(D) Tissue expression enrichment analysis

Category	Term	Count	%	P-Value	Genes	Fold Enrichment	Bonferroni	Benjamini
GNF_U133A_QUARTILE	Whole Brain_3rd	18	90	0.0000	NM_000804, NM_145699, NM_005621, NM_000607, NM_000045, NM_006418, NM_000570, NM_000717, NM_002923, NM_015364, NM_004994, NM_004385, NM_012387, NM_001838, NM_031476, NM_001995, NM_003870, NR_002201	3.0952	0.0000	0.0000
GNF_U133A_QUARTILE	bronchialepithelialcells_3rd	13	65	0.0000	NM_000804, NM_145699, NM_005621, NM_000607, NM_000045, NM_006418, NM_000570, NM_015364, NM_002923, NM_004994, NM_012387, NM_001838, NM_001995	5.1254	0.0000	0.0000
GNF_U133A_QUARTILE	globuspallidus_3rd	14	70	0.0000	NM_002612, NM_005621, NM_000607, NM_000045, NM_006418, NM_000570, NM_000717, NM_002923, NM_004994, NM_004385, NM_031476, NM_001995, NR_002201, NM_003870	4.1566	0.0001	0.0000
GNF_U133A_QUARTILE	skin_3rd	7	35	0.0171	NM_004994, NM_001838, NM_016377, NM_001995, NM_006418, NM_015364, NM_003870	3.0229	0.7108	0.2667

Table 3. Validation of identified DEGs with the PubMed literature using publication enrichment analysis

PMID	Description	Gene Count(Observed / Background)	FDR	Matching Genes in the Query
28446746	Peripheral blood AKAP7expression as an early marker for lymphocyte-mediated post-stroke blood brain barrier disruption (2017)	15 / 27	0.00	ACSL1,AKAP7,APOBEC3A,CA4,CCR7,CRISPLD2,FCGR3B,FOLR3,IQGAP1,LY96,MMP9,ORM1,PADI4,S
27672514	Integrated analysis of ischemic stroke datasets revealed sex and age difference in anti-stroke targets (2016)	8 / 30	0.00	ARG1,CA4,CCR7,MMP9,ORM1,PDK4,RGS2,TLR4
29637127	Peripheral blood RNA gene expression in children with pneumococcal meningitis: a prospective case-control study (2017)	8 / 40	0.00	ACSL1,AGER,ARG1,CA4,FOLR3,MMP9,S100A12,TLR4
25972779	Rational modulation of the innate immune system for neuroprotection in ischemic stroke (2015)	7 / 52	0.00	AGER,ARG1,CA4,LY96,MMP9,S100A12,TLR4
26515089	The Role of Arginase 1 (ARG1) in Post-Stroke Immunosuppression and Ischemic Stroke Severity (2016)	5 / 7	0.00	ARG1,CCR7,LY96,MMP9,S100A12
23533687	Triggers and effectors of oxidative stress at blood-brain barrier level: relevance for brain ageing and neurodegeneration(2013)	5 / 41	0.00	AGER,CCL19,CDC42,MMP9,TIMP2
28222567	Immune biomarkers for the diagnosis of mild traumatic brain injury(2017)	4 / 4	0.00	CCR7,LY96,MMP9,S100A12
21631912	A dual role for microglia in promoting tissue inhibitor of metalloproteinase (TIMP) expression in glial cells in response to neuroinflammatory stimuli (2011)	4 / 9	0.00	MMP9,TIMP1,TIMP2,TLR4
26074872	Unbalanced Metalloproteinase-9 and Tissue Inhibitors of Metalloproteinases Ratios Predict Hemorrhagic Transformation of Lesion in Ischemic Stroke Patients Treated with Thrombolysis: Results from the MAGIC Study (2015)	4 / 9	0.00	ACSL1,MMP9,TIMP1,TIMP2

Figures

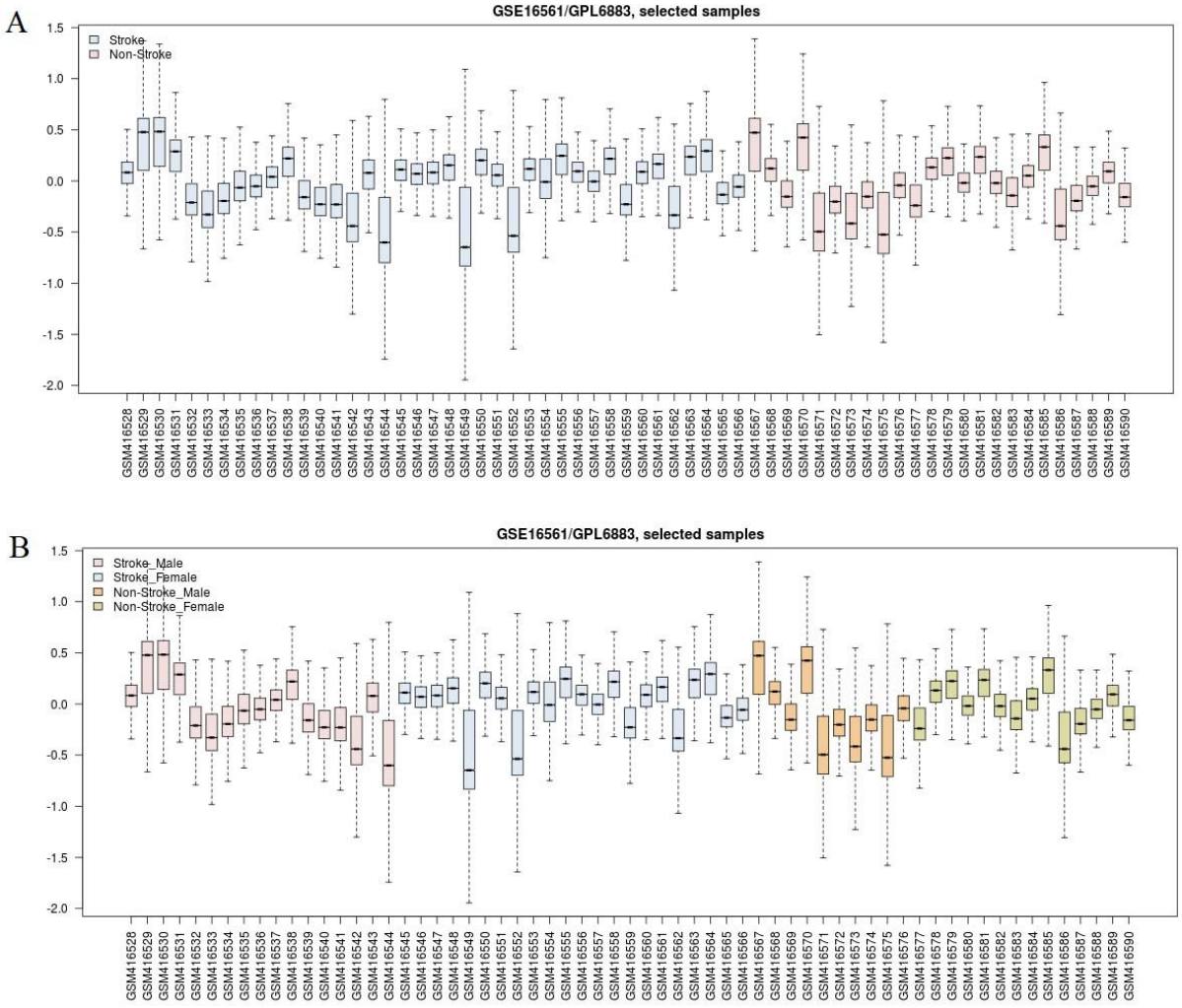


Figure 1

A: Boxplot of the gene expression profile of all 63 samples (stroke =39, control=24). B: Boxplot of the gene expression profile of stroke and control in both males and females.

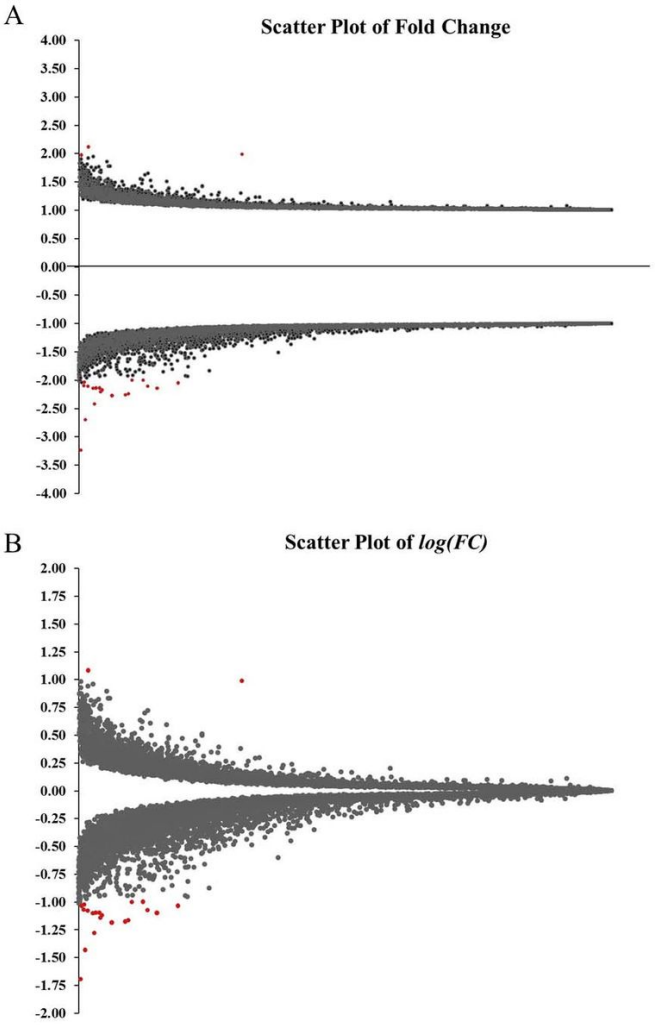


Figure 2

A: Scatter plot of Fold Change (FC) between stroke and control. B: Scatter plot of log(FC) between stroke and control.

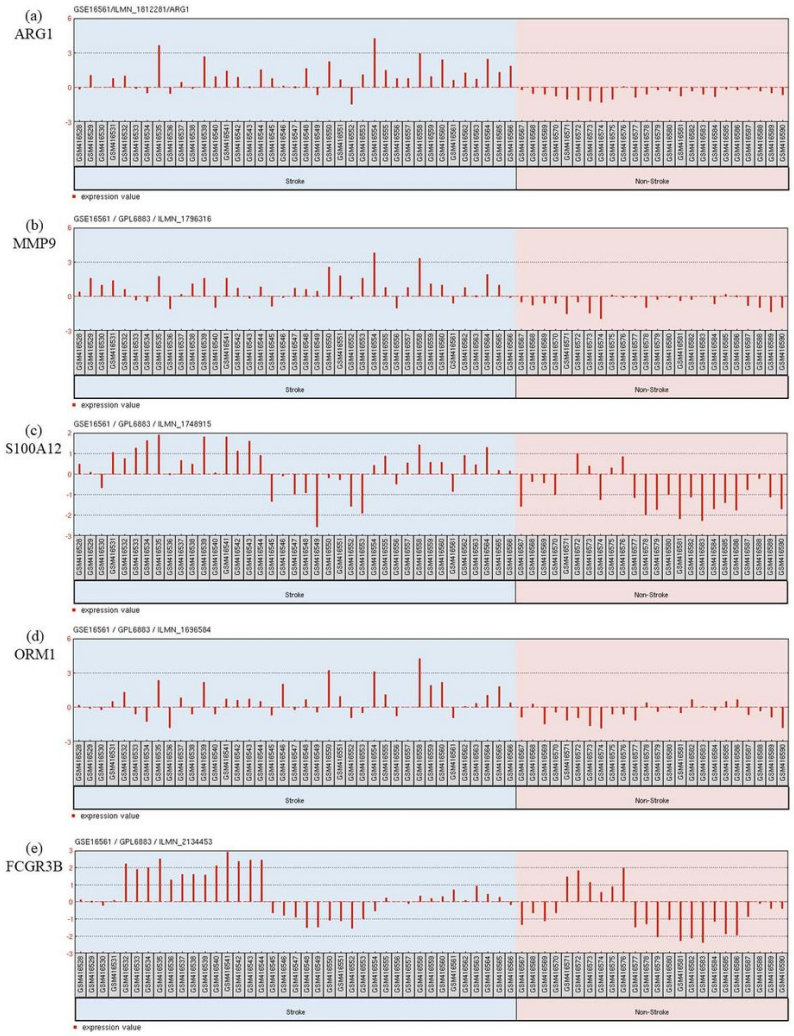
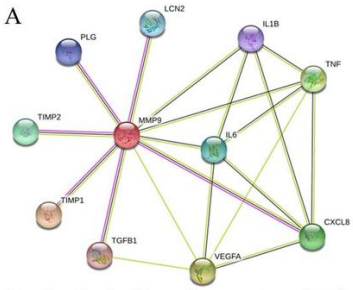


Figure 3

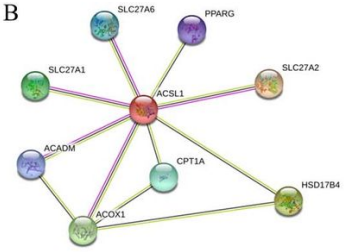
Profile graph of the top five highly DEGs – ARG1, MMP9, S100A12, ORM1, and FCGR3B.



Interactions legends: Magenta line represents experimentally determined, blue line co-expression and yellow line text-mining.

KEGG Pathways				
Pathway	*Description*	*Count in gene set*	*FDR*	*Proteins in the network*
hsa04657	IL-17 signaling pathway	6 out of 92	0.00	CXCL8,IL1B,IL6,LCN2, MMP9,TNF
hsa04933	AGE-RAGE signaling pathway in diabetic complications	6 out of 98	0.00	CXCL8,IL1B,IL6,TGFB 1,TNF,VEGFA
hsa05323	Rheumatoid arthritis	6 out of 84	0.00	CXCL8,IL1B,IL6,TGFB 1,TNF,VEGFA
hsa04060	Cytokine-cytokine receptor interaction	6 out of 263	0.00	CXCL8,IL1B,IL6,TGFB 1,TNF,VEGFA

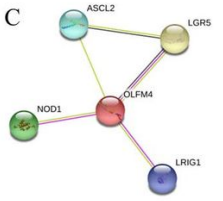
Reactome Pathways				
Pathway	*Description*	*Count in gene set*	*FDR*	*Proteins in the network*
hsa168256	Immune System	10 out of 1925	0.00	CXCL8, IL1B, IL6, LCN2, MMP9, TGFB1, TIMP1, TIMP2, TNF, VEGFA
hsa6785807	Interleukin-4 and Interleukin-13 signaling	9 out of 106	0.00	CXCL8, IL1B, IL6, LCN2, MMP9, TGFB1, TIMP1, TNF, VEGFA
hsa162582	Signal Transduction	7 out of 2605	0.00	CXCL8, IL6, MMP9, PLG, TGFB1, TNF, VEGFA



Interactions legends: Magenta line represents experimentally determined, blue line co-expression and yellow line text-mining

KEGG Pathways				
Pathway	*Description*	*Count in gene set*	*FDR*	*Proteins in the network*
hsa03320	PPAR signaling pathway	8 out of 72	0.00	ACADM,ACOX1,ACSL1, CPT1A,PPARG,SLC27A1, SLC27A2,SLC27A6
hsa00071	Fatty acid degradation	4 out of 44	0.00	ACADM,ACOX1,ACSL1, CPT1A
hsa01212	Fatty acid metabolism	4 out of 48	0.00	ACADM,ACOX1,ACSL1, CPT1A
hsa04146	Peroxisome	4 out of 81	0.00	ACOX1,ACSL1,HSD17B4, SLC27A2
hsa04931	Insulin resistance	4 out of 107	0.00	CPT1A,SLC27A1,SLC27A 2,SLC27A6

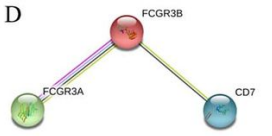
Reactome Pathways				
Pathway	*Description*	*Count in gene set*	*FDR*	*Proteins in the network*
HSA-1989781	PPARA activates gene expression	5 out of 114	0.00	ACADM,ACSL1,CPT1A,P PARG,SLC27A1
HSA-556833	Metabolism of lipids	7 out of 721	0.00	ACADM,ACSL1,CPT1A, HSD17B4,PPARG,SLC27 A1,SLC27A2
HSA-8978868	Fatty acid metabolism	5 out of 171	0.00	ACADM,ACSL1,CPT1A, HSD17B4,SLC27A2



KEGG Pathways
Not available with sufficient gene count

Reactome Pathways
Not available with sufficient gene count

Interactions legends: Magenta line represents experimentally determined, blue line co-expression and yellow line text-mining



KEGG Pathways
Not available with sufficient gene count

Reactome Pathways
Not available with sufficient gene count

Interactions legends: Magenta line represents experimentally determined, blue line co-expression and yellow line text-mining

Figure 4

A: Protein-protein interaction of MMP9, along with their KEGG and Reactome pathways descriptions. B: Protein-protein interaction of ACSL1, along with their KEGG and Reactome pathways descriptions. C: Protein-protein interaction of OLFM4, along with their KEGG and Reactome pathways descriptions. D: Protein-protein interaction of FCGR3B, along with their KEGG and Reactome pathways descriptions.

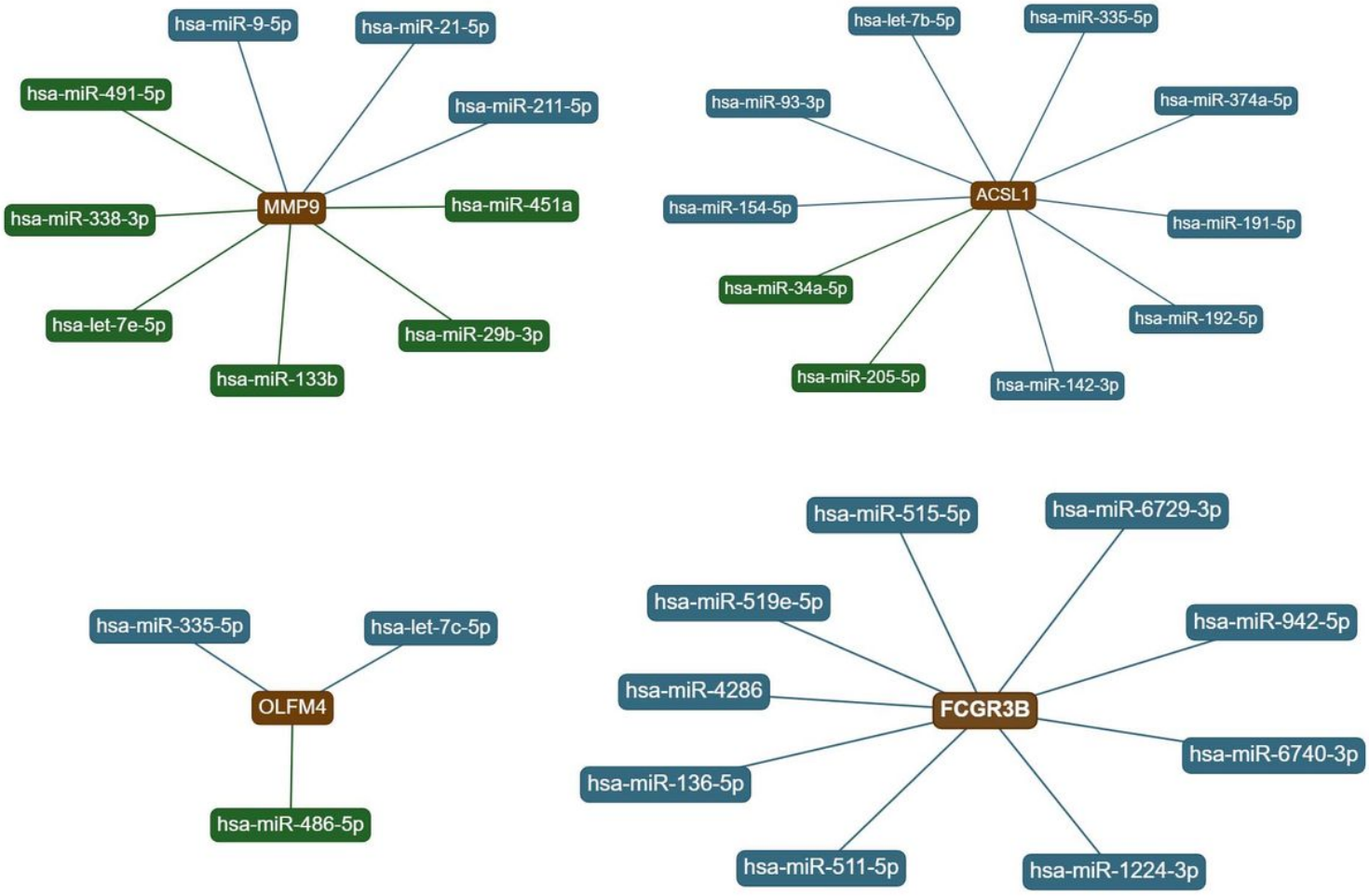


Figure 5

MicroRNAs that target MMP9, ACSL1, OLFM4, and FCGR3B. Green represents strong interactions, while blu represents weak interactions