Discovery and validation of molecular patterns and immune characteristics in the peripheral blood of ischemic stroke patients

Lin Cong [1], **Yijie He** [1], **Yun Wu** [1], **Ze Li** [1], **Siwen Ding** [1], **Weiwei Liang** [1], **Xingjun Xiao** [1], **Huixue Zhang** [1], **Lihua Wang** Corresp. [1]

[1] Department of Neurology, The Second Affiliated Hospital of Harbin Medical University, Harbin Medical University, Harbin, China

Corresponding Author: Lihua Wang
Email address: wang_lihua0910@163.com

Background: Stroke is a disease with high morbidity, disability, and mortality. Immune factors play a crucial role in the occurrence of ischemic stroke (IS), but their exact mechanism is not clear. This study aims to identify possible immunological mechanisms by recognizing immune-related biomarkers and evaluating the infiltration pattern of immune cells.

Methods: We downloaded datasets of IS patients from GEO, applied R language to discover differentially expressed genes, and elucidated their biological functions using GO, KEGG analysis, and GSEA analysis. The hub genes were then obtained using two machine learning algorithms (least absolute shrinkage and selection operator (LASSO) and support vector machine-recursive feature elimination (SVM-REF)) and the immune cell infiltration pattern was revealed by CIBERSORT. Gene-drug target networks and mRNA-miRNA-lncRNA regulatory networks were constructed using Cytoscape. Finally, we used RT-qPCR to validate the hub genes and applied logistic regression methods to build diagnostic models validated with ROC curves.

Results: We screened 188 differentially expressed genes whose functional analysis was enriched to multiple immune-related pathways. Six hub genes (ANTXR2, BAZ2B, C5AR1, PDK4, PPIH, and STK3) were identified using LASSO and SVM-REF. ANTXR2, BAZ2B, C5AR1, PDK4, and STK3 were positively correlated with neutrophils and gamma delta T cells, and negatively correlated with T follicular helper cells and CD8, while PPIH showed the exact opposite trend. Immune infiltration indicated increased activity of monocytes, macrophages M0, neutrophils, and mast cells, and decreased infiltration of T follicular helper cells and CD8 in the IS group. The ceRNA network consisted of 306 miRNA-mRNA interacting pairs and 285 miRNA-lncRNA interacting pairs. RT-qPCR results indicated that the expression levels of BAZ2B, C5AR1, PDK4, and STK3 were significantly increased in patients with IS. Finally, we developed a diagnostic model based on these four genes. The AUC value of the model was verified to be 0.999 in the training set and 0.940 in the validation set.

Conclusion: Our research explored the immune-related gene expression modules and provided a specific basis for further study of immunomodulatory therapy of IS.

Discovery and validation of molecular patterns and immune characteristics in the peripheral blood of ischemic stroke patients

Lin Cong[1], Yijie He[1], Yun Wu[1], Ze Li[1], Siwen Ding[1], Weiwei Liang[1], Xingjun Xiao[1], Huixue Zhang[1], Lihua Wang[1]

[1]Department of Neurology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

Corresponding Author:
Lihua Wang [1]
Harbin, Heilongjiang, China
Email address: wang_lihua0910@163.com

Abstract

Background: Stroke is a disease with high morbidity, disability, and mortality. Immune factors play a crucial role in the occurrence of ischemic stroke (IS), but their exact mechanism is not clear. This study aims to identify possible immunological mechanisms by recognizing immune-related biomarkers and evaluating the infiltration pattern of immune cells.

Methods: We downloaded datasets of IS patients from GEO, applied R language to discover differentially expressed genes, and elucidated their biological functions using GO, KEGG analysis, and GSEA analysis. The hub genes were then obtained using two machine learning algorithms (least absolute shrinkage and selection operator (LASSO) and support vector machine-recursive feature elimination (SVM-REF)) and the immune cell infiltration pattern was revealed by CIBERSORT. Gene-drug target networks and mRNA-miRNA-lncRNA regulatory networks were constructed using Cytoscape. Finally, we used RT-qPCR to validate the hub genes and applied logistic regression methods to build diagnostic models validated with ROC curves.

Results: We screened 188 differentially expressed genes whose functional analysis was enriched to multiple immune-related pathways. Six hub genes (ANTXR2, BAZ2B, C5AR1, PDK4, PPIH, and STK3) were identified using LASSO and SVM-REF. ANTXR2, BAZ2B, C5AR1, PDK4, and STK3 were positively correlated with neutrophils and gamma delta T cells, and negatively correlated with T follicular helper cells and CD8, while PPIH showed the exact opposite trend. Immune infiltration indicated increased activity of monocytes, macrophages M0, neutrophils, and mast cells, and decreased infiltration of T follicular helper cells and CD8 in the IS group. The ceRNA network consisted of 306 miRNA-mRNA interacting pairs and 285 miRNA-

lncRNA interacting pairs. RT-qPCR results indicated that the expression levels of BAZ2B, C5AR1, PDK4, and STK3 were significantly increased in patients with IS. Finally, we developed a diagnostic model based on these four genes. The AUC value of the model was verified to be 0.999 in the training set and 0.940 in the validation set.

Conclusion: Our research explored the immune-related gene expression modules and provided a specific basis for further study of immunomodulatory therapy of IS.

Introduction

Stroke is the second leading cause of death and the third leading cause of disability in adults (Campbell & Khatri, 2020), and more than 80% of strokes are ischemic (Hasan et al., 2018). Ischemic stroke (IS) seriously affects patients' quality of life and causes a heavy burden to families and society. Therefore, our study focuses on discovering novel targets for diagnosing and treating IS.

Currently, IS diagnosis mainly depends on clinical manifestations and neuroimaging procedures, such as computed tomography (CT) (Harston et al., 2015) and magnetic resonance imaging (MRI) (Zameer et al., 2021). The most effective early treatment methods are thrombolysis and intravascular therapy (Cassella & Jagoda, 2017; Hasan et al., 2018), both of which have time limitations when administering. The clinical manifestations of the disease are variable, and the cost of administering CT and MRI scans is expensive and time-consuming. If the symptoms are not typical or the above examinations cannot be performed promptly, diagnosis and treatment may be delayed. Additionally, the above treatment methods carry the risk of bleeding. This has led us to search for new diagnostics and treatments.

Genetic factors are related to the occurrence and prognosis of stroke (Malik et al., 2018; Torres-Aguila et al., 2019). Possible stroke-related gene loci, such as ANK2, CDK6, and KCNK3, were found using whole genome sequencing (Malik et al., 2018). However, the specific mechanism of action was not clear. CYP450, COX-2, PTGIS, TBXAS1, P2RY1, TGB3, and GPIIa are candidate genes that may be associated with stroke prognosis (Torres-Aguila et al., 2019). Presently, bioinformatics technology has been widely applied to various diseases, so we used bioinformatics technology to screen and identify key genes associated with IS and build a classification diagnosis model based on these genes, in order to provide certain auxiliary methods for the diagnosis of IS and potential therapeutic targets for the management of diseases.

Immune factors are also closely associated with IS (Iadecola et al., 2020; Levard et al., 2021). In the acute phase, the immune response is activated immediately when the brain is in ischemia. Damaged or dead brain cells release damage-associated molecular patterns (DAMPS) (Wu et al., 2022). These substances quickly activate intracranial immune cells (such as microglia), release chemokines and cytokines, and trigger a series of inflammatory cascades (Anrather & Iadecola, 2016; Levard et al., 2021). More importantly, immune cells in the peripheral blood can penetrate

the central nervous system (CNS) through the damaged blood-brain barrier (BBB) (Jayaraj et al., 2019), further exacerbating damage. At the same time, these DAMPS and cytokines can enter the bloodstream and activate the systemic immune system and inflammatory response, which can lead to severe immunosuppression and even fatal infections (Iadecola et al., 2020). In the chronic phase, adaptive immune responses to the brain are mobilized due to antigen presentation, which may be the basis of neuropsychiatric sequelae, poststroke dementia/fatigue/depression, and the root cause of post-stroke morbidity (Iadecola et al., 2020). Epidemic-modulated therapy may be a good alternative to IS therapy. Several studies have shown that immune regulation can delay disease progression and improve neurological function and prognosis. These changes highlight the need to maintain a stable immune microenvironment in the CNS (Javidi & Magnus, 2019; Jayaraj et al., 2019). TGF-β1 has a significant neuroprotective effect (Taylor et al., 2017) and can reduce post-stroke infection (Cekanaviciute et al., 2014; Doll et al., 2014). Relevant targeted drugs are now being developed. Specific inhibitors of IL-1β can delay atherosclerosis by interfering with immune pathways associated with atherosclerotic plaque formation (Khambhati et al., 2018). Meanwhile, IL-1 receptor antagonists have been found to effectively reduce peripheral inflammation in patients with acute IS and improve clinical outcomes in these patients (Smith et al., 2018). In conclusion, the analysis of immune infiltration analysis and study of the relationship between target genes and immune cell distribution may help to elucidate the immune-related molecular mechanism of IS and provide the specific basis for immunoregulatory therapy. At the same time, the prediction of drug target gene action network may provide a new target for the treatment of IS.

Competing endogenous RNA (ceRNA) does not represent a specific type of RNA, but a regulatory mechanism. By combining with microRNA response elements (MREs) existing on mRNA, miRNAs inhibit mRNA translation or lead to its degradation, thus achieving the function of regulating gene expression after transcription (Salmena et al., 2011). Different mRNAs can compete to bind to the same miRNA and thus participate in the regulation of gene expression. Transcriptome studies have found that MREs exist not only on mRNA, but also on lncRNA, circRNA, and other types of RNA, which means that the same miRNA can bind to several and multiple types of RNA. A competitive relationship is formed between RNA molecules that bind to the same miRNA, which forms the ceRNA network. If the regulation of ceRNA is abnormal, it may cause the occurrence of diseases, as shown in the research of tumors (Karreth et al., 2015), stroke (Li et al., 2020), and other complex diseases. It has been reported that miRNAs intervene in the occurrence and prognosis of stroke by modulating immune and inflammatory factors. Huang et al. (2018) found that the inhibition of microRNA-210 suppresses pro-inflammatory response and reduces acute brain injury of IS in mice. Boldin et al. (2011) reported that miR-146a inhibits the production of inflammatory cytokines. In addition, the occurrence, progression, and prognosis of acute IS are also related to abnormal lncRNA expression (Bai et al., 2014; Feng et al., 2019; Fu et al., 2021; Li et al., 2020; Wang et al., 2020). Therefore, they are potential candidates for stroke diagnosis. Meanwhile accumulating evidence demonstrates

that lncRNAs regulate target gene expression by acting as ceRNAs for miRNAs and participate in immunoinflammation in cerebral ischemia-reperfusion injury and various diseases. Knockdown metastasis-associated lung adenocarcinoma transcript 1 (MALAT1), a ceRNA for miR-181c-5p, attenuated inflammatory injury after cerebral ischemia (Cao et al., 2020). lncRNA SNHG14 promotes inflammatory response induced by cerebral ischemia /reperfusion injury by regulating miR-136-5p/ROCK1 (Zhong et al., 2019). The lncRNA small nucleolar RNA host gene 15 (SNHG15) was shown to regulate the expression of programmed death ligand 1 (PD-L1) by inhibiting miR-141, which in turn promoted the resistance of stomach cancer cells to the immune response (Dang et al., 2020). The lncRNA HIX003209 functioned as a ceRNA and enhanced inflammation by sponging miR-6089 via the Toll-like receptor 4 (TLR4)/nuclear factor (NF)-κB pathway in macrophages in rheumatoid arthritis (Yan et al., 2019). The construction of a target gene-miRNA-LncRNA-based CERNA network may provide a molecular basis for understanding stroke and help predict the possible immunological pathways of IS.

In this study, the least absolute shrinkage and selection operator (LASSO) regression and support vector machine-recursive feature elimination (SVM-RFE) algorithm were applied to screen essential genes, and their relationship was subsequently analyzed with immune cells. The combined use of two algorithms made the process more rigorous and standard. A ceRNA network of hub genes-miRNA-lncRNA was then established. Finally, a logistic regression diagnostic model was built based on hub genes to provide an auxiliary method for the diagnosis of IS. The technical route is shown in Figure 1.

Materials & Methods

Obtaining and processing profile datasets

We obtained and downloaded the IS gene expression profile from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) (Table 1). The criteria for screening were that the data in the dataset contained both IS and control groups (CTL), and the number of cases in each group was no less than 20. Patients in the IS group were those diagnosed with acute IS. In addition, the data were based on whole blood samples with full gene expression profiles. After screening, GSE16561 (Barr et al., 2010) and GSE58294 (Stamova et al., 2014) matched the criteria. GSE16561 was used as a training set. GSE58294 was used for the validation of biomarker genes. We applied the Perl language script (https://www.perl.org/) to get the gene name corresponding to each probe from the platform annotation file, and the names were then converted. In this process, we used the average of all probes' expressions if there were multiple probes corresponding to the same gene name.

Differential expression analysis

We applied R package limma (Ritchie et al., 2015) (version 3.6.0) to perform differential expression analysis to identify DEGs between groups in the dataset GSE16561. FoldChange >

1.5 and adjusted *P*-value < 0.05 were taken as significant criteria. DEGs were represented by heatmap, designed by the pheatmap package (Kolde, 2012) (version 1.0.12), and volcano maps, created by the ggplot2 package (Wickham, 2009) (version 3.3.5).

Function enrichment analysis

We not only annotated the function of the differentially expressed genes themselves using Gene Ontology (GO) enrichment, but also studied the pathways through Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. GO Enrichment and KEGG pathways analysis results were obtained using clusterProfiler (Yu et al., 2012). GSEA software was used to analyze the potential molecular mechanisms and signaling pathways of Hub genes in IS patients. C2 (c2.cp. Kegg. v7.4.symbols. gmt) was selected as the reference gene set for analysis. The R package clusterProfiler also conducted the GSEA analysis of the IS and CTL groups, with an adjusted *P*-value<0.05.

Screening for hub genes

Two machine learning methods (LASSO and SVM-REF) were used to identify the hub genes of IS. These two methods were first applied to analyze breast tumors (Wang & Liu, 2015), and then also to neurological diseases such as AD (Liu et al., 2021). LASSO regression is characterized by variable selection and complexity adjustment while fitting a generalized linear model. Thus, the target-dependent/response variable can be modelled and predicted whether it is continuous, binary, or multivariate discrete (Friedman et al., 2010). The LASSO analysis was undertaken using the glmnet package (Engebretsen & Bohlin, 2019) with the following parameters: response type set to binomial, alpha set to 1, and 10-fold cross-validation to adjust the optimal value of the parameter λ. SVM has shown many unique advantages in solving small samples and in nonlinear and high-dimensional pattern recognition, and can be extended to other machine learning problems such as function fitting. SVM-RFE, on the other hand, is used as a practical feature selection technique to find the best variables by removing the feature vectors generated by SVM (Wang & Liu, 2015). The algorithm was constructed using the e1071 package and halve.above=100 and k=5 were used as parameter criteria. The hub genes were the intersecting genes obtained by the two machine learning algorithms.

Evaluation of immune cell infiltration

Analysis of immune cell infiltration relied on CIBERSORT, which is a widespread tool to research the proportion of 22 immune cells in specimens. Furthermore, we analysed the Spearman correlation between immune cells and key genes. The corrplot software package (Friendly, 2002) calculated the Pearson correlation coefficient between hub genes and each immune cell.

Construction of the ceRNA network

First, the miRanda, MicroRNADB, and TargetScan databases were used to simultaneously predict the target miRNAs of hub genes and determine the intersection. We obtained the miRNA-mRNA interaction pair file. Second, the spongeScan database (http://spongescan.rc.ufl.edu) was used to predict the lncRNA based on the resulting file from the previous step. Finally, the ceRNA network was formed, and the feature files were jointly built based on the lncRNA-miRNA-mRNA files. We visualised ceRNA regulatory networks using Cytoscape (Shannon et al., 2003).

Gene-drug target network analysis

The Drug-Gene Interaction Database (DGIDB, http://www.dgidb.org/), an online database that collects data on drug and gene interactions, was the source of material for building the network. The hub gene-drug interaction pair was compared with the DGIDB database to identify immunomodulatory drugs that may be related to IS. The gene-drug target network was constructed based on the above information.

Sample collection

Our samples were all obtained from patients in the Department of Neurology of the Second Affiliated Hospital of Harbin Medical University, approved by the Medical Ethics Committee of the Second Affiliated Hospital of Harbin Medical University (NO. KY2022-285) (Document S1), and all enrolled individuals signed an informed consent form (Document S2). There were five cases in the disease group and five cases in the healthy CTL. The disease group's (IS group) inclusion criteria: patients were suffering from their first acute IS confirmed by MRI, meeting the diagnostic criteria in the International Classification of Diseases (9th Revision). Exclusion criteria: patients with a history of hematologic disorders, type 1 diabetes, autoimmune, thyroid, neoplastic, renal, or liver diseases were excluded. We collected 5 ml fasting venous blood samples across both groups. Peripheral blood for the disease group was required to be collected within 24 hours of onset.

Real-time quantitative PCR (RT-qPCR)

RT-qPCR was used to detect the expression of six hub genes in the peripheral whole blood clinical samples. Total RNA was isolated from each sample using a Trizol reagent according to the manufacturer's instructions. Reverse transcription from RNA to cDNA was performed using HiScript Q RT SuperMix for qPCR (+gDNA wiper) (Vazyme Biotech, Nanjing, China). RT-qPCR was performed on cDNA samples using ChamQ SYBR Color qPCR Master Mix (2X) (Vazyme Biotech, Nanjing, China). Results were analyzed using the 2-ΔΔCt method and expressed as fold changes, with GAPDH selected as the internal reference gene. The PCR primers were designed by Majorbio (Shanghai, China) (Table S1). A *P*-value less than 0.05 was considered statistically significant.

Statistical analysis

The Shapiro-Wilk test was used to assess whether the continuous data were normally distributed in the CTL and IS groups. Normally distributed data were counted using the independent samples t-test. Non-normally distributed data were evaluated using the Wilcoxon-Mann Whitney test. A *P*-value < 0.05 was considered statistically significant.

Development and validation of a diagnostic model

Boxplots were used to show the expression of hub genes in GSE58294 as a validation dataset. We considered the difference statistically significant at a *P*-value < 0.05. A logistic regression algorithm was used to build a diagnostic model for IS classification based on crucial genes. The Proc R package was applied to generate the ROC curve and calculate the area under the ROC curve (AUC) (Stamova et al., 2014) to verify the accuracy of the genes and the model in dataset GSE16561. In addition, we calculated the AUC values of the key genes and logistic regression models in dataset GSE58294 to validate the diagnostic models' classification performance. To avoid the occurrence of overfitting, we chose to reduce the complexity of the model and reduce the parameters to make the model simpler. When training the model, the dataset was partitioned in a 7:3 ratio for training the model and predicting the accuracy of the created model, respectively, to balance the accuracy. Finally, the visual analysis of the above process was completed using ggplot2.

Results

Differential expression analysis

The volcano plot (Fig. 2A) showed that by using R package limma, we obtained 188 differential expression genes (Table S2), 85 genes of which were down-regulated in expression and 103 were up-regulated in expression. The top 50 differentially expressed genes are shown in a heatmap (Fig. 2B) created by the pheatmap package.

Function enrichment analysis

The GO analysis results (Table S3) were presented in a barplot (Figure 3A), a bubble plot (Figure 3B), and a circle plot (Figure 3C). GO-Biological Process (BP) shows that DEGs were primarily involved in immune-related biological processes, including immune response−regulating signaling pathway, positive regulation of cytokine production, and immune response−regulating cell surface receptor signaling pathway; GO- Molecular Function (MF) enrichment results showed DEGs were primarily associated with integrin binding oxidoreductase activity, acting on NAD(P)H quinone or a similar compound as acceptor and immunoglobulin binding; and GO-Cellular Component (CC) analysis was significantly enriched in platelet alpha granule and secretory granule lumen. KEGG pathways of each module (Figure 4 and Table S4) were mainly enriched in the MAPK signaling pathway, cell adhesion molecules, complement

and coagulation, and the cascades. The GO and KEGG pathway analysis results suggested that the immune system played a crucial role in IS.

We further conducted a GSEA analysis to avoid missing the functions of some insignificant but biologically significant genes, biological characteristics, and regulatory networks. GSEA results (Figure 5, Table S5) showed that the pathway involved axon guidance, complement and coagulation cascades, Fc gamma r mediated phagocytosis, focal adhesion, and regulation of actin cytoskeleton in the IS group (Figure 5A). Complement and coagulation cascades, the identified pathway of IS, were consistent with the result of KEGG, and the other pathways are related to immunity or cellular metabolism. Genes in the treatment group (IS group) were enriched at the top, which means that this gene set was upregulated. Conversely, genes in the CTL group (Figure 5B) were enriched at the bottom with a down-regulation trend.

Identification of the hub gene

To explore the biomarkers of IS, we performed feature screening through LASSO regression and the SVM-RFE algorithm. Twelve genes were identified as the most associated with IS by LASSO regression (Fig. 6A). The SVM-RFE algorithm evaluated 34 characteristic genes in IS (Fig. 6B). Both machine algorithms were subjected to 10-fold cross-validation to ensure the accuracy of the results. Six differential expression genes (ANTXR2, BAZ2B, C5AR1, PDK4, PPIH, and STK3) (Fig. 6C) were then identified as the hub genes by these two joint algorithms for subsequent research.

Results of immune cell infiltration analysis

The immune cell infiltration analysis results (Table S6) were presented by barplot (Figure 7A) and heatmap (Figure 7B). Barplot showed a difference in the percentage of immune cells between the IS group and CTL. Compared with the CTL, monocyte, macrophage M0, neutrophil, and mast cell activated infiltration increased, but T follicular helper cells and CD8 T cell infiltration decreased in the IS group. The relationship between immune cells and key genes is presented in Figure 8. ANTXR2, BAZ2B, C5AR1, PDK4, and STK3 were positively correlated with neutrophils and negatively correlated with T follicular helper and CD8 T cells, but the PPIH gene was positively associated with T follicular helper and CD8 T cells and negatively correlated with mast cell activation and neutrophils. Only PPIH was a down-regulated expression gene, while the other five genes (ANTXR2, BAZ2B, C5AR1, PDK4, and STK3) were all up-regulated. Meanwhile, the correlation between individual hub genes and immune cells is shown in a scatter plot (Figures S1-8). R-value indicates positive and negative correlation. P-value<0.05 is statistically significant. This evidence suggests that changes in the immune microenvironment of IS patients may be related to these six hub genes.

Construction of ceRNA networks for hub genes

Many studies have confirmed that ceRNA regulatory networks play a role in the biology and pathophysiology of various diseases. We also constructed networks (Figure 9, Table S7) to determine whether the six hub genes have similar regulatory relationships. The six hub genes were separately entered into the database, yielding a total of 306 miRNA-mRNA interaction pairs, of which 117 were ANTXR2-miRNA interaction pairs, 87 were BAZ2B-miRNA interaction pairs, and 285 were miRNA-lncRNA interaction pairs. We hypothesised that ANTXR2 and BAZ2B might be involved in the regulation of ceRNA. According to the value of a degree (Table S8), hsa-miR-766-3p, hsa-miR-149-3p, hsa-miR-1972, hsa-miR-186-5p, hsa-miR-1207-5p, MUC19, and LINC01002 may play a vital role in the ceRNA network.

Establishment of drug-hub genes regulatory networks

Drugs related to the six essential genes were screened out, and three genes (BAZ2B, C5AR1, and STK3) had target drugs. The gene-drug network is shown in Figure 10 and Table S9. We retrieved a total of 58 drugs acting on the Hub gene. BAZ2B is predicted to have the most targets of drug action, and a total of 30.25 drugs targeted STK3. Only one drug targeted C5AR1.

Validation of hub genes and diagnosis model

The RT-qPCR results showed that the expression of BAZ2B, C5AR1, PDK4, and STK3 was significantly higher in the IS group (Figure 11 and Table S10). A diagnosis model for IS classification was established using a logistic regression algorithm based on the four key genes. The results show that the diagnostic model could distinguish patients from normal samples. The ROC curves for the HUB genes and models are shown in Fig 12A and Fig12B. The AUC values are as follows: BAZ2B (AUC=0.892), C5AR1 (AUC=0.966), PDK4 (AUC=0.891), STK3 (AUC=0.966), and Model (AUC=0.999). Finally, we verified the accuracy of hub genes (Figure 12C) and the model (Figure 12D) using another dataset, GSE58294. AUC values for genes and models were as follows: BAZ2B (AUC=0.936), C5AR1 (AUC=0.734), PDK4 (AUC=0.697), STK3 (AUC=0.751), and Model (AUC=0.940). The boxplot (Figure 13) shows the differential expression of hub genes in the CTL and IS groups in the validation dataset. Our established model performed well in distinguishing between IS and normal samples.

Discussion

IS is a disease that threatens people's health worldwide, with high incidence, disability rate, and mortality. It is urgent to develop more convenient diagnosis and treatment programs in addition to the current methods in order to help solve this problem. Primary ischemic brain injury is combined with multiple mechanisms, including excitotoxicity, oxidative stress, apoptosis, and inflammation (Moskowitz et al., 2010). After that, a series of immune cascade reactions can further aggravate the damage to the brain tissue, such as the production of proinflammatory cytokines and activation of destructive serine proteases (Anrather & Iadecola, 2016; Xu et al., 2011). In our research, we explored possible biomarkers of IS, mechanisms of action, and potential drug targets in terms of inflammation and immunity.

In this study, 188 DEGs were observed through R package limma. By applying the joint LASSO and SVM algorithm, we finally obtained six genes (ANTXR2, BAZ2B, C5AR1, PDK4, PPIH, and STK3) most associated with IS. However, only BAZ2B, C5AR1, PDK4, and STK3 were upregulated expression in IS patients as verified by RT-PCR. C5AR1 has been proven to play a crucial role in regulating inflammatory and neurocognitive functions in IS, Alzheimer's disease, malaria, and neuropathic pain (Brandolini et al., 2019; McDonald et al., 2015; Moriconi et al., 2014). A marked increase in C5AR1 expression was observed in the MCAO- and OGD-induced models. Meanwhile, C5AR1 inhibitors have significant neuroprotective effects and notably inhibit neuro-inflammation and apoptosis in primary cortical neurons and MCAO-induced stroke models (Brandolini et al., 2019; Shi et al., 2017). TLR4 and C5AR1 promote apoptosis and inflammation by activating the cAMP/PKA/I-κB/NF-κB signaling pathway during brain ischemia-reperfusion (Kim & Jang, 2017; Shi et al., 2017; Zaal et al., 2017b). Previous studies showed that C5AR1 can inhibit human monocyte-derived dendritic cells and potentially exacerbate inflammatory responses (Zaal et al., 2017a; Zaal et al., 2017b).

There have been no studies on the correlation between IS and BAZ2B, PDK4, and STK3 genes. The biological function of BAZ2B remains unclear, besides its involvement in nucleosome remodelling (Oppikofer et al., 2017). However, it includes at least four functional domains that could encode or bind multiple proteins or DNA to perform various roles. Due to this characteristic, we speculate that BAZ2B may be associated with many diseases, including IS. However, previous studies were limited to suggesting that it is associated with neurodevelopment, its functional loss and haploinsufficiency are one of the causes of autism, and it may act through transcriptional regulation (Krupp et al., 2017; Scott et al., 2020). Another study found that BAZ2B activates M2 macrophages to participate in the inflammatory response (Xia et al., 2021), indicating its research value in immunity. PDK4 is a mitochondrial matrix enzyme essential in cellular energy regulation, which regulates the pyruvate dehydrogenase complex in the CNS and has important implications for neuron-glia metabolic interactions (Jha et al., 2012). Atherosclerosis is the result of cholesterol and lipid deposition in the arterial wall, usually associated with calcification, and is the most common cause of IS. Ma et al. (2020) found that PDK4 could promote vascular calcification by interfering with autophagic activity and metabolic reprogramming, contributing to the development of atherosclerosis. STK3 (MST2) is a component of the MAPK module and Hippo signaling pathway. In cardiovascular disease studies, STK3 was found to mediate mir155 to initiate inflammation and redox stress, leading to vascular smooth muscle cell proliferation and remodeling. It regulates the ERK1/2 signaling pathway by competing with MAP2K of the MAPK pathway, which in turn triggers inflammation and oxidative stress after vascular injury (Thiriet, 2018). In IS, the MAPK pathway also acts by regulating cytokines, inflammation, apoptosis, and death (Qin et al., 2022; Sun & Nan, 2016). STK3 is highly expressed in most cell types in the brain and may play a role in stroke by regulating the MAPK pathway. It has also been described as a substrate for CASP6 to intervene in neurodegeneration and apoptosis (Riechers et al., 2016; van Raam et al., 2013). Cho et al. (2021) found that STK3 increases phagocytosis of adipocytes, leading to obesity due to reduced

367 catabolic function of adipocytes. In obese human patients, STK3 expression levels were
368 elevated. STK3 inhibitors improved metabolic patterns in obese mouse models, suggesting that
369 there may be a viable pathway to investigate and develop drugs targeting STK3 to treat obesity-
370 related diseases including stroke. All of the above suggest that STK3 may be relevant to the
371 onset of IS, but the exact mechanism needs to be further investigated.

372 The critical pathways enriched by GO and KEGG analysis were related to immunity, including
373 the immune response−regulating, immune response−regulating cell surface receptor, and MAPK
374 signaling pathways. Notably, C5AR1 and STK3 were enriched in the positive regulatory
375 pathway of the MAPK cascade. Previous studies have found that the MAPK signaling pathway
376 in inflammation and BBB dysfunction MAPK is comprised of three main effectors: ERK1/2,
377 JNK, and p38 (Qin et al., 2022; Sun & Nan, 2016). C5AR1 affects inflammation in bone by
378 activating the MAPK pathway and regulating gene expression in pathways associated with
379 insulin and transforming growth factor-β (Modinger et al., 2018). STK3, as described above,
380 mediates cardiovascular injury through the MAPK pathway and also promotes apoptosis by
381 inducing the activation of JNK (Chen et al., 2018). These results are consistent with our
382 bioinformatics analysis and suggest that the MAPK signaling pathway plays an important role in
383 these key gene-mediated biological processes. Additionally, the results of the GSEA enrichment
384 analysis showed that C5AR1 was involved in the complement and coagulation cascades
385 pathway, which has been proven to correlate with IS (Berkowitz et al., 2021). The functions of
386 PDK4 are all focused on the involvement of lipid metabolism in the process of atherosclerosis.
387 These results are also in accordance with the role of C5AR1 and PDK4 found in some studies
388 (Berkowitz et al., 2021; Ma et al., 2020).

389 Many stroke models have confirmed that neutrophils in the acute phase of IS are among the
390 earliest immune cells recruited into the brain tissue (Gokhan et al. 2013; Jickling et al., 2015;
391 Kaito et al., 2013). Activated neutrophils may contribute to the development of IS by stimulating
392 the systemic inflammatory response and destroying the BBB. Van Duijn et al. (2018) found that
393 CD8 T cells can exacerbate the inflammatory response by secreting various inflammatory
394 cytokines, resulting in increased instability of atherosclerotic plaques. But CD8 T cells also have
395 an anti-atherosclerotic effect, which is mediated by stimulation of inhibitory receptor production
396 and cytolytic killing of antigen-presenting cells. The results of immunological infiltration
397 analysis by CIBERSORT revealed that monocyte, macrophage M0, neutrophil, and mast cell
398 activation increased, while T follicular helper and CD8 cells decreased infiltration in the IS
399 group, which was consistent with previous studies (Zheng et al., 2022). Our research also
400 showed that neutrophils were significantly and positively correlated with genes ANTXR2,
401 BAZ2B, C5AR1, PDK4, and STK3, negatively correlated with PPIH. CD8 was significantly and
402 negatively correlated with genes ANTXR2, BAZ2B, C5AR1, PDK4, and STK3, and positively
403 correlated with PPIH. This evidence further suggests that neutrophils play a significant role, and
404 hub genes are involved in regulating the immune microenvironment in IS patients.

The ceRNA network we conducted included nodes (six mRNAs, 249 miRNAs, and 218 lncRNAs) with 591 edges. In these nodes, the miRNAs with high degrees, hsa-miR-766-3p, hsa-miR-149-3p, and hsa-miR-186-5p, have been reported as essential molecules in cerebral ischemic/reperfusion injury (Cai et al., 2019; Hu et al., 2020; You et al., 2022). Of these miRNAs, hsa-miR-766-3p is known to be involved in immune responses or inflammation in various diseases. It contributes to anti-inflammatory responses through the indirect inhibition of NF-κB signaling in rheumatoid arthritis (RA) (Hayakawa et al., 2019). The miR-766-3p/NR3C2 axis is participating in the protection against cerebral ischemia and reperfusion (Cai et al., 2019). LINC00689, one of the highest degree lncRnas in the ceRNA network, is a key regulator of the toll-like receptor (TLR) signaling pathway controlling innate immunity (Liu et al., 2020). However, the mode of regulation between them and hub genes still needs further validation. Meanwhile, we constructed a hub gene-based gene-drug regulatory network and predicted possible targeted drugs. The result reflected that BAZ2B had the most targets of drug action, likely due to its multiple functional areas. The network provides a theoretical basis for developing targeted immunotherapies for IS.

Finally, based on the four key genes mentioned above, we constructed an IS diagnostic model using logistic regression methods. We also verified the validity of the model in a publicly available dataset. Bioinformatics analysis combined with machine learning methods achieved good results.

However, there are some limitations to our study. First, although we analyzed clinical data and performed RT-qPCR validation, the sample size was relatively limited. Second, more in vivo and in vitro studies are needed to identify and validate the underlying mechanisms between these genes and stroke. These will be the most critical aspects of our future research to gain insight into the immunological pathogenesis of IS and to provide additional options for clinical treatment.

Conclusion

We explored six hub genes (ANTXR2, BAZ2B, C5AR1, PDK4, PPIH, and STK3) for IS and revealed the immune cell infiltration pattern via bioinformatics analysis and machine learning algorithm. We successfully constructed the ceRNA and drug-gene networks, which provide new ideas to explore the possible molecular mechanisms and develop targeted drugs of IS. Finally, BAZ2B, C5AR1, PDK4, and STK3 were verified to be the biological markers of IS. Based on the four genes, we developed a diagnostic model and validated its effectiveness, providing additional tools for screening and diagnosing IS.

Acknowledgements

We want to thank the GEO database for the data support.

References

Anrather J, and Iadecola C. 2016. Inflammation and Stroke: An Overview. *Neurotherapeutics* 13:661-670. 10.1007/s13311-016-0483-x

Bai Y, Nie S, Jiang G, Zhou Y, Zhou M, Zhao Y, Li S, Wang F, Lv Q, Huang Y, Yang Q, Li Q, Li Y, Xia Y, Liu Y, Liu J, Qian J, Li B, Wu G, Wu Y, Wang B, Cheng X, Yang Y, Ke T, Li H, Ren X, Ma X, Liao Y, Xu C, Tu X, and Wang QK. 2014. Regulation of CARD8 expression by ANRIL and association of CARD8 single nucleotide polymorphism rs2043211 (p.C10X) with ischemic stroke. *Stroke* 45:383-388. 10.1161/STROKEAHA.113.003393

Barr TL, Conley Y, Ding J, Dillman A, Warach S, Singleton A, and Matarin M. 2010. Genomic biomarkers and cellular pathways of ischemic stroke by RNA gene expression profiling. *Neurology* 75:1009-1014. 10.1212/WNL.0b013e3181f2b37f

Berkowitz S, Chapman J, Dori A, Gofrit SG, Maggio N, and Shavit-Stein E. 2021. Complement and Coagulation System Crosstalk in Synaptic and Neural Conduction in the Central and Peripheral Nervous Systems. *Biomedicines* 9. 10.3390/biomedicines9121950

Boldin MP, Taganov KD, Rao DS, Yang L, Zhao JL, Kalwani M, Garcia-Flores Y, Luong M, Devrekanli A, Xu J, Sun G, Tay J, Linsley PS, and Baltimore D. 2011. miR-146a is a significant brake on autoimmunity, myeloproliferation, and cancer in mice. *J Exp Med* 208:1189-1201. 10.1084/jem.20101823

Brandolini L, Grannonico M, Bianchini G, Colanardi A, Sebastiani P, Paladini A, Piroli A, Allegretti M, Varrassi G, and Di Loreto S. 2019. The Novel C5aR Antagonist DF3016A Protects Neurons Against Ischemic Neuroinflammatory Injury. *Neurotox Res* 36:163-174. 10.1007/s12640-019-00026-w

Cai J, Shangguan S, Li G, Cai Y, Chen Y, Ma G, Miao Z, Liu L, and Deng Y. 2019. Knockdown of lncRNA Gm11974 protect against cerebral ischemic reperfusion through miR-766-3p/NR3C2 axis. *Artif Cells Nanomed Biotechnol* 47:3847-3853. 10.1080/21691401.2019.1666859

Campbell BCV, and Khatri P. 2020. Stroke. *Lancet* 396:129-142. 10.1016/S0140-6736(20)31179-X

Cao DW, Liu MM, Duan R, Tao YF, Zhou JS, Fang WR, Zhu JR, Niu L, and Sun JG. 2020. The lncRNA Malat1 functions as a ceRNA to contribute to berberine-mediated inhibition of HMGB1 by sponging miR-181c-5p in poststroke inflammation. *Acta Pharmacol Sin* 41:22-33. 10.1038/s41401-019-0284-y

Cassella CR, and Jagoda A. 2017. Ischemic Stroke: Advances in Diagnosis and Management. *Emerg Med Clin North Am* 35:911-930. 10.1016/j.emc.2017.07.007

Cekanaviciute E, Fathali N, Doyle KP, Williams AM, Han J, and Buckwalter MS. 2014. Astrocytic transforming growth factor-beta signaling reduces subacute neuroinflammation after stroke in mice. *Glia* 62:1227-1240. 10.1002/glia.22675

Chen S, Fang Y, Xu S, Reis C, and Zhang J. 2018. Mammalian Sterile20-like Kinases: Signalings and Roles in Central Nervous System. *Aging Dis* 9:537-552. 10.14336/AD.2017.0702

Cho YK, Son Y, Saha A, Kim D, Choi C, Kim M, Park JH, Im H, Han J, Kim K, Jung YS, Yun J, Bae EJ, Seong JK, Lee MO, Lee S, Granneman JG, and Lee YH. 2021. STK3/STK4 signalling in adipocytes regulates mitophagy and energy expenditure. *Nat Metab* 3:428-441. 10.1038/s42255-021-00362-2

Dang S, Malik A, Chen J, Qu J, Yin K, Cui L, and Gu M. 2020. LncRNA SNHG15 Contributes to Immuno-Escape of Gastric Cancer Through Targeting miR141/PD-L1. *Onco Targets Ther* 13:8547-8556. 10.2147/OTT.S251625

Doll DN, Barr TL, and Simpkins JW. 2014. Cytokines: their role in stroke and potential use as biomarkers and therapeutic targets. *Aging Dis* 5:294-306. 10.14336/AD.2014.0500294

491 Engebretsen S, and Bohlin J. 2019. Statistical predictions with glmnet. *Clin Epigenetics* 11:123.
492 10.1186/s13148-019-0730-1
493 Feng L, Guo J, and Ai F. 2019. Circulating long noncoding RNA ANRIL downregulation
494 correlates with increased risk, higher disease severity and elevated pro-inflammatory
495 cytokines in patients with acute ischemic stroke. *J Clin Lab Anal* 33:e22629.
496 10.1002/jcla.22629
497 Friedman J, Hastie T, and Tibshirani R. 2010. Regularization Paths for Generalized Linear
498 Models via Coordinate Descent. *J Stat Softw* 33:1-22.
499 Friendly M. 2002. Corrgrams. *The American Statistician* 56:316-324. 10.1198/000313002533
500 Fu J, Yu Q, Xiao J, and Li S. 2021. Long noncoding RNA as a biomarker for the prognosis of
501 ischemic stroke: A protocol for meta-analysis and bioinformatics analysis. *Medicine*
502 *(Baltimore)* 100:e25596. 10.1097/MD.0000000000025596
503 Gokhan S, Ozhasenekler A, Mansur Durgun H, Akil E, Ustundag M, and Orak M. 2013.
504 Neutrophil lymphocyte ratios in stroke subtypes and transient ischemic attack. *Eur Rev*
505 *Med Pharmacol Sci* 17:653-657.
506 Harston GW, Rane N, Shaya G, Thandeswaran S, Cellerini M, Sheerin F, and Kennedy J. 2015.
507 Imaging biomarkers in acute ischemic stroke trials: a systematic review. *AJNR Am J*
508 *Neuroradiol* 36:839-843. 10.3174/ajnr.A4208
509 Hasan TF, Rabinstein AA, Middlebrooks EH, Haranhalli N, Silliman SL, Meschia JF, and Tawk
510 RG. 2018. Diagnosis and Management of Acute Ischemic Stroke. *Mayo Clin Proc*
511 93:523-538. 10.1016/j.mayocp.2018.02.013
512 Hayakawa K, Kawasaki M, Hirai T, Yoshida Y, Tsushima H, Fujishiro M, Ikeda K, Morimoto S,
513 Takamori K, and Sekigawa I. 2019. MicroRNA-766-3p Contributes to Anti-Inflammatory
514 Responses through the Indirect Inhibition of NF-kappaB Signaling. *Int J Mol Sci* 20.
515 10.3390/ijms20040809
516 Hu X, Xiang Z, Zhang W, Yu Z, Xin X, Zhang R, Deng Y, and Yuan Q. 2020. Protective effect of
517 DLX6-AS1 silencing against cerebral ischemia/reperfusion induced impairments. *Aging*
518 *(Albany NY)* 12:23096-23113. 10.18632/aging.104070
519 Huang L, Ma Q, Li Y, Li B, and Zhang L. 2018. Inhibition of microRNA-210 suppresses pro-
520 inflammatory response and reduces acute brain injury of ischemic stroke in mice. *Exp*
521 *Neurol* 300:41-50. 10.1016/j.expneurol.2017.10.024
522 Iadecola C, Buckwalter MS, and Anrather J. 2020. Immune responses to stroke: mechanisms,
523 modulation, and therapeutic potential. *J Clin Invest* 130:2777-2788. 10.1172/JCI135530
524 Javidi E, and Magnus T. 2019. Autoimmunity After Ischemic Stroke and Brain Injury. *Front*
525 *Immunol* 10:686. 10.3389/fimmu.2019.00686
526 Jayaraj RL, Azimullah S, Beiram R, Jalal FY, and Rosenberg GA. 2019. Neuroinflammation:
527 friend and foe for ischemic stroke. *J Neuroinflammation* 16:142. 10.1186/s12974-019-
528 1516-2
529 Jha MK, Jeon S, and Suk K. 2012. Pyruvate Dehydrogenase Kinases in the Nervous System:
530 Their Principal Functions in Neuronal-glial Metabolic Interaction and Neuro-metabolic
531 Disorders. *Curr Neuropharmacol* 10:393-403. 10.2174/157015912804143586
532 Jickling GC, Liu D, Ander BP, Stamova B, Zhan X, and Sharp FR. 2015. Targeting neutrophils
533 in ischemic stroke: translational insights from experimental studies. *J Cereb Blood Flow*
534 *Metab* 35:888-901. 10.1038/jcbfm.2015.45
535 Kaito M, Araya S, Gondo Y, Fujita M, Minato N, Nakanishi M, and Matsui M. 2013. Relevance of
536 distinct monocyte subsets to clinical course of ischemic stroke patients. *PLoS One*
537 8:e69409. 10.1371/journal.pone.0069409
538 Karreth FA, Reschke M, Ruocco A, Ng C, Chapuy B, Leopold V, Sjoberg M, Keane TM, Verma
539 A, Ala U, Tay Y, Wu D, Seitzer N, Velasco-Herrera Mdel C, Bothmer A, Fung J,
540 Langellotto F, Rodig SJ, Elemento O, Shipp MA, Adams DJ, Chiarle R, and Pandolfi PP.

2015. The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. *Cell* 161:319-332. 10.1016/j.cell.2015.02.043

Khambhati J, Engels M, Allard-Ratick M, Sandesara PB, Quyyumi AA, and Sperling L. 2018. Immunotherapy for the prevention of atherosclerotic cardiovascular disease: Promise and possibilities. *Atherosclerosis* 276:1-9. 10.1016/j.atherosclerosis.2018.07.007

Kim SH, and Jang YS. 2017. Yersinia enterocolitica Exploits Signal Crosstalk between Complement 5a Receptor and Toll-like Receptor 1/2 and 4 to Avoid the Bacterial Clearance in M cells. *Immune Netw* 17:228-236. 10.4110/in.2017.17.4.228

Kolde R. 2012. Pheatmap: pretty heatmaps. *R package version* 1:726.

Krupp DR, Barnard RA, Duffourd Y, Evans SA, Mulqueen RM, Bernier R, Riviere JB, Fombonne E, and O'Roak BJ. 2017. Exonic Mosaic Mutations Contribute Risk for Autism Spectrum Disorder. *Am J Hum Genet* 101:369-390. 10.1016/j.ajhg.2017.07.016

Levard D, Buendia I, Lanquetin A, Glavan M, Vivien D, and Rubio M. 2021. Filling the gaps on stroke research: Focus on inflammation and immunity. *Brain Behav Immun* 91:649-667. 10.1016/j.bbi.2020.09.025

Li P, Duan S, and Fu A. 2020. Long noncoding RNA NEAT1 correlates with higher disease risk, worse disease condition, decreased miR-124 and miR-125a and predicts poor recurrence-free survival of acute ischemic stroke. *J Clin Lab Anal* 34:e23056. 10.1002/jcla.23056

Liu J, Wang Y, Chu Y, Xu R, Zhang D, and Wang X. 2020. Identification of a TLR-Induced Four-lncRNA Signature as a Novel Prognostic Biomarker in Esophageal Carcinoma. *Front Cell Dev Biol* 8:649. 10.3389/fcell.2020.00649

Liu Z, Li H, and Pan S. 2021. Discovery and Validation of Key Biomarkers Based on Immune Infiltrates in Alzheimer's Disease. *Front Genet* 12:658323. 10.3389/fgene.2021.658323

Ma WQ, Sun XJ, Zhu Y, and Liu NF. 2020. PDK4 promotes vascular calcification by interfering with autophagic activity and metabolic reprogramming. *Cell Death Dis* 11:991. 10.1038/s41419-020-03162-w

Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese AK, van der Laan SW, Gretarsdottir S, Anderson CD, Chong M, Adams HHH, Ago T, Almgren P, Amouyel P, Ay H, Bartz TM, Benavente OR, Bevan S, Boncoraglio GB, Brown RD, Jr., Butterworth AS, Carrera C, Carty CL, Chasman DI, Chen WM, Cole JW, Correa A, Cotlarciuc I, Cruchaga C, Danesh J, de Bakker PIW, DeStefano AL, den Hoed M, Duan Q, Engelter ST, Falcone GJ, Gottesman RF, Grewal RP, Gudnason V, Gustafsson S, Haessler J, Harris TB, Hassan A, Havulinna AS, Heckbert SR, Holliday EG, Howard G, Hsu FC, Hyacinth HI, Ikram MA, Ingelsson E, Irvin MR, Jian X, Jimenez-Conde J, Johnson JA, Jukema JW, Kanai M, Keene KL, Kissela BM, Kleindorfer DO, Kooperberg C, Kubo M, Lange LA, Langefeld CD, Langenberg C, Launer LJ, Lee JM, Lemmens R, Leys D, Lewis CM, Lin WY, Lindgren AG, Lorentzen E, Magnusson PK, Maguire J, Manichaikul A, McArdle PF, Meschia JF, Mitchell BD, Mosley TH, Nalls MA, Ninomiya T, O'Donnell MJ, Psaty BM, Pulit SL, Rannikmae K, Reiner AP, Rexrode KM, Rice K, Rich SS, Ridker PM, Rost NS, Rothwell PM, Rotter JI, Rundek T, Sacco RL, Sakaue S, Sale MM, Salomaa V, Sapkota BR, Schmidt R, Schmidt CO, Schminke U, Sharma P, Slowik A, Sudlow CLM, Tanislav C, Tatlisumak T, Taylor KD, Thijs VNS, Thorleifsson G, Thorsteinsdottir U, Tiedt S, Trompet S, Tzourio C, van Duijn CM, Walters M, Wareham NJ, Wassertheil-Smoller S, Wilson JG, Wiggins KL, Yang Q, Yusuf S, Consortium AF, Cohorts for H, Aging Research in Genomic Epidemiology C, International Genomics of Blood Pressure C, Consortium I, Starnet, Bis JC, Pastinen T, Ruusalepp A, Schadt EE, Koplev S, Bjorkegren JLM, Codoni V, Civelek M, Smith NL, Tregouet DA, Christophersen IE, Roselli C, Lubitz SA, Ellinor PT, Tai ES, Kooner JS, Kato N, He J, van der Harst P, Elliott P, Chambers JC, Takeuchi F, Johnson AD, BioBank Japan Cooperative Hospital G, Consortium C, Consortium E-C, Consortium

EP-I, International Stroke Genetics C, Consortium M, Neurology Working Group of the CC, Network NSG, Study UKYLD, Consortium M, Sanghera DK, Melander O, Jern C, Strbian D, Fernandez-Cadenas I, Longstreth WT, Jr., Rolfs A, Hata J, Woo D, Rosand J, Pare G, Hopewell JC, Saleheen D, Stefansson K, Worrall BB, Kittner SJ, Seshadri S, Fornage M, Markus HS, Howson JMM, Kamatani Y, Debette S, and Dichgans M. 2018. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 50:524-537. 10.1038/s41588-018-0058-3

McDonald CR, Cahill LS, Ho KT, Yang J, Kim H, Silver KL, Ward PA, Mount HT, Liles WC, Sled JG, and Kain KC. 2015. Experimental Malaria in Pregnancy Induces Neurocognitive Injury in Uninfected Offspring via a C5a-C5a Receptor Dependent Pathway. *PLoS Pathog* 11:e1005140. 10.1371/journal.ppat.1005140

Modinger Y, Rapp A, Pazmandi J, Vikman A, Holzmann K, Haffner-Luntzer M, Huber-Lang M, and Ignatius A. 2018. C5aR1 interacts with TLR2 in osteoblasts and stimulates the osteoclast-inducing chemokine CXCL10. *J Cell Mol Med* 22:6002-6014. 10.1111/jcmm.13873

Moriconi A, Cunha TM, Souza GR, Lopes AH, Cunha FQ, Carneiro VL, Pinto LG, Brandolini L, Aramini A, Bizzarri C, Bianchini G, Beccari AR, Fanton M, Bruno A, Costantino G, Bertini R, Galliera E, Locati M, Ferreira SH, Teixeira MM, and Allegretti M. 2014. Targeting the minor pocket of C5aR for the rational design of an oral allosteric inhibitor for inflammatory and neuropathic pain relief. *Proc Natl Acad Sci U S A* 111:16937-16942. 10.1073/pnas.1417365111

Moskowitz MA, Lo EH, and Iadecola C. 2010. The science of stroke: mechanisms in search of treatments. *Neuron* 67:181-198. 10.1016/j.neuron.2010.07.002

Oppikofer M, Bai T, Gan Y, Haley B, Liu P, Sandoval W, Ciferri C, and Cochran AG. 2017. Expansion of the ISWI chromatin remodeler family with new active complexes. *EMBO Rep* 18:1697-1706. 10.15252/embr.201744011

Qin C, Yang S, Chu YH, Zhang H, Pang XW, Chen L, Zhou LQ, Chen M, Tian DS, and Wang W. 2022. Signaling pathways involved in ischemic stroke: molecular mechanisms and therapeutic interventions. *Signal Transduct Target Ther* 7:215. 10.1038/s41392-022-01064-1

Riechers SP, Butland S, Deng Y, Skotte N, Ehrnhoefer DE, Russ J, Laine J, Laroche M, Pouladi MA, Wanker EE, Hayden MR, and Graham RK. 2016. Interactome network analysis identifies multiple caspase-6 interactors involved in the pathogenesis of HD. *Hum Mol Genet* 25:1600-1618. 10.1093/hmg/ddw036

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47. 10.1093/nar/gkv007

Salmena L, Poliseno L, Tay Y, Kats L, and Pandolfi PP. 2011. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146:353-358. 10.1016/j.cell.2011.07.014

Scott TM, Guo H, Eichler EE, Rosenfeld JA, Pang K, Liu Z, Lalani S, Bi W, Yang Y, Bacino CA, Streff H, Lewis AM, Koenig MK, Thiffault I, Bellomo A, Everman DB, Jones JR, Stevenson RE, Bernier R, Gilissen C, Pfundt R, Hiatt SM, Cooper GM, Holder JL, and Scott DA. 2020. BAZ2B haploinsufficiency as a cause of developmental delay, intellectual disability, and autism spectrum disorder. *Hum Mutat* 41:921-925. 10.1002/humu.23992

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498-2504. 10.1101/gr.1239303

Shi YW, Zhang XC, Chen C, Tang M, Wang ZW, Liang XM, Ding F, and Wang CP. 2017. Schisantherin A attenuates ischemia/reperfusion-induced neuronal injury in rats via

regulation of TLR4 and C5aR1 signaling pathways. *Brain Behav Immun* 66:244-256. 10.1016/j.bbi.2017.07.004

Smith CJ, Hulme S, Vail A, Heal C, Parry-Jones AR, Scarth S, Hopkins K, Hoadley M, Allan SM, Rothwell NJ, Hopkins SJ, and Tyrrell PJ. 2018. SCIL-STROKE (Subcutaneous Interleukin-1 Receptor Antagonist in Ischemic Stroke): A Randomized Controlled Phase 2 Trial. *Stroke* 49:1210-1216. 10.1161/STROKEAHA.118.020750

Stamova B, Jickling GC, Ander BP, Zhan X, Liu D, Turner R, Ho C, Khoury JC, Bushnell C, Pancioli A, Jauch EC, Broderick JP, and Sharp FR. 2014. Gene expression in peripheral immune cells following cardioembolic stroke is sexually dimorphic. *PLoS One* 9:e102550. 10.1371/journal.pone.0102550

Sun J, and Nan G. 2016. The Mitogen-Activated Protein Kinase (MAPK) Signaling Pathway as a Discovery Target in Stroke. *J Mol Neurosci* 59:90-98. 10.1007/s12031-016-0717-8

Taylor RA, Chang CF, Goods BA, Hammond MD, Mac Grory B, Ai Y, Steinschneider AF, Renfroe SC, Askenase MH, McCullough LD, Kasner SE, Mullen MT, Hafler DA, Love JC, and Sansing LH. 2017. TGF-beta1 modulates microglial phenotype and promotes recovery after intracerebral hemorrhage. *J Clin Invest* 127:280-292. 10.1172/JCI88647

Thiriet M. 2018. Cardiovascular Risk Factors and Markers. *Vasculopathies*, 91-198.

Torres-Aguila NP, Carrera C, Muino E, Cullell N, Carcel-Marquez J, Gallego-Fabrega C, Gonzalez-Sanchez J, Bustamante A, Delgado P, Ibanez L, Heitsch L, Krupinski J, Montaner J, Marti-Fabregas J, Cruchaga C, Lee JM, Fernandez-Cadenas I, and Acute Endophenotypes Group of the International Stroke Genetics C. 2019. Clinical Variables and Genetic Risk Factors Associated with the Acute Outcome of Ischemic Stroke: A Systematic Review. *J Stroke* 21:276-289. 10.5853/jos.2019.01522

van Duijn J, Kuiper J, and Slutter B. 2018. The many faces of CD8+ T cells in atherosclerosis. *Curr Opin Lipidol* 29:411-416. 10.1097/MOL.0000000000000541

van Raam BJ, Ehrnhoefer DE, Hayden MR, and Salvesen GS. 2013. Intrinsic cleavage of receptor-interacting protein kinase-1 by caspase-6. *Cell Death Differ* 20:86-96. 10.1038/cdd.2012.98

Wang M, Chen W, Geng Y, Xu C, Tao X, and Zhang Y. 2020. Long Non-Coding RNA MEG3 Promotes Apoptosis of Vascular Cells and is Associated with Poor Prognosis in Ischemic Stroke. *J Atheroscler Thromb* 27:718-726. 10.5551/jat.50674

Wang Q, and Liu X. 2015. Screening of feature genes in distinguishing different types of breast cancer using support vector machine. *Onco Targets Ther* 8:2311-2317. 10.2147/OTT.S85271

Wickham H. 2009. ggplot2: Elegant graphics for data analysis (Use R) Springer-Verlag. *New York*.

Wu F, Liu Z, Zhou L, Ye D, Zhu Y, Huang K, Weng Y, Xiong X, Zhan R, and Shen J. 2022. Systemic immune responses after ischemic stroke: From the center to the periphery. *Front Immunol* 13:911661. 10.3389/fimmu.2022.911661

Xia L, Wang X, Liu L, Fu J, Xiao W, Liang Q, Han X, Huang S, Sun L, Gao Y, Zhang C, Yang L, Wang L, Qian L, and Zhou Y. 2021. lnc-BAZ2B promotes M2 macrophage activation and inflammation in children with asthma through stabilizing BAZ2B pre-mRNA. *J Allergy Clin Immunol* 147:921-932 e929. 10.1016/j.jaci.2020.06.034

Xu J, Zhang X, Monestier M, Esmon NL, and Esmon CT. 2011. Extracellular histones are mediators of death through TLR2 and TLR4 in mouse fatal liver injury. *J Immunol* 187:2626-2631. 10.4049/jimmunol.1003930

Yan S, Wang P, Wang J, Yang J, Lu H, Jin C, Cheng M, and Xu D. 2019. Long Non-coding RNA HIX003209 Promotes Inflammation by Sponging miR-6089 via TLR4/NF-kappaB Signaling Pathway in Rheumatoid Arthritis. *Front Immunol* 10:2218. 10.3389/fimmu.2019.02218

693 You J, Qian F, Huang Y, Guo Y, Lv Y, Yang Y, Lu X, Guo T, Wang J, and Gu B. 2022. lncRNA
694 WT1-AS attenuates hypoxia/ischemia-induced neuronal injury during cerebral ischemic
695 stroke via miR-186-5p/XIAP axis. *Open Med (Wars)* 17:1338-1349. 10.1515/med-2022-
696 0528
697 Yu G, Wang LG, Han Y, and He QY. 2012. clusterProfiler: an R package for comparing
698 biological themes among gene clusters. *OMICS* 16:284-287. 10.1089/omi.2011.0118
699 Zaal A, Dieker M, Oudenampsen M, Turksma AW, Lissenberg-Thunnissen SN, Wouters D, van
700 Ham SM, and Ten Brinke A. 2017a. Anaphylatoxin C5a Regulates 6-Sulfo-LacNAc
701 Dendritic Cell Function in Human through Crosstalk with Toll-Like Receptor-Induced
702 CREB Signaling. *Front Immunol* 8:818. 10.3389/fimmu.2017.00818
703 Zaal A, Nota B, Moore KS, Dieker M, van Ham SM, and Ten Brinke A. 2017b. TLR4 and C5aR
704 crosstalk in dendritic cells induces a core regulatory network of RSK2, PI3Kbeta, SGK1,
705 and FOXO transcription factors. *J Leukoc Biol* 102:1035-1054. 10.1189/jlb.2MA0217-
706 058R
707 Zameer S, Siddiqui AS, and Riaz R. 2021. Multimodality Imaging in Acute Ischemic Stroke. *Curr*
708 *Med Imaging* 17:567-577. 10.2174/1573405616666201130094948
709 Zheng PF, Chen LZ, Liu P, Pan HW, Fan WJ, and Liu ZY. 2022. Identification of immune-
710 related key genes in the peripheral blood of ischaemic stroke patients using a weighted
711 gene coexpression network analysis and machine learning. *J Transl Med* 20:361.
712 10.1186/s12967-022-03562-w
713 Zhong Y, Yu C, and Qin W. 2019. LncRNA SNHG14 promotes inflammatory response induced
714 by cerebral ischemia/reperfusion injury through regulating miR-136-5p /ROCK1. *Cancer*
715 *Gene Ther* 26:234-247. 10.1038/s41417-018-0067-5
716

Table 1(on next page)

Basic information ongene expression profiling

Table 1 Basic information on gene expression profiling

1

Location	Dataset ID	Platform	Type	Number
Whole blood samples	GSE16561	GPL6883	Microarray	39IS VS 24 CTL
Whole blood samples	GSE58294	GPL570	Microarray	69IS VS 23CTL

2

Figure 1

The technical route

Abbreviations: IS, ischemic stroke; CTL, healthy control

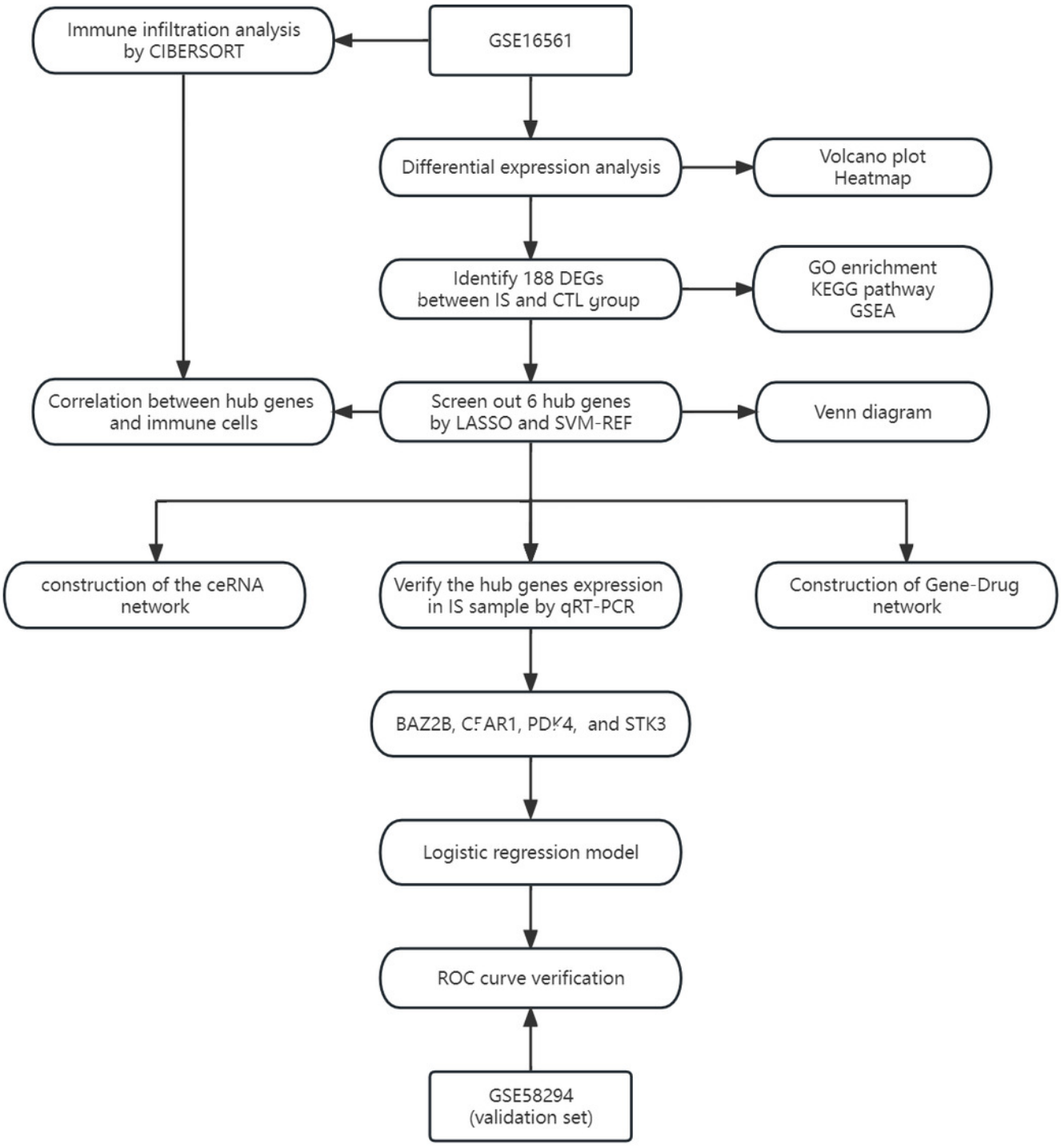
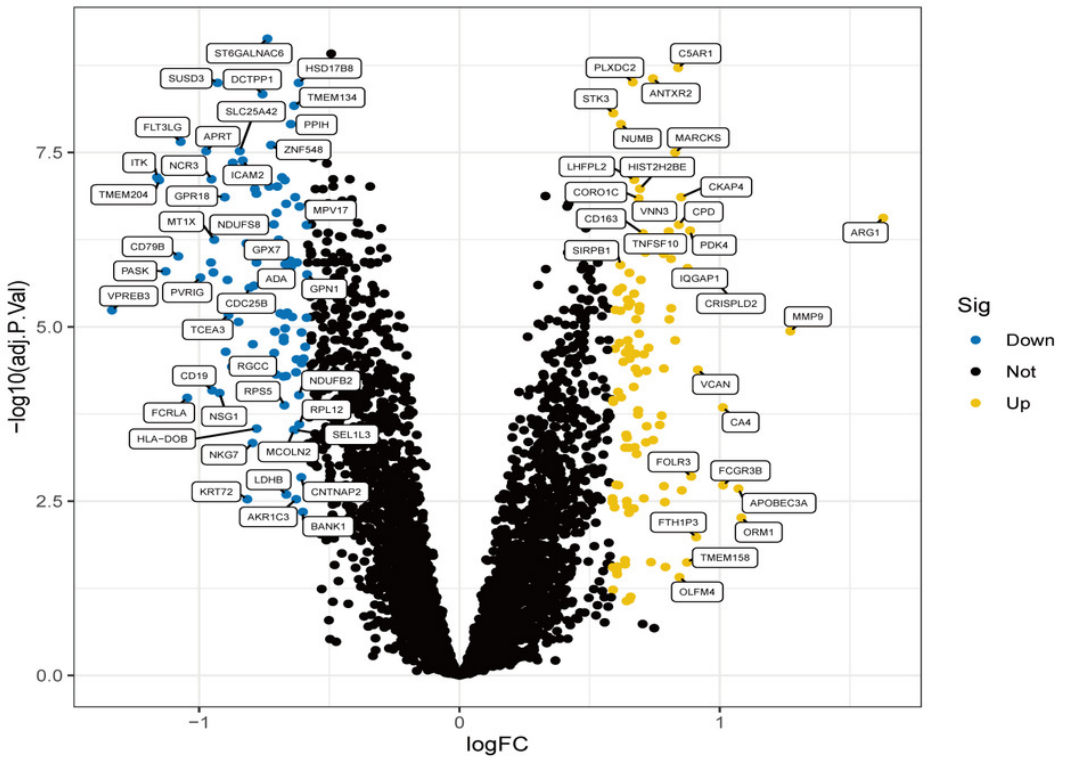

Immune infiltration analysis by CIBERSORT
GSE16561
Differential expression analysis
Volcano plot
Heatmap
Identify 188 DEGs between IS and CTL group
GO enrichment
KEGG pathway
GSEA
Correlation between hub genes and immune cells
Screen out 6 hub genes by LASSO and SVM-REF
Venn diagram
construction of the ceRNA network
Verify the hub genes expression in IS sample by qRT-PCR
Construction of Gene-Drug network
BAZ2B, CFAR1, PDK4, and STK3
Logistic regression model
ROC curve verification
GSE58294
(validation set)

Figure 2

Differential expression analysis

(A)The differential expression analysis results are shown in the volcano plot. The x-axis represents log1.5 (fold change), and the y-axis represents -log10 (adjust p. value). Green dots represent downregulated genes, red dots represent upregulated genes, and black dots represent genes with no evident differential expression. (B) Heatmap of DEGs. Each column in the graph represents a sample, each row represents a gene, and the expression status of the gene is indicated from high to low in orange to blue, respectively, and at the top of the heatmap, red/blue represents the IS group/CTL group, respectively. Abbreviations: ischemic stroke (IS), healthy control (CTL), differentially expressed genes (DEGs)

A


ST6GALNAC6
SUSD3
DCTPP1
HSD17B8
TMEM134
SLC25A42
PPIH
FLT3LG
APRT
ZNF548
ITK
NCR3
ICAM2
TMEM204
GPR18
MPV17
MT1X
NDUFS8
CD79B
GPX7
PASK
ADA
VPREB3
PVRIG
GPN1
CDC25B
TCEA3
RGCC
CD19
NDUFB2
RPS5
FCRLA
NSG1
RPL12
HLA-DOB
SEL1L3
NKG7
MCOLN2
LDHB
KRT72
CNTNAP2
AKR1C3
BANK1
C5AR1
PLXDC2
ANTXR2
STK3
NUMB
MARCKS
LHFPL2
HIST2H2BE
CKAP4
CORO1C
CD163
VNN3
CPD
TNFSF10
PDK4
SIRPB1
IQGAP1
CRISPLD2
MMP9
ARG1
VCAN
CA4
FOLR3
FCGR3B
APOBEC3A
FTH1P3
ORM1
TMEM158
OLFM4
−log10(adj.P.Val)
7.5
5.0
2.5
0.0
−1
0
1
logFC
Sig
Down
Not
Up

B

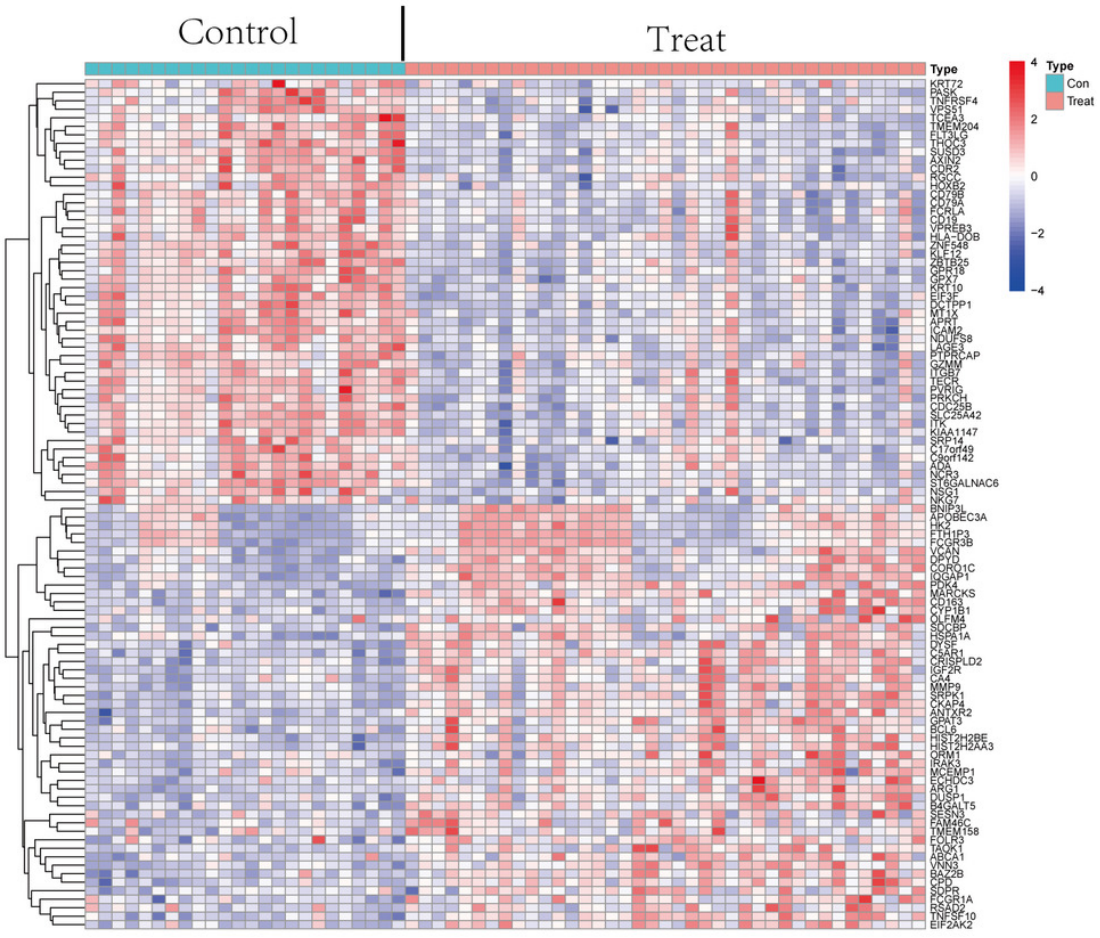

Control
Treat
Type
Con
Treat
4
2
0
−2
−4

Figure 3

The top 10 significantly enriched biological process (BP), cellular component (CC), and molecular function (MF)

(A) In the barplot, the bar size represents the number of genes, and the colour shades represent the p-value. (B) In the bubble plot, the x-axis represents the gene ratio, the y-axis represents the -log10 (FDR) value, the bubble size represents the number of genes, and the colour shades represent the size of the p-value. (A) and (B) The p-value ≤ 0.001 is shown in red; 0.001 < p-value ≤ 0.002 is shown in pink; 0.002 < p-value ≤ 0.003 is shown in purple; p-value□0.003 is shown in blue. Immunoglobulin binding and amide binding are shown in pink. Oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor and oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor are shown in purple. Lipoprotein particle binding, protein-lipid complex binding and hydro-lyase activity are shown in blue. Others are shown in red. (C) The Gene Ontology (GO) enrichment analysis results of DEGs are shown as circle plots. First circle: enriched classification, outside the process, is the sitting scale for the number of genes. Different colours represent different classifications; Second circle: the number of genes in the category in the background and the Q or P value. The more genes, the longer the bar and the smaller the value the redder the colour; Third circle: the total number of foreground genes; Fourth circle: RichFactor value for each classification (the number of foreground genes divided by the number of background genes in that classification), with each cell of the auxiliary background line indicating 0.1.

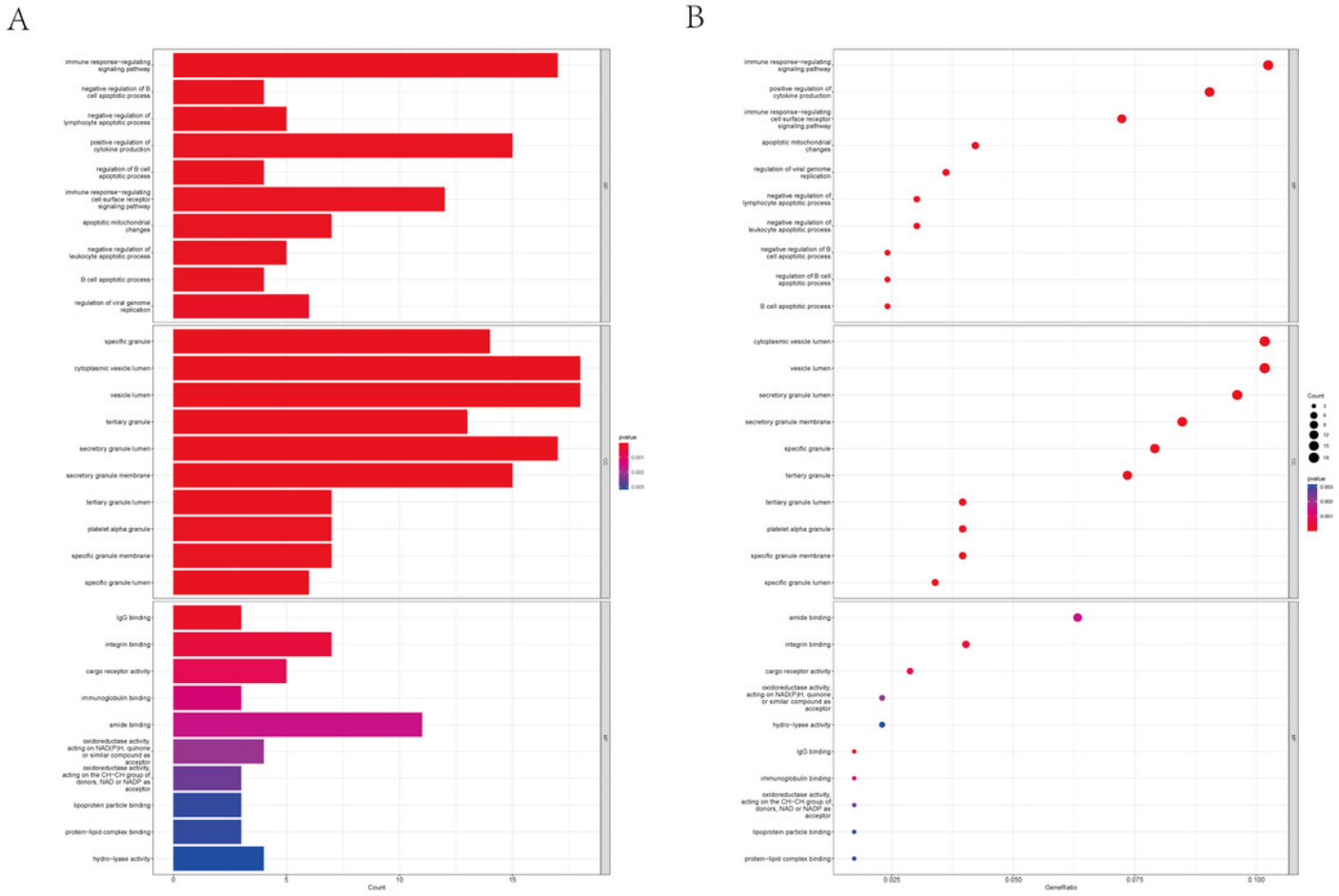
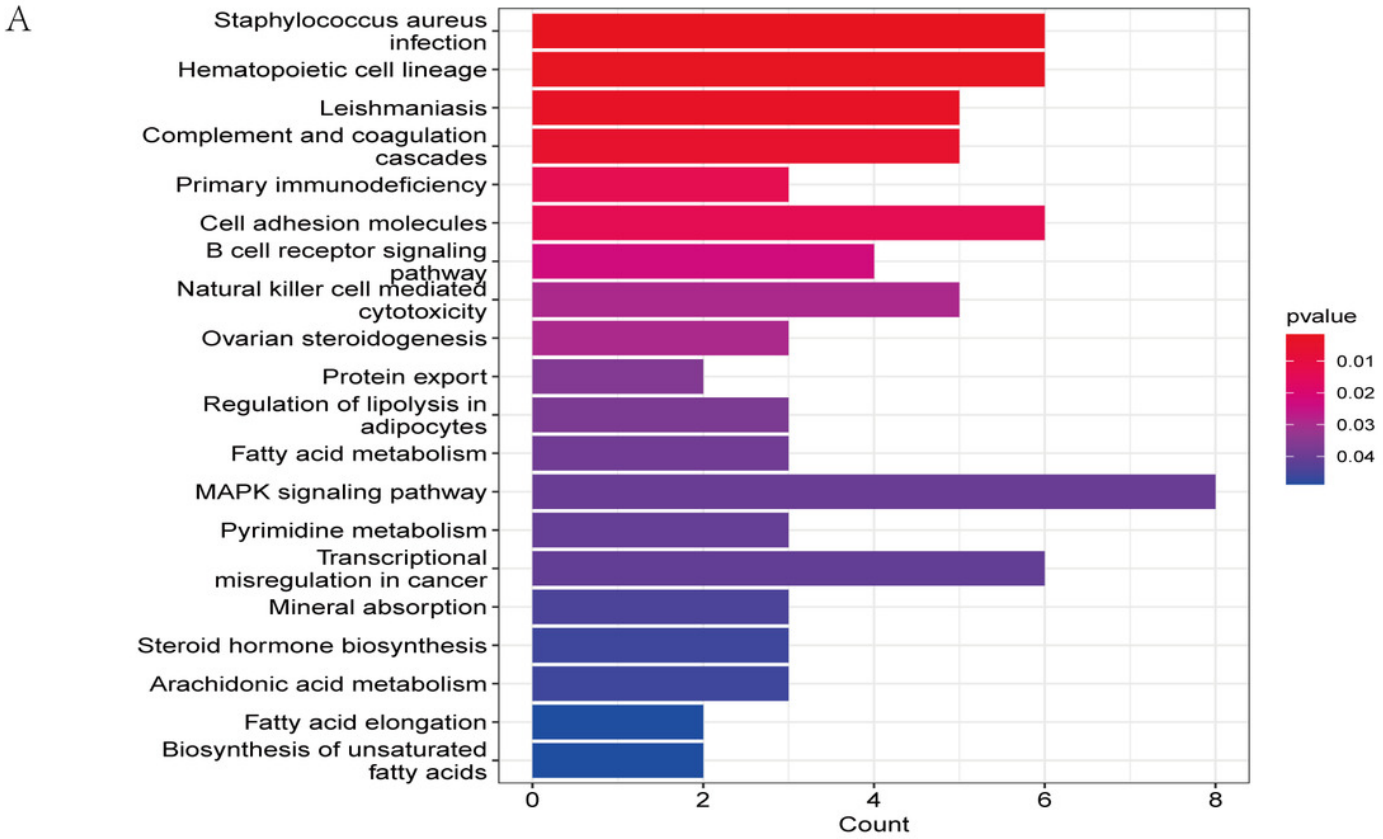

A
immune response−regulating signaling pathway
negative regulation of B cell apoptotic process
negative regulation of lymphocyte apoptotic process
positive regulation of cytokine production
regulation of B cell apoptotic process
immune response−regulating cell surface receptor signaling pathway
apoptotic mitochondrial changes
negative regulation of leukocyte apoptotic process
B cell apoptotic process
regulation of viral genome replication
specific granule
cytoplasmic vesicle lumen
vesicle lumen
tertiary granule
secretory granule lumen
secretory granule membrane
tertiary granule lumen
platelet alpha granule
specific granule membrane
specific granule lumen
IgG binding
integrin binding
cargo receptor activity
immunoglobulin binding
amide binding
oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor
oxidoreductase activity, acting on the CH−CH group of donors, NAD or NADP as acceptor
lipoprotein particle binding
protein−lipid complex binding
hydro−lyase activity
Count
pvalue
B
GeneRatio
Count
pvalue
C
GO:0031983
GO:0070820
GO:0034774
GO:0030667
GO:0072764
GO:0002903
GO:0070229
GO:0001819
GO:0002002
GO:0072788
GO:0042581
GO:0060205
327
164
322
311
469
467
315
160
325
Number of Genes
Number of Select
Rich Factor(0−1)
ONTOLOGY
Biological Process
Molecular Function
Cellular Component
−log10(Pvalue)
(0,2]
(2,4]
(4,6]
(6,8]
(8,10]
(10,15]
(15,20]
>=20

Figure 4

The KEGG pathway analysis results of DEGs

(A) In the barplot, the bar size represents the number of genes, and the colour shades represent the p-value. (B) In the bubble plot, the x-axis represents the gene ratio, the y-axis represents the -log10 (FDR) value, the bubble size represents the number of genes, and the colour shades represent the size of the p-value. The p-value ≤ 0.01 is shown in red; 0.01 < p-value ≤ 0.02 is shown in light red; 0.02 < p-value ≤ 0.03 is shown in light purple; 0.03 < p-value ≤ 0.04 is shown in purple; p-value> 0.04 is shown in blue. Staphylococcus aureus infection, hematopoietic cell lineage, leishmaniasis, and complement and coagulation cascades are shown in red. Primary immunodeficiency and cell adhesion molecules are shown in light red.


A
Staphylococcus aureus infection
Hematopoietic cell lineage
Leishmaniasis
Complement and coagulation cascades
Primary immunodeficiency
Cell adhesion molecules
B cell receptor signaling pathway
Natural killer cell mediated cytotoxicity
Ovarian steroidogenesis
Protein export
Regulation of lipolysis in adipocytes
Fatty acid metabolism
MAPK signaling pathway
Pyrimidine metabolism
Transcriptional misregulation in cancer
Mineral absorption
Steroid hormone biosynthesis
Arachidonic acid metabolism
Fatty acid elongation
Biosynthesis of unsaturated fatty acids
0
2
4
6
8
Count
pvalue
0.01
0.02
0.03
0.04

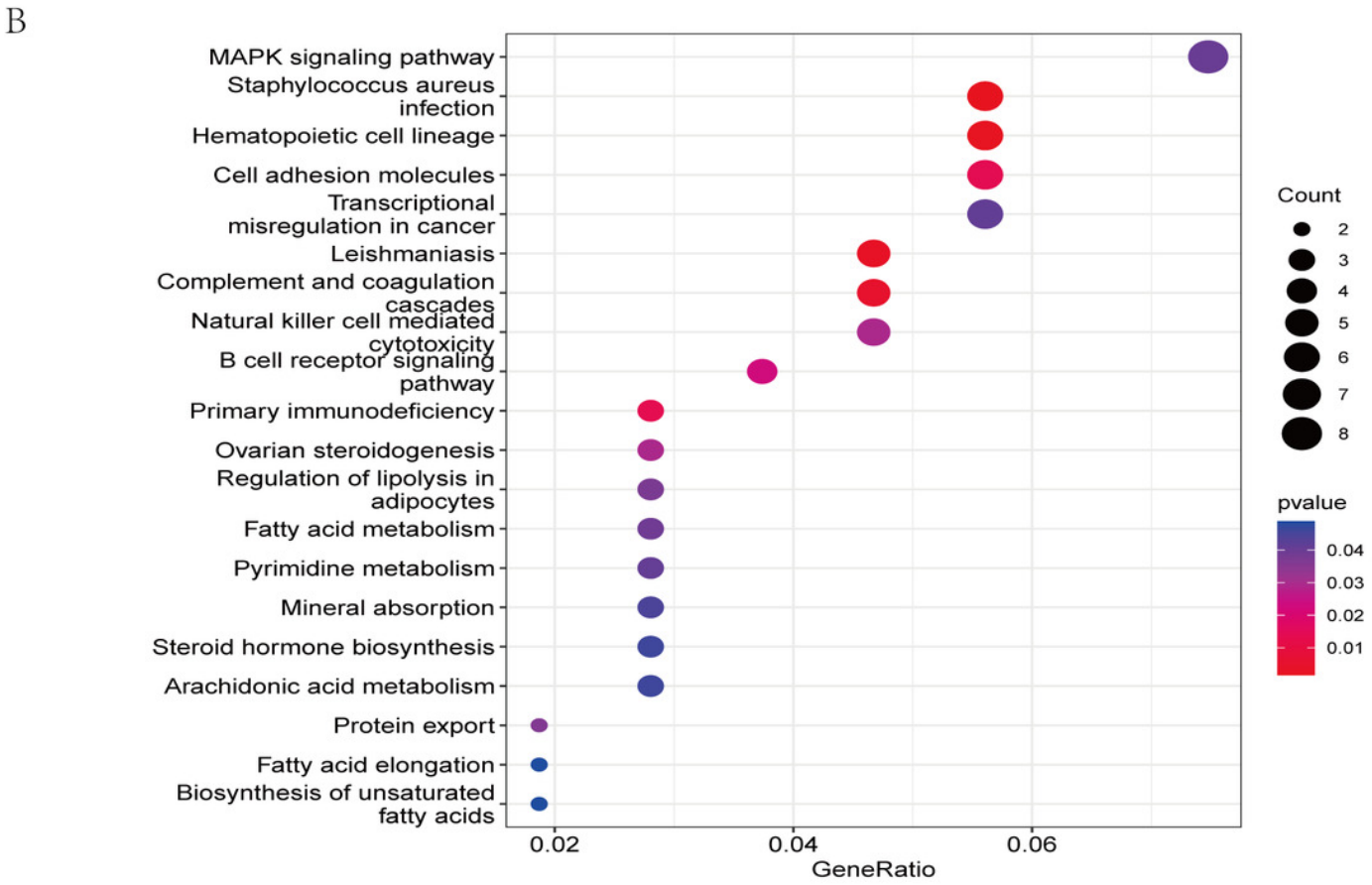
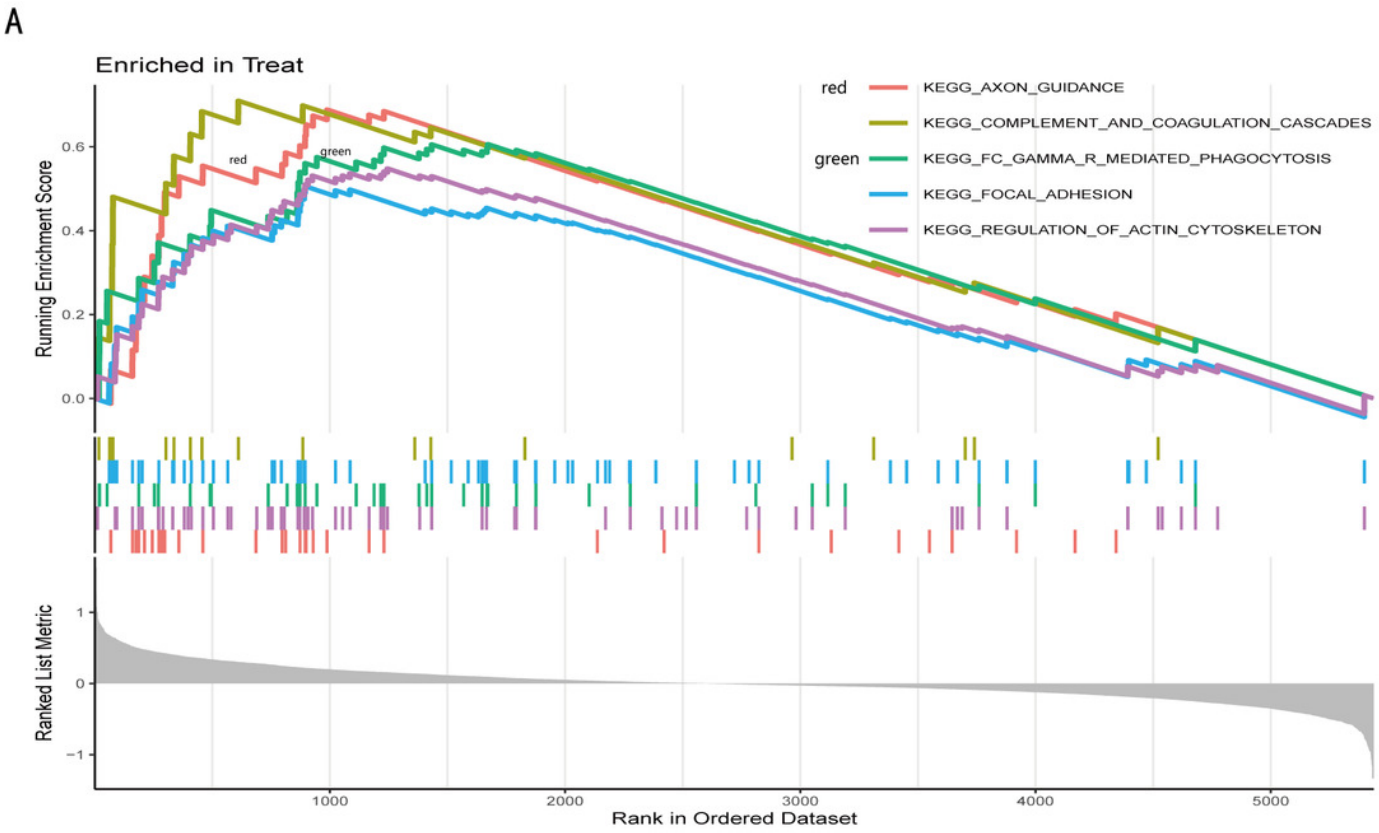

B
MAPK signaling pathway
Staphylococcus aureus infection
Hematopoietic cell lineage
Cell adhesion molecules
Transcriptional misregulation in cancer
Leishmaniasis
Complement and coagulation cascades
Natural killer cell mediated cytotoxicity
B cell receptor signaling pathway
Primary immunodeficiency
Ovarian steroidogenesis
Regulation of lipolysis in adipocytes
Fatty acid metabolism
Pyrimidine metabolism
Mineral absorption
Steroid hormone biosynthesis
Arachidonic acid metabolism
Protein export
Fatty acid elongation
Biosynthesis of unsaturated fatty acids
0.02
0.04
0.06
GeneRatio
Count
2
3
4
5
6
7
8
pvalue
0.04
0.03
0.02
0.01

Figure 5

Gene set enrichment analysis of IS and CTL groups in the GSE dataset.

The different colours represent the different pathways obtained by enrichment. The figure has three parts. Part I: A line graph of the gene enrichment score (ES). The vertical axis is the corresponding running ES. There is a peak at the top or bottom of the line. The core genes are those before the peak for the treatment group with a positive ES, while the core genes are those after the peak for the CTL group with a negative ES. The horizontal axis represents each gene under this gene set and corresponds to the barcode-like vertical line in the second part. Part II: Hits, each vertical line corresponds to a gene under this gene set. Part III: The distribution of rank values for all genes, and the vertical coordinate is the ranked list metric, which is the value of the rated amount of that gene. (A) The GSEA results of the IS group. (B) The GSEA results of the CTL group.


A
Enriched in Treat
red
green
KEGG_AXON_GUIDANCE
KEGG_COMPLEMENT_AND_COAGULATION_CASCADES
KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS
KEGG_FOCAL_ADHESION
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset

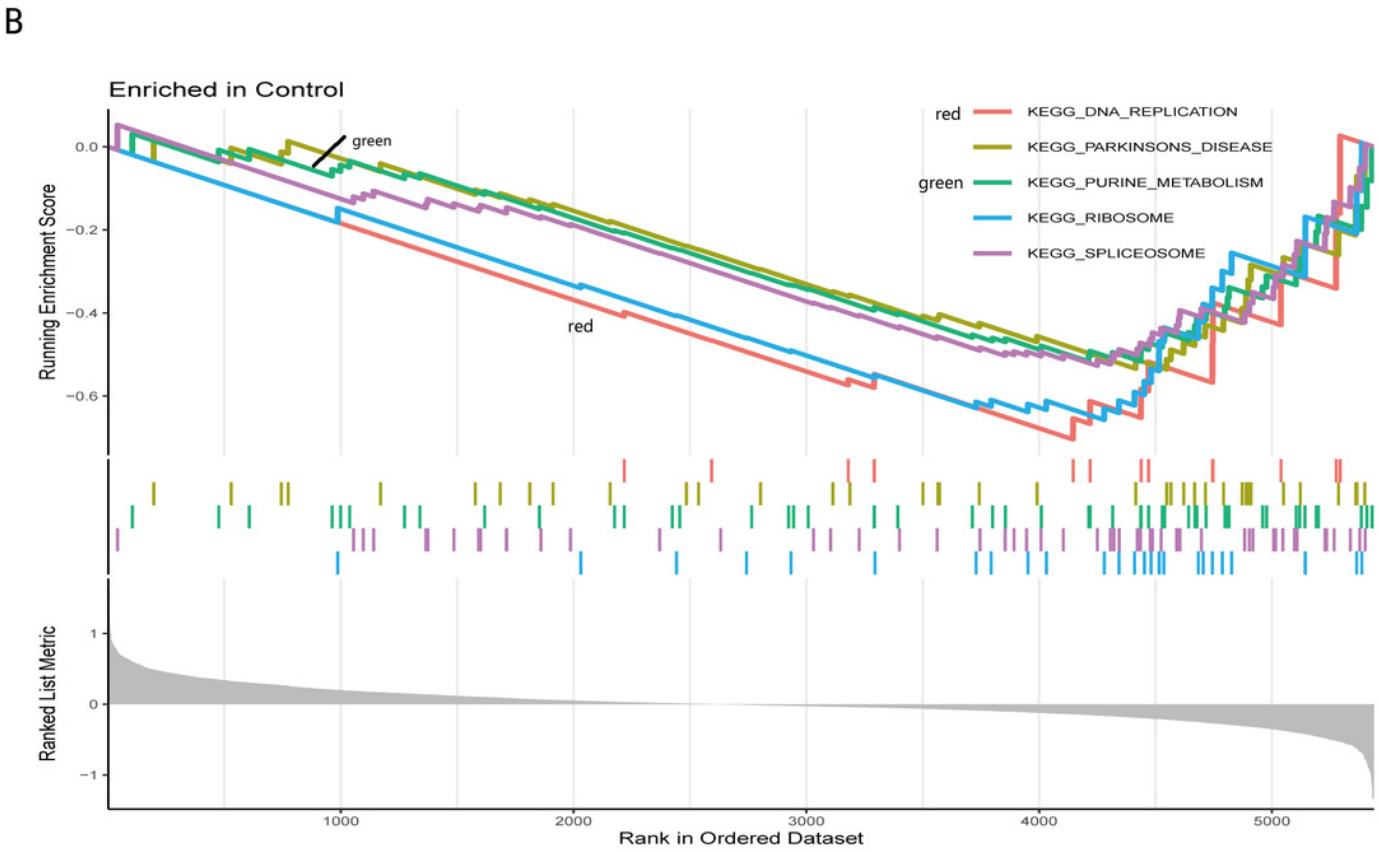
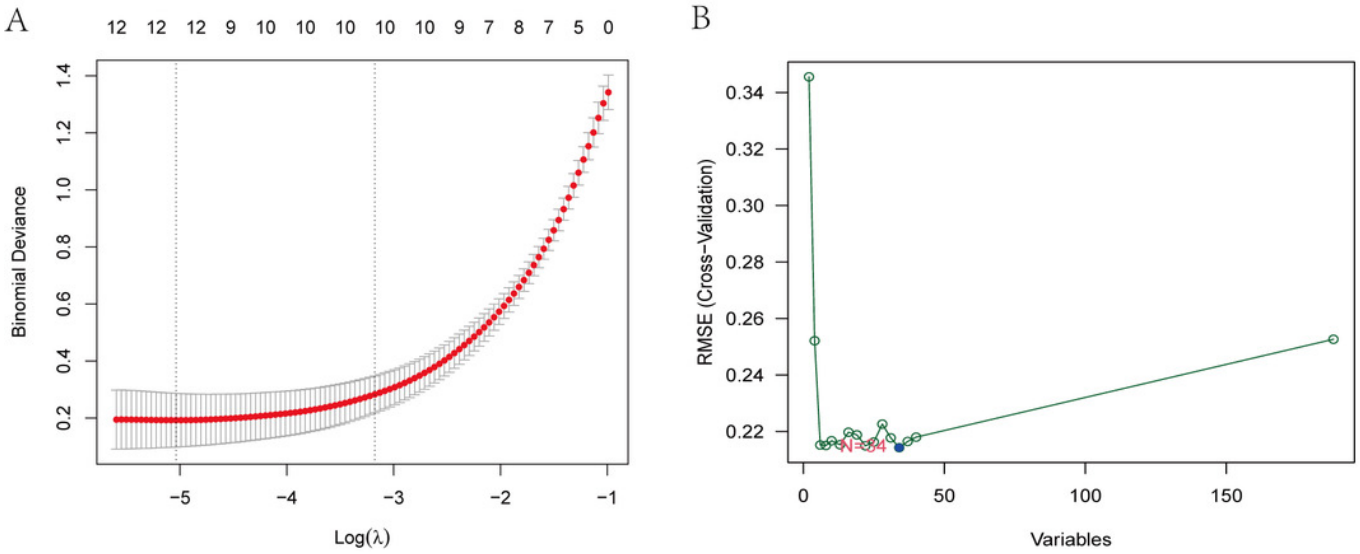

B
Enriched in Control
red
green
KEGG_DNA_REPLICATION
KEGG_PARKINSONS_DISEASE
KEGG_PURINE_METABOLISM
KEGG_RIBOSOME
KEGG_SPLICEOSOME
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset

Figure 6

The identification of six hub genes.

(A) The LASSO logistic regression algorithm identified 12 IS-related features with a 10-fold cross-validation set for selecting the penalty parameter to determine the optimal lambda value. (B) A total of 34 feature genes were filtered out using the SVM-RFE algorithm. (C) Venn diagram of genes extracted from LASSO and SVM-RFE methods. Abbreviations: least absolute shrinkage and selection operator (LASSO), support vector machine (SVM), recursive feature elimination (RFE).


A
12 12 12 9 10 10 10 10 10 9 7 8 7 5 0
Binomial Deviance
0.2 0.4 0.6 0.8 1.0 1.2 1.4
−5 −4 −3 −2 −1
Log(λ)
B
RMSE (Cross−Validation)
0.22 0.24 0.26 0.28 0.30 0.32 0.34
N=24
0 50 100 150
Variables

C

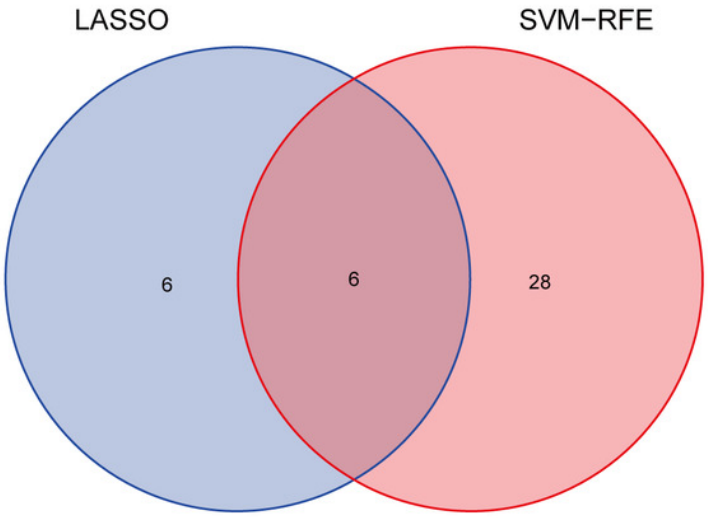

LASSO
SVM−RFE
6
6
28

Figure 7

Immune infiltration landscape between the IS and CTL groups obtained by CIBERSORT analysis.

(A) The bar-plot diagram indicates the relative percentage of different types of immune cells between IS and CTL. (B) The heatmap shows the correlation in the infiltration of innate immune cells. (C) Violin plot showing the difference in immune cell infiltration between IS (red) and Control (purple), P< 0.05, was considered statistically significant.

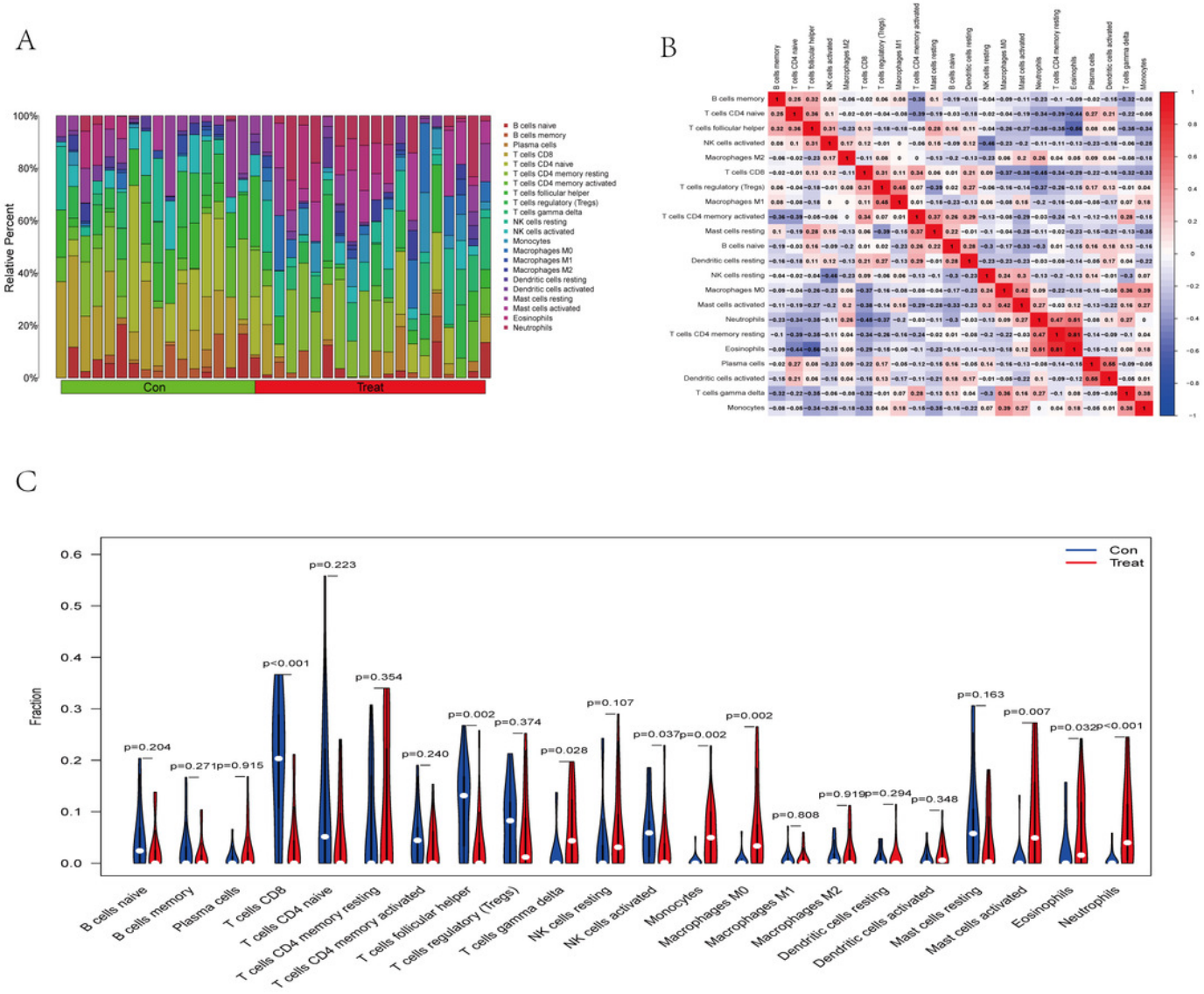

A
Relative Percent
100%
80%
60%
40%
20%
0%
Con
Treat
B cells naive
B cells memory
Plasma cells
T cells CD8
T cells CD4 naive
T cells CD4 memory resting
T cells CD4 memory activated
T cells follicular helper
T cells regulatory (Tregs)
T cells gamma delta
NK cells resting
NK cells activated
Monocytes
Macrophages M0
Macrophages M1
Macrophages M2
Dendritic cells resting
Dendritic cells activated
Mast cells resting
Mast cells activated
Eosinophils
Neutrophils
B
C
Fraction
Con
Treat
p=0.204
p=0.271
p=0.915
p<0.001
p=0.223
p=0.354
p=0.240
p=0.002
p=0.374
p=0.028
p=0.107
p=0.037
p=0.002
p=0.002
p=0.808
p=0.919
p=0.294
p=0.348
p=0.163
p=0.007
p=0.032
p<0.001
B cells naive
B cells memory
Plasma cells
T cells CD8
T cells CD4 naive
T cells CD4 memory resting
T cells CD4 memory activated
T cells follicular helper
T cells regulatory (Tregs)
T cells gamma delta
NK cells resting
NK cells activated
Monocytes
Macrophages M0
Macrophages M1
Macrophages M2
Dendritic cells resting
Dendritic cells activated
Mast cells resting
Mast cells activated
Eosinophils
Neutrophils

Figure 8

Lollipop charts show the correlations between hub genes and infiltration level.

The p-value ≤ 0.2 is shown in green; 0.2 < p-value ≤ 0.4 is shown in lime green; 0.4 < p-value ≤ 0.6 is shown in orange; 0.6 < p-value ≤ 0.8 is shown in pink; 0.8 < p-value ≤ 1 is shown in yellow. (A) ANTXR2, (B) BAZ2B, (C) C5AR1, (D) PDK4, (E) PPIH, (F) STK3.

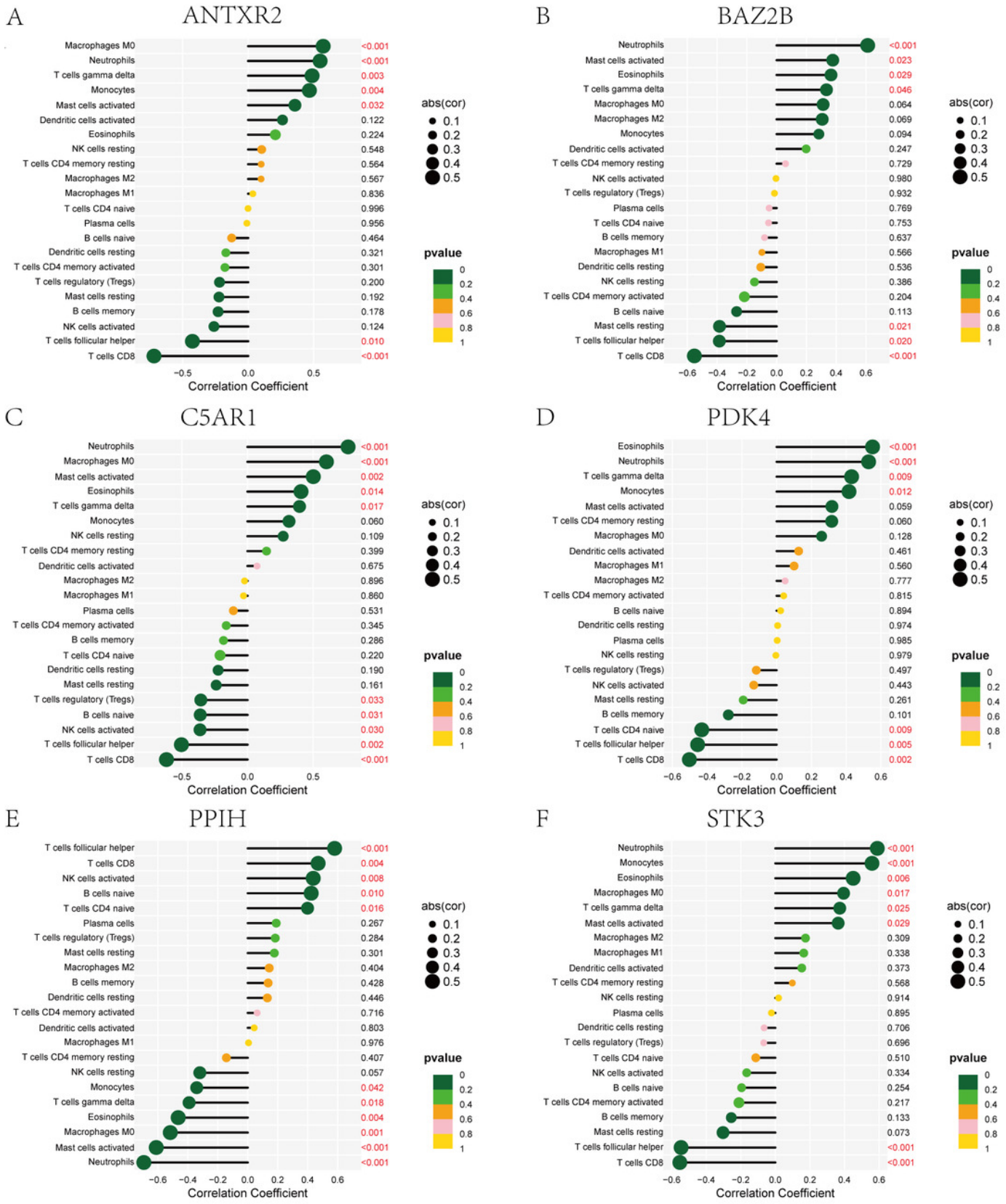

A
ANTXR2
Macrophages M0 <0.001
Neutrophils <0.001
T cells gamma delta 0.003
Monocytes 0.004
Mast cells activated 0.032
Dendritic cells activated 0.122
Eosinophils 0.224
NK cells resting 0.548
T cells CD4 memory resting 0.564
Macrophages M2 0.567
Macrophages M1 0.836
T cells CD4 naive 0.996
Plasma cells 0.956
B cells naive 0.464
Dendritic cells resting 0.321
T cells CD4 memory activated 0.301
T cells regulatory (Tregs) 0.200
Mast cells resting 0.192
B cells memory 0.178
NK cells activated 0.124
T cells follicular helper 0.010
T cells CD8 <0.001
−0.5 0.0 0.5
Correlation Coefficient
abs(cor)
0.1
0.2
0.3
0.4
0.5
pvalue
0
0.2
0.4
0.6
0.8
1
B
BAZ2B
Neutrophils <0.001
Mast cells activated 0.023
Eosinophils 0.029
T cells gamma delta 0.046
Macrophages M0 0.064
Macrophages M2 0.069
Monocytes 0.094
Dendritic cells activated 0.247
T cells CD4 memory resting 0.729
NK cells activated 0.980
T cells regulatory (Tregs) 0.932
Plasma cells 0.769
T cells CD4 naive 0.753
B cells memory 0.637
Macrophages M1 0.566
Dendritic cells resting 0.536
NK cells resting 0.386
T cells CD4 memory activated 0.204
B cells naive 0.113
Mast cells resting 0.021
T cells follicular helper 0.020
T cells CD8 <0.001
−0.6 −0.4 −0.2 0.0 0.2 0.4 0.6
Correlation Coefficient
abs(cor)
0.1
0.2
0.3
0.4
0.5
pvalue
0
0.2
0.4
0.6
0.8
1
C
C5AR1
Neutrophils <0.001
Macrophages M0 <0.001
Mast cells activated 0.002
Eosinophils 0.014
T cells gamma delta 0.017
Monocytes 0.060
NK cells resting 0.109
T cells CD4 memory resting 0.399
Dendritic cells activated 0.675
Macrophages M2 0.896
Macrophages M1 0.860
Plasma cells 0.531
T cells CD4 memory activated 0.345
B cells memory 0.286
T cells CD4 naive 0.220
Dendritic cells resting 0.190
Mast cells resting 0.161
T cells regulatory (Tregs) 0.033
B cells naive 0.031
NK cells activated 0.030
T cells follicular helper 0.002
T cells CD8 <0.001
−0.5 0.0 0.5
Correlation Coefficient
abs(cor)
0.1
0.2
0.3
0.4
0.5
pvalue
0
0.2
0.4
0.6
0.8
1
D
PDK4
Eosinophils <0.001
Neutrophils <0.001
T cells gamma delta 0.009
Monocytes 0.012
Mast cells activated 0.059
T cells CD4 memory resting 0.060
Macrophages M0 0.128
Dendritic cells activated 0.461
Macrophages M1 0.560
Macrophages M2 0.777
T cells CD4 memory activated 0.815
B cells naive 0.894
Dendritic cells resting 0.974
Plasma cells 0.985
NK cells resting 0.979
T cells regulatory (Tregs) 0.497
NK cells activated 0.443
Mast cells resting 0.261
B cells memory 0.101
T cells CD4 naive 0.009
T cells follicular helper 0.005
T cells CD8 0.002
−0.6 −0.4 −0.2 0.0 0.2 0.4 0.6
Correlation Coefficient
abs(cor)
0.1
0.2
0.3
0.4
0.5
pvalue
0
0.2
0.4
0.6
0.8
1
E
PPIH
T cells follicular helper <0.001
T cells CD8 0.004
NK cells activated 0.008
B cells naive 0.010
T cells CD4 naive 0.016
Plasma cells 0.267
T cells regulatory (Tregs) 0.284
Mast cells resting 0.301
Macrophages M2 0.404
B cells memory 0.428
Dendritic cells resting 0.446
T cells CD4 memory activated 0.716
Dendritic cells activated 0.803
Macrophages M1 0.976
T cells CD4 memory resting 0.407
NK cells resting 0.057
Monocytes 0.042
T cells gamma delta 0.018
Eosinophils 0.004
Macrophages M0 0.001
Mast cells activated <0.001
Neutrophils <0.001
−0.6 −0.4 −0.2 0.0 0.2 0.4 0.6
Correlation Coefficient
abs(cor)
0.1
0.2
0.3
0.4
0.5
pvalue
0
0.2
0.4
0.6
0.8
1
F
STK3
Neutrophils <0.001
Monocytes <0.001
Eosinophils 0.006
Macrophages M0 0.017
T cells gamma delta 0.025
Mast cells activated 0.029
Macrophages M2 0.309
Macrophages M1 0.338
Dendritic cells activated 0.373
T cells CD4 memory resting 0.568
NK cells resting 0.914
Plasma cells 0.895
Dendritic cells resting 0.706
T cells regulatory (Tregs) 0.696
T cells CD4 naive 0.510
NK cells activated 0.334
B cells naive 0.254
T cells CD4 memory activated 0.217
B cells memory 0.133
Mast cells resting 0.073
T cells follicular helper <0.001
T cells CD8 <0.001
−0.6 −0.4 −0.2 0.0 0.2 0.4 0.6
Correlation Coefficient
abs(cor)
0.1
0.2
0.3
0.4
0.5
pvalue
0
0.2
0.4
0.6
0.8
1

Figure 9

A ceRNA network based on hub genes.

The network includes nodes (six mRNAs, 249 miRNAs, 218 lncRNAs) with 591 edges. The most central part uses red orbs to represent the hub genes. The outermost two laps are shown with green orbs representing the miRNAs, and purple orbs represent lncRNAs.

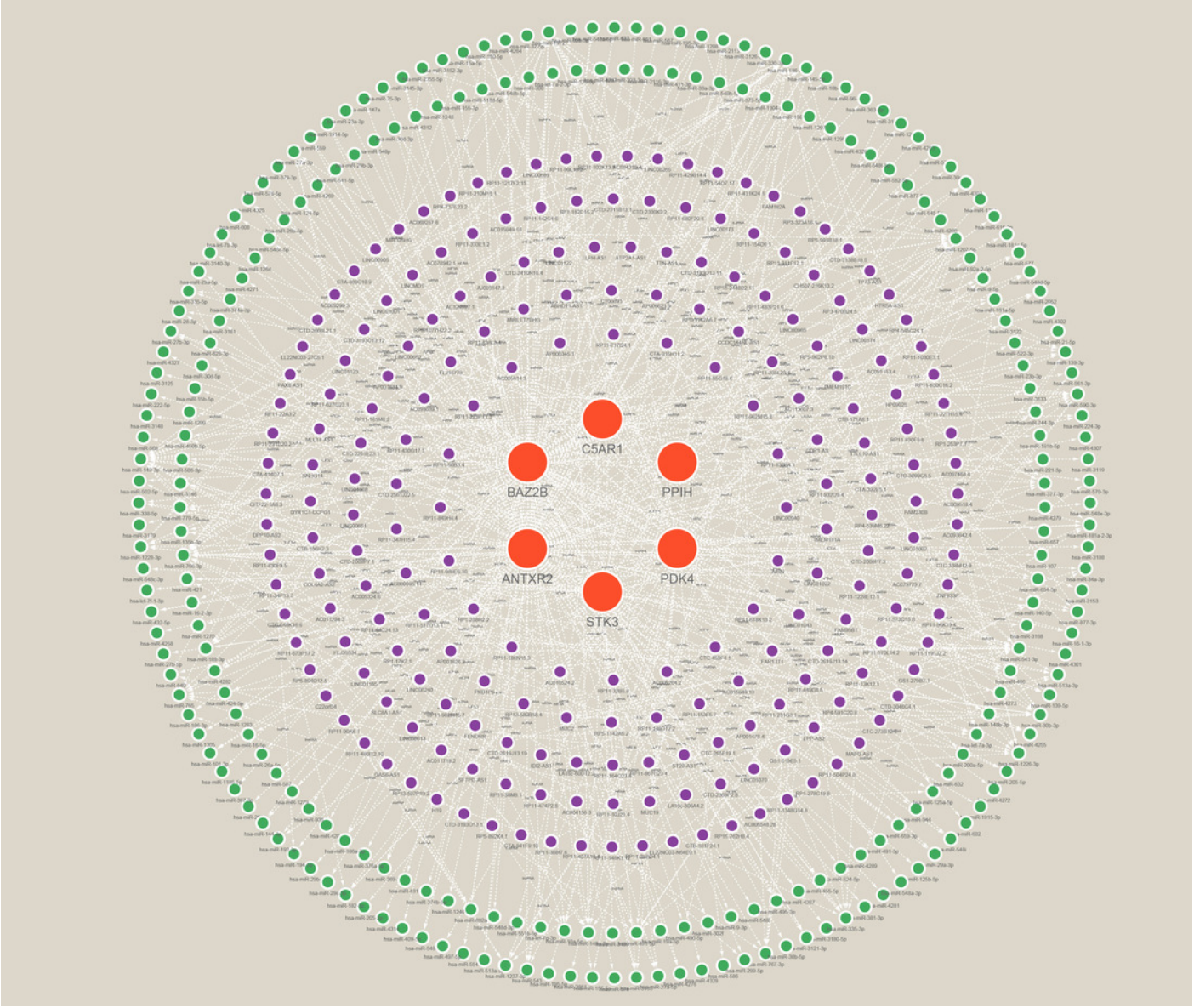


Figure 10

Prediction of marker gene-targeted drugs

Red orbs represent up-regulated mRNA and purple orbs represent drugs.

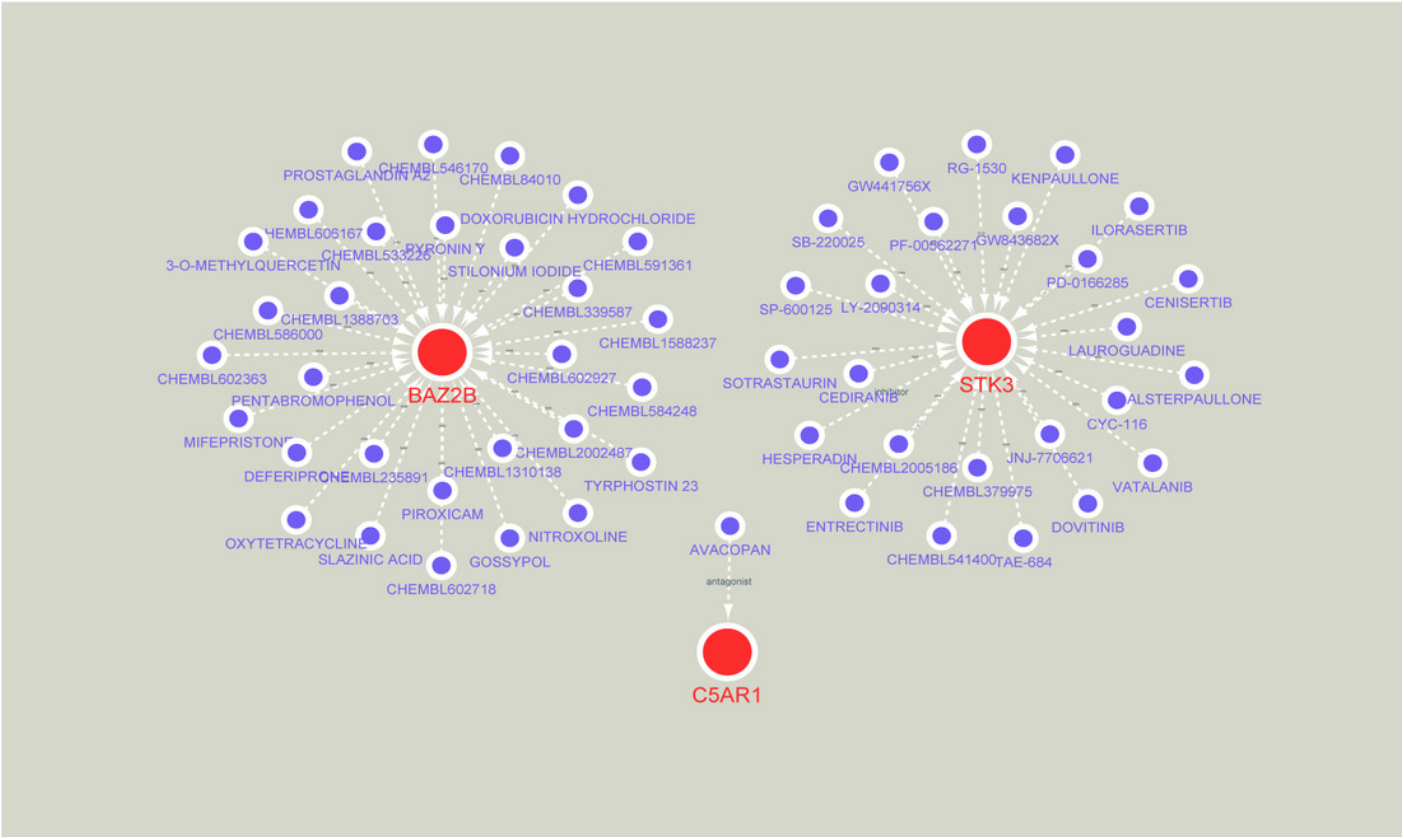


Figure 11

Prediction of marker gene-targeted drugs.

The relative expression levels of (A) ANTXR2, P=0.2233; (B) BAZ2B, P=0.0034; (C) C5AR1, P=0.0018; (D) PDK4, P=0.0026; (E) PPIH, P=0.9792; (F) STK3, P=0.0226 in IS and CTL. #P>0.05,* 0.01 ≤ p-value ≤ 0.05 **P<0.01. Results for the IS group are shown in red. Results for the CTL group are shown in blue.

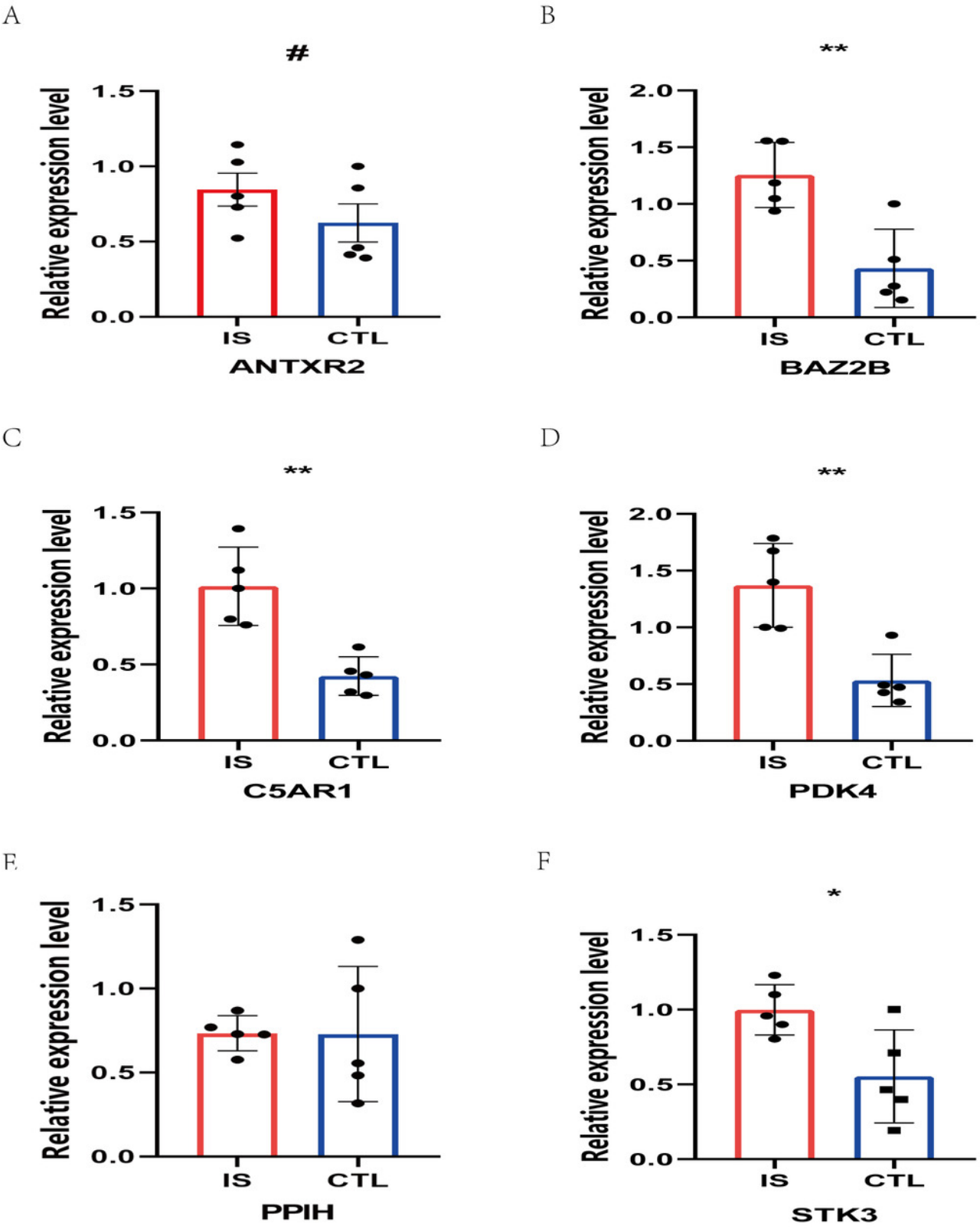

A
#
Relative expression level
1.5
1.0
0.5
0.0
IS
CTL
ANTXR2
B
**
Relative expression level
2.0
1.5
1.0
0.5
0.0
IS
CTL
BAZ2B
C
**
Relative expression level
1.5
1.0
0.5
0.0
IS
CTL
C5AR1
D
**
Relative expression level
2.0
1.5
1.0
0.5
0.0
IS
CTL
PDK4
E
Relative expression level
1.5
1.0
0.5
0.0
IS
CTL
PPIH
F
*
Relative expression level
1.5
1.0
0.5
0.0
IS
CTL
STK3

Figure 12

Diagnostic model building and ROC curve validation.

(A) ROC curves for evaluating the accuracy of hub genes of the training set. (B) ROC curves for evaluating the accuracy of the model of the training set. (C) ROC curves for assessing the accuracy of hub genes using the validation set. (D) ROC curves for assessing the accuracy of the model using the validation set.

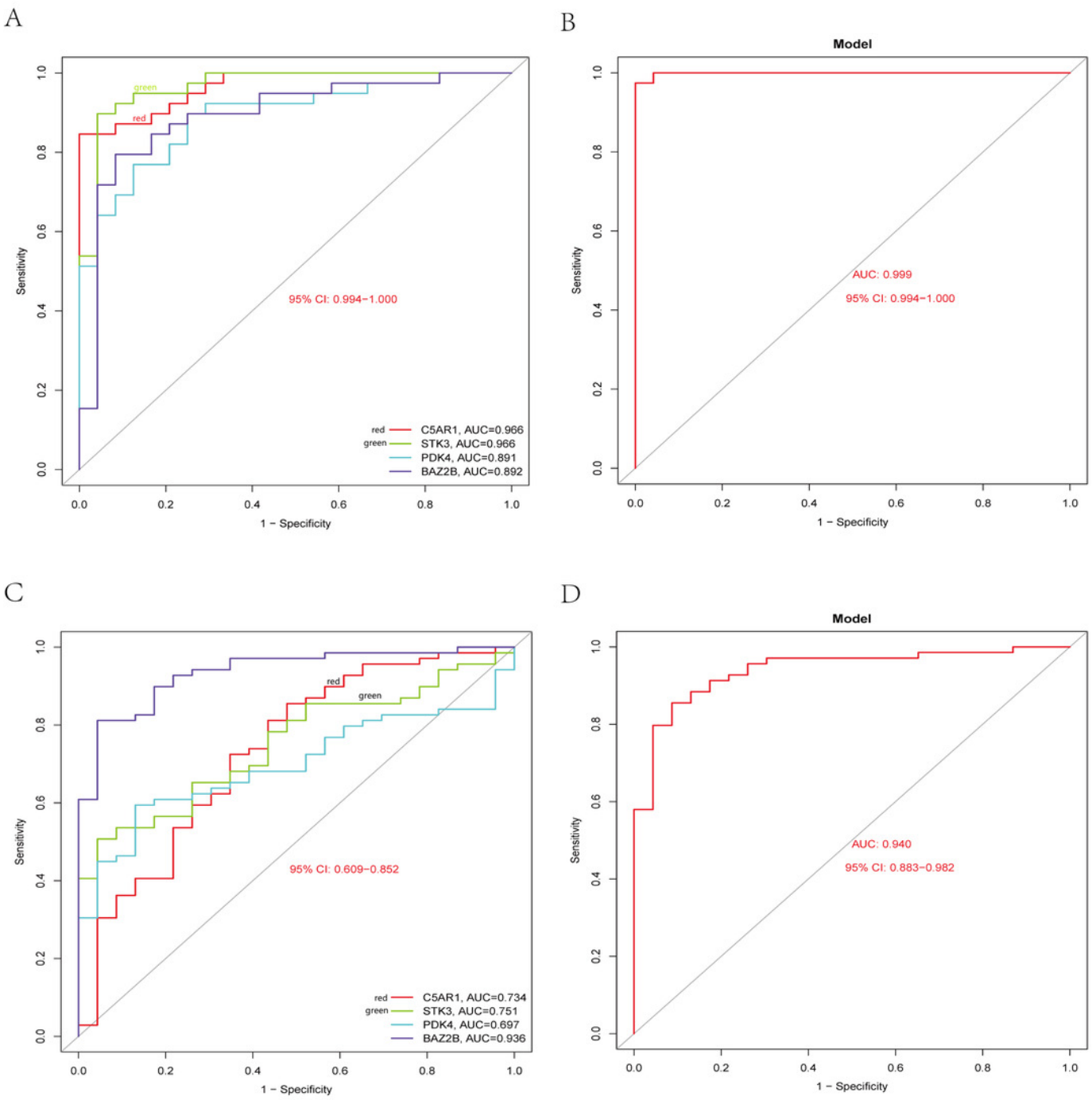

A
green
red
95% CI: 0.994−1.000
red C5AR1, AUC=0.966
green STK3, AUC=0.966
PDK4, AUC=0.891
BAZ2B, AUC=0.892
Sensitivity
1 − Specificity
B
Model
AUC: 0.999
95% CI: 0.994−1.000
Sensitivity
1 − Specificity
C
red
green
95% CI: 0.609−0.852
red C5AR1, AUC=0.734
green STK3, AUC=0.751
PDK4, AUC=0.697
BAZ2B, AUC=0.936
Sensitivity
1 − Specificity
D
Model
AUC: 0.940
95% CI: 0.883−0.982
Sensitivity
1 − Specificity

Figure 13

External validation of the key genes.

The expression levels of hub genes (A) BAZ2B, (B) C5AR1, (C) PDK4, (D) STK3 in the GSE58294 testing set. The results for the treatment group are shown in red. The results for the CTL group are shown in blue.