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

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, 中国

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

Background: Stroke is a disease with high morbidity, disability, and mortality. The immune factor plays a crucial role in the progression of ischemic stroke, but the exact mechanism is not clear. The study aims to identify immune-related biomarkers and evaluate the infiltration pattern of immune cells.

Methods: We downloaded datasets of ischemic stroke patients from GEO, applied R language to discover differential expressed genes, and elucidated their biological functions by GO, KEGG analysis, and GSEA analysis. Then the hub genes were obtained using two machine learning algorithms (LASSO and 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, which were validated with ROC curves.

Results: We screened 188 differentially expressed genes whose functional analysis was enriched to multiple immune-related pathways. Then 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, T cells gamma delta, negatively correlated with T cell follicular helper cells and CD8, while PPIH showed the exact opposite trend. Immunoinfiltration indicated increased activity of monocytes, macrophages M0, neutrophils and mast cells and decreased infiltration of T cell follicular helper cells and CD8 in the IS group. The ceRNA network consists 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 ischemic stroke. 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

Lin Cong[1], Yijie He[1], Yun Wu[1], Ze Li[1], Siwen Ding[1], Weiwei Liang[1], Xinjun 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]

Xuefu Road, Nangang District, Harbin, Heilongjiang, China

Email address: wang_lihua0910@163.com

Abstract

Background: Stroke is a disease with high morbidity, disability, and mortality. The immune factor plays a crucial role in the progression of ischemic stroke, but the exact mechanism is not clear. The study aims to identify immune-related biomarkers and evaluate the infiltration pattern of immune cells.

Methods: We downloaded datasets of ischemic stroke patients from GEO, applied R language to discover differential expressed genes, and elucidated their biological functions by GO, KEGG analysis, and GSEA analysis. Then the hub genes were obtained using two machine learning algorithms (LASSO and 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, which were validated with ROC curves.

Results: We screened 188 differentially expressed genes whose functional analysis was enriched to multiple immune-related pathways. Then 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, T cells gamma delta, negatively correlated with T cell follicular helper cells and CD8, while PPIH showed the exact opposite trend. Immunoinfiltration indicated increased activity of monocytes, macrophages M0, neutrophils and mast cells and decreased infiltration of T cell follicular helper cells and CD8 in the IS group. The ceRNA network consists of 306 miRNA-mRNA interacting pairs and 285 miRNA-lncRNA interacting pairs. RT-qPCR results indicated that the expression levels of

38 BAZ2B, C5AR1, PDK4 and STK3 were significantly increased in patients with ischemic stroke.
39 Finally, we developed a diagnostic model based on these four genes. The AUC value of the
40 model was verified to be 0.999 in the training set and 0.940 in the validation set.

41 **Conclusion**:   Our research explored the immune-related gene expression modules and
42 provided a specific basis for further study of immunomodulatory therapy of IS.
43

Introduction

45 Stroke, of which more than 80% are ischemic (Hasan et al., 2018), is the second leading cause of
46 death and the third leading cause of disability in adults (Campbell and Khatri, 2020). It seriously
47 affects patients' quality of life and causes a heavy burden to the family and society. Therefore,
48 our study focuses on discovering novel targets for diagnosing and treating ischemic stroke (IS).
49 At present, the diagnosis of ischemic stroke (IS) mainly depends on clinical manifestations and
50 neuroimaging, such as computed tomography (CT) (Harston et al., 2015) and magnetic
51 resonance imaging (MRI) (Zameer et al., 2021). The most effective early treatment methods are
52 thrombolysis and intravascular therapy (Cassella and Jagoda, 2017, Hasan et al., 2018), both of
53 which have time limitations to administer. Nonetheless, the clinical manifestations of the disease
54 are variable, and the cost of administering CT and MRI scans is expensive and time-consuming.
55 If the symptoms are not typical or the above examinations cannot be performed in a timely
56 manner, diagnosis and treatment may be delayed. At the same time, the above treatment methods
57 carry the risk of bleeding. This has led us to search for new diagnostics and treatments.
58 Genetic factors are related to the occurrence and prognosis of stroke (Malik et al., 2018, Torres-
59 Aguila et al., 2019). Possible stroke-related gene loci, such as ANK2, CDK6 and KCNK3, were
60 found by whole genome sequencing (Malik et al., 2018). However, the specific mechanism of
61 action is not clear. CYP450, COX-2, PTGIS, TBXAS1, P2RY1, and TGB3, or GPIIa are
62 candidate genes that may be associated with stroke prognosis (Torres-Aguila et al., 2019).
63 Presently, bioinformatics technology has been widely applied to various diseases, so we use
64 bioinformatics technology to screen and identify key genes associated with IS, and build a
65 classification diagnosis model based on these genes, to provide certain auxiliary methods for the
66 diagnosis of ischemic stroke and potential therapeutic targets for the management of diseases.
67 Immune factors are also closely associated with ischemic stroke (Iadecola et al., 2020, Levard et
68 al., 2021). In the acute phase, the immune response is activated immediately when the brain is in
69 ischemia. Damaged or dead brain cells release damage-associated molecular patterns (DAMPS)
70 (Wu et al., 2022). These substances quickly activate intracranial immune cells (microglia, etc.),
71 release chemokines and cytokines, and trigger a series of inflammatory cascades (Anrather and
72 Iadecola, 2016, Levard et al., 2021). More importantly, immune cells in the peripheral blood can
73 penetrate the central nervous system (CNS) through the damaged blood-brain barrier (BBB)
74 (Jayaraj et al., 2019), further exacerbating the damage. At the same time, these DAMPS and
75 cytokines can enter the bloodstream and activate the systemic immune system and inflammatory
76 response, which can lead to severe immunosuppression and even fatal infections (Iadecola et al.,

77   2020). In the chronic phase, adaptive immune responses to the brain are mobilized due to antigen
78   presentation, which may be the basis of neuropsychiatric sequelae, poststroke
79   dementia/fatigue/depression, and the root cause of post-stroke morbidity (Iadecola et al., 2020).
80   Epidemic-modulated therapy may be a good alternative to IS therapy. Several studies have
81   shown that immune regulation can delay disease progression and improve neurological function
82   and prognosis. These changes highlight the need to maintain a stable immune microenvironment
83   in the central nervous system (Javidi and Magnus, 2019, Jayaraj et al., 2019). TGF-β1 has a
84   significant neuroprotective effect (Taylor et al., 2017) and can reduce post-stroke infection
85   (Cekanaviciute et al., 2014, Doll et al., 2014). Relevant targeted drugs are now being developed.
86   Specific inhibitors of IL-1β can delay atherosclerosis by interfering with immune pathways
87   associated with atherosclerotic plaque formation (Khambhati et al., 2018). Meanwhile, IL-1
88   receptor antagonists have been found to effectively reduce peripheral inflammation in patients
89   with acute ischemic stroke and improve clinical outcomes in these patients (Smith et al., 2018).
90   In conclusion, the analysis of immune infiltration analysis and the study of the relationship
91   between target genes and immune cell distribution may help to elucidate the immune-related
92   molecular mechanism of IS and provide the specific basis for immunoregulatory therapy. At the
93   same time, the prediction of drug target gene action network may provide a new target for the
94   treatment of IS.
95   Competing endogenous RNA (ceRNA) does not represent a specific type of RNA, but a
96   regulatory mechanism. By combining with microRNA response elements (MREs) existing on
97   mRNA, miRNAs inhibit mRNA translation or lead to its degradation, thus achieving the function
98   of regulating gene expression after transcription (Salmena et al., 2011). Different mRNAs can
99   compete to bind to the same miRNA and thus participate in the regulation of gene expression.
100 With the deepening of transcriptome studies, it is found that MREs exist not only on mRNA, but
101 also on lncRNA, circRNA and other types of RNA, which means that the same miRNA can bind
102 to several and multiple types of RNA, and a competitive relationship is formed between RNA
103 molecules that bind to the same miRNA. This forms the ceRNA network. If the regulation of
104 ceRNA is abnormal, it may cause the occurrence of diseases, as shown in the research fields of
105 tumors (Karreth et al., 2015), stroke (Li et al., 2020) and other complex diseases. It has been
106 reported that miRNA regulates gene expression by silencing target genes and plays an important
107 role in neurological diseases such as IS (Bulygin et al., 2020, Chi et al., 2020). In addition, the
108 occurrence, progression and prognosis of acute ischemic stroke are also related to abnormal
109 lncRNA expression (Bai et al., 2014, Feng et al., 2019, Li et al., 2020, Wang et al., 2020, Fu et
110 al., 2021). Therefore, they are potential candidates for stroke diagnosis. lncRNAs are also
111 involved in the progression of inflammation (Feng et al., 2019, Zhang and Niu, 2020) and
112 apoptosis (Chen et al., 2017, Liu et al., 2019, Zhong et al., 2020). Therefore, the construction of
113 ceRNA network may help to predict the possible molecular mechanism related to the onset of IS.
114 Most studies have used databases to predict miRNA target genes and lncRNAs. Our analysis is
115 based on hub genes to predict their targeted miRNAs, and then construct a ceRNA network of
116 mRNA-miRNA-lncRNA.

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

Materials & Methods

Obtaining and processing profile datasets

126 We obtained and downloaded the ischemic stroke gene expression profile from Gene Expression
127 Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) (Table 1). The criteria for screening were that
128 the data in the dataset contained both IS and control groups (CTL), and the number of cases in
129 each group was no less than 20. Patients in IS group were those diagnosed with acute ischemic
130 stroke. In addition, the data were based on whole blood samples with full gene expression
131 profiles. After screening, GSE16561 (Barr et al., 2010) and GSE58264 (Stamova et al., 2014)
132 matched the criteria. GSE16561 was used as a training set. GSE58294 was for the validation of
133 biomarker genes. We applied the Perl language script (https://www.perl.org/) to get the gene
134 name corresponding to each probe from the platform annotation file; the names were then
135 converted. In this process, we used the average of all probes' expressions if there were multiple
136 probes corresponding to the same gene name.

Differential Expression Analysis

138 We applied R package limma (Ritchie et al., 2015) (version 3.6.0) to perform differential
139 expression analysis to identify DEGs between groups in the dataset GSE16561. FoldChange >
140 1.5 and adjusted P-value < 0.05 were taken as significant criteria. DEGs were represented by
141 heatmap, designed by the pheatmap package (Kolde, 2012) (version 1.0.12), and volcano maps,
142 created by the ggplot2 package (Wickham, 2009) (version 3.3.5).

Function enrichment analysis

144 We not only annotated the function of the differential expression genes themselves,Gene
145 Ontology (GO) enrichment, but we also studied the pathway by Kyoto Encyclopedia of Genes
146 and Genomes (KEGG) analysis. GO Enrichment and KEGG pathways analysis results were
147 obtained using clusterProfiler (Yu et al., 2012). GSEA software was used to analyze the potential
148 molecular mechanisms and signaling pathways of Hub genes in IS patients. C2 (c2.cp. Kegg.
149 v7.4.symbols. gmt) was selected as the reference gene set for analysis. The R package
150 clusterProfiler also conducted the GSEA analysis of the IS and CTL group, adjusted P-
151 value<0.05.

152 **Screening for hub genes**

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

168 **Evaluation of immune cell infiltration**

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

174 **Construction of the ceRNA Network**

175 Firstly, the miRanda, MicroRNADB, and TargetScan databases were used to simultaneously
176 predict the target miRNAs of hub genes and get the intersection. Obtained miRNA-mRNA
177 interaction pair file. Secondly, the spongeScan database (http://spongescan.rc.ufl.edu), was used
178 to predict the lncRNA based on the result file from the previous step. Finally, the ceRNA
179 network was formed, and the feature files were jointly built based on the lncRNA-miRNA-
180 mRNA files. Visualise ceRNA regulatory networks using Cytoscape (Shannon et al., 2003).

181 **Gene- Drug Target Network Analysis**

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

187 **Samples collection**

188 Our samples were all obtained from patients attending the Department of Neurology of the
189 Second Affiliated Hospital of Harbin Medical University, approved by the Medical Ethics
190 Committee of the Second Affiliated Hospital of Harbin Medical University (NO.KY2022-285)
191 (Figure S1), and all enrolled individuals signed an informed consent form (Document S2). 5
192 cases in the disease group and 5 cases in the healthy control group. Disease group (IS group)
193 inclusion criteria: Acute ischemic stroke confirmed by MRI, meeting the diagnostic criteria in
194 the International Classification of Diseases (9th Revision). Exclusion criteria: Patients with a
195 history of haematologic, type 1 diabetes, autoimmune, thyroid, neoplastic, renal or liver diseases
196 were excluded. Fasting venous blood sample of 5ml were collected from both groups. Peripheral
197 blood for the disease group was required to be collected within 24 hours of onset.

198 **Real-Time Quantitative PCR (RT-qPCR)**

199 RT-qPCR was used to detect the expression of six hub genes in whole peripheral blood of
200 clinical samples. Total RNA was isolated from each sample using Trizol reagent according to the
201 manufacturer's instructions. Reverse transcription from RNA to cDNA was performed using
202 HiScript Q RT SuperMix for qPCR (+gDNA wiper) (Vazyme Biotech,Nanjing). RT-qPCR
203 was performed on cDNA samples using ChamQ SYBR Color qPCR Master Mix (2X) (Vazyme
204 Biotech,Nanjing). Results were analyzed by the 2-$\Delta\Delta$Ct method and expressed as fold changes,
205 with GAPDH selected as the internal reference gene. The PCR primers were designed by
206 Majorbio (Majorbio, Shanghai) (Table S3). A *P*-value less than 0.05 was considered statistically
207 significant.

208 **Statistical analysis**

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

213 **Development and Validation of a Diagnostic Model**

214 On the one hand, boxplots were used to show the expression of hub genes in GSE58294 as a
215 validation dataset. We considered the difference statistically significant at a P-value < 0.05. On
216 the other hand, a logistic regression algorithm was used to build a diagnostic model for IS
217 classification based on crucial genes. The Proc R package was applied to generate the ROC
218 curve and calculate the area under the ROC curve (AUC) (Stamova et al., 2014) to verify the
219 accuracy of the genes and the model in dataset GSE16561. In addition, we calculated the AUC
220 values of the key genes and logistic regression models in dataset GSE58294 to validate the
221 diagnostic models' classification performance. To avoid the occurrence of overfitting, we chose

222 to reduce the complexity of the model and reduce the parameters to make the model simpler.
223 When training the model, the dataset was partitioned in a 7:3 ratio for training the model and
224 predicting the accuracy of the created model, respectively, to balance the accuracy. Finally, the
225 visual analysis of the above process was completed using ggplot2.

Results

Differential Expression Analysis

228 The volcano plot (Fig. 2A) showed that by using R package limma, we obtained 188 differential
229 expression genes (Table S4), of which 85 genes were down-regulated in expression and 103
230 genes were up-regulated in expression. The top 50 differentially expressed genes are shown in a
231 heatmap (Fig. 2B), created by the pheatmap package.

Function enrichment analysis

233 The GO analysis results (Table S5) were presented in a barplot (Figure 3A), bubble plot (Figure
234 3B), and circle plot (Figure 3C). GO-Biological Process (BP) shows that DEGs were primarily
235 involved in immune-related biological processes, including immune response−regulating
236 signaling pathway, positive regulation of cytokine production, and immune response−regulating
237 cell surface receptor signaling pathway; GO- Molecular Function (MF) enrichment results
238 showed DEGs were primarily associated with integrin binding oxidoreductase activity, acting on
239 NAD(P)H quinone or similar compound as acceptor and immunoglobulin binding and GO-
240 Cellular Component (CC) analysis was significantly enriched in platelet alpha granule, secretory
241 granule lumen, etc. KEGG pathways of each module were listed in Figure 4 and Table S6,
242 mainly enriched in MAPK signaling pathway, Cell adhesion molecules, Complement and
243 coagulation, The Cascades, and so on. The GO and KEGG pathway analysis results suggested
244 that the immune system played a crucial role in IS.

245 We further conducted GSEA analysis to avoid missing the functions of some insignificant but
246 biologically significant genes, biological characteristics, and regulatory networks. GSEA results
247 (Figure 5; Table S7) showed that the pathway involved axon guidance, complement and
248 coagulation cascades, Fc gamma r mediated phagocytosis, focal adhesion, and regulation of actin
249 cytoskeleton in IS group (Figure 5A). Complement and coagulation cascades, the identified
250 pathway of IS, were consistent with the result of KEGG, and the other pathways are related to
251 immunity or cellular metabolism. Genes in the treatment group (IS group) were enriched at the
252 top, which means that this gene set was upregulated. Conversely, genes in the control group
253 (Figure 5B) were enriched at the bottom with a down-regulation trend.

Identification of the hub gene

255 To explore the biomarkers of IS, we performed feature screening through LASSO regression and
256 the SVM-RFE algorithm. 12 genes were identified as the most associated with IS by LASSO

257   regression (Fig. 6A). The SVM-RFE algorithm evaluated 34 characteristic genes in IS (Fig. 6B).
258   Both machine algorithms were subjected to 10-fold cross-validation to ensure the accuracy of the
259   results. Then, 6 differential expression genes (ANTXR2, BAZ2B, C5AR1, PDK4, PPIH, and
260   STK3) (Fig. 6C) were identified as the hub genes by these two joint algorithms for subsequent
261   research.

262   **Results of Immune Cell Infiltration Analysis**

263   The immune cell infiltration analysis results (Table S8) were presented by barplot (Figure 7A)
264   and heatmap (Figure 7B). Barplot showed a difference in the percentage of immune cells
265   between IS group and CTL. Compared with the CTL, Monocytes, macrophage M0, Neutrophils,
266   and Mast cells activated infiltration increased, but T cells follicular helper, CD8 T cells
267   infiltration decreased in IS group. The relationship between immune cells and key genes is
268   presented in Figure 8. ANTXR2, BAZ2B, C5AR1, PDK4, and STK3 were positively correlated
269   with neutrophils and negatively correlated with T cells follicular helper and CD8 T cells; but the
270   PPIH gene was positively associated with T cells follicular helper and CD8 T cells and
271   negatively correlated with Mast cells activated and neutrophils; Only PPIH was a down-
272   regulated expression gene, while the other five genes (ANTXR2, BAZ2B, C5AR1, PDK4, and
273   STK3) were all up-regulated expression. Meanwhile, the correlation between individual hub
274   genes and immune cells is shown by a Scatter plot (Figure S9-16). R-value indicates positive and
275   negative correlation. P-value<0.05 is statistically significant. These evidence suggest that
276   changes in the immune microenvironment of IS patients may be related to these 6 Hub genes.

277   **Construction of ceRNA Networks for Hub Genes**

278   Many studies have confirmed that ceRNA regulatory networks play a role in the biology and
279   pathophysiology of various diseases. We also constructed networks (Figure 9; Table S17) to
280   determine whether 6 hub genes have similar regulatory relationships. 6 hub genes were
281   separately entered into the database, yielding a total of 306 miRNA-mRNA interaction pairs, of
282   which there were 117 ANTXR2-miRNA interaction pairs and 87 BAZ2B-miRNA interaction
283   pairs, and 285 miRNA-lncRNA interaction pairs. We hypothesised that ANTXR2 and BAZ2B
284   might be involved in the regulation of ceRNA. According to the value of a degree (Table S18),
285   hsa-miR-766-3p, hsa-miR-149-3p, hsa-miR-1972, hsa-miR-186-5p, hsa-miR-1207-5p, MUC19
286   and LINC01002 may play a vital role in the ceRNA network.

287   **Establishment of Drug-Hub Genes Regulatory Networks**

288   Drugs related to the 6 essential genes were screened out, and 3 genes, including BAZ2B,
289   C5AR1, and STK3, had target drugs. A gene-drug network is shown in Figure 10 and Table S19.
290   We retrieved a total of 58 drugs acting on the Hub gene. BAZ2B is predicted to have the most
291   targets of drug action, a total of 30.25 drugs targeted STK3. Only one drug targeted C5AR1.

292   **Validation of Hub Genes and Diagnosis Model**

293 The results of RT-qPCR showed that the expression of BAZ2B, C5AR1, PDK4, and STK3 was
294 significantly higher in the ischemic stroke group (Figure 11 and Table S20). A diagnosis model
295 for IS classification was established using a logistic regression algorithm based on the 4 key
296 genes. The results show that the diagnostic model could distinguish patients from normal
297 samples. The ROC curves for the HUB genes and models are shown in Fig 12A and Fig12B. The
298 AUC values are as follows: BAZ2B (AUC=0.892), C5AR1 (AUC=0.966), PDK4 (AUC=0.891),
299 STK3 (AUC=0.966), and Model (AUC=0.999). Finally, we verified the accuracy of hub genes
300 (Figure 12C) and the model (Figure 12D) using another dataset, GSE58294. AUC values for
301 genes and models were as follows: BAZ2B (AUC=0.936), C5AR1 (AUC=0.734), PDK4
302 (AUC=0.697), STK3 (AUC=0.751), and Model (AUC=0.940). The boxplot (Figure 13) shows
303 the differential expression of hub genes in the CTL and IS groups in the validation dataset. Our
304 established model performs well in distinguishing between IS and normal samples.

Discussion

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

315 In this study, 188 DEGs were observed through R package limma. By applying the joint LASSO
316 and SVM algorithm, we finally obtained 6 genes (ANTXR2, BAZ2B, C5AR1, PDK4, PPIH, and
317 STK3) most associated with IS. But only BAZ2B, C5AR1, PDK4, and STK3 were upregulated
318 expression in ischemic stroke patients as verified by RT-PCR. C5AR1 has been proven to play a
319 crucial role in regulating inflammatory and neurocognitive functions in ischemic stroke,
320 Alzheimer's disease, malaria, and neuropathic pain (Brandolini et al., 2019, McDonald et al.,
321 2015, Moriconi et al., 2014). A markedly increase in C5AR1 expression was observed in the
322 MCAO- and OGD-induced models. Meanwhile, C5AR1 inhibitors have significant
323 neuroprotective effects and remarkably inhibit neuro-inflammation and apoptosis in primary
324 cortical neurons and MCAO-induced stroke models (Shi et al., 2017, Brandolini et al., 2019).
325 TLR4 and C5AR1 promote apoptosis and inflammation by activating of the cAMP/PKA/I-
326 κB/NF-κB signaling pathway during brain ischemia-reperfusion (Zaal et al., 2017a, Kim and
327 Jang, 2017, Shi et al., 2021). Some study shows that C5AR1 can inhibit human monocyte-
328 derived dendritic cells, potentially exacerbating inflammatory responses (Zaal et al., 2017a, Zaal
329 et al., 2017b).

330   However, there are no studies on the correlation between BAZ2B, PDK4, and STK3 genes and
331   ischemic stroke. The biological function of BAZ2B, remains unclear, besides the involvement in
332   nucleosome remodelling (Oppikofer et al., 2017). But it includes at least four functional domains
333   that could encode or bind multiple proteins or DNA to perform various roles. Due to this
334   characteristic, we speculate that BAZ2B may be associated with many diseases, including
335   ischemic stroke. But current studies are limited to suggesting that it is associated with
336   neurodevelopment and that its functional loss and haploinsufficiency is one of the causes of
337   autism and may act through transcriptional regulation (Krupp et al., 2017, Scott et al., 2020).
338   Another study found that BAZ2B activates M2 macrophages to participate in the inflammatory
339   response (Xia et al., 2021), indicating its research value in immunity. PDK4 is a mitochondrial
340   matrix enzyme essential in cellular energy regulation, which regulates the pyruvate
341   dehydrogenase complex in the central nervous system and has important implications for neuron-
342   glia metabolic interactions (Jha et al., 2012). Atherosclerosis is the result of cholesterol and lipid
343   deposition in the arterial wall, usually associated with calcification, and is the most common
344   cause of ischemic stroke. Ma et al found that PDK4 could promote vascular calcification by
345   interfering with autophagic activity and metabolic reprogramming (Ma et al., 2020), contributing
346   to the development of atherosclerosis. STK3 (MST2) is a component of the MAPK module and
347   Hippo signaling pathway. In studies of cardiovascular disease, STK3 was found to mediate
348   mir155 to initiate inflammation and redox stress, leading to vascular smooth muscle cell
349   proliferation and remodeling. It regulates the ERK1/2 signaling pathway by competing with
350   MAP2K of the MAPK pathway, which in turn triggers inflammation and oxidative stress after
351   vascular injury (Thiriet, 2018). And in ischemic stroke, the MAPK pathway also acts by
352   regulating cytokines, inflammation, apoptosis and death (Sun and Nan, 2016, Qin et al., 2022),
353   STK3 is highly expressed in most cell types in the brain and may play a role in stroke by
354   regulating the MAPK pathway. It has also been described as a substrate for CASP6 to intervene
355   in neurodegeneration and apoptosis. (van Raam et al., 2013, Riechers et al., 2016). Cho et al.
356   found that STK3 increases phagocytosis of adipocytes, leading to obesity due to reduced
357   catabolic function of adipocytes. In obese human patients, STK3 expression levels were
358   elevated. STK3 inhibitors improved metabolic patterns in obese mouse models, suggesting that
359   there may be a viable pathway to investigate and develop drugs targeting STK3 to treat obesity-
360   related diseases including stroke (Cho et al., 2021). All of the above suggest that STK3 may be
361   relevant to the onset of ischemic stroke, but the exact mechanism needs to be further
362   investigated.

363   The critical pathways enriched by GO and KEGG analysis were related to immunity, including
364   the immune response−regulating signaling pathway, immune response−regulating cell surface
365   receptor signaling pathway, and MAPK signaling pathway. Notably, C5AR1 and STK3 were
366   enriched in the positive regulatory pathway of the MAPK cascade. Previous studies have found
367   that MAPK signaling pathway in inflammation and BBB dysfunction MAPK comprises three
368   main effectors: ERK1/2, JNK, and p38 (Sun and Nan, 2016, Qin et al., 2022). C5AR1 affects
369   inflammation in bone by activating the MAPK pathway and regulating gene expression in

370 pathways associated with insulin and transforming growth factor-β (Modinger et al., 2018).
371 STK3, as described above, mediates cardiovascular injury through the MAPK pathway and also
372 promotes apoptosis by inducing the activation of JNK (Chen et al., 2018). These results are
373 consistent with our bioinformatics analysis and suggest that the MAPK signaling pathway plays
374 an important role in these key gene-mediated biological processes. In addition, the results of the
375 GSEA enrichment analysis showed that C5AR1 was involved in complement, and coagulation
376 cascades pathway which had been proven to correlate with IS (Berkowitz et al., 2021). And the
377 functions of PDK4 are all focused on the involvement of lipid metabolism in the process of
378 atherosclerosis. These results are also in accordance with the role of C5AR1 and PDK4 found in
379 some studies (Berkowitz et al., 2021, Ma et al., 2020).

380 Many stroke models have confirmed that neutrophils in the acute phase of ischemic stroke are
381 among the earliest immune cells recruited into the brain tissue (Gokhan et al., 2013, Kaito et al.,
382 2013, Jickling et al., 2015). Activated neutrophils may contribute to the development of ischemic
383 stroke by stimulating the systemic inflammatory response and destroying the BBB. Van Duijn et
384 al. found that CD8 T cells can exacerbate the inflammatory response by secreting various
385 inflammatory cytokines, resulting in increased instability of atherosclerotic plaques. But CD8 T
386 cells also have an anti-atherosclerotic effect, which is mediated by stimulation of inhibitory
387 receptor production and cytolytic killing of antigen-presenting cells (van Duijn et al., 2018). The
388 results of immunological infiltration analysis by CIBERSORT revealed that Monocytes,
389 macrophage M0, Neutrophils, and Mast cells activated were increased, while T cells follicular
390 helper and CD8 decreased infiltration in IS group, which was consistent with the previous
391 studies (Zheng et al., 2022). Our research also showed that neutrophils were significantly and
392 positively correlated with gene ANTXR2, BAZ2B, C5AR1, PDK4, and STK3, negatively
393 correlated with PPIH; CD8was significantly and negatively correlated with gene ANTXR2,
394 BAZ2B, C5AR1, PDK4, and STK3, positively correlated with PPIH. This evidence further
395 suggests that neutrophils play a significant role, and hub genes are involved in regulating the
396 immune microenvironment in IS patients.

397 The ceRNA network we conducted included nodes (6 mRNAs, 249 miRNAs, 218 lncRNAs)
398 with 591 edges. In these nodes, the miRNAs with high degrees, hsa-miR-766-3p, hsa-miR-149-
399 3p, and hsa-miR-186-5p, have been reported as essential molecules in cerebral
400 ischemic/reperfusion injury(Cai et al., 2019, Hu et al., 2020, You et al., 2022). But the mode of
401 regulation between them and hub genes still needs further validation. Meanwhile, we constructed
402 a Hub gene-based Gene-Drug regulatory network and predicted possible targeted drugs. The
403 result reflected that BAZ2B had the most targets of drug action and the likely reason was that it
404 had multiple functional areas. The network provides a theoretical basis for developing targeted
405 immunotherapies for IS.

406 Finally, based on the four key genes mentioned above, we constructed an IS diagnostic model
407 using logistic regression methods. We also verified the validity of the model in a publicly

408 available dataset. Bioinformatics analysis combined with machine learning methods achieved
409 good results.

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

Conclusions

417 We explored 6 hub genes (ANTXR2, BAZ2B, C5AR1, PDK4, PPIH, and STK3) for ischemic
418 stroke and revealed the immune cell infiltration pattern via bioinformatics analysis and machine
419 learning algorithm. Secondly, we successfully constructed the ceRNA and Drug-Gene networks,
420 which provide new ideas to explore the possible molecular mechanisms and develop targeted
421 drugs of IS. Finally, BAZ2B, C5AR1, PDK4, and STK3 were verified to be the biological
422 markers of ischemic stroke. Based on the 4 genes, we developed a diagnostic model and
423 validated its effectiveness, providing additional tools for screening and diagnosing of ischemic
424 stroke.

Acknowledgements

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

References

429 Anrather J, and Iadecola C. 2016. Inflammation and Stroke: An Overview. *Neurotherapeutics*
430   13:661-670. 10.1007/s13311-016-0483-x

431 Ardestani A, Lupse B, and Maedler K. 2018. Hippo Signaling: Key Emerging Pathway in
432   Cellular and Whole-Body Metabolism. *Trends Endocrinol Metab* 29:492-509.
433   10.1016/j.tem.2018.04.006

434 Arevalo MT, Navarro A, Arico CD, Li J, Alkhatib O, Chen S, Diaz-Arevalo D, and Zeng M.
435   2014. Targeted silencing of anthrax toxin receptors protects against anthrax toxins. *J Biol*
436   *Chem* 289:15730-15738. 10.1074/jbc.M113.538587

437 Arking DE, Junttila MJ, Goyette P, Huertas-Vazquez A, Eijgelsheim M, Blom MT, Newton-
438   Cheh C, Reinier K, Teodorescu C, Uy-Evanado A, Carter-Monroe N, Kaikkonen KS,
439   Kortelainen ML, Boucher G, Lagace C, Moes A, Zhao X, Kolodgie F, Rivadeneira F,
440   Hofman A, Witteman JC, Uitterlinden AG, Marsman RF, Pazoki R, Bardai A, Koster
441   RW, Dehghan A, Hwang SJ, Bhatnagar P, Post W, Hilton G, Prineas RJ, Li M, Kottgen
442   A, Ehret G, Boerwinkle E, Coresh J, Kao WH, Psaty BM, Tomaselli GF, Sotoodehnia N,
443   Siscovick DS, Burke GL, Marban E, Spooner PM, Cupples LA, Jui J, Gunson K,

444   Kesaniemi YA, Wilde AA, Tardif JC, O'Donnell CJ, Bezzina CR, Virmani R, Stricker
445   BH, Tan HL, Albert CM, Chakravarti A, Rioux JD, Huikuri HV, and Chugh SS. 2011.
446   Identification of a sudden cardiac death susceptibility locus at 2q24.2 through genome-
447   wide association in European ancestry individuals. *PLoS Genet* 7:e1002158.
448   10.1371/journal.pgen.1002158

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

452 Brandolini L, Grannonico M, Bianchini G, Colanardi A, Sebastiani P, Paladini A, Piroli A,
453   Allegretti M, Varrassi G, and Di Loreto S. 2019. The Novel C5aR Antagonist DF3016A
454   Protects Neurons Against Ischemic Neuroinflammatory Injury. *Neurotox Res* 36:163-174.
455   10.1007/s12640-019-00026-w

456 Cai J, Shangguan S, Li G, Cai Y, Chen Y, Ma G, Miao Z, Liu L, and Deng Y. 2019. Knockdown
457   of lncRNA Gm11974 protect against cerebral ischemic reperfusion through miR-766-
458   3p/NR3C2 axis. *Artif Cells Nanomed Biotechnol* 47:3847-3853.
459   10.1080/21691401.2019.1666859

460 Campbell BCV, and Khatri P. 2020. Stroke. *Lancet* 396:129-142. 10.1016/S0140-
461   6736(20)31179-X

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

465 Chackerian B, and Remaley A. 2016. Vaccine strategies for lowering LDL by immunization
466   against proprotein convertase subtilisin/kexin type 9. *Curr Opin Lipidol* 27:345-350.
467   10.1097/MOL.0000000000000312

468 Cho YK, Son Y, Saha A, Kim D, Choi C, Kim M, Park JH, Im H, Han J, Kim K, Jung YS, Yun
469   J, Bae EJ, Seong JK, Lee MO, Lee S, Granneman JG, and Lee YH. 2021. STK3/STK4
470   signalling in adipocytes regulates mitophagy and energy expenditure. *Nat Metab* 3:428-
471   441. 10.1038/s42255-021-00362-2

472 Cohen JC, Boerwinkle E, Mosley TH, Jr., and Hobbs HH. 2006. Sequence variations in PCSK9,
473   low LDL, and protection against coronary heart disease. *N Engl J Med* 354:1264-1272.
474   10.1056/NEJMoa054013

475 Dennis MK, and Mogridge J. 2014. A protective antigen mutation increases the pH threshold of
476   anthrax toxin receptor 2-mediated pore formation. *Biochemistry* 53:2166-2171.
477   10.1021/bi5000756

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

480 Feng Y, Chen S, Zhou M, Zhang J, Min J, and Liu K. 2022. Crystal structure of the BAZ2B
481   TAM domain. *Heliyon* 8:e09873. 10.1016/j.heliyon.2022.e09873

482 Friedman J, Hastie T, and Tibshirani R. 2010. Regularization Paths for Generalized Linear
483   Models via Coordinate Descent. *J Stat Softw* 33:1-22.

484   Gokhan S, Ozhasenekler A, Mansur Durgun H, Akil E, Ustundag M, and Orak M. 2013.
485   Neutrophil lymphocyte ratios in stroke subtypes and transient ischemic attack. *Eur Rev*
486   *Med Pharmacol Sci* 17:653-657.

487   Hasan TF, Rabinstein AA, Middlebrooks EH, Haranhalli N, Silliman SL, Meschia JF, and Tawk
488   RG. 2018. Diagnosis and Management of Acute Ischemic Stroke. *Mayo Clin Proc*
489   93:523-538. 10.1016/j.mayocp.2018.02.013

490   He J, Li X, Zhou J, and Hu R. 2022. BATF2 and PDK4 as diagnostic molecular markers of
491   sarcoidosis and their relationship with immune infiltration. *Ann Transl Med* 10:106.
492   10.21037/atm-22-180

493   Horowitz DS, Lee EJ, Mabon SA, and Misteli T. 2002. A cyclophilin functions in pre-mRNA
494   splicing. *EMBO J* 21:470-480. 10.1093/emboj/21.3.470

495   Hu X, Xiang Z, Zhang W, Yu Z, Xin X, Zhang R, Deng Y, and Yuan Q. 2020. Protective effect
496   of DLX6-AS1 silencing against cerebral ischemia/reperfusion induced impairments.
497   *Aging (Albany NY)* 12:23096-23113. 10.18632/aging.104070

498   Iadecola C, Buckwalter MS, and Anrather J. 2020. Immune responses to stroke: mechanisms,
499   modulation, and therapeutic potential. *J Clin Invest* 130:2777-2788. 10.1172/JCI135530

500   Javidi E, and Magnus T. 2019. Autoimmunity After Ischemic Stroke and Brain Injury. *Front*
501   *Immunol* 10:686. 10.3389/fimmu.2019.00686

502   Jayaraj RL, Azimullah S, Beiram R, Jalal FY, and Rosenberg GA. 2019. Neuroinflammation:
503   friend and foe for ischemic stroke. *J Neuroinflammation* 16:142. 10.1186/s12974-019-
504   1516-2

505   Jickling GC, Liu D, Ander BP, Stamova B, Zhan X, and Sharp FR. 2015. Targeting neutrophils
506   in ischemic stroke: translational insights from experimental studies. *J Cereb Blood Flow*
507   *Metab* 35:888-901. 10.1038/jcbfm.2015.45

508   Kaito M, Araya S, Gondo Y, Fujita M, Minato N, Nakanishi M, and Matsui M. 2013. Relevance
509   of distinct monocyte subsets to clinical course of ischemic stroke patients. *PLoS One*
510   8:e69409. 10.1371/journal.pone.0069409

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

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

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

520   Levard D, Buendia I, Lanquetin A, Glavan M, Vivien D, and Rubio M. 2021. Filling the gaps on
521   stroke research: Focus on inflammation and immunity. *Brain Behav Immun* 91:649-667.
522   10.1016/j.bbi.2020.09.025

523 Ma WQ, Sun XJ, Zhu Y, and Liu NF. 2020. PDK4 promotes vascular calcification by interfering
524   with autophagic activity and metabolic reprogramming. *Cell Death Dis* 11:991.
525   10.1038/s41419-020-03162-w

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

530 Moriconi A, Cunha TM, Souza GR, Lopes AH, Cunha FQ, Carneiro VL, Pinto LG, Brandolini
531   L, Aramini A, Bizzarri C, Bianchini G, Beccari AR, Fanton M, Bruno A, Costantino G,
532   Bertini R, Galliera E, Locati M, Ferreira SH, Teixeira MM, and Allegretti M. 2014.
533   Targeting the minor pocket of C5aR for the rational design of an oral allosteric inhibitor
534   for inflammatory and neuropathic pain relief. *Proc Natl Acad Sci U S A* 111:16937-
535   16942. 10.1073/pnas.1417365111

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

538 O'Connell GC, Chantler PD, and Barr TL. 2017. Stroke-associated pattern of gene expression
539   previously identified by machine-learning is diagnostically robust in an independent
540   patient population. *Genom Data* 14:47-52. 10.1016/j.gdata.2017.08.006

541 O'Connell GC, Petrone AB, Treadway MB, Tennant CS, Lucke-Wold N, Chantler PD, and Barr
542   TL. 2016. Machine-learning approach identifies a pattern of gene expression in
543   peripheral blood that can accurately detect ischaemic stroke. *NPJ Genom Med* 1:16038.
544   10.1038/npjgenmed.2016.38

545 Oberbach A, Adams V, Schlichting N, Heinrich M, Kullnick Y, Lehmann S, Lehmann S, Feder
546   S, Correia JC, Mohr FW, Volker U, and Jehmlich N. 2016. Proteome profiles of HDL
547   particles of patients with chronic heart failure are associated with immune response and
548   also include bacteria proteins. *Clin Chim Acta* 453:114-122. 10.1016/j.cca.2015.12.005

549 Okada T, Ohama T, Takafuji K, Kanno K, Matsuda H, Sairyo M, Zhu Y, Saga A, Kobayashi T,
550   Masuda D, Koseki M, Nishida M, Sakata Y, and Yamashita S. 2019. Shotgun proteomic
551   analysis reveals proteome alterations in HDL of patients with cholesteryl ester transfer
552   protein deficiency. *J Clin Lipidol* 13:317-325. 10.1016/j.jacl.2019.01.002

553 Pamir N, Pan C, Plubell DL, Hutchins PM, Tang C, Wimberger J, Irwin A, Vallim TQA,
554   Heinecke JW, and Lusis AJ. 2019. Genetic control of the mouse HDL proteome defines
555   HDL traits, function, and heterogeneity. *J Lipid Res* 60:594-608. 10.1194/jlr.M090555

556 Papasaikas P, Tejedor JR, Vigevani L, and Valcarcel J. 2015. Functional splicing network
557   reveals extensive regulatory potential of the core spliceosomal machinery. *Mol Cell* 57:7-
558   22. 10.1016/j.molcel.2014.10.030

559 Rajiv C, and Davis TL. 2018. Structural and Functional Insights into Human Nuclear
560   Cyclophilins. *Biomolecules* 8. 10.3390/biom8040161

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

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

568   Riwanto M, Rohrer L, Roschitzki B, Besler C, Mocharla P, Mueller M, Perisa D, Heinrich K,
569   Altwegg L, von Eckardstein A, Luscher TF, and Landmesser U. 2013. Altered activation
570   of endothelial anti- and proapoptotic pathways by high-density lipoprotein from patients
571   with coronary artery disease: role of high-density lipoprotein-proteome remodeling.
572   *Circulation* 127:891-904. 10.1161/CIRCULATIONAHA.112.108753

573   Scott TM, Guo H, Eichler EE, Rosenfeld JA, Pang K, Liu Z, Lalani S, Bi W, Yang Y, Bacino
574   CA, Streff H, Lewis AM, Koenig MK, Thiffault I, Bellomo A, Everman DB, Jones JR,
575   Stevenson RE, Bernier R, Gilissen C, Pfundt R, Hiatt SM, Cooper GM, Holder JL, and
576   Scott DA. 2020. BAZ2B haploinsufficiency as a cause of developmental delay,
577   intellectual disability, and autism spectrum disorder. *Hum Mutat* 41:921-925.
578   10.1002/humu.23992

579   Shi Y, Chen X, Liu J, Fan X, Jin Y, Gu J, Liang J, Liang X, and Wang C. 2021. Isoquercetin
580   Improves Inflammatory Response in Rats Following Ischemic Stroke. *Front Neurosci*
581   15:555543. 10.3389/fnins.2021.555543

582   Shi YW, Zhang XC, Chen C, Tang M, Wang ZW, Liang XM, Ding F, and Wang CP. 2017.
583   Schisantherin A attenuates ischemia/reperfusion-induced neuronal injury in rats via
584   regulation of TLR4 and C5aR1 signaling pathways. *Brain Behav Immun* 66:244-256.
585   10.1016/j.bbi.2017.07.004

586   Smith CJ, Hulme S, Vail A, Heal C, Parry-Jones AR, Scarth S, Hopkins K, Hoadley M, Allan
587   SM, Rothwell NJ, Hopkins SJ, and Tyrrell PJ. 2018. SCIL-STROKE (Subcutaneous
588   Interleukin-1 Receptor Antagonist in Ischemic Stroke): A Randomized Controlled Phase
589   2 Trial. *Stroke* 49:1210-1216. 10.1161/STROKEAHA.118.020750

590   Stamova B, Jickling GC, Ander BP, Zhan X, Liu D, Turner R, Ho C, Khoury JC, Bushnell C,
591   Pancioli A, Jauch EC, Broderick JP, and Sharp FR. 2014. Gene expression in peripheral
592   immune cells following cardioembolic stroke is sexually dimorphic. *PLoS One*
593   9:e102550. 10.1371/journal.pone.0102550

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

598   Tejedor JR, Papasaikas P, and Valcarcel J. 2015. Genome-wide identification of Fas/CD95
599   alternative splicing regulators reveals links with iron homeostasis. *Mol Cell* 57:23-38.
600   10.1016/j.molcel.2014.10.029

601   Vaisar T, Tang C, Babenko I, Hutchins P, Wimberger J, Suffredini AF, and Heinecke JW. 2015.
602   Inflammatory remodeling of the HDL proteome impairs cholesterol efflux capacity. *J*
603   *Lipid Res* 56:1519-1530. 10.1194/jlr.M059089

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

606 van Raam BJ, Ehrnhoefer DE, Hayden MR, and Salvesen GS. 2013. Intrinsic cleavage of
607   receptor-interacting protein kinase-1 by caspase-6. *Cell Death Differ* 20:86-96.
608   10.1038/cdd.2012.98

609 Wang Q, and Liu X. 2015. Screening of feature genes in distinguishing different types of breast
610   cancer using support vector machine. *Onco Targets Ther* 8:2311-2317.
611   10.2147/OTT.S85271

612 Xu J, Zhang X, Monestier M, Esmon NL, and Esmon CT. 2011. Extracellular histones are
613   mediators of death through TLR2 and TLR4 in mouse fatal liver injury. *J Immunol*
614   187:2626-2631. 10.4049/jimmunol.1003930

615 You J, Qian F, Huang Y, Guo Y, Lv Y, Yang Y, Lu X, Guo T, Wang J, and Gu B. 2022.
616   lncRNA WT1-AS attenuates hypoxia/ischemia-induced neuronal injury during cerebral
617   ischemic stroke via miR-186-5p/XIAP axis. *Open Med (Wars)* 17:1338-1349.
618   10.1515/med-2022-0528

619 Zaal A, Dieker M, Oudenampsen M, Turksma AW, Lissenberg-Thunnissen SN, Wouters D, van
620   Ham SM, and Ten Brinke A. 2017a. Anaphylatoxin C5a Regulates 6-Sulfo-LacNAc
621   Dendritic Cell Function in Human through Crosstalk with Toll-Like Receptor-Induced
622   CREB Signaling. *Front Immunol* 8:818. 10.3389/fimmu.2017.00818

623 Zaal A, Nota B, Moore KS, Dieker M, van Ham SM, and Ten Brinke A. 2017b. TLR4 and C5aR
624   crosstalk in dendritic cells induces a core regulatory network of RSK2, PI3Kbeta, SGK1,
625   and FOXO transcription factors. *J Leukoc Biol* 102:1035-1054. 10.1189/jlb.2MA0217-
626   058R

627 Zhang Q, Chen W, Chen S, Li S, Wei D, and He W. 2019. Identification of key genes and
628   upstream regulators in ischemic stroke. *Brain Behav* 9:e01319. 10.1002/brb3.1319

629 Zheng PF, Chen LZ, Liu P, Pan HW, Fan WJ, and Liu ZY. 2022. Identification of immune-
630   related key genes in the peripheral blood of ischaemic stroke patients using a weighted
631   gene coexpression network analysis and machine learning. *J Transl Med* 20:361.
632   10.1186/s12967-022-03562-w

633

Basic information ongene expression profiling

Table 1 Basic information on gene expression profiling

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

Figure 1

The technical route

Abbreviations: IS, ischemic stroke; CTL, Healthy Control

```
                                    ┌──────────────────┐      ┌─────────────────────────┐
                                    │ Immune infiltration │◄─────┤        GSE16561          │
                                    │  analysis by       │      │                          │
                                    │    CIBERSORT       │      └─────────────────────────┘
                                    └──────────────────┘                   │
                                                                           ▼
                                             ┌──────────────────────────┐      ┌─────────────────┐
                                             │ Differential expression  │─────►│  Volcano plot   │
                                             │        analysis          │      │    Heatmap      │
                                             └──────────────────────────┘      └─────────────────┘
                                                           │
                                                           ▼
                                             ┌──────────────────────────┐      ┌─────────────────┐
                                             │  Identify 188 DEGs       │─────►│ GO enrichment   │
                                             │ between IS and CTL group │      │ KEGG pathway    │
                                             └──────────────────────────┘      │     GSEA        │
                                                           │                    └─────────────────┘
                                                           ▼
   ┌──────────────────────────┐        ┌──────────────────────────┐      ┌─────────────────┐
   │ Correlation between hub  │◄───────┤  Screen out 6 hub genes  │─────►│  Venn diagram   │
   │ genes and immune cells   │        │  by LASSO and SVM-REF    │      │                 │
   └──────────────────────────┘        └──────────────────────────┘      └─────────────────┘
```

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/control group, respectively. Abbreviations: IS (ischemic stroke), Control (Healthy Control), DEGs (differential expression genes)

A

B

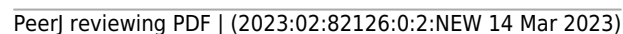



Figure 3

Show the top 10 significantly enriched BP (biological process), CC (cellular component) considerably, and MF (molecular function) in the barplot and the bubble plot.

(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. (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.

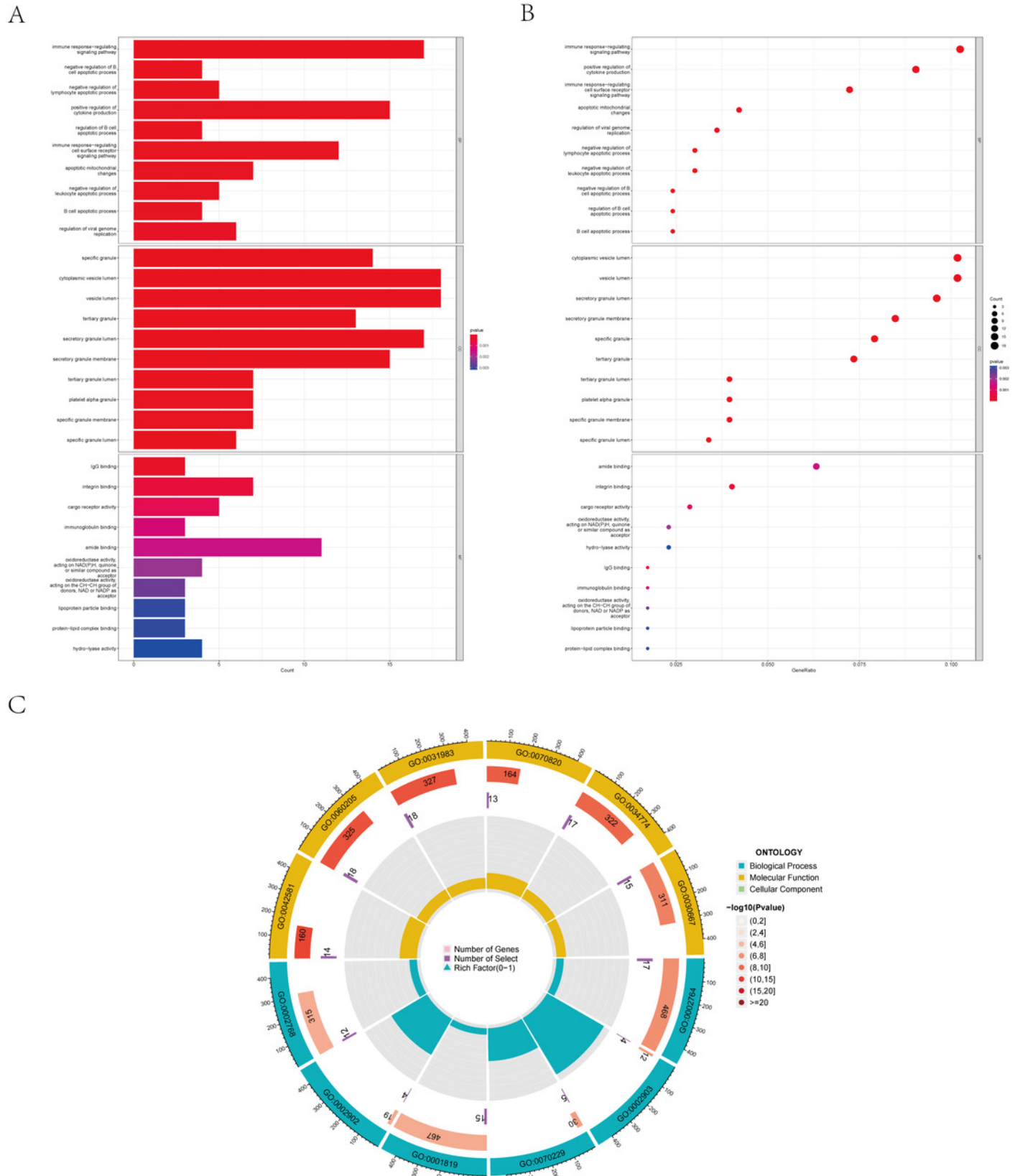



Figure 4

The KEGG pathway analysis results of DEGs are shown as the barplot and the bubble plot.

(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

B

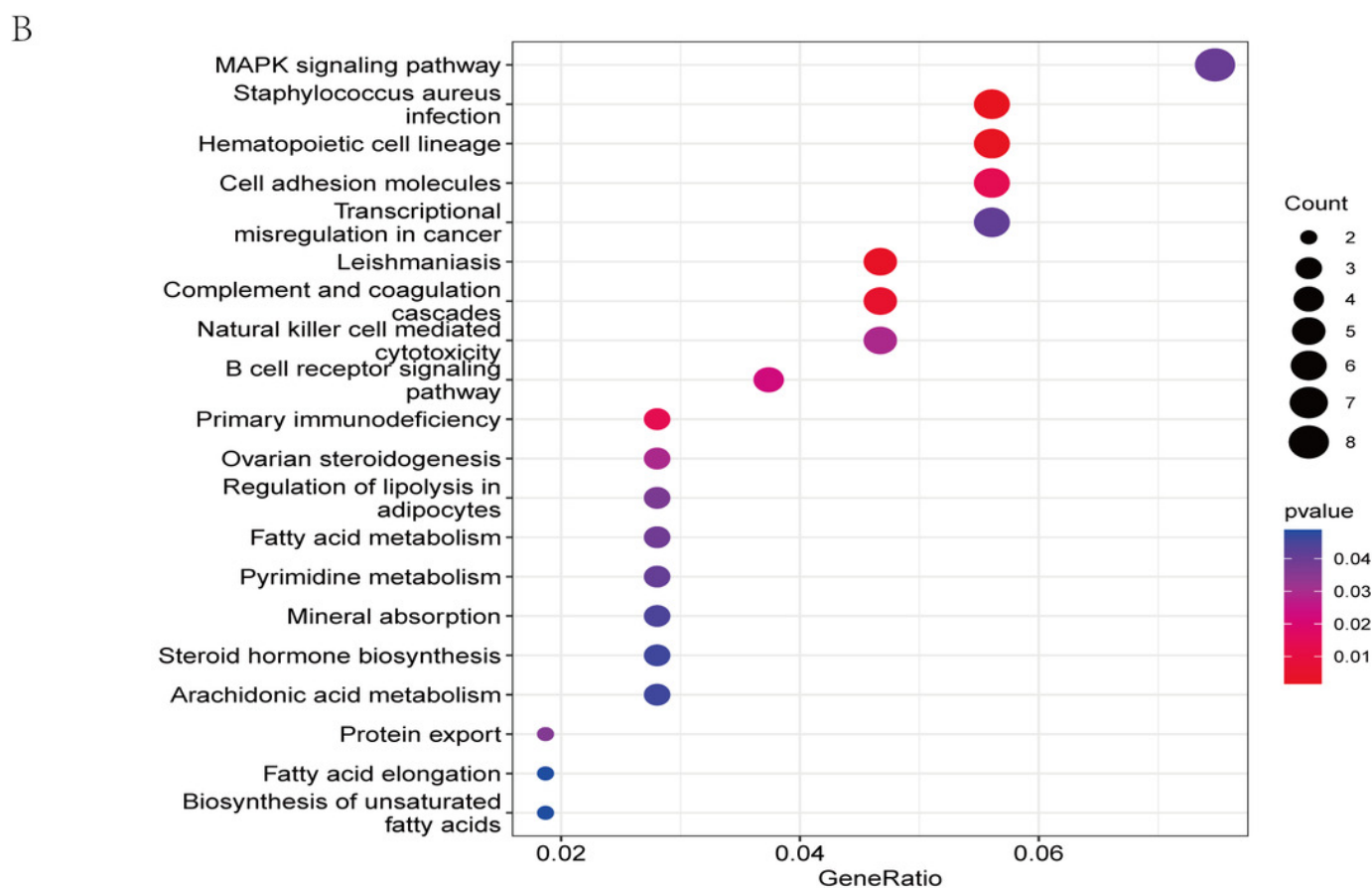
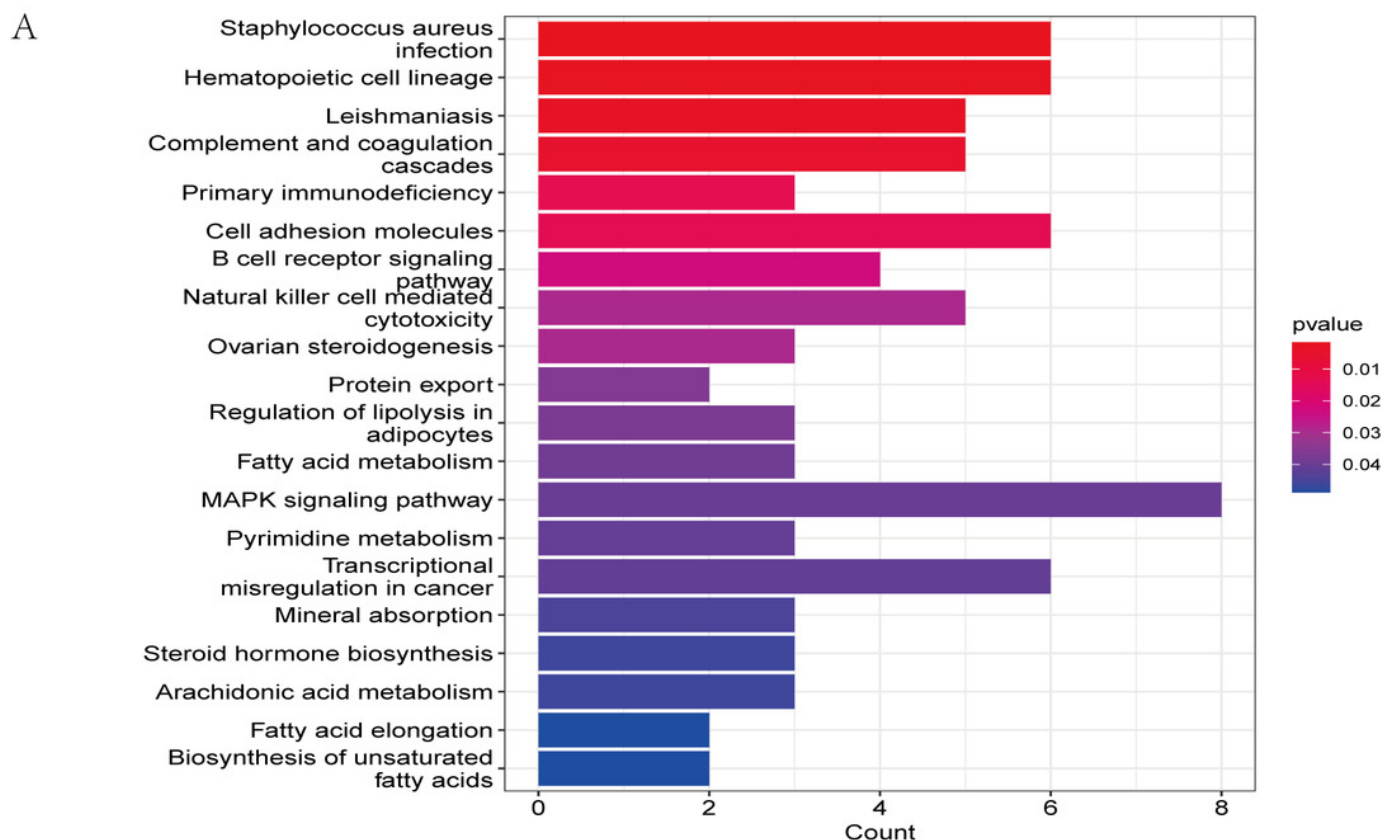



Figure 5

Gene Set Enrichment Analysis of IS and CTL group in the GSE dataset.

The different colours represent the different pathways obtained by enrichment. The figure has three parts. Part I: The top part is a line graph of the Gene Enrichment Score. 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 Enrichment score, while for the CTL group with a negative Enrichment score, the core genes are those after the peak. The horizontal axis represents each gene under this gene set and corresponds to the barcode-like vertical line in the second part. The second part: Hits, each vertical line corresponds to a gene under this gene set. The third part: is 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 IS group. (B) the GSEA results of the CTL group.

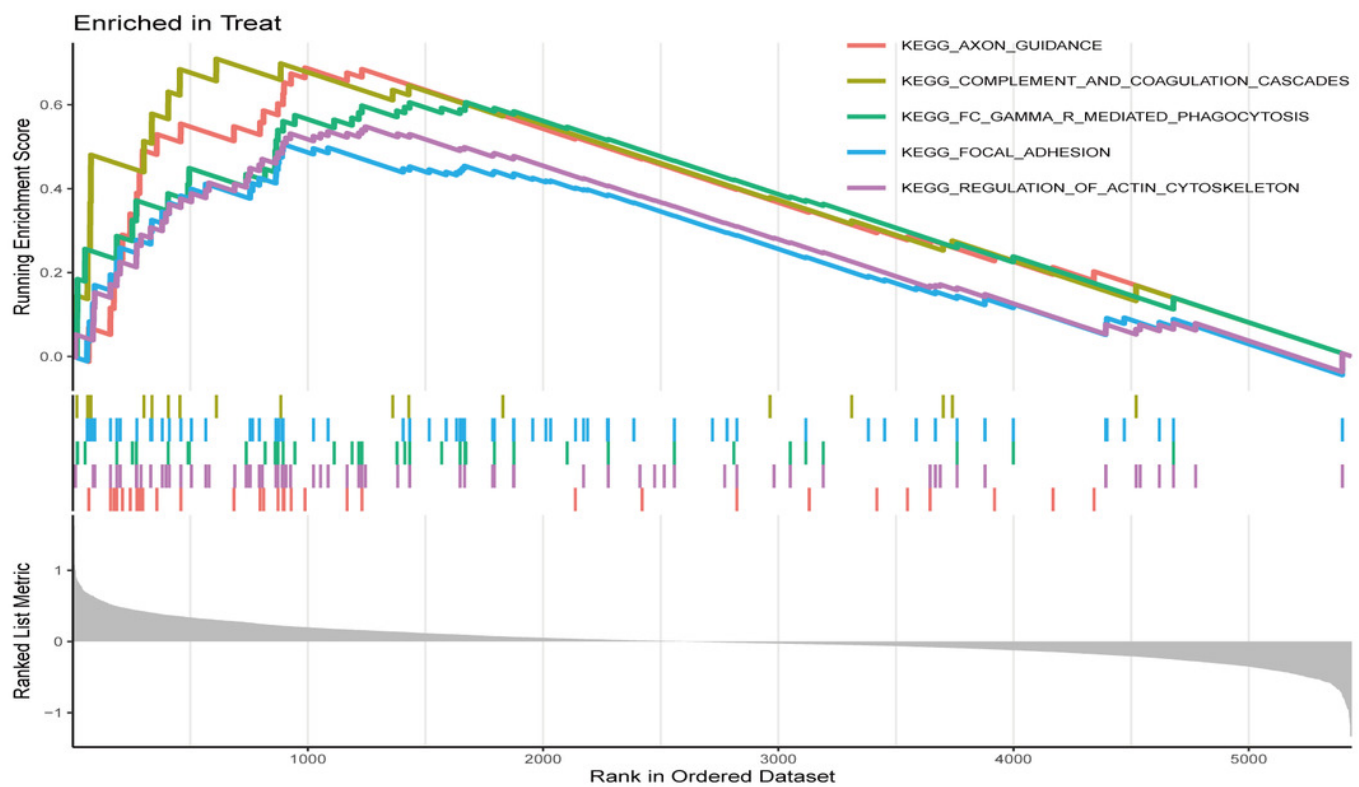
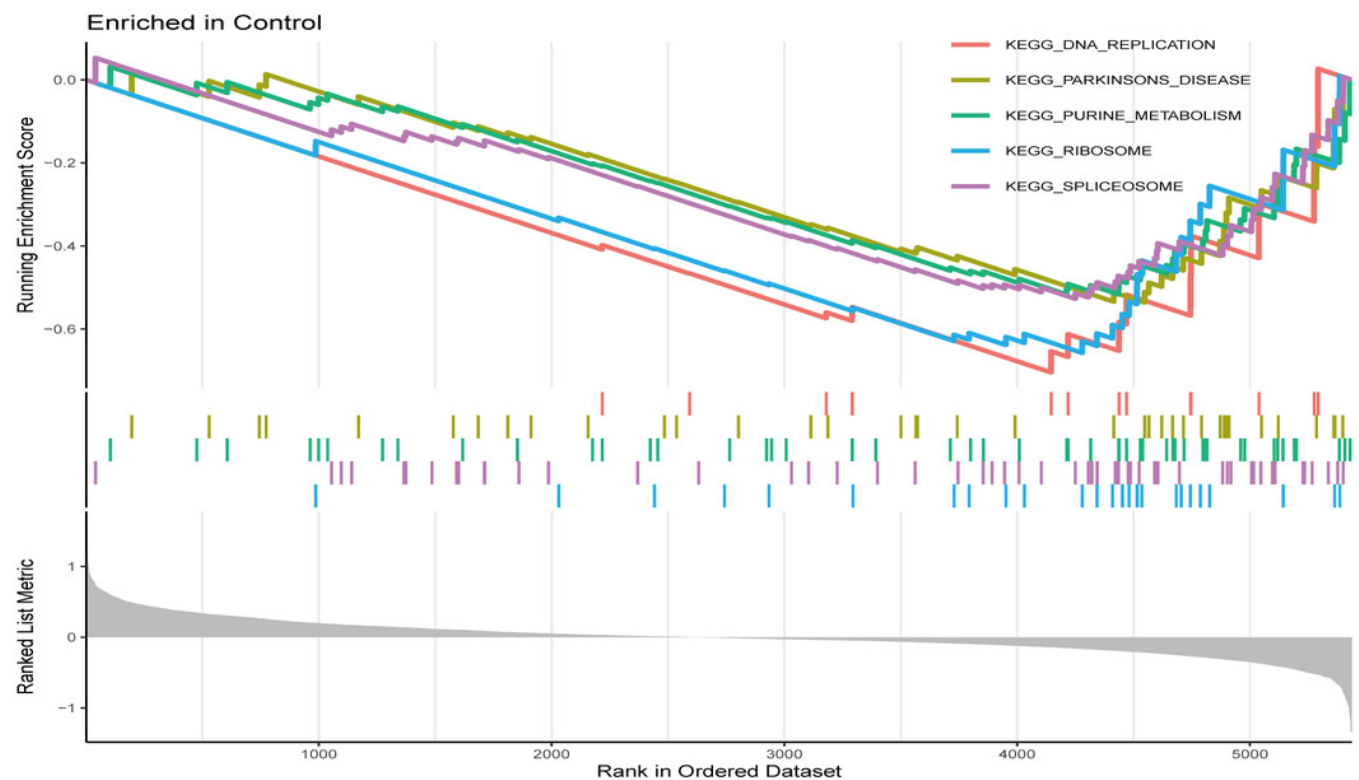



Figure 6

Immune infiltration landscape between IS and Control 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.

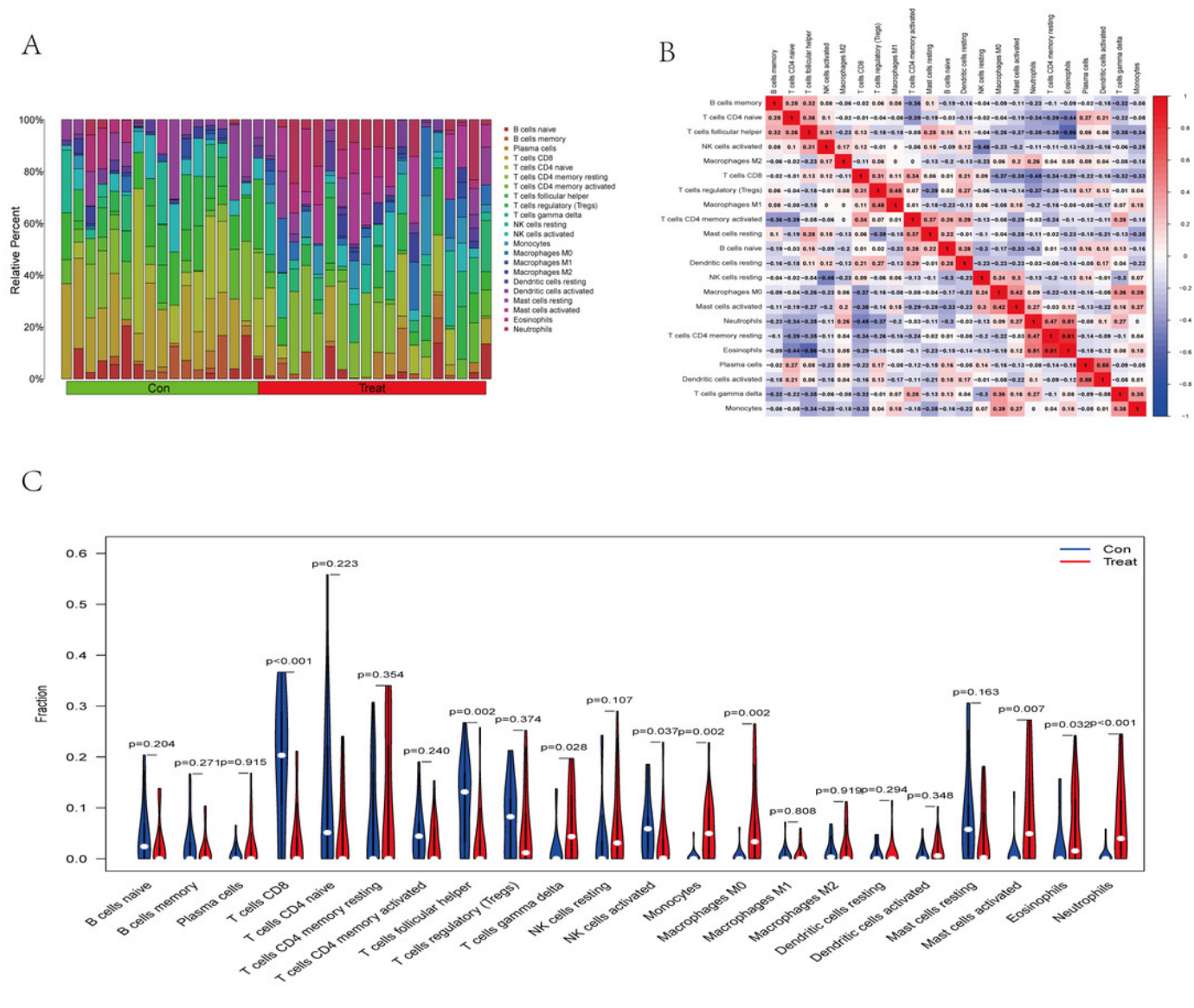



Figure 7

The identification of 6 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: LASSO(least absolute shrinkage and selection operator), SVM(support vector machine⬚, RFE (recursive feature elimination).

A

B

C

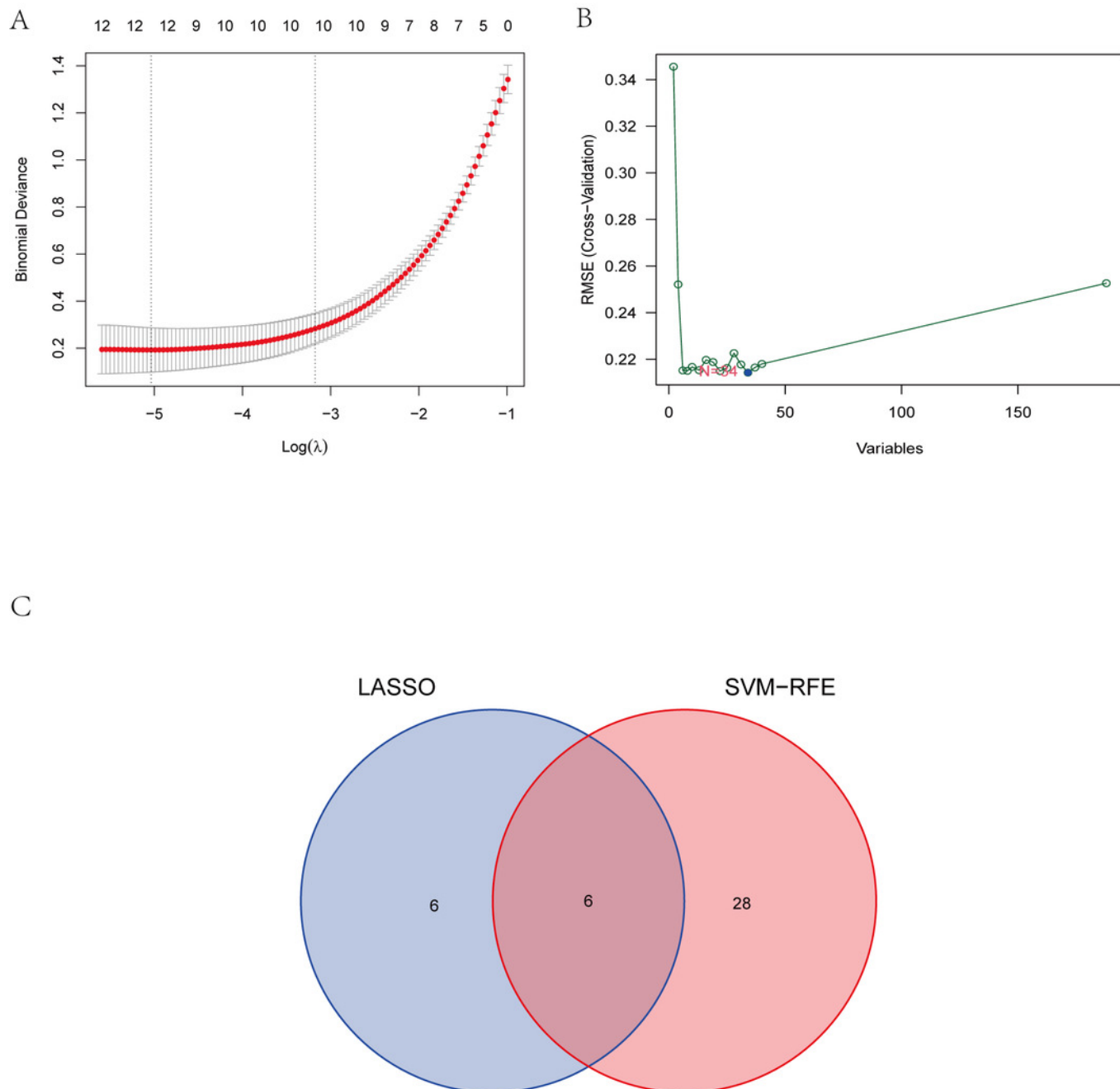



Figure 8

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

(A) *ANTXR2* **(B)** *BAZ2B* **(C)***C5AR1* **(D)***PDK4* **(E)** *PPIH* **(F)** *STK3*

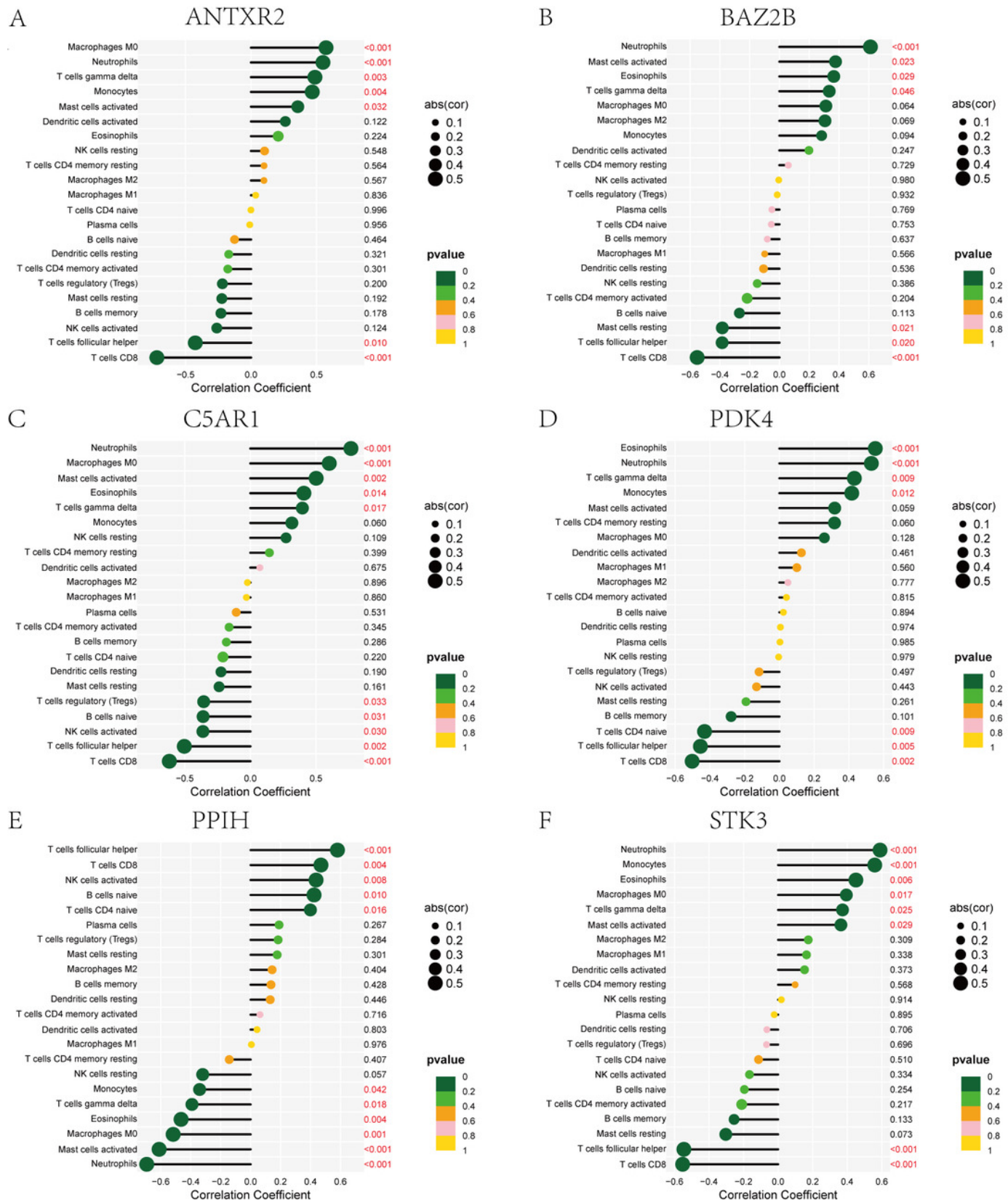



Figure 9

A ceRNA network based on Hub genes.

The network includes nodes (6 mRNAs, 249 miRNAs, 218 lncRNAs) with 591 edges. Red orbs represent Hub genes, green orbs represent miRNAs, and purple orbs represent lncRNAs.

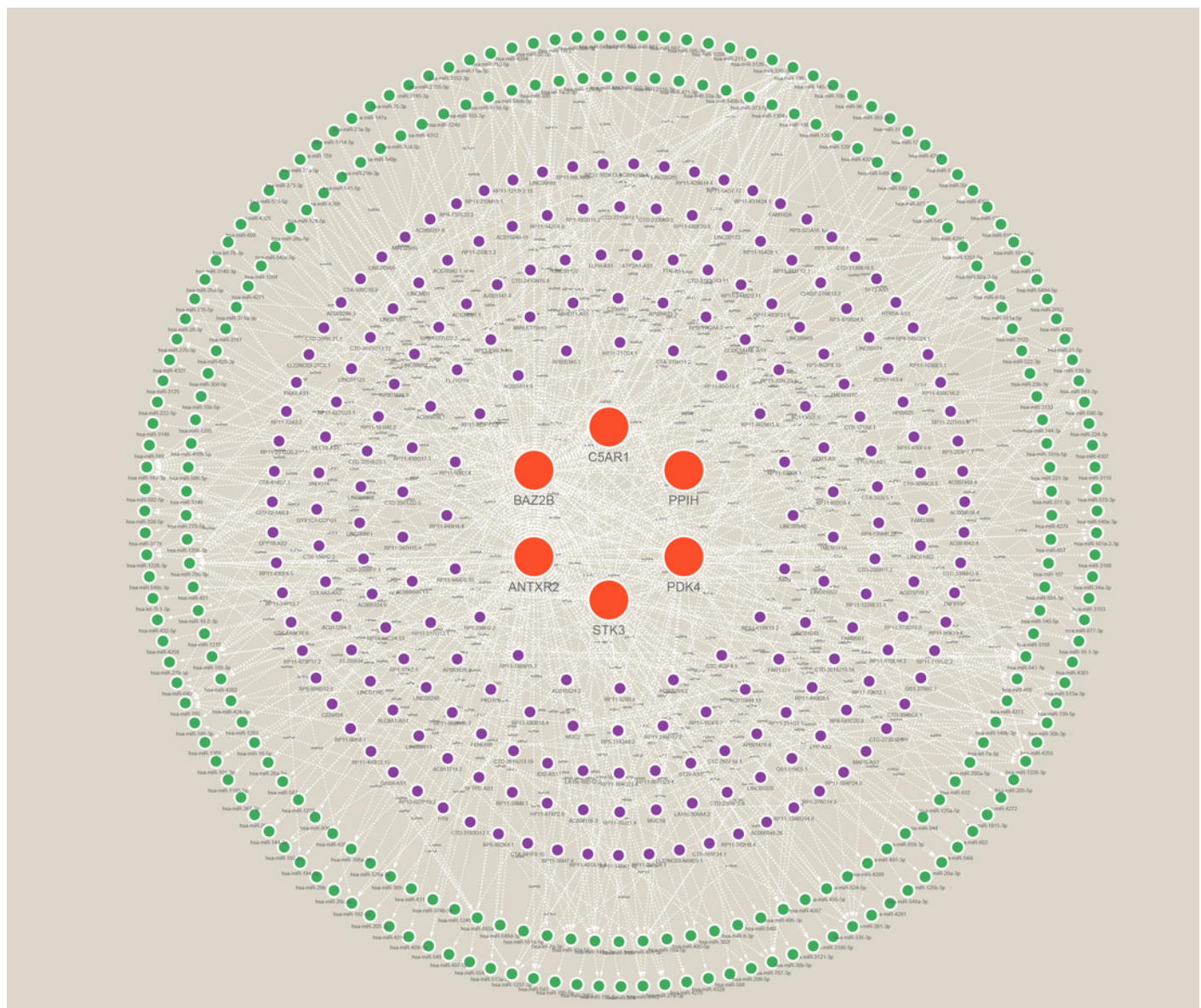



Figure 10

Prediction of marker gene-targeted drugs.

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

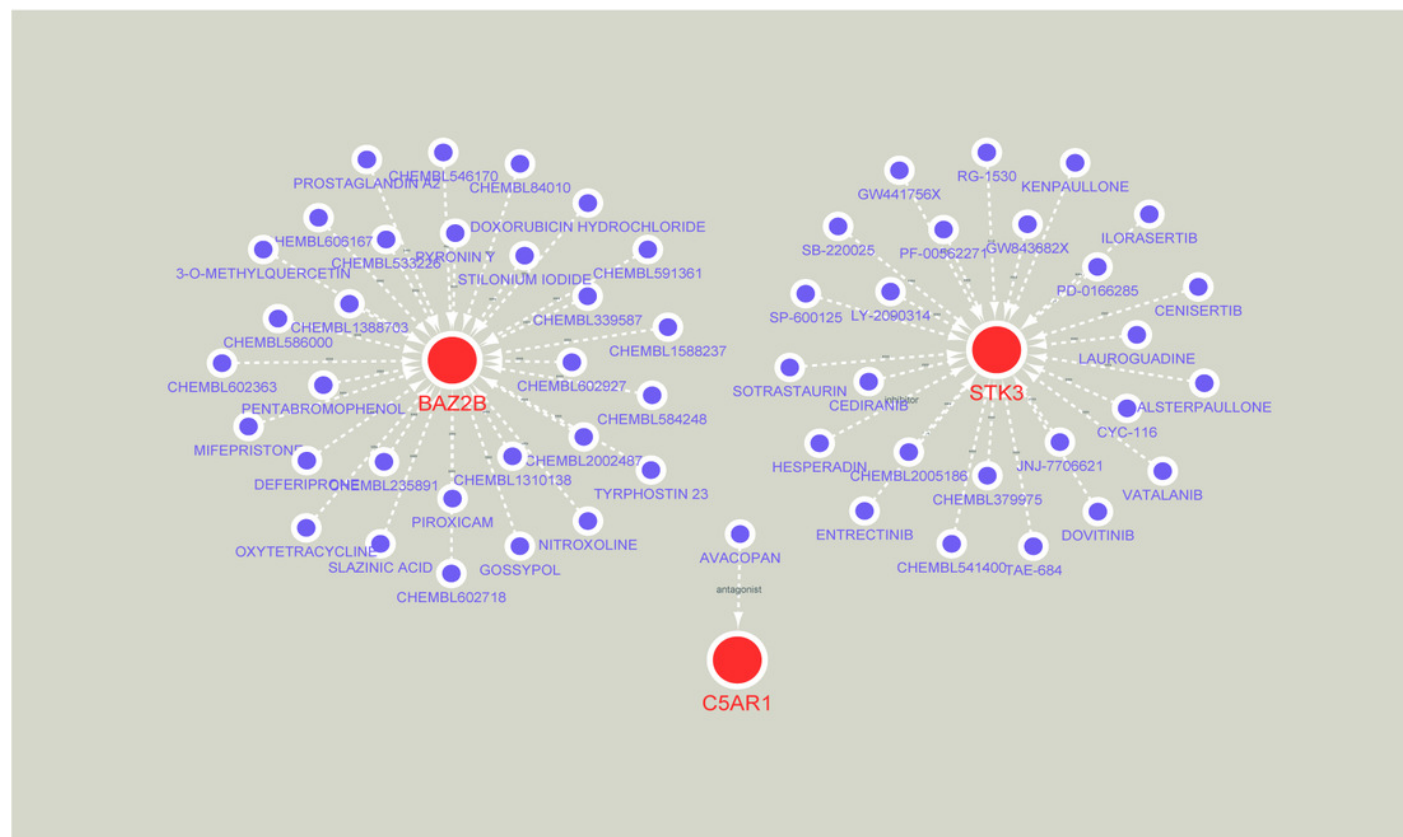



Figure 11

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,* P<0.05⬜ P≥0.01, **P<0.01

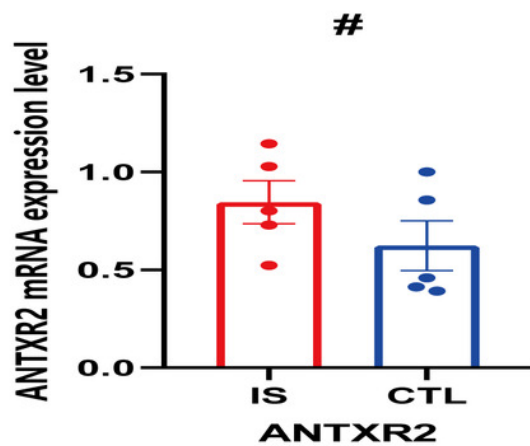
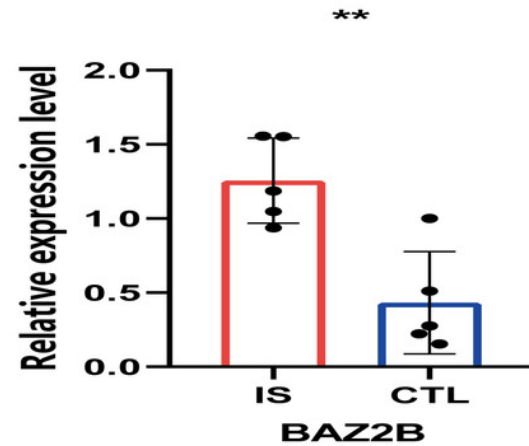
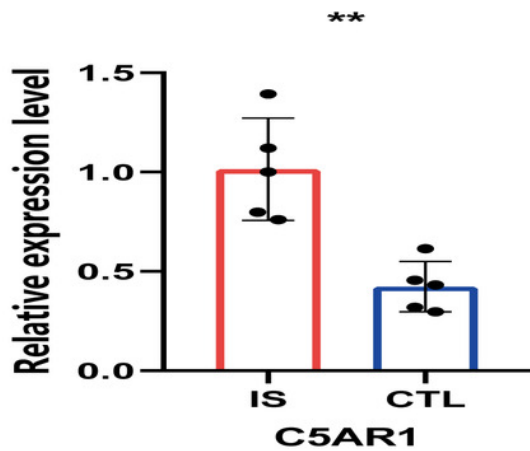
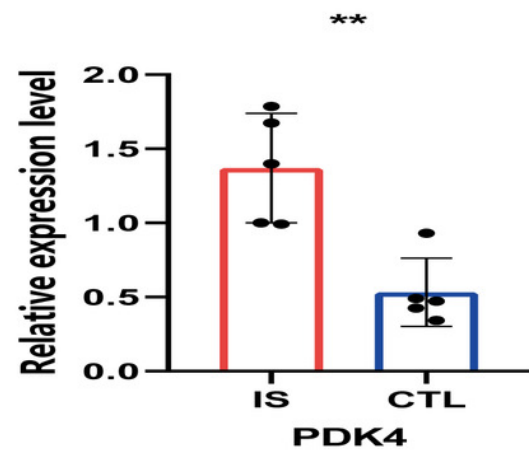
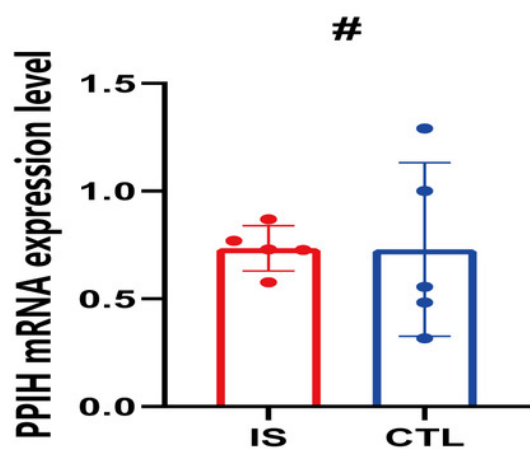
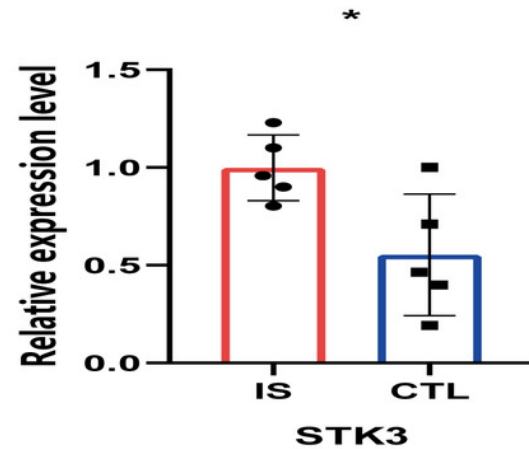
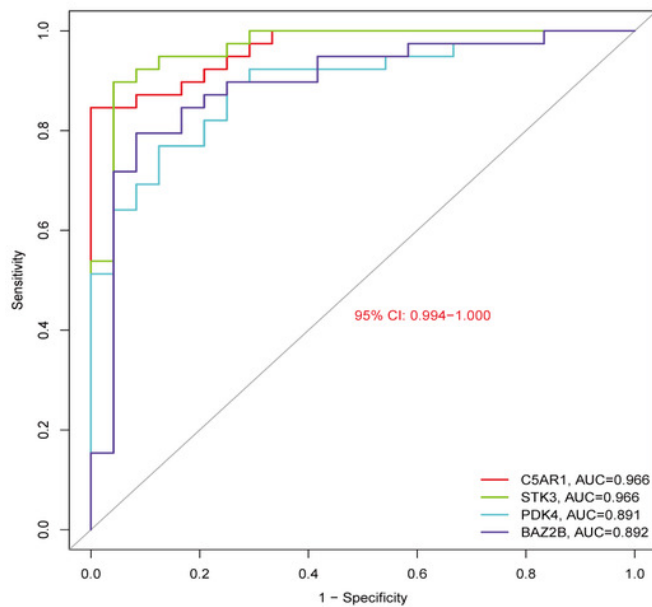
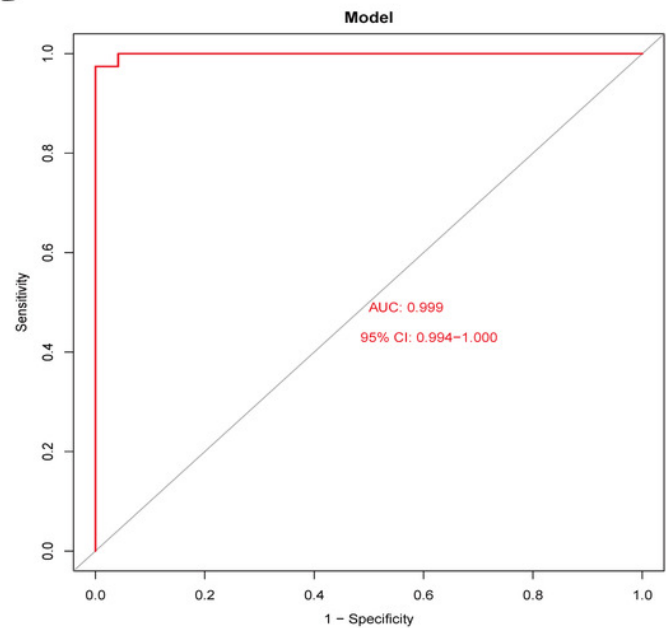
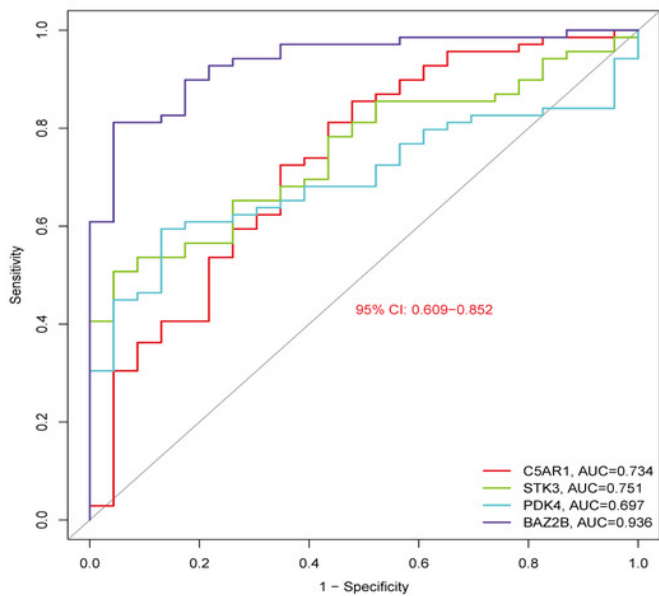



Figure 12

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

B

C

D

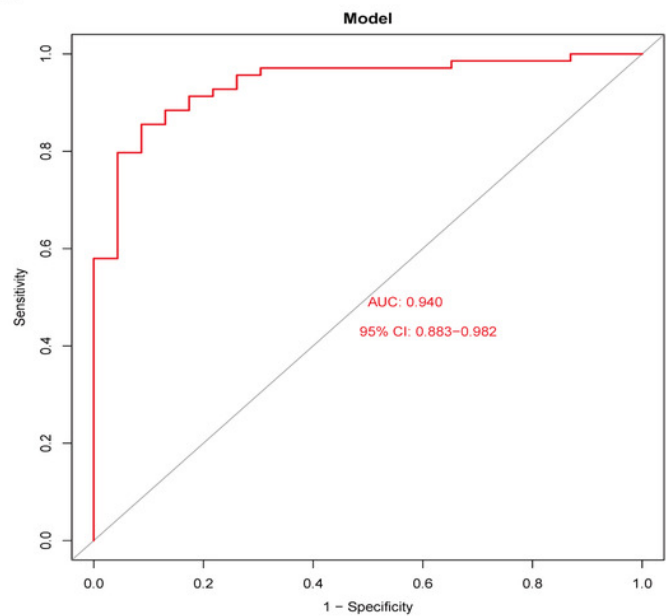



Figure 13

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.

A

B

C

D