Identification and analysis of DNA methylation inflammation related key genes in intracerebral hemorrhage

Sanpeng Xu
Changchun University of Chinese Medicine
Qiong Wu
Xin Yang Central Hospital
Ping Li (lp452664199@163.com)
Changchun University of Chinese Medicine

Research Article

Keywords: Biomarkers, DNA methylation, GSEA, Intracerebral hemorrhage, Inflammation-related genes

Posted Date: February 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2568701/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Biochemical Genetics on June 24th, 2023.
See the published version at https://doi.org/10.1007/s10528-023-10430-9.

Abstract

Background: Inflammation and DNA methylation have been reported to play key roles in intracerebral hemorrhage (ICH). The proposed study intended to investigate new diagnostic biomarkers associated with inflammation and DNA methylation through comprehensive bioinformatics approaches.

Methods: GSE179759 and GSE125512 were sourced via the Gene Expression Omnibus (GEO) database, and 3222 inflammation-related genes (IFRGs) were downloaded from the Molecular Signatures Database (MSigDB). Key differentially expressed methylation-regulated and inflammation-related genes (DE-MIRGs) were achieved by overlapping methylation-regulated differentially expressed genes (MeDEGs) between ICH patients and control samples, module genes from Weighted Correlation Network Analysis (WGCNA), and the IFRGs. The functional annotation of DE-MIRGswas performed by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) resources. A protein-protein interaction (PPI) network was further constructed to clarify the interrelationships between the different DE-MIRGs. The key genes were categorized by Least Absolute Shrinkage Selection Operator (LASSO), and support vector machine recursive feature elimination (SVM-RFE), and subsequently performed Gene Set Enrichment Analysis (GSEA).

Results: A number of 22 DE-MIRGs were acquired among 451 MeDEGs, 3222 IFRGs and 302 module genes, and they were mainly enriched in GO terms of wound healing, blood coagulation and hemostasis; KEGG pathways of PI3K-AKT signaling pathway, Focal adhesion, and Regulation of actin cytoskeleton. A PPI network with 22 nodes and 87 edges was constructed based on the 22 DE-MIRGs, and 11 of them were selected for the following key gene selection. Moreover, 2 key genes (SELP and S100A4) were obtained according to LASSO and SVM-RFE. Finally, SELP was mainly enriched in Cell morphogenesis involved in differentiation, Cytoplasm translation, and Actin binding of GO terms, and the KEGG pathway including Edocytosis, Focal adhesion, and Platelet activation. S100A4 was major enriched in GO terms including Mitochondrial inner membrane, Mitochondrial respirasome, and Lysosomal membrane; Oxidative phosphorylation, Regulation of actin cytoskeleton, and Chemical carcinogensis-reactive oxygen species in KEGG pathways.

Conclusion: 22 DE-MIRGs were identified associated with inflammation and DNA methylation between ICH patients and normal controls, and 2 key genes (SELP and S100A4) were obtained and regarded as the biomarker for ICH, which could provide the research foundation for the further pathological mechanism investigation of ICH.

Highlight

In this study, we observed the differential expression of methylation regulation and inflammation related genes of key genes in patients with intracerebral hemorrhage, indicating that when intracerebral hemorrhage occurs, the methylation modified gene pattern can activate multiple inflammation regulated related pathways, and abnormal methylation and dysregulation of inflammatory pathways jointly

participate in the occurrence of intracerebral hemorrhage.The purpose of this study is to study the changes of DNA methylation patterns in patients with cerebral hemorrhage and healthy people with the involvement of inflammation related genes by integrating genes of patients with cerebral hemorrhage, control samples and inflammation related pathways, so as to obtain key gene differences and activated inflammation pathways. The purpose of this study is to obtain the key genes of inflammation and DNA methylation modification as potential biomarkers, and provide basic research for the diagnosis and treatment of cerebral hemorrhage in genetic genetics.

1. Introduction

Intracerebral hemorrhage (ICH) is one of the common diseases of stroke, and its incidence rate is considered to be the second cerebrovascular disease subtype of stroke. It mainly refers to the hemorrhage caused by non-traumatic rupture of blood vessels in brain tissue, including primary ICH and secondary ICH[1].Relevant research reports show that the incidence rate of cerebral hemorrhage in developing countries is much higher than that in developed countries in Europe and America[2].At present, stroke disease in China is at a high incidence segment. According to the epidemiological survey, the mortality rate caused by cerebrovascular diseases in China in 2018 was 149.49/100000, accounting for about 22% of the national resident mortality rate.The incidence rate of patients with ICH was 14.9%, and the proportion of patients with ICH who died in hospital / left hospital without medical advice was 19.5% [3] Additionally, Nearly half of patients with ICH died within 30 days after onset, and only 12% ~ 39% of patients could live independently for a long time[4–5]. With the continuous in-depth study of the pathogenesis of ICH, it has been identified that cerebral atherosclerosis and arteriolaregeneration caused by hypertension are the main risk factors of ICH. Therefore, effective control of hypertension in the early stage can significantly reduce the occurrence of ICH[6]. The injury of ICH to the body mainly includes primary injury and secondary injury. The incidence of primary cerebral hemorrhage accounts for about 10% – 20% of all strokes[7]. It mainly refers to the space occupying effect of blood clot, hematoma increasing intracranial pressure, compressing brain tissue, causing cerebral ischemia and even cerebral hernia[8]. Secondary injury is mainly caused by hematoma pressing the brain tissue, resulting in the body releasing a large amount of thrombin, inflammatory reaction, complement reaction, released components of blood clot, free radicals and other waterfall reactions[9]. Head computed tomography (CTH) can clearly diagnose the volume of hemorrhage and the location of nerve injury, but it fails to improve the clinical symptoms after ICH. Therefore, how to accurately prevent ICH and more effectively improve the repair of injured nerves are particularly important. It has been reported that genetic factors account for 37.9% of the stroke pathogenesis. Other studies have shown that genes in the brain may change in expression after ICH[10–11]. However, the mechanism by which ICH causes these changes remains unknown[12].

The formation of inflammation is a complex process, which is mediated by cellular and molecular components. The cellular components mainly include leukocytes, macrophages, astrocytes, T cells and microglia, while the molecular components include prostaglandins, chemokines, cytokines, extracellular proteases and reactive oxygen species.After the occurrence of intracerebral hemorrhage, the hematoma

forms, and then the secondary damage after the hemorrhage produces a large number of cytokines and plasma proteins. When these substances are released into the brain tissue, the human complement is activated, the body's stress function is generated, hemostasis and immune system are activated, causing inflammation. Inflammation runs through the whole process of brain tissue damage and repair after intracerebral hemorrhage[13].After intracerebral hemorrhage, the inflammatory reaction can be regarded as a "double-edged sword". On the one hand, it can clear the necrotic brain cells and their metabolic components, and provide a good self-healing environment for the repair of brain tissue. On the other hand, a large number of inflammatory factors are produced to induce immune response and further aggravate the damage of brain tissue. Therefore, the inflammatory response has become a research hotspot of the pathological mechanism of secondary injury after intracerebral hemorrhage[14–15].

DNA methylation is an important epigenetic modification, the selective hypermethylation or demethylation of genes to regulate gene expression. DNA hypermethylation can directly inhibit transcription or indirectly inhibit gene expression through transcriptional silencing[16]. DNA methylation changes are associated with cardiovascular and cerebrovascular diseases[17]. RNA sequencing (RNA SEQ) is a powerful transcriptome analysis method that can provide global unbiased transcriptomic analysis with high sensitivity and specificity. Compared with other methods to characterize gene expression, it can identify more overall differentially expressed genes (DEGs)[18]. Weighted gene coexpression network analysis (WGCNA) is a powerful method for identifying coexpressed genomes from large heterogeneous messenger RNA expression datasets[19]. It is widely used to elucidate some transcriptomic alterations[20]. Therefore, this study integrates the whole gene expression profile and DNA methylation data for analysis, and obtains the key genes regulated by DNA methylation through differential analysis, WGCNA analysis, machine learning and functional enrichment analysis, providing a theoretical basis for exploring potential biomarkers of cerebral hemorrhage.

2. Materials And Methods

2.1 Data Source

GSE179759, including the DNA methylation profiles of peripheral blood samples from 30 ICH patients and 34 normal controls, were acquired from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo). GSE125512 containing genome-wide expression profiles of 11 pairs of peripheral blood samples, which were draw within 24 hours of symptom onset and 72 hours after the initial collection from 11 ICH patients, were obtained from the GEO as well. Additionally, 3222 inflammation-related genes (IFRGs) were downloaded from the Molecular Signatures Database (MSigDB) (https://www.gsea-msigdb.org/).

2.2 Screening of DMPs and DEGs

The raw data of GSE179759 were imported into ChAMP (version 2.24.0)[21] to conduct quality control and standardization. Moreover, clustering, multidimensional scaling (MDS), and methylation distribution

density analyses were employed based on the beta values on the methylation probes.

In the next phase, the site-level analysis was utilized based on the ChAMP to screen the differentially methylation probes (DMPs) between ICH and matched control samples, with adj. $p \leq 0.05$ as thresholds, and the results were visualized into a volcano map by ggplot2 (version 3.3.5) and a heat map by pheatmap (version 1.0.12). In addition, the screened DMPs were separated into hypomethylation (Hypo-DMPs) and hypermethylation (Hyper-DMPs) according to whether the delta beta was greater than 0. RIdeogram (version 0.2.2) [22]was used to present the most significant Hypo- and Hyper-DMPs (|delta beta|>0.1), and UpSetR (version 1.4.0) was employed to visualize the number of genes located in different functional regions in DMP.

In terms of the gene expression dataset (GSE125512), the raw data were introduced into DESeq2 (version 1.32.0) to obtain differentially expressed genes (DEGs) between cases and controls (control: blood first draw from the 11 ICH patients; case: blood draw from the 11 ICH patients 72 hours after first collection) with the criteria of $|log2FC| > 0.5$ and $adj.p < 0.05$.

2.3 Overlap and Functional Enrichment Analyses

VennDiagram package was used to investigate overlapping genes between DMPs and DEGs, which were considered as differentially expressed methylation-regulated genes (DEMGs). In specific, hyper-methylation-low expression genes were acquired based on the intersection of Hyper-DMPs and downregulated DEGs. Similarly, hypo-methylation-high expression genes were the superimposition of hypomethylated genes and upregulated genes.

Functional enrichment of DEMGs was evaluated by Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) through clusterProfiler (version 4.2.2) based on the gene set in org.Hs.eg.db package, with p adj.< 0.05 and count > 2 as cutoff values, and the results were visualized by enrichplot (version 4.2.2).

2.4 Identification and Functional Enrichment Analyses of Key DE-MIRGs

Firstly, a sample clustering tree map of all the genes in GSE125512 was constructed to detect and eliminate outliers. In order to identify the module genes highly associated with ICH, genes whose expression level were lower than 1 were removed from GSE125512, and the remained genes were selected and grouped to WGCNA (version 1.7-3)[23] based on ICH or normal as the trait. In network construction processes, soft thresholding power β was selected as the lowest power. The tree was cut into different modules by the dynamic cutting method. The minimum module size was set to 100, and modules with similar gene expressions were clustered and displayed in a tree diagram. To identify modules associated ICH, a heat map of module-feature relationships with correlation coefficients and p-values was drawn. Modules with a strong correlation with the phenotypes were identified as modules of

interest. Genes in the selected module were considered as module genes. Furthermore, the overlap analysis was applied to DEMGs, module genes, and 3222 IFRGs, and the intersected genes were regarded as key differentially expressed methylation-regulated and inflammation-related genes (DE-MIRGs).

The clusterProfiler (version 4.2.2)[24] was used on key DE-MIRGs to perform functional enrichment based on the gene set in org.Hs.eg.db package again, with p adj.< 0.05 and count > 2 as significant thresholds.

2.5 Construction of a Protein-Protein Interaction (PPI) Network

STRING (https://string-db.org/)[25] was employed to explore interactions between the DE-MIRGs and plot the interactions was a PPI network with the setting of Confidence = 0.15, and the interaction network was visualize by Cytoscape (version 3.8.2). Then Cytoscape and MCODE subnetwork discovery algorithm were applied to cluster the protein interaction network graphs and predict the subclusters.

2.6 Identification of Key Genes

GSE125512 was used as the training set to build the Lasso regression model. In order to reduce the feature dimension, glmnet package (version4.0-2)[26]in R was performed with the set of parameter famil = binomial and type.measure = Class, to achieve Lasso logistic regression. Further, 10 fold cross validation was finally performed to calculate the error rate under different genes. The corresponding genes were selected according to the minimum ramada (λmin) and defined as Lasso-feature genes.

Simultaneously, in the SVM-RFE model of R package e1071 (version 1.7-9)[27] , the key DE-MIRGs in cluster 1 were ranked by the SVM algorithm, and the importance and ranking of each gene were obtained using the RFE method, and the error rate and accuracy rate of each combination of iterations were also obtained. The lowest error rate and the highest accuracy were selected as the optimal combinations, and the corresponding genes were taken as candidate feature genes. After that, cross-tabulation analysis was used to extract the common genes in the Lasso-candidate feature and SVM-RFE-candidate feature genes, which were deemed as key genes.

2.7 Gene Set Enrichment Analysis (GSEA) on Key Genes

The correlations between key genes and other genes in GSE142153 were computed based on the default background gene set in “org.Hs.eg.db” package, and the correlation sequencing list L of key genes and other genes was obtained. GSEA was performed on the key genes with the thresholds of |NES| > 1, p < 0.05, q < 0.2.

3. Results

3.1 Identification of DMPs and DEGs

At the beginning, the DNA methylation levels of the data in GSE179759 were plotted, and it can be found that the levels of most data were around 0.9 (**Figure 1A**). After quality control and standardization,

746333 out of 844668 probes were filtered out, and the beta values of the 746333 samples were all around 0.1 or 0.9, and there were no abnormal samples, indicating that the methylation degrees of the samples were similar (**Figure 1B, 1C**).

Moreover, 90089 Hypo-DMPs and 116900 Hyper-DMPs were identified between 30 ICH patients and 34 controls, as shown by the volcano plot (**Figure 1D**).

The top 1000 Hypo-DMPs and Hyper-DMPs CpG positions are shown by the heatmap in **Figure 1E.** Furthermore, the locations of the most significant Hypo- and Hyper-DMPs were shown in **Figure 1F**, and the majority of both Hypo-DMPs Hyper-DMPs located in the Body region (**Figure 1G, 1H**). Additionally, 456 upregulated DEGs and 183 downregulated DEGs were identified between 11 pairs of samples in GSE1255125 (**Figure 1I, 1J**).

3.2 Identification and functional Enrichment analyses of MeDEGs

Venn diagrams showed a total of 451 MeDEGs, containing 340 hyper-methylation low-expression gene and 111 hypo-methylation high-expression genes, were identified (**Figure 2A, 2B).**

The GO enrichment results revealed that MeDEGs were significantly enriched in 134 GO terms, including 101 in BP, 27 in CC, and 6 in MF. For instance, wound healing, coagulation, blood coagulation, hemostasis, etc. were the significant enriched terms in BP category. In view of the CC, collagen-containing extracellular matrix, secretory granule lumen, cytoplasmic vesicle lumen, and vesicle lumen were the major enriched terms. In terms of the MF, heparin binding, lipopolysaccharide binding, actin binding, cytokine binding, growth factor binding, platelet-derived growth factor receptor binding were the 6 enriched terms (Figure 2C). Moreover, MeDEGs were mainly enriched in 3 KEGG pathways including PI3K-Akt signaling pathway, Regulation of actin cytoskeleton, and Malaria (Figure 2D).

3.3 Identification and Functional Enrichment Analyses of Key DE-MIRGs

A number of 12130 genes in GSE122512 were screened out based on their expression levels. After clustering the genes in the data set, no outliers were detected, and subsequently, a sample clustering tree was drawn (**Figure 3A**). The soft threshold was set to 12 (R^2 = 0.8) to construct a scale-free network (**Figure 3B**). In addition, 17 modules were identified based on the setting of minimum module size = 150 (**Figure 3C**). The cluster of genes in the modules were shown in the dendrogram (**Figure 3D**). The lightyellow module was the most relevant module to ICH (cor = 0.52 and P = 0.01; **Figure 3E**), and therefore the 302 genes in this module were identified as module genes. Further, 22 intersection genes were acquired among 451 MeDEGs, 3222 IFRGs and 302 module genes by the overlap analysis (**Figure 3F**), which were defined as key DE-MIRGs, namely F13A1, MGLL, ITGB3, ITGB5, PDGFA, SPARC, PDGFRA, IGFBP2, MMRN1, CD9, PDLIM1, SIGLEC10, S100A4, LRRC32, CLU, PPBP, TFPI, SELP, JAM3, CDKN1A, PRKAR2B, and GNG11.

From the perspective of functional enrichment analyses, key DE-MIRGs were significantly enriched in 105 GO terms, (67 in BP, 18 in CC, 20 in MF) and 12 KEGG pathways. For instance, wound healing, coagulation, blood coagulation, and hemostasis were the significant enriched terms in BP category, which were similar with MeDEGs. In view of the CC, platelet alpha granule, platelet alpha granule lumen, secretory granule lumen, cytoplasmic vesicle lumen, vesicle lumen, etc.. In terms of the MF, integrin binding, platelet–derived growth factor receptor binding, growth factor binding, etc. were the mainly enriched terms (**Figure 3G**). Besides, key DE-MIRGs were mainly enriched in KEGG pathways of PI3K-Akt signaling pathway, Focal adhesion, Regulation of actin cytoskeleton, Human cytomegalovirus infection, MicroRNAs in cancer, etc.. (**Figure 3H**).

The 22 key DE-MIRGs were uploaded to the STRING, after removing discrete nodes, a PPI network of these 22 DE-MIRGs were plotted, which included 22 nodes and 87 edges. Moreover, a rank chart of the node Degree was plotted, which revealed that ITGB3, SELP, and SPARC were the top 3 Degree score DE-MIRGs (**Figure 4A**). Next, the PPI network was divided into 2 clusters by Cytoscape and MCODE subnetwork discovery algorithm, which can be observed that the cluster 1 included 11 nodes and 46 edges, and there were 4 nodes and 5 edges in cluster 2 (**Figure 4B**). The 11 DE-MIRGs in cluster 1 (ITGB3, SELP, S100A4, SPARC, PDGFRA, CD9, ITGB5, PDGFA, CLU, IGFBP2, CDKN1A) were included in the subsequent analyses.

3.5 Identification of Key Genes

LASSO and SVM-RFE were performed to further downsize the dimensionality of the 11 DE-MIRGs in cluster 1, and integrated them with the characterized variables to filter the key genes. The LASSO results showed that 2 feature genes were screened out based on λ min = 0.1001, namely SELP and S100A4 (**Figure 5A**). Meanwhile, in the SVM-RFE model, 10-fold cross-validation ensured that the combination of SELP and S100A4 was equipped with high accuracy (accuracy = 0.745), and both of the two genes were with the average rank of 2 (**Figure 5B**).Subsequently, the Venn graph highlighted that SELP and S100A4 were the overlapped feature genes in both algorithms, and they were regarded as the key genes (**Figure 5C**).

3.6 Functional Analysis of the 2 Key Genes

The GSEA results demonstrated that in GO enrichment, both SELP and S100A4 were enriched in focal adhesion. Beside focal adhesion, SELP was mainly enriched in actin binding, actin filament organization, cell cortex, cell morphogenesis involved in differentiation, etc., and S100A4 was mostly enriched in electron transfer activity, generation of precursor metabolites and energy, lysosomal membrane, mitochondrial electron transport, NADH to ubiquinone, etc.. Furthermore, SELP and S100A4 were both enriched in KEGG pathways of Herpes simplex virus 1 infection. To be more specific, SELP was enriched in Axon guidance, Endocytosis, Focal adhesion, Pathways in cancer, Platelet activation, Proteoglycans in cancer, etc.. In terms of S100A4, Alzheimer disease, Chemical carcinogenesis – reactive oxygen species, Diabetic cardiomyopathy, Huntington disease, Oxidative phosphorylation, and Parkinson disease were the major enriched pathways (**Figure 6A-6B**).

4. Discussion

In recent years, some scholars have explored the pathogenesis and prevention of ICH from the regulation of DNA methylation in epigenetics. Yufeng GAO found that the up-regulation of dock1 expression after ICH may be related to its DNA hypomethylation by analyzing the gene database of patients with ICH and combining with clinical verification[28]. Zhang confirmed that the sites modified by DNA methylation after ICH were different from those of normal people[29]. After ICH, the blood vessel rupture, hematoma formation, and brain parenchyma damage, lead to the primary and secondary brain injury. Hypertension and body decline are the maifactors of cerebral hemorrhage[30]. The occurrence of ICH not only causes damage to the limb function and psychological function of the patients, but also brings a heavy burden to the patients' families, nursing staff and society. However, with the development of medicine, the prognosis has not been effectively improved.In recent years, with the development of epigenetics, a new idea has been found for the early prevention and treatment of cerebral hemorrhage. DNA methylation can change chromatin structure, DNA conformation, stability and protein interaction without changing DNA sequence. At present, it has been confirmed in biological research such as transcriptional regulation, transposition factor silencing, gene imprinting and X chromosome inactivation. However, little is known about the involvement of DNA methylation in the pathogenesis of intracerebral hemorrhage. How to explore the pathogenesis of ICH from the perspective of epigenetics and how to prevent and treat the occurrence of ICH at an early stage also provide new ideas for clinical and experimental research in the future.

In this study, the enrichment analysis of GO and KEGG of 451 intersecting genes showed that the GO terms such as wound healing, blood coagulation and hemostasis were mainly enriched; PI3K Akt signal pathway, malaria, actin cytoskeleton regulate KEGG pathway.Some researchers observed the release of inflammatory factors and the damage of brain nerve cells in the brain tissues of patients with cerebral hemorrhage, which confirmed that the inflammatory reaction and the damage of nerve cells after cerebral hemorrhage play an important role in the formation of secondary injury[31].Therefore, inhibition of inflammation and anti apoptosis of nerve cells after intracerebral hemorrhage may be the key to prevent and treat the disease.PI3K/Akt is a transduction protein of intracellular phosphorylation process. This signal pathway can protect brain tissue by regulating apoptosis protein and apoptosis gene[32].Bao confirmed through research that after intracerebral hemorrhage, PI3K and p-Akt increased in brain tissue of rats, and reached the peak at 24 hours. After 3 days, the expression of PI3K and p-Akt decreased, indicating that PI3K/Akt signal pathway participated in the pathological process of intracerebral hemorrhage.It has also been confirmed that PI3K/Akt signaling pathway is involved in the repair and metabolism of nerve cells after middle cerebral artery injury, and GSK-3 is an important downstream target of the pathway β It is the convergence point of cell protection signals and plays a positive role in brain protection, cell repair, growth, survival and recovery of neural function after intracerebral hemorrhage[33].At present, some researchers use recombinant human erythropoietin (rhEPO) to intervene in the rat model of intracerebral hemorrhage. They found that after treatment, the expression levels of PI3K and p-Akt in the brain tissue of intracerebral hemorrhage rats increased. The activation of PI3K/Akt

signal pathway can inhibit the apoptosis of neural cells and reduce the damage to neurons after hemorrhage[34].In addition, some researchers also activated PI3K/Akt signal pathway through progesterone to reduce the degree of neuroinflammatory reaction, blood brain barrier damage and lipid peroxidation, thus protecting neurons[35].Although more and more researchers have found that PI3K/Akt signaling pathway is involved in the pathogenesis of cerebral hemorrhage, and related drug research has been carried out, its specificity in cerebral hemorrhage remains to be further confirmed.

The long arm of chromosome 1 (1q21-24) of SELP locus is a 50kb DNA sequence, which contains 17 exons and 16 introns.There are many kinds of gene polymorphisms in SELP. So far, 13 kinds of gene polymorphisms have been detected[36].SELP is a component of resting platelet granular membrane and endothelial cell Weibel Prade body. There are about 10000 SELP on activated platelet surface, which is a disease with multiple causes.The complex genetic and environmental factors can be used as the pathogenic factors of stroke, and atherosclerosis is the basis of its pathological changes[37].In recent years, the role of SELP gene in inflammatory response and the pathogenesis of atherosclerotic plaque has become a hot spot in the study of stroke. As a member of the selectin family of adhesion molecules, SELP is mainly involved in the recognition and adhesion between cells, especially in mediating the adhesion of platelets, endothelial cells and white blood cells, and the movement of white blood cells to inflammatory sites, and participating in thrombosis, accelerating the generation of atherosclerosis, It is closely related to the destruction of body immunity and tumor metastasis[38].Studies have shown that when vascular endothelium is damaged, the expression of SELP in platelets increases rapidly, and the expression of SELP in endothelial cells also increases. The combination of SELP and PSGL-1 on the surface of leukocytes can activate signal transduction in leukocytes, make leukocytes release inflammatory factors, aggravate inflammatory reaction, and induce cardiovascular and cerebrovascular diseases[39].Wang confirmed through research that when cerebral infarction occurred in rats, the level of SELP in peripheral blood increased, and the expression of SELP in plaque endothelial cells increased[40].Some researchers also mentioned that the polymorphism of SELP gene can also be one of the risk factors for preventing and treating stroke[41].

S100A4 belongs to the S100 protein family. Studies have shown that S100B can reflect the severity of brain injury, which is of certain value for the evaluation of the condition of patients with cerebral hemorrhage[42].S100A4 has a protein molecule expressed by 4 effective coding sequences and contains 101 amino acid units. It has a strong binding ability to calcium ions and plays an important role in inflammatory reaction, fibrosis, tumor metastasis and autoimmune diseases.S100A4 was first found in tumor cells and highly expressed in metastatic cancer cells[43].However, some studies have shown that S100A4 can play a protective role in myocardial damage, and even some studies have proved that it lacks specificity in cardiac remodeling and fibrosis, and its role in the heart is still controversial[44].Liu found that the serum S100A4 protein level in patients with stroke was highly expressed, which was significantly related to the severity of stroke and short-term prognosis[45].Wang alleviates oxidative damage of vascular endothelial cells induced by oxidized low density lipoprotein (ox LDL) and inhibits apoptosis by targeting S100A4 expression with Mi R193a-3p[51].S100A4 is also considered as an alarm

element, a ligand with multiple receptors. In the past many years, S100A4 has been considered as a fibroblast specific marker, participating in the fibrosis process of liver, lung, kidney and other organs[46].

Declarations

ACKNOWLEDGEMENTS

I grateful thanks to Ms. Li Ping for my guidance in writing articles and teaching me bioinformatics.Thanks to Wu Qiong for his help in statistical analysis and consult relevant literature.

Data Availability Statemen

In this study, GSE179759 and GSE125512 were downloaded from the GEO database. The data obtained through analysis are in the article, supporting the public publication of this journal.

Funding Statemen

The research was supported by Professor Li Ping from the National Natural Science Foundation of China, project number: 81603557.

Conflict of Interest

There is no conflict of interest between this article and the fund support. All authors have no interest relationship.

References

1. Diener H-C,Hankey G.J.Primary and secondary prevention of ischemic stroke and cerebral hemorrhage.Journal of the American College of Cardiology[J].2022,75(15): 1804–1818. doi: 10.1016/j.jacc.2019.12.072.
2. TirschwellDL.Intracerebral haemorrhage is hard to stop,and must be attacked before,during and after[J].BMJE Vid Based Med,2019,24(5):e4.
3. Wang YJ,Lzx,Guhq,etal.China stroke statistics 2019;areport from the national center for heal thcare quality mana-ment inneurological diseases,chinanation alclinical resarchenter forneurolog-ical diseases,thechinese strokeas sociation,national center for disaese neurological diseases,thechinese storkeas sociation,national center for chronic and non-communicable disease contral and prevention and institute for glorbal neuroscience and storke collaborations[J].Strokeand Vasculair Neurology,2020,5(3):211-239.
4. SaccoS,MariniC,ToniD,etal,Incidence and 10-year survival of intracerebral hemprrhage inapopulation based registry[J].Stroke,2009,40(2):394-393.
5. Van AschCJ,luitseMJ,RinkelGJ,etal.Incidence,case fatality,and functional outcome of intraceebal haemorrhage overtome,according toage,sex,and ethaic origin: asystem aticreview and

metaanalysis[J].lance Neurol,2010,9(2):167-176.

6. QureshiAI,PaleschYY,MartinR,ETAL.Effect of systolic blood pressure reduction on hematomaex pansion,perihematomaledema,and3-Month outcome among pastients with intracerebral hemorrhageresults from the attihypertensive traetment of acutecerbral hemorrhage stufy[J].Arch N eurol,2010,67(5):570-576.
7. Fernando S.M., Qureshi D., Talarico R.,et al. Intracerebral Hemorrhage Incidence, Mortality, and Association With Oral Anticoagulation Use: A Population Study. Stroke. 2021;52:1673–1681. doi: 10.1161/STROKEAHA.120.032550.
8. Van Asch C.J., Luitse M.J., Rinkel G.J., van der Tweel I., Algra A., Klijn C.J. Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: A systematic review and meta-analysis. Lancet Neurol. 2010;9:167–176. doi: 10.1016/S1474-4422(09)70340-0.
9. T. Apostolaki-Hansson, T. Ullberg, B. Norrving, and J. Petersson, “Prognosis for intracerebral hemorrhage during ongoing oral anticoagulant treatment,” *Acta Neurologica Scandinavica*, vol. 139, no. 5, pp. 415–421, 2019.
10. Carmichael ST, Vespa PM, Saver JL, Coppola G, Geschwind DH, Starkman S, et al. . Genomic Profiles of Damage and Protection in Human Intracerebral Hemorrhage. J Cereb Blood Flow Metab (2008) 28:1860–75. 10.1038/jcbfm.2008.77
11. Rosell A, Vilalta A, García-Berrocoso T, Fernández-Cadenas I, Montaner J. Brain Perihematoma Genomic Profile Following Spontaneous Human Intracerebral Hemorrhage. PloS One (2012) 6:e16750. 10.1371/journal.pone.0016750
12. Zhang Y, Long H, Wang S, et al. Genome-Wide DNA Methylation Pattern in Whole Blood Associated With Primary Intracerebral Hemorrhage. Front Immunol. 2021;12:702244. Published 2021 Aug 13. doi:10.3389/fimmu.2021.702244
13. Ma Z. Research Progress on the Mechanism of Inflammatory Response in Secondary Brain Injury after Intracerebral Hemorrhage[J]. Chinese Journal of Clinical Neurosurgery,2020,25(2):124-126.DOI:10.13798/j.issn.1009-153X.2020.02.022.
14. nnunziato F,Cosmi L,Liotta F,Maggi E,Romagnani S.Type17t helper cells—origins,features and possible roles in rheumatic disease[J].Nature reviews,Rheumatology.2009;5:325—331.
15. Zuniga LA Jain R Haines C Cua DJ Thl7 cell development From the cradle to the grave Immunological reviews 2013 252 78—88.
16. Qureshi IA, Mehler MF. Epigenetic Mechanisms Underlying Human Epileptic Disorders and the Process of Epileptogenesis. Neurobiol Dis (2010) 39:53–60. 10.1016/j.nbd.2010.02.005
17. Sharma P, Kumar J, Garg G, Kumar A, Patowary A, Karthikeyan G, et al. . Detection of Altered Global DNA Methylation in Coronary Artery Disease Patients. DNA Cell Biol (2008) 27:357–65. 10.1089/dna.2007.0694.
18. Li J., Hou R., Niu X., Liu R., Wang Q., Wang C., et al. (2016). Comparison of microarray and RNA-Seq analysis of mRNA expression in dermal mesenchymal stem cells. Biotechnol Lett 38, 33–41.

19. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis.[J]. Carcinogenesis, 2013.
20. Tao W , He X , Liu X , et al. Weighted Gene Co-expression Network Analysis Identifies FKBP11 as a Key Regulator in Acute Aortic Dissection through a NF-kB Dependent Pathway.
21. Yuan Tian Tiffany J Morris, et al.ChAMP: updated methylation analysis pipeline for Illumina BeadChips[J].Bioinformatics, 33(24), 2017, 3982–3984.
22. Guangchuang Yu, Li-Gen Wang,et al.clusterProfifiler: an R Package for Comparing Biological Themes Among Gene Clusters[J].Technical Communication,2012,5(16):284-287.
23. Peter Langfelder,Steve Horvath.WGCNA: an R package for weighted correlation network analysis[J].BMC Bioinformatics 2008, 9:559.
24. Jerome Friedman, Trevor Hastie et al.Regularization Paths for Generalized Linear Models via Coordinate Descent[J].J Stat Softw. 2010 ; 33(1): 1–22.
25. Damian,Szklarczyk,Annika L.Gable,et al.The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets[J].Nucleic Acids Research, 2021,49:605-612
26. Damian Szklarczyk,Annika L.Gable,etal.The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement Sets[J].Nucleic Acids Research, 2020,1:1-20.
27. David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel and Friedrich Leisch (2021). e1071:Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-9.
28. Yufeng Gao,Xiaojie Fu. et al. DNA Hypomethylation of DOCK1 Leading to High Expression Correlates with Neurologic Deterioration and Poor Function Outcomes after Spontaneous Intracerebral Hemorrhage[J].Evidence-Based Complementary and Alternative Medicine,2021,10:
29. Yupeng Zhang,Hongyu Long,etal.Genome-Wide DNA Methyiation Patter in Whole Blood Associated With Primary Intracerebral Hemorrhage[J].2021,7(26):1-26
30. Kim S,Chen J,Cheng T,etal,Pub Chem in2021:new data content and improved web interfaces[J].Nucleic Acids Res,2021,49(D1):D1388-D1395.
31. Zhu H, Wang Z, Yu J, et al. Role and mechanisms of cyto-kines in the secondary brain injury after intracerebral hemorrhage [J]. Prog Neurobiol, 2019, 178: 101610
32. Kim J,Lee S,Choi B R,et al. Sulforaphane epigenetically enhances neuronal BDNF expression and Trk B signaling pathways . Mol Nutr Food Res,2017,61 2 .DOI: 10.1002/mnfr.201600194
33. Wang C, Wei Z, Jiang G, et al. Neuroprotective mechanisms of miR-124 activating PI3K/Akt signaling pathway in ischemic stroke[J]. Experimental & Therapeutic Medicine, 2017, 13(6): 3315-3318.
34. XIANG Yong,ZHU Jianping,LIU Danrong.Beneficial effect of recombinant human erythropoietin (rh EPO) on neuronalinjury in intracerebral hemorrhage rats via PI3K / AKT signaling pathway[J].Medical Science Journal of Central South China,2017,9(45):454-458

35. Yang J, Arima H,Wu G,et al. Prognostic significance of perihematomal edema in acute intracerebral hemorrhage pooled analysis from the intensive blood pressure reduction in acute cerebral hemorrhage trial studies. Stroke 2015 46 4 1009-1013.
36. Zhao Y Yang X en L et al Correlations of SELE and SELPgenetic polymorphisms with myocardial infarction risk: A meta-analysis and meta-regression J].Mol Biol Rep,2014,41(7) :4521-4532
37. Kinsella JA Tobin WO Tierney S et al Increased platelet activa-tion in early symptomatic vs asymptomatic carotid stenosis andrelationship with microembolic status: Results from the Plateletsand Carotid Stenosis Study[J]. J Thromb Haemost,2013,11(7):1407-1416
38. Vlachadis N Tsamadias V Vrachnis N et al Associations ofcombined polymorphisms of the platelet membrane glycoproteinsIa and IIIa and the platelet-endothelial cell adhesion molecule-1and P-Selectin genes with IVF implantation failures J J ObstetGvnaecol 2017 37(3) : 363-369
39. Wang H Kleiman K Wang J et al Deficiency of P-selectin glyco-protein ligand-1 is protective against the prothrombotic effects ofinterleukin-1β[J].J Thromb Haemost 2015 13(12) : 2273-2276
40. Bielinski SJ,Berardi C,Decker PA,et al.P-selectin and subclinical and clinical atherosclerosis: The Multi-Ethnic Study of Atheroscle- rosis (MESA)[J].Atherosclerosis 2015 240(1) : 3-9
41. Abbasi M Sajjadi M Fathi M et al Serum S100B Protein as an out-come prediction tool in emergency department patients with traumaticbrain injury J .Turk J Emerg Med,2014,14(4) : 147-152
42. Chen L, Li J, Zhang J, et al. S100A4 promotes liver fibrosis via activation of hepatic stellate cells[J] .J Hepatol. 2015 ,62(1):156-64.
43. Österreicher CH, Penz-Österreicher M, Grivennikov SI, et al. Fibroblast-specific protein 1 identifies an inflammatory subpopulation of macrophages in the liver[J] . Proc Natl Acad Sci U S A. 2011,108(1):308-1
44. Liu Saibing,Zhang Hongxing.Correlation analysis of serru S100A4 protein level with stroke severity and short-term outcome in patients with acute ischemic stroke[J].Journal of Brain and Nervous Diseases.2019,29(12):734-736.
45. Wang Bin, Qi Xianwei, Zhang Xianliang.Influence of mi R-193a-3p on oxidized low density lipoprotein-induced vascular endothelial cell injury through regulating S100A4 expression[J].Chin J Evid Based Cardiovasc Med,2021,13(6):742-751.
46. Zhang W, Ohno S, Steer B,et al. S100a4 Is Secreted by Alternatively Activated Alveolar Macrophages and Promotes Activation of Lung Fibroblasts in Pulmonary Fibrosis[J] . Front Immunol. 2018 ,9:1216.

Figures

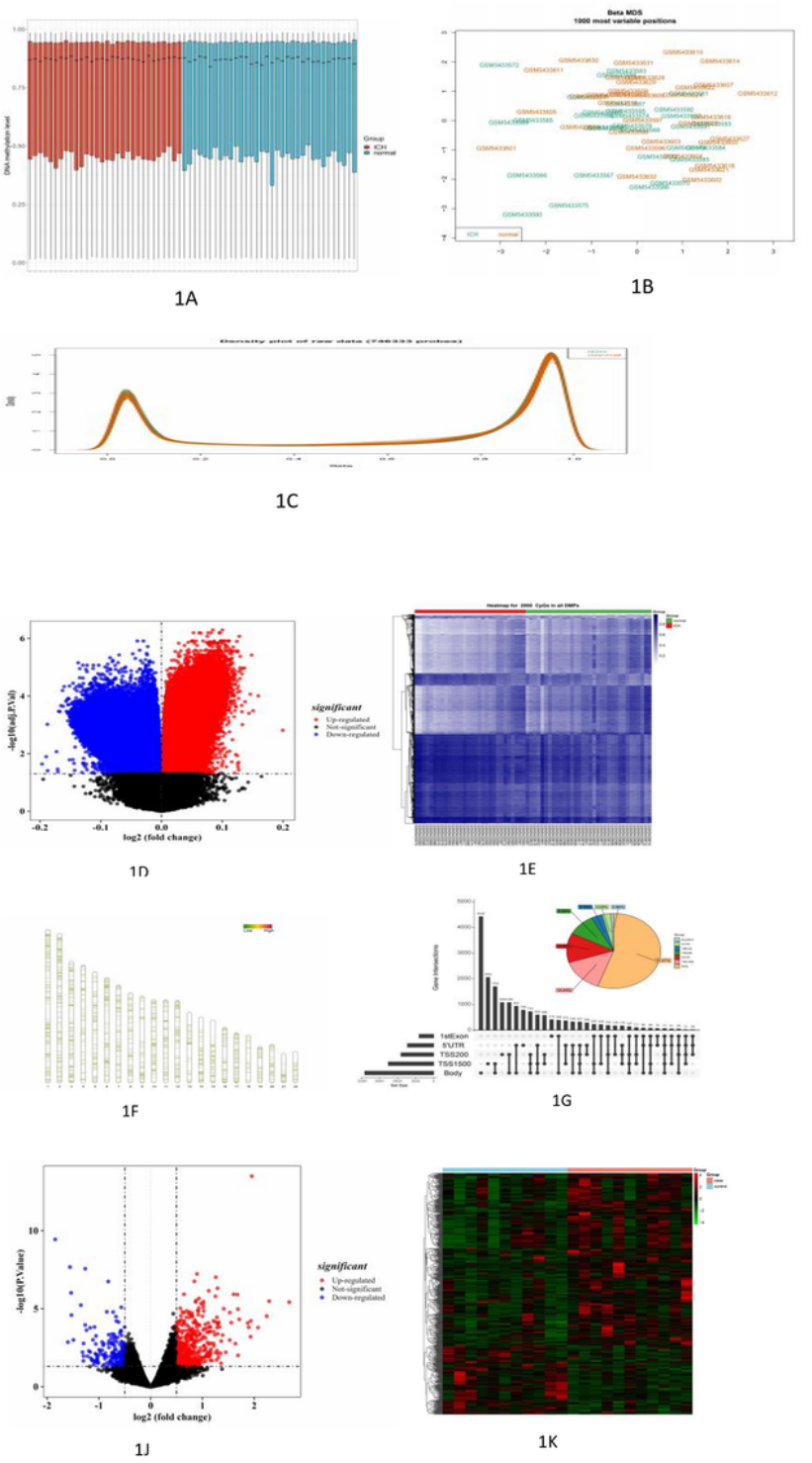


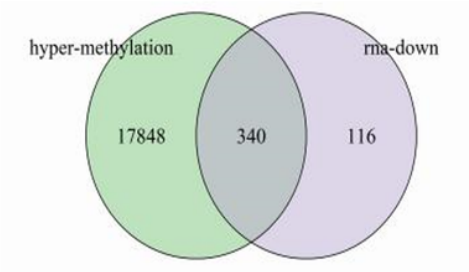
Figure 1

Methylation data processing. (1A The sample methylation level.(1B)The MDS diagram of methylation sample. (1C) The density map of beta value of methylated samples

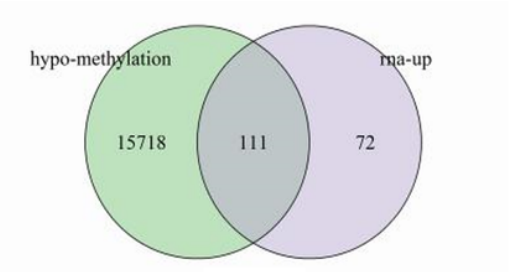
Analysis of differential methylation sites. (1D)The volcano map of differential methylation sites. (1E) The differential methylation site Heatmap. (1F) The differential DMP (| logfc |> 0.1) chromosome coordinates.

(1G)The differential DMP (| logfc |> 0.1) chromosome coordinates. (1H) The different methyl site genes of hypo DMR.

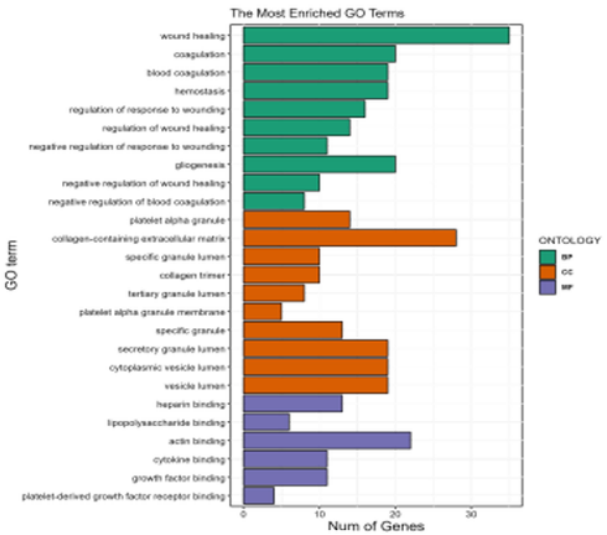
(1J)Each dot in the volcano map represents a gene, and blue and red dots represent significantly differentially expressed genes. Red dots indicate that the basal expression level is up-regulated (disease samples vs. control samples), blue dots indicate that the gene expression level is down regulated (disease samples vs. control samples), and black dots indicate that there is no significant difference between these genes. (1K) Each small square in the heat map represents each gene, and its color represents the expression amount of the gene. The higher the expression amount, the more red the color is, and the lower the expression amount, the more green the color is. The first row indicates the sample grouping, the blue indicates the disease sample, and the red indicates the control sample. Each row represents the expression of each gene in different samples, and each column represents the expression of all differential genes in each sample. The tree on the left shows the results of clustering analysis of different genes from different samples.



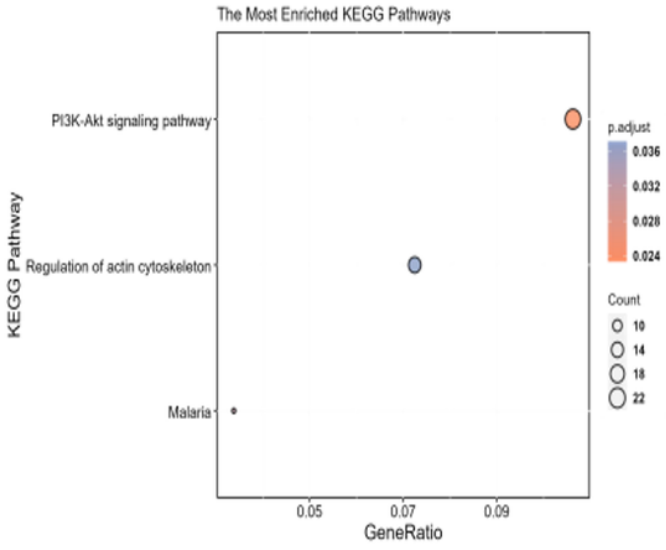
2A



2B



2C

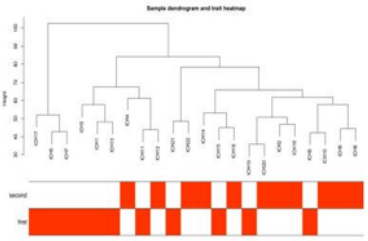


2D

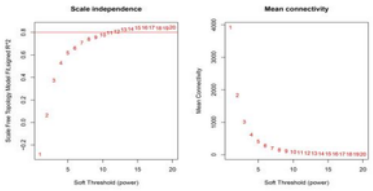
Figure 2

(2A, 2B) 340 genes were obtained from high methylation low expression genes, 111 genes were obtained from low methylation high expression, and 451 differentially methylated genes were obtained in total

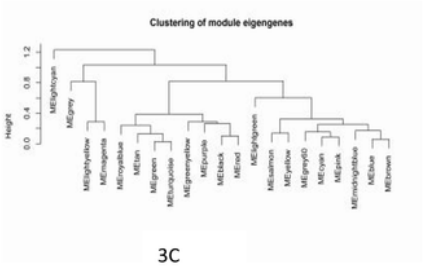
(2C, 2D):Through go analysis, 134 terms were enriched, including 101 BD terms, 27 CC terms and 6 MF terms. The main terms are wound healing, blood coagulation and hemostasis.2C:Medeg is mainly enriched in 3 KEGG pathways, including PI3K Akt signaling pathway, regulation of actin cytoskeleton and malaria.



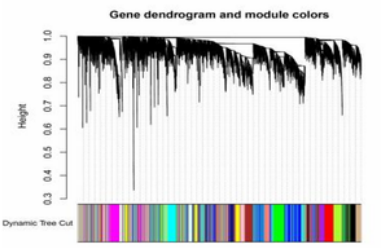
3A



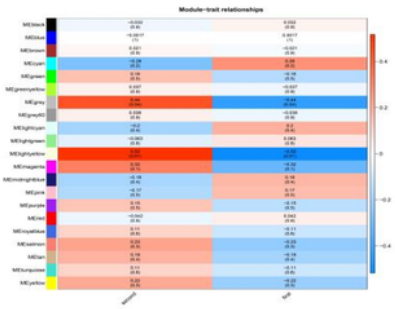
3B



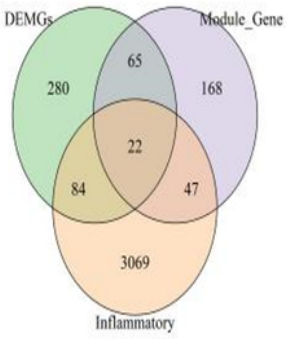
3C



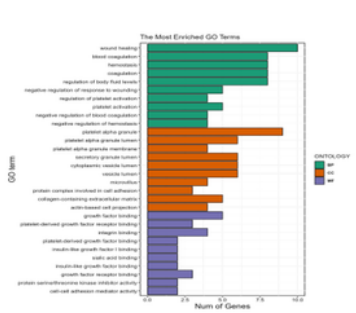
3D



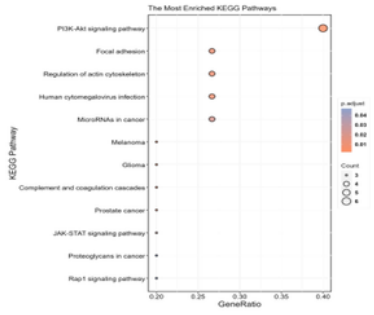
3E



3F



3G



3H

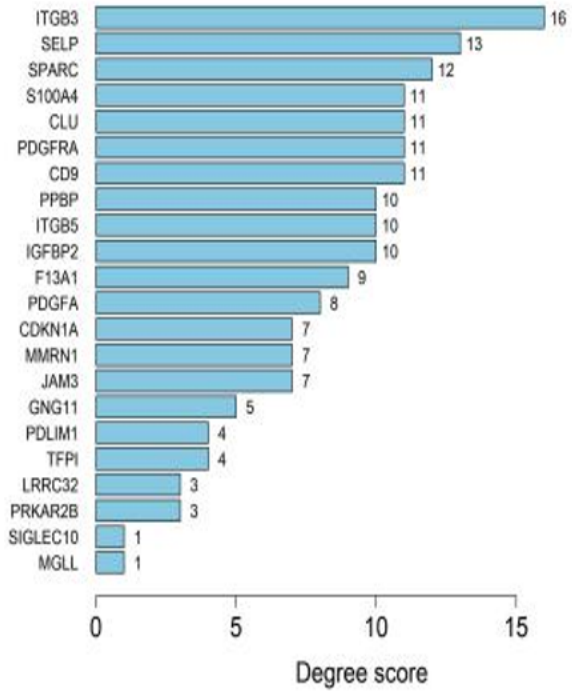
Figure 3

(3A) The clustering and phenotypic information of merged data samples. Left figure: the horizontal axis of the above figure represents the weight

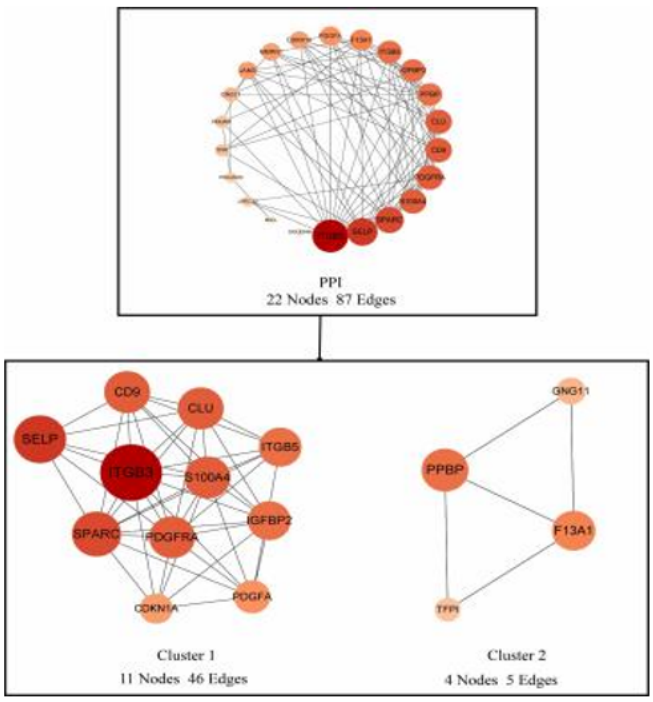
(3C) Construct a coexpression network to obtain 21 modules, of which the genes in the grey module cannot be classified into a certain class according to the similarity between genes. Choose not to merge

the modules, and use 21 modules to continue the analysis. (3D) Module gene identification. (3E) The heatmap of correlationb etween modules and clinical traits. (3F) The key module genes and differentially expressed methylation genes intersected with inflammation related genes, and a total of 22 differentially methylated genes related to cerebral hemorrhage were obtained.

(3G) The histogram of go enrichment of differentially methylated genes in intracerebral hemorrhage. (3H)The bubble Diagram of KEGG enrichment of differentially methylated gene in intracerebral hemorrhage.



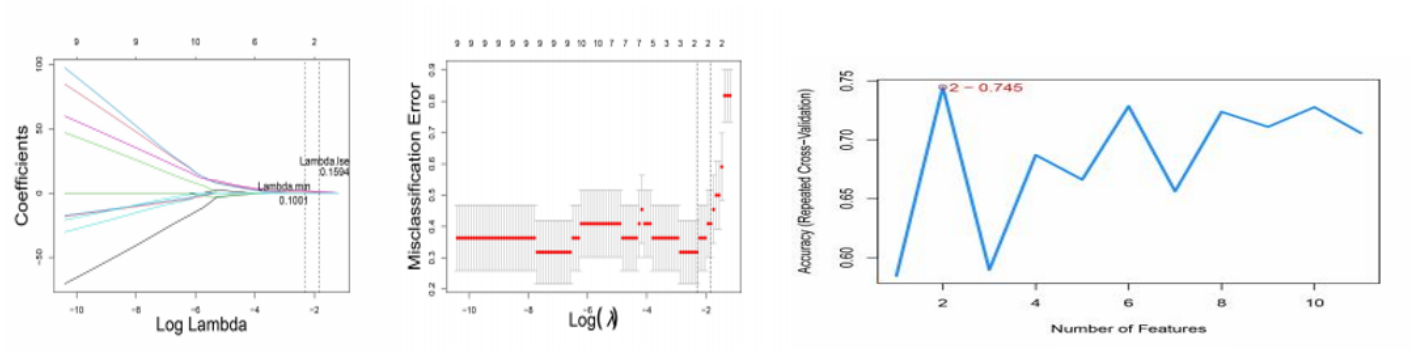
4A



4B

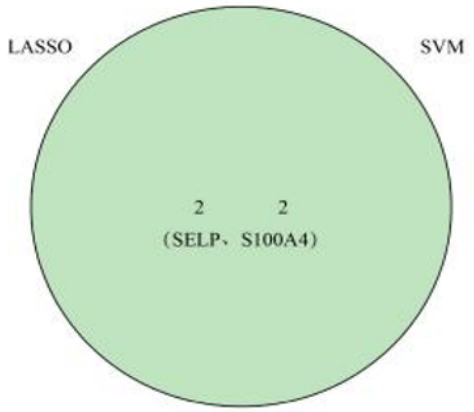
Figure 4

(4A) The degree ranking chart. (4B) The protein interaction network of differentially methylated genes.



5A

5B



5C

Figure 5

(5A) Each curve in the left figure represents the change trajectory of each independent variable coefficient, the ordinate is the coefficient value, and the upper abscissa is the number of non-zero coefficients in the model at this time. The abscissa in the right figure is log (lambda), and the ordinate represents the error of cross validation. Red dots represent the mean square error and the upper and lower standard deviation. The smaller the mean square error, the better the model; The number above indicates the number of independent variables that still exist in the model (not necessarily monotonically decreasing). The first dashed line indicates the minimum value of mean square error; The second dashed line marks the position of the double standard deviation of the lowest point, indicating the simplest model that can be obtained at the expense of the double standard deviation. As the penalty coefficient lambda changes, the coefficients of most variables are finally compressed to 0, and the best lambda value is selected when the cross validation error of 10 fold is the smallest, as well as the error rate of the model. (5B) The accuracy of support vector machine model. (5C) The wayne diagram of Characteristic Genes.