Heliyon 10 (2024) e33846

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

# FCER1G as a novel immune-associated blood biomarker in cardiogenic stroke

Yuanzheng Hu [a,1], Xiangxin Li [b,1], Kaiqi Hou [c], Shoudu Zhang [a], Siyi Zhong [a], Qian Ding [a], Wuyang Xi [a], Zongqing Wang [a], Juan Xing [b,**], Fanghui Bai [b,***], Qian Xu [a,*]

[a] Henan Provincial Engineering Laboratory of Insects Bio-Reactor, Nanyang Normal University, Nanyang, 473061, China
[b] Henan Provincial Key Laboratory of Stroke Prevention and Treatment, Nanyang Central Hospital, Nanyang, 473000, China
[c] School of Computer Science and Technology, Nanyang Normal University, Nanyang, 473061, China

ARTICLE INFO

*Keywords:*
Cardiogenic stroke
Atrial fibrillation
*FCER1G*
Blood biomarkers
Machine-learning algorithm

ABSTRACT

*Background:* Cardioembolic stroke (CE) exhibits the highest recurrence rate and mortality rate among all subtypes of cerebral ischemic stroke (CIS), yet its pathogenesis remains uncertain. The immune system plays a pivotal role in the progression of CE. Growing evidence indicates that several immune-associated blood biomarkers may inform the causes of stroke. The study aimed to identify new immune-associated blood biomarkers in patients with CE and create an online predictive tool in distinguishing CE from noncardioembolic stroke (non-CE) in CIS.

*Methods:* Gene expression profiles that were publicly available were obtained from the Gene Expression Omnibus (GEO). The identification of differentially expressed genes (DEGs) was conducted using the Limma package. The hub module and hub genes were identified through the application of weighted gene coexpression network analysis (WGCNA). In order to identify potential diagnostic biomarkers for CE, both the random forest (RF) model and least absolute shrinkage and selection operator (LASSO) regression analysis were employed. Concurrently, the CIBERSORT algorithm was employed to evaluate the infiltration of immune cells in CE samples and examine the correlation between the biomarkers and the infiltrating immune cells. The diagnostic gene expression in blood samples was confirmed using qRT-PCR in a self-constructed dataset. Univariate and multiple logistic regression analyses were used to identify the risk factors for CE. Subsequently, the mathematical model of the nomogram was employed via Java's "Spring Boot" framework to develop the corresponding online tool, which was then deployed on a cloud server utilizing "nginx".

*Results:* Eleven differentially expressed genes (DEGs) that were upregulated and seven DEGs that were downregulated were identified. Through bioinformatics analysis and clinical sample

*Abbreviations:* AF, Atrial fibrillation; AUC, Area under the ROC curve; CE, Cardioembolic stroke; CIS, Cerebral ischemic stroke; DEG, Differentially expressed genes; DO, Human Disease Ontology database; Fc epsilon receptor Ig, FCER1G; GEO, Gene Expression Omnibus; LAA, Large artery atherosclerosis; LASSO, Least absolute shrinkage and selection operator; MMD, Moyamoya Disease; ROC, Receiver operating characteristic; RF, Random Forest; TOAST, Trial of org 10172 in acute stroke treatment; WGCNA, Weighted gene co-expression network analysis.

* Corresponding author.
** Corresponding author.
*** Corresponding author.
*E-mail addresses:* 13333600266@189.cn (J. Xing), bfh1986@163.com (F. Bai), xuqian7666@nynu.edu.cn (Q. Xu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e33846
Received 24 June 2024; Accepted 27 June 2024
Available online 28 June 2024

**Table 1**
Characteristics of GEO datasets in this study.

| GEO datasets | GSE66724 | GSE41177 | GSE58294 | GSE20129 | GSE146882 | GSE20680 | GSE20681 | GSE42148 |
|---|---|---|---|---|---|---|---|---|
| Platform | GPL570 | GPL570 | GPL570 | GPL10558 | GPL23178 | GPL4133 | GPL4133 | GPL13607 |
| Case/control | 8/8 | 3/16 | 23/23 | 57/78 | 10/10 | 143/52 | 99/99 | 13/11 |
| Year | 2016 | 2013 | 2014 | 2012 | 2020 | 2018 | 2018 | 2018 |
| Region | Spain | Taiwan, China | USA | USA | China | USA | USA | India |
| Reference | Allende M et al., 2019 | Yeh YH et al., 2019 | Stamova B et al., 2019 | Huang CC et al., 2018 | Yaxuan Sun et al., 2020 | Wingrove JA et al., 2018 | Wingrove JA et al., 2018 | Arvind P et al., 2018 |

verification, it was discovered that Fc Fragment of IgE Receptor Ig (*FCER1G*) could serve as a novel potential blood biomarker for CE. *FCER1G*, along with other risk factors associated with CE, were utilized to develop a nomogram. The training and validation sets, which consisted of 65 CIS patients, yielded areas under the curve (AUCs) of 0.9722 and 0.9689, respectively. These results indicate a high level of precision in risk delineation by the nomogram. Furthermore, the associated online predictive platform has the potential to serve as a more efficacious and appropriate predictive instrument (https://www.origingenetic.com/CardiogenicStroke-FCER1G) for distinguishing between CE and non-CE.

*Conclusion:* Blood biomarker FCER1G has the potential to identify patients who are at a higher risk of cardioembolism and direct the search for occult AF.The utilization of this online tool is anticipated to yield significant implications in terms of distinguishing between CE and non-CE, as well as enhancing the optimization of treatment decision support.

## 1. Introduction

Stroke is a heterogeneous disease that has multiple risk factors and causes [1]. Cerebral ischemic stroke (CIS) can be attributed to various factors, including cardioembolic stroke (CE), large artery atherosclerosis (LAA), small vessel disease (SAA), and other known causes [2]. Among the different subtypes of CIS, CE accounts for approximately 20–30 % of all CIS cases, while LAA is observed in approximately 23 % of cases [3]. As a result of the distinct etiologies and components of emboli in these two subtypes of CIS, the treatment strategies employed for each subtype differ accordingly [4]. Cerebral embolism (CE) exhibits a strong positive correlation with atrial fibrillation (AF) and manifests with more severe symptoms, a higher prevalence of disability, and an increased likelihood of recurrence compared to non-CE cases [5]. Notably, strokes induced by AF are associated with medical costs 1.5 times higher than strokes not attributed to AF [6]. Furthermore, there exist cryptogenic strokes caused by paroxysmal AF, which pose challenges in detection through routine electrocardiogram and other diagnostic procedures, thereby depriving these patients of appropriate secondary prevention measures and leading to stroke recurrence [7]. Due to the asymptomatic nature of subclinical AF, a significant proportion of patients (40 %) remain undiagnosed [8], resulting in a potential missed diagnosis rate exceeding 10 % in Asia [9]. The diagnosis of cardioembolic stroke necessitates the consideration of both clinical and radiological indicators of cerebral infarction, as well as the evaluation of potential cardiac factors. However, certain diagnostic tests associated with this condition are costly, time-consuming, and occasionally irrelevant to the specific needs of the patient. Consequently, the accurate diagnosis of cardioembolic stroke demands considerable effort and clinical expertise, yet it may still prove elusive in some cases [10].

Biomarkers play a crucial role in providing objective assessments of both normal and abnormal physiological processes, as well as in evaluating the efficacy of therapeutic interventions and predicting future prognoses [11]. The clinical application of blood biomarkers is already widely adopted in diverse medical decision-making contexts [12]. The domain of stroke research is continuously expanding, with an increasing repertoire of potential biomarkers encompassing proteins, ribonucleic acids, lipids, and metabolites [13]. Although these biomarkers have the potential to improve the diagnosis and treatment of stroke patients, their current levels of sensitivity, specificity, and cost-effectiveness do not meet the necessary criteria for routine implementation in stroke management. This underscores the importance of further research in this area. The utilization of a combination of weighted gene coexpression network analysis (WGCNA) and machine learning algorithms has demonstrated its effectiveness in identifying disease-related biomarkers. By employing bioinformatic analysis and machine learning techniques, researchers have effectively identified diagnostic signatures pertaining to diseases of the nervous system [14]. Machine learning algorithms have identified *IL1R2*, *IRAK3*, and *THBD* as the most suitable reference biomarker genes for acute myocardial infarction [15]. This methodology has exhibited substantial effectiveness in identifying essential biomarkers linked to diverse diseases [16].

The objective of this study was to identify differentially expressed genes (DEGs) by utilizing publicly available datasets. Machine learning algorithms and WGCNA were employed to screen potential biomarkers for CE. Following this, the expression of the identified biomarkers was validated using clinical samples. Furthermore, a nomogram was constructed, accompanied by an online predictive website, to facilitate the differentiation between CE and non-CE cases. This research will offer a unique perspective and foundation for investigating the pathogenesis of CE and developing innovative diagnostic tools.

## 2. Materials and methods

### 2.1. Data information and processing

The keywords "blood", "cerebral ischemic stroke", "atrial fibrillation", and "coronary artery disease," were used to obtain datasets from the Gene Expression Omnibus (GEO) public database (https://www.ncbi.nlm.nih.gov/gds/) [17]. The details of the datasets GSE66724 [18], GSE41177 [19], GSE58294 [20], GSE20129 [21], GSE146882, GSE20680 [22], GSE20681 [22], and GSE42148 for this study is described in Table 1. Among them , GSE66724、GSE41177 and GSE58294 annotated using GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. Two datasets of GSE20680 and GSE20681, both of which were annotated using GPL4133 Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F (Feature Number version). GSE20129 annotated using GPL10558 Illumina HumanHT-12 V4.0 expression beadchip. GSE146882 annotated using GPL23178[OElncRNAs520855F] Affymetrix Human Custom lncRNA Array. GSE42148 annotated using GPL13607 Agilent-028004 SurePrint G3 Human GE 8 × 60K

Microarray (Feature Number version). The GSE66724 dataset contains the peripheral blood of eight stroke patients with AF and eight stroke-free patients with AF. The GSE41177 dataset includes 19 samples, including 16 tissues from the left atrial junction and 3 tissue control samples from patients with sinus rhythm (SR). The GSE58294 dataset consists of 23 blood samples obtained from patients diagnosed with CE and an additional 23 control blood samples from individuals without symptomatic vascular diseases. Furthermore, the dataset includes findings at three time points (<3 h, 5 h, and 24 h) following the occurrence of the stroke event in the patients. The GSE20129 dataset comprises 57 samples with atherosclerosis and 78 normal controls,The dataset is all of female peripheral blood data. The GSE146882 dataset includes blood samples from 10 patients with atherosclerosis-induced ischemic stroke and 10 healthy controls. The GSE20680 dataset includes 143 coronary artery disease and 52 healthy control blood samples. The GSE20681 dataset includes 99 coronary artery disease and 99 healthy control blood samples. The GSE42148 dataset contains the peripheral blood of 13 patients with angiographically confirmed coronary artery disease and 11 population-based asymptomatic controls. Two datasets, GSE66724 and GSE41177, were merged to create the training set, and batch effects were corrected using the combat function in SVA (sva function in the R package sva). The GSE58294 dataset was used as the verification set, while the GSE20129, GSE146882, GSE20680, and GSE20681 datasets were utilized as the reverse validation sets. All datasets underwent standardized data preprocessing procedures. This approach allowed us to confirm the robustness and reproducibility of our results across different datasets.

### 2.2. Blood sample collection and preparation

The study included 20 healthy controls and a total of 78 patients (36 CE patients, including 2 mitral regurgitation patients, 2 patients with patent foramen oval, and 32 patients with AF, 42 noncardioembolic stroke patients) who were enrolled consecutively from Nanyang Central Hospital (Nanyang, China). Diagnosis of disease entailed gathering information regarding the patient's illness history, clinical manifestations, and intracranial angiography and the review of imaging and case notes by a minimum of two specialized neurologists. The classification of stroke subtypes adhered to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification system. In patients with CIS, confirmation was obtained through intracranial angiography and imaging. Demographic data and clinical information were obtained by analyzing patients' medical records and conducting a comprehensive investigation of their medical history. The inclusion criteria for CE were as follows: (1) cerebral infarction caused by cerebral artery obstruction derived from a thrombus originating from the heart and (2) at least one cardiac-related risk factor. The inclusion criteria for LAA stroke were as follows: (1) imaging indicating >50 % occlusion or stenosis of the common carotid artery, anterior and posterior cerebral artery, or vertebrobasilar artery caused by atherosclerosis; (2) audible murmur detected during neck auscultation; and (3) lesion diameter >1.5 cm according to imaging results. Patients with cardioembolic stroke or stroke from other causes were excluded. The inclusion criterion for SAA stroke was an infarct diameter measured by brain CT < 1.5 cm; strokes caused by other reasons such as patients diagnosed with moyamoya disease. Twenty volunteers of similar age and sex were recruited as healthy controls. The volunteers included in the study had no prior history of neurological events (cerebrovascular stroke or transient ischemic attack). As part of this study, informed consent was obtained from all patients or their relatives. The study protocol was reviewed and approved by the Ethics Committee of Nanyang Central Hospital. Peripheral venous blood samples were collected from subjects for RNA extraction and subsequent analysis.

### 2.3. Identification and functional enrichment analysis of differentially expressed genes

The Limma package was used to compare gene expression levels between different samples, and the filtering criteria ($|\log2FC| > 1.5$, $P < 0.05$) were set to identify DEGs. Subsequently, DEGs were visualized using the gplots package to show their distribution in a volcano plot and display changes in gene expression levels across different samples in a heatmap. To perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional annotation for DEGs, we utilized Metascape (http://metascape.org) [23], which is a commonly used online tool in biological research. Additionally, an analysis of disease ontology term enrichment was conducted using the Human Disease Ontology (DO) database (http://www.disease-ontology.org) [24], a community-driven standards-based ontology that annotates genes based on human diseases. To determine the statistical enrichment of DEGs in KEGG pathways, we used the Cluster Profiler R package. The DEGs were evaluated for protein–protein interactions (PPIs) using the STRING database v12.0 (http://www.string-db.org/), which calculates a combined score to assess the reliability of the interaction, with a threshold >0.4 considered significant [25]. In addition, we created a network of coexpressed genes using GeneMANIA (http://www.genemania.org) and single gene gene-set enrichment analysis(GSEA) were performed using R package clusterProfiler [26].

### 2.4. Selection and evaluation of the candidate diagnostic biomarkers

To identify candidate diagnostic biomarkers for CE, we used random forest (RF) logistic regression, least absolute shrinkage and selection operator (LASSO) regression analysis, and WGCNA to screen diagnostic genes. The WGCNA package was used to construct the weighted gene coexpression network [16]. To ensure the construction of scale-free coexpression gene networks, we determined the soft threshold parameters, setting the scale-free $R^2$ value at 0.85 and the power value ($\beta$) at 19. Subsequently, a cluster dendrogram was generated based on the topological overlap matrix, with a minimum cluster size of 100 genes. Based on the WGCNA, hub modules were generated, and the most biologically significant hub genes in each module were obtained. We also employed two machine learning approaches (LASSO and RF algorithms) to screen candidate biomarkers. LASSO regression analysis was carried out using the "glmnet" R package, and the RF model was generated using the "randomforest" R package. Finally, the overlapping genes among the RF and LASSO algorithms and the WGCNA were considered potential key biomarkers for CE. Finally, receiver operating characteristic (ROC) curves were drawn to evaluate the diagnostic value according to the screening results.

### 2.5. Total RNA extraction and quantitative real-time PCR analysis

Upon collection, the peripheral blood samples were promptly preserved in blood RNA storage tubes (Bioteke Corporation, Beijing, China). Total RNA was extracted from these preserved samples. Subsequently, cDNA synthesis was carried out using an M5 Super plus qPCR RT kit (Mei5 Biotechnology Co. Ltd., Beijing, China). To quantify the expression levels of the target genes, qRT−PCR was performed using the CFX96TM Real-Time System (*Bio-Rad*, Hercules, CA, USA). The $2^{-\Delta\Delta CT}$ method was utilized for relative quantification. *GAPDH* served as the internal reference. The sequences of the primers used were as follows: *FCER1G*-F: CTCTGCTATATCCTGGATGC, *FCER1G*-R: CTAAAGCTACTGTGGTGGTT.

### 2.6. Construction and assessment of the prediction model

The risk factors linked to CE were determined by univariate and multivariate logistic regression analyses. The selected risk factors were then utilized by logistic regression to construct the nomogram. To assess the reliability of the binary logistic regression, the Hosmer−Lemeshow test was employed. The "rms" package was employed to visualize the nomogram. To randomly divide the dataset into training and test sets, the "caTools" package was used. Finally, the "pROC" package was utilized to draw the ROC curves. The "ggplot2" package was used to draw a waterfall plot and calibration curve. The calibration curve was used to evaluate the validity of the nomogram. Subsequently, we used the mathematical model of the nomogram through Java's "Spring Boot" to construct the accompanying online tool and deployed it to the cloud server using "nginx".

### 2.7. Statistical analysis

GraphPad Prism 8 (https://www.graphpad.com/scientific-software/prism/) and R software (http://www.R-project.org/, version 4.2.1) were used for all statistical analyses in this study. The "corrplot" package was used to create lollipop charts to visualize the correlation between infiltrating immune cells and diagnostic genes. The resulting associations were visualized using the chart technique with the "ggplot2" package. To assess the correlation between diagnostic gene expression levels and clinical factors, unpaired Student's *t* tests were used for continuous variables, while Fisher's exact tests were employed for categorical variables. $P < 0.05$ was considered statistically significant.

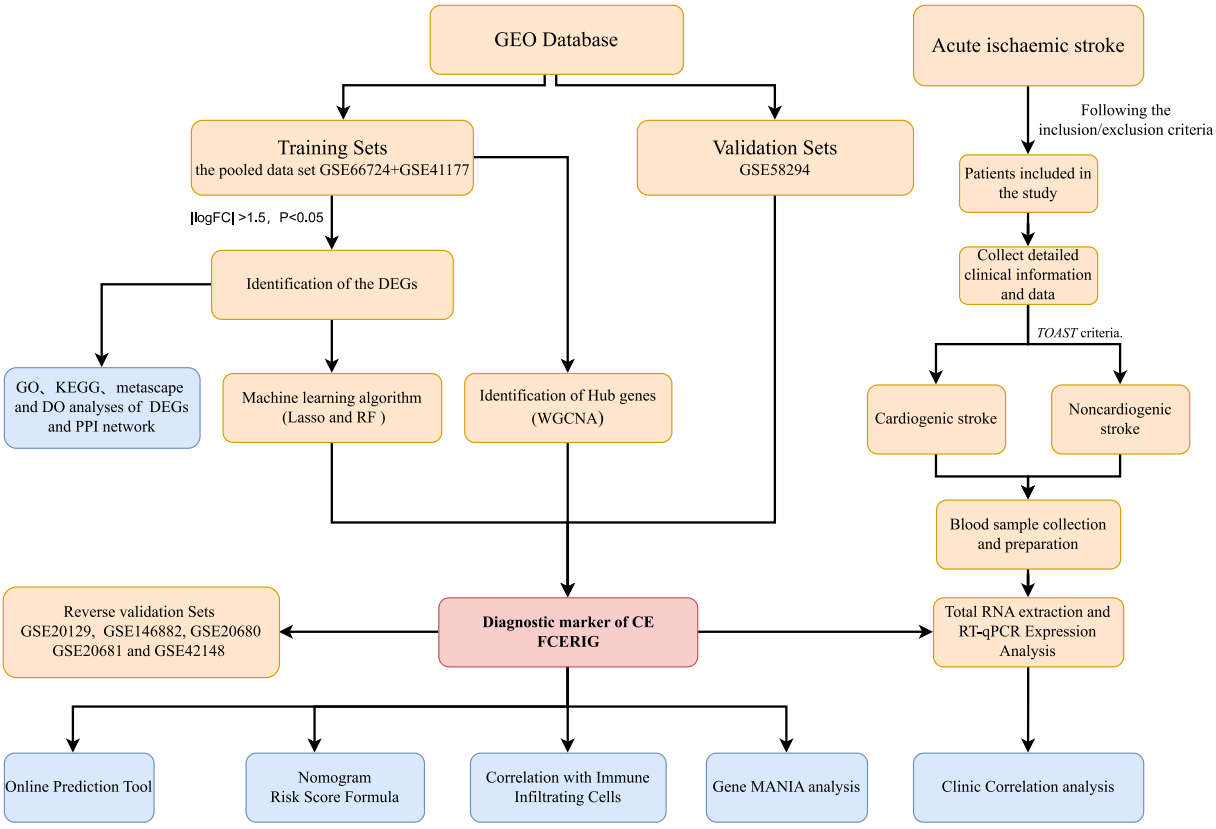


**Fig. 1.** Flow chart of the study.

## 3. Results

### 3.1. Determination of DEGs

The flow chart presented in Fig. 1 illustrates the study design. CE is intricately associated with numerous cardiovascular diseases, with AF-related stroke constituting over 79 % of cases, thereby emerging as the foremost risk factor in the prevention of CE. Initially, we merged two AF-related datasets, GSE66724 and GSE41177, and batch effects were corrected using the combat function in SVA. Subsequently, differential expression analysis was performed on the merged datasets utilizing the Limma package in the R programming language. Ultimately, a total of 18 genes were identified as differentially expressed genes (DEGs) specifically associated with AF. In Fig. 2A–B, 11 genes were upregulated (*S100A12*, *FCGR3B*, *S100A9*, *CXCR2*, *S100A8*, *BCL2A1*, *TYROBP*, *FCER1G, CLC*, *CFD*, *PROK2*), while seven genes were downregulated (*ADIPOQ*, *SLC35G3*, *NMU*, *RNF128*, *SOD2*, *GPR22*, *DNER*).

### 3.2. Functional enrichment analysis and PPI network of DEGs

To reveal the underlying biological functions and pathways correlated with DEGs, Metascape, KEGG pathway analysis, and GO enrichment and DO enrichment analysis of DEGs were performed. According to the KEGG pathway analysis, the DEGs were mainly involved in natural killer cell-mediated cytotoxicity, longevity-regulating pathway, IL-17 signaling pathway, Staphylococcus aureus infection, osteoclast differentiation, phospholipase D signaling pathway, and type II diabetes mellitus (Fig. 3C). Metascape was used to annotate the functions of DEGs. The results revealed that DEGs were markedly enriched in neutrophil degranulation, positive regulation of apoptotic process, neutrophil activation, peptide ligand-binding receptors, apoptotic signaling pathways, and negative regulation of cytokine production (Fig. 3D). According to GO enrichment analysis, DEGs were involved in several immune response pathways, such as response to neutrophil chemotaxis and activation, as well as cell chemotaxis. In addition, DEGs were also involved in receptor and immunoglobulin binding through interactions with RAGE receptors, Toll-like receptors, and antioxidant activity (Fig. 3A). DO enrichment analysis showed that the DEGs were mainly associated with arteriosclerotic cardiovascular disease, atherosclerosis, psoriatic arthritis, lymph node disease, and lymphadenitis (Fig. 3B). Using the STRING online database, we constructed a PPI network to mine 18 DEG-related proteins (Fig. 3E).

### 3.3. FCER1G as a novel potential biomarker for cardiogenic stroke

Gene coexpression network analysis was conducted on obtained from the merged GSE41177 and GSE66724 datasets. The optimized soft-thresholding power ($\beta = 19$ with $R^2 = 0.85$) was selected as the scale-free topology criterion, as illustrated in Fig. 4A–B. Subsequently, coexpression networks were constructed, resulting in the identification of five coexpressed modules represented by the colors blue, brown, gray, turquoise, and yellow. The cluster dendrogram of these five modules is presented in Fig. 4C. We observed 476, 408, 3444, 874, and 212 genes for the blue, brown, gray, turquoise, and yellow modules, respectively. Notably, the gray module exhibited the highest correlation coefficient with CIS(cor = 0.6, $P < 0.001$, Fig. 4D) , which was deemed crucial for further analysis. Then, the set gene importance (geneSigFilter = 0.5) and the minimum threshold of gene-module correlation (moduleSigFilter = 0.8) were filtered to obtain the core genes of the module: *PRPF6, TRIM55, C11orf74, RAB8A, UNC45B, LAPTM5, PILRA, NEB* and *FCER1G*. Then, we performed functional network analysis of the hub genes by using GeneMANIA. Interestingly, these genes were significantly enriched in cardiac muscle cell development, cardiac cell development, receptor metabolic process, and cardiac muscle cell

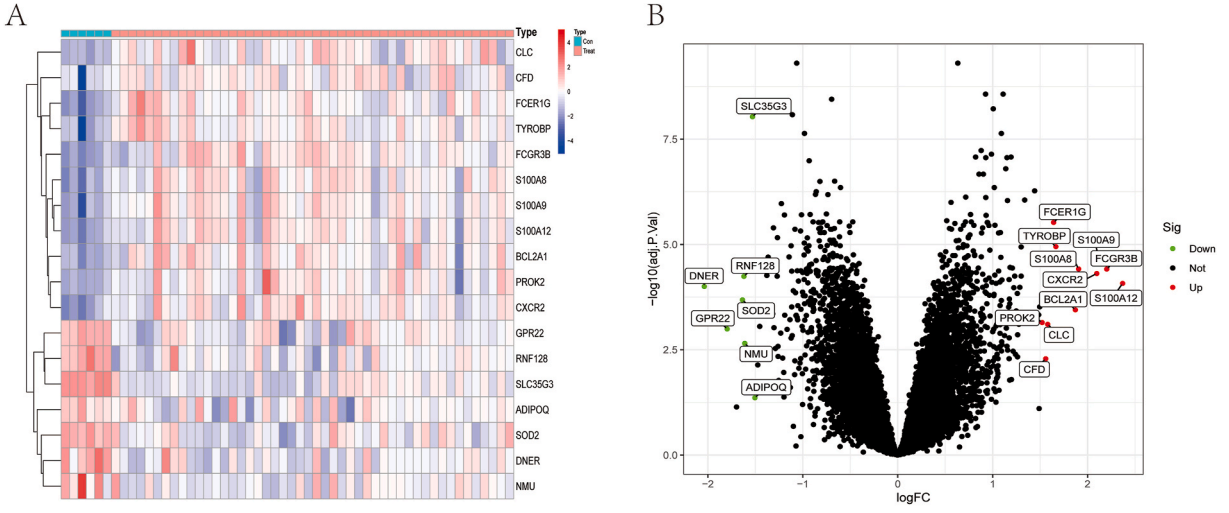


**Fig. 2.** Results of differentially expressed gene analysis
(A) Heatmaps and (B) volcano plots of the differentially expressed genes (DEGs).

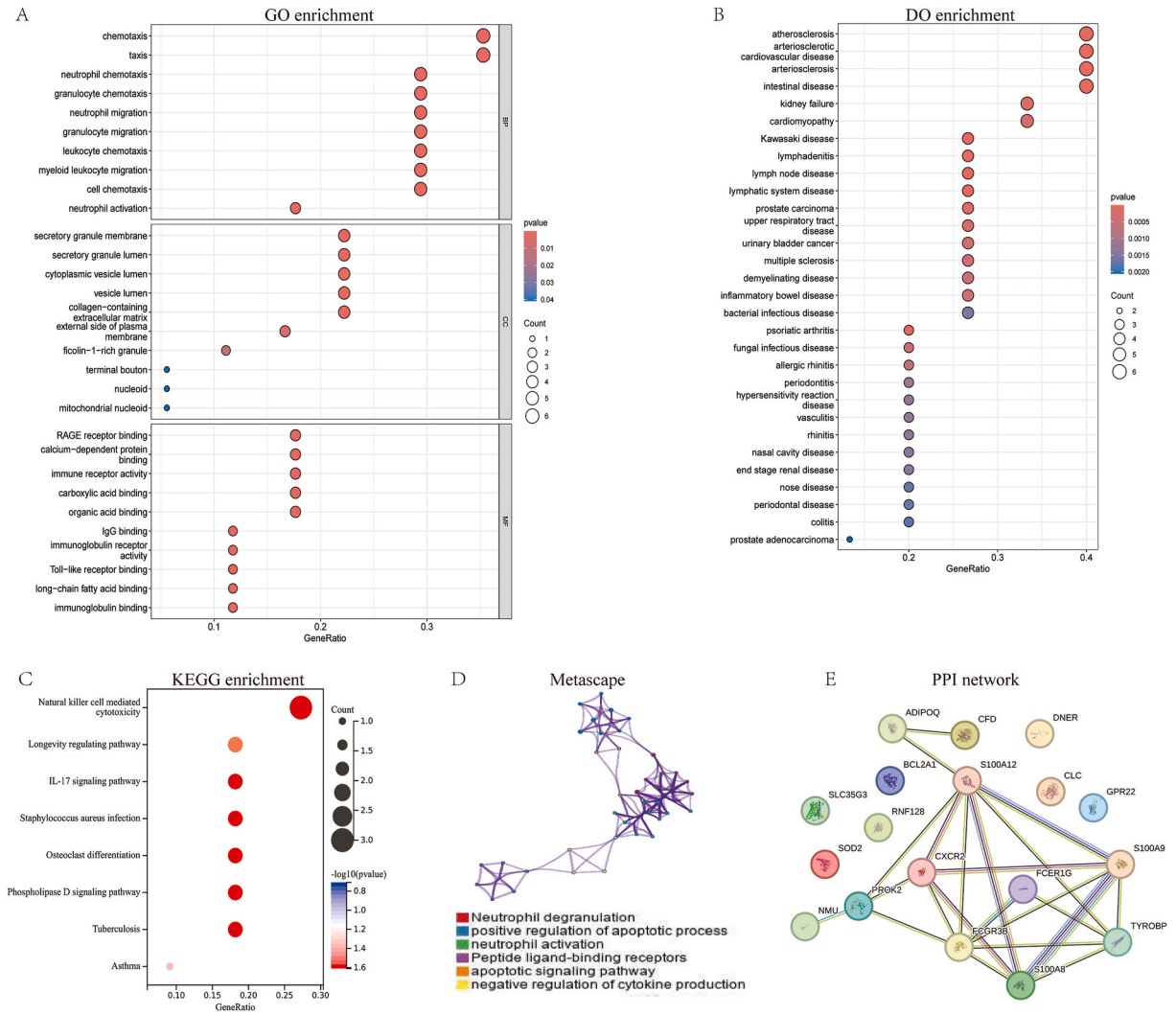


**Fig. 3.** Enrichment analysis.
(A) Bubble plot of the GO terms with DEG enrichment and (B) DO enrichment analysis of DEGs. (C) Bubble plot of the main KEGG signaling pathways. (D) Enrichment analysis of DEGs by Metascape. (E) PPI network of DEGs.

differentiation (Supplementary Fig. 1).

Next, we utilized two distinct algorithms, LASSO and RF, to select feature genes by combining the training sets (GSE66724 merge GSE41177). Following a tenfold cross-validation analysis using the LASSO algorithm, the optimal Lambda value (lambda.min) was determined to be 0.0002244043. At this stage, a total of six genes were identified, namely *SLC35G3*, *FCER1G*, *CXCR2*, *RNF128*, *DNER*, and *SOD2*(Fig. 5A–B). In the RF algorithm, the optimal tree is determined to be 29, which corresponds to a minimum error rate of 0.0185185185185185. Genes with a score greater than 2, including *SLC35G3* and *FCER1G*, were selected for further analysis (Fig. 5C–D). Overall, *FCER1G* and *SLC35G3* were shared between the LASSO and RF algorithms. Only FCER1G was screened and identified as a candidate biomarker for further analysis after overlapping the genes selected by the RF and LASSO algorithms and the WGCNA (Fig. 5E). The ROC curves in the training group indicated the predictive power of *FCER1G* (AUC = 0.997, Fig. 6B). CE had the highest proportions of AF-related stroke To further confirm the reliability and reproducibility of *FCER1G* in the blood of CE patients, we selected the GSE58294 dataset for validation. As shown in Fig. 6A, *FCER1G* was differentially expressed in blood between the CE patients and healthy controls ($P < 0.001$). In addition, as shown in Fig. 6C, the ROC curve analysis provided robust evidence for the excellent predictive ability of *FCER1G* (AUC = 0.962). In addition, we did not observe a clear difference in *FCER1G* expression in samples from other corresponding datasets that were blood based, including the atherosclerosis (GSE20129 and GSE146882) and coronary artery disease (GSE20680, GSE20681 and GSE42148) datasets ($P > 0.05$, Fig. 6A).

For further validation of the expression of *FCER1G*, qRT−PCR was performed in self-collected clinical samples (98 blood samples from 36 CE patients, 25 LAA stroke patients, 4 SAA patients, 13 MDD patients and 20 healthy controls). Table 2 illustrates the clinical characteristics of all patients, and Fig. 7 shows the clinical representative image characteristics of two cases pathologically confirmed

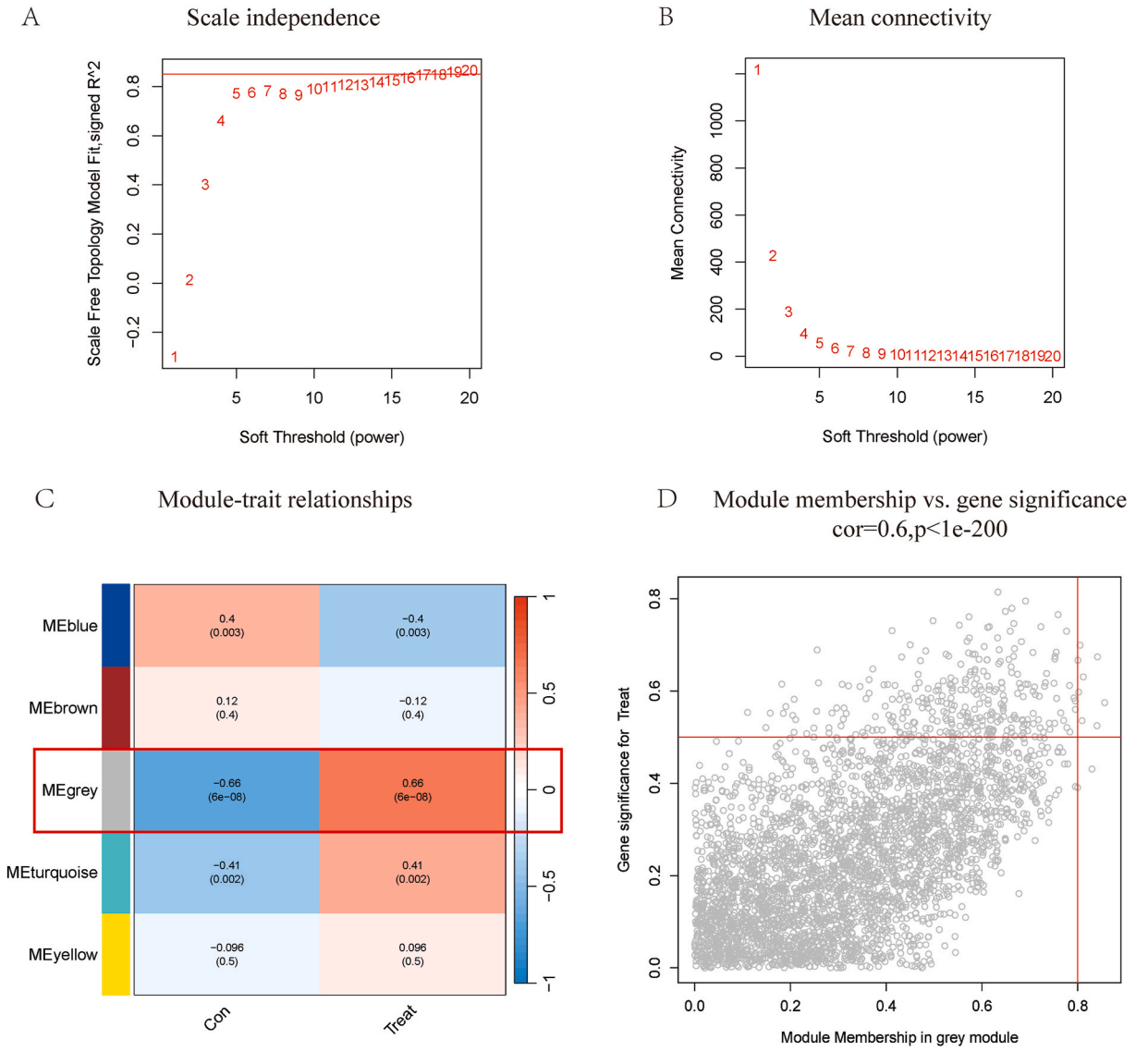


**Fig. 4.** WGCNA of module eigengenes.
(A) Analysis of the scale-free fit index and (B) the mean connectivity for various soft-thresholding powers. (C) Five modules significantly associated with AF were identified, with the gray module being the most significant. (D) The correlation coefficient of 3444 genes and AF-related genes in the gray module was 0.66. WGCNA, weighted gene coexpression network analysis.

as CE and LAA stroke. According to the qRT−PCR analysis, the relative expression of *FCER1G* in the CE groups was significantly upregulated compared with that in the noncardioembolic stroke group and heathy control group ($P < 0.05$, Fig. 8A). The AUC of *FCER1G* was 0.9431 when comparing CE patients with healthy controls (95 % CI: 0.8804–1.0000; $P < 0.0001$; Fig. 8B). The AUC was 0.6900 when comparing LAA patients with healthy controls (95 % CI: 0.5359–0.8441, $P < 0.05$; Fig. 8C). The AUC was 0.8444 when comparing CE with LAA (95 % CI: 0.7480–0.9408, $P < 0.0001$; Fig. 8D). The AUC was 0.8592 when comparing CE with no-CE (95 % CI: 0.7700–0.9484, $P < 0.0001$; Fig. 8E)Taken together, these data provide strong evidence that the expression of *FCER1G* can effectively distinguish between healthy individuals and patients with CE and LAA. The results also indicate that *FCER1G* could differentiate between CE and non-CE in patients with CIS. Furthermore, the specificity of *FCER1G* expression for these two subtypes of CIS indicates its potential as a diagnostic biomarker for CE.

### *3.4. Expression of FCER1G was associated with clinicopathological parameters in CIS patients*

Through univariate and multivariate analyses, our study discovered that the presence of cardiogenic diseases (Cds, $P < 0.05$), *FC1R1G* ($P < 0.05$), and N-terminal pro-brain natriuretic peptide (NT-proBNP, $P < 0.05$) may serve as independent indicators for

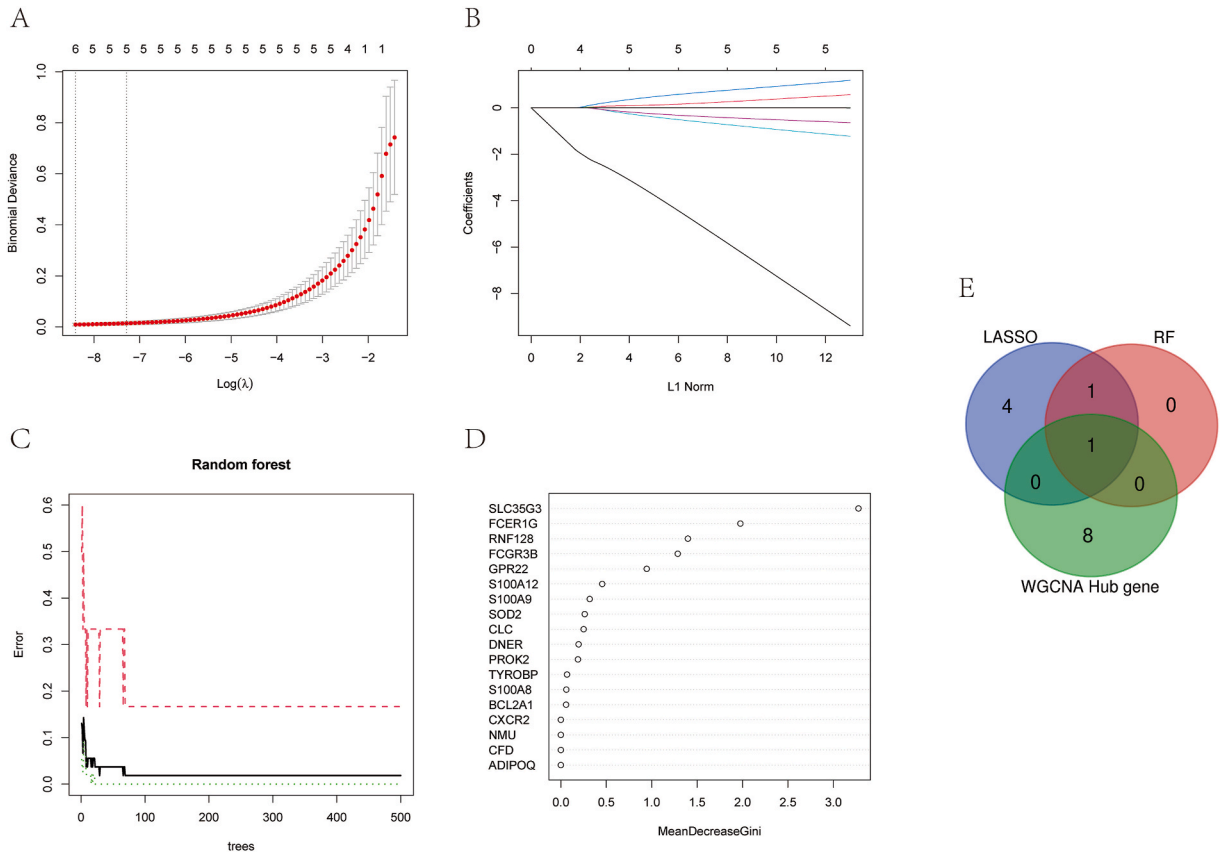


**Fig. 5.** Machine learning algorithms for screening diagnostic genes.
(A–B) The LASSO regression algorithm and (C–D) RF algorithm were applied to screen diagnostic biomarkers of CE. (E) Venn diagram showing that *FCER1G* was a novel biomarker for CE identification based on the previous WGCNA and two algorithms.

predicting the risk of CE (as shown in Table 3). Subsequently, we proceeded to examine the correlation between the expression levels of *FCER1G* and various clinical pathological features in CIS patients, including sex, age, AF, cardiogenic diseases, diabetes, hypertension, alcohol consumption, and smoking. Based on the comparative analysis of general data presented in Tables 2 and it was observed that the expression of *FCER1G* exhibited a positive association with age ($P = 0.0074$), AF ($P < 0.001$), existence of cardiogenic stroke etiology ($P < 0.001$), diabetes ($P = 0.0495$), drinking ($P < 0.001$) and smoking ($P < 0.001$) and was different across sex groups ($P < 0.001$), but it showed no association with the remaining indicators, such as hypertension I ($P > 0.05$) and hypertension III ($P > 0.05$).

### 3.5. Potential functions and molecular mechanisms of FCER1G

Next, the relationship between *FCER1G* expression and immune cell infiltration was assessed (Fig. 9A). Specifically, as shown in Fig. 7B–F, *FCER1G* was positively correlated with resting mast cells ($P < 0.001$) but negatively correlated with plasma cells ($P < 0.05$), activated mast cells ($P < 0.05$), naive B cells ($P < 0.05$), and follicular helper T cells ($P < 0.001$). Furthermore, GeneMANIA and Single-gene GSEA was utilized to predict the regulatory partners and potential biological functions of *FCER1G*. As shown in Fig. 9B, *FCER1G* was coexpressed with the Fc family genes, which belongs to the immunoglobulin superfamily and is involved in mediating and executing antibody-mediated immune responses. For the functions of *FCER1G* in ncardioembolic stroke, GSEA was performed in GSE58294 and the results revealed that the major biological processes and KEGG pathways where *FCER1G* Fc gamma R−mediated phagocytosis, Apoptosis, Necroptosis, and Complement and coagulation cascades (Fig. 9C). In addition, *FCER1G* was found to play a crucial role in various signaling pathways, such as the FC receptor signaling pathway, Fc-gamma receptor signaling pathway, Fc receptor-mediated stimulatory signaling pathway, immune response-regulating cell surface receptor signaling pathway involved in phagocytosis, phagocytosis, regulation of innate immune response, and regulation of phosphatidylinositol 3-kinase signaling. Collectively, these results suggested that the potential influences of FCER1G on CE progression may involve FCER1G-mediated regulation of immune responses.

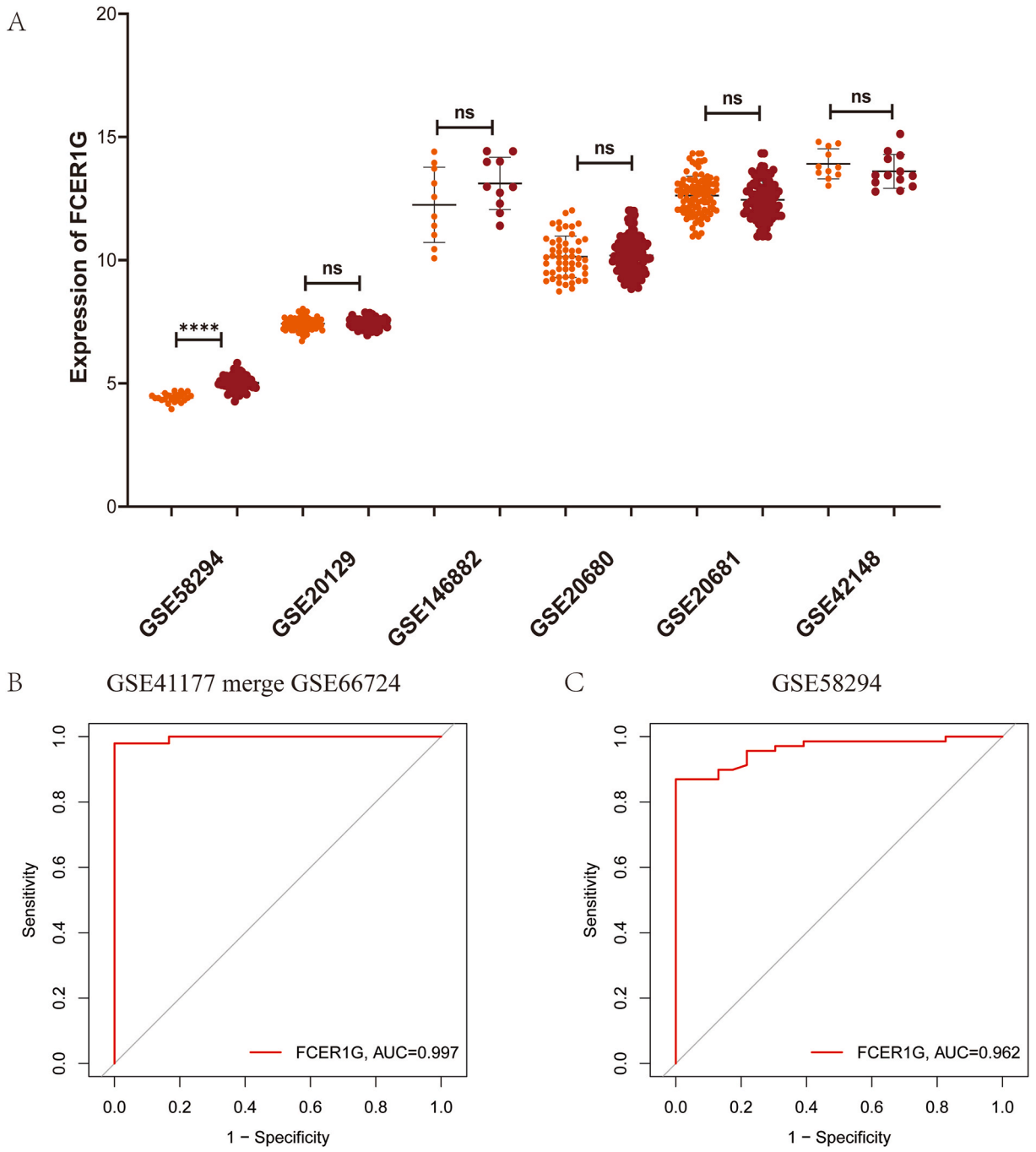


**Fig. 6.** Validation of FCER1G in a datasets.
(A) *FCER1G* expression level in the validation dataset and reverse validation sets. (B) ROC curves of *FCER1G* analysis for evaluating the diagnostic efficiency of *FCER1G* in training datasets (GSE41177 merge GSE66724). (C) ROC curve analysis for evaluating the diagnostic efficiency of *FCER1G* in the validation set (GSE58294).

### *3.6. Construction and evaluation of the nomogram and online prediction tool*

Based on the findings from the logistic regression analysis, a nomogram was developed for the prediction of CE risk for stroke patients that incorporated the aforementioned three indicators (existence of cardiogenic diseases, *FCER1G* expression level, and NT-proBNP) as significant predictors (Fig. 10A). The model risk score formula was as follows: risk of cardiogenic stroke = −0.000487142 * points ^3 + 0.015944345 * points ^2 + −0.074330208 * points +0.084079519, points = 0.222222222 * *FCER1G* + 0, existence of

**Table 2**
Association between FCER1G expression and clinicopathologic characteristics.

| Indexes | Group | No. of cases | FCER1G | | |
|---|---|---|---|---|---|
| | | | 95%CI | $R^2$ | *P-value* |
| Sex | Female | 37 | −2.358 - 57.15 | 0.04237 | **<0.0001** |
| | Male | 41 | | | |
| Age | > 60 | 51 | −54.18 - 8.803 | 0.02638 | **0.0074** |
| | ≤60 | 27 | | | |
| AF | YES | 36 | 27.18–82.68 | 0.1698 | **<0.0001** |
| | NO | 42 | | | |
| Cds | YES | 37 | 19.34–76.11 | 0.1286 | **<0.0001** |
| | NO | 41 | | | |
| Diabetes | YES | 23 | −19.34–46.96 | 0.008978 | **0.0495** |
| | NO | 55 | | | |
| Hypertension | NO | 33 | | | |
| | I | 8 | −38.27–40.16 | 6.089e-005 | > 0.05 |
| | II | 14 | −27.13–63.49 | 0.01430 | **0.0002** |
| | III | 23 | −16.02–44.57 | 0.01626 | > 0.05 |
| Driking | YES | 19 | −61.42 - 8.273 | 0.02946 | **<0.0001** |
| | NO | 59 | | | |
| Smoking | YES | 15 | −59.08–17.38 | 0.01529 | **<0.0001** |
| | NO | 63 | | | |

Cds: Existence of cardiogenic stroke etiology, AF:Atrial fibrillation.

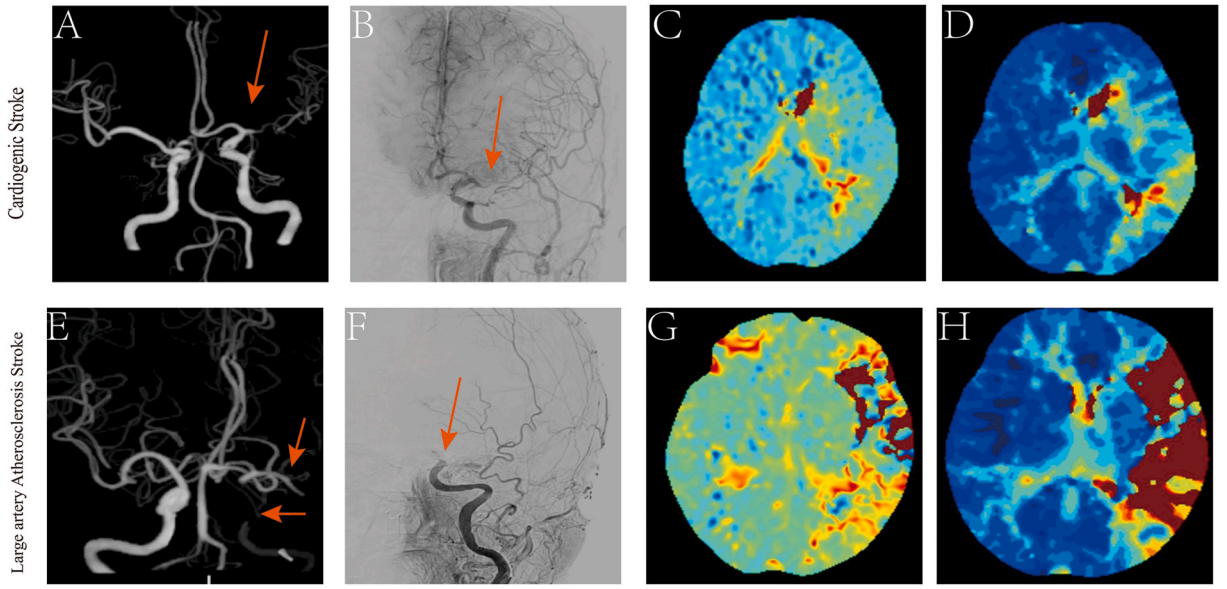


**Fig. 7.** Imaging examination of a CE patient and a LAA stroke patient.
(A–D) Representative radiographic images of the CE patient. (A) Head CTA shows cerebrovascular atherosclerosis and severe M1 stenosis in the left middle cerebral artery. (B) DSA shows severe M1 stenosis of the left middle cerebral artery, distal slenderness and better compensation. (C) Prolonged mean cerebral blood flow transit time in the left hemisphere (MTT). (D) Prolonged peak time of cerebral blood flow in the left hemisphere (Tmax). (E–H) Representative radiographic images of the LAA patient. (E) Head CTA shows M2 occlusion of the end of the left internal carotid artery and the left middle cerebral artery, indicating complete embolus blockage. (F) Safe terminal occlusion of the left internal carotid artery in DSA. (G) Prolonged mean transit time (MTT) in the left hemisphere. (H) Prolonged peak time of cerebral blood flow (Tmax) in the left hemisphere.

cardiogenic stroke etiology = 6.290814, normal existence of cardiogenic stroke etiology level = 0; normal NT-proBNP = 0 (0–300 pg/ml), and abnormal NT-proBNP = 25.578767 (NT-proBNP>300 pg/ml). Decision curve analysis (DCA) showed that the "impact factor" curve was much higher than the gray line, which indicated that the accuracy of the line graph model was higher (Fig. 10B). To assess the clinical impact of the nomogram more visually, a clinical impact curve was generated based on the DCA curve. The high-risk population represented by the red curve and the true positive patients represented by the blue curve were very close, indicating that the diagnostic model has excellent predictive ability (Fig. 10C). The calibration curves indicated that the nomogram has high accuracy in predicting the risk of CE (Fig. 10D). We also divided the clinical data into training and validation sets, and the AUC of the training set was 0.9722 (Fig. 10E), and the AUC of the validation set was 0.9689 (Fig. 10F), indicating a high precision in risk

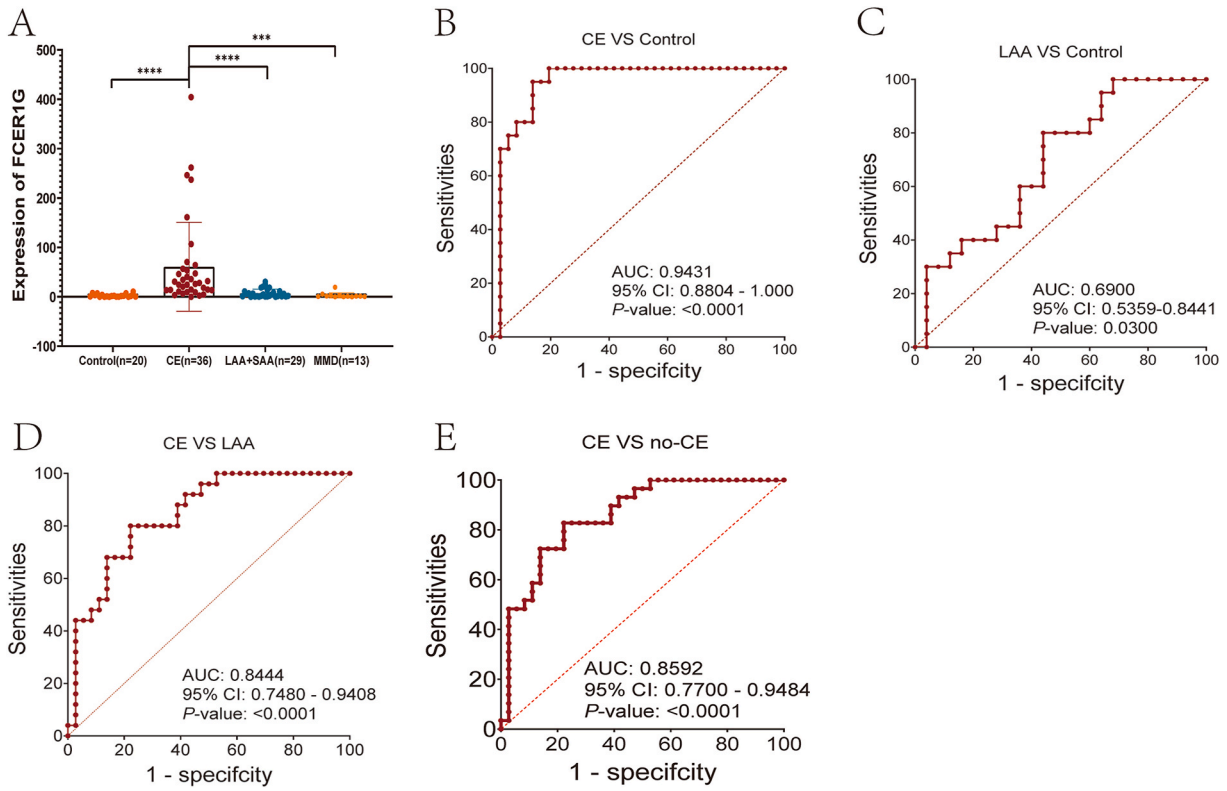


**Fig. 8.** Diagnostic value of *FCER1G*.
(A) The expression levels of *FCER1G* were analyzed by RT−qPCR. (B–E) ROC curves of FCER1G: *P < 0:05, **P < 0:01, ***P < 0:001. Horizontal lines represent median levels and interquartile ranges. CT: control; CE: cardioembolic stroke; LAA: large artery atherosclerosis stroke; SAA: small vessel disease; MMD: moyamoya disease.

**Table 3**
Univariate and multivariate logistic regression for independent risk factor analysis.

| Variables | Univariate analysis | | Multivariate analysis | |
|---|---|---|---|---|
| | *p* | OR (95 % CI) | *p* | OR (95 % CI) |
| Sex | 0.394 | 1.458(0.612–3.473) | – | – |
| Age | 0.998 | – | – | – |
| Cds | **0.000** | 44.433(12.435–158.774) | **0.001** | 39.837(4.533–350.075) |
| Diabetes | 0.957 | 1.028(0.379–2.784) | – | – |
| Smoking | 0.438 | 0.629(0.195–2.032) | – | – |
| *B*-type natriuretic | **0.000** | 35.556(10.023–126.129) | **0.004** | 26.251(2.923–235.724) |
| Drinking | 0.758 | 0.827(0.246–2.774) | – | – |
| *FCER1G* | **0.000** | 1.148(1.076–1.226) | **0.011** | 1.139(1.031–1.259) |
| hypertension | 0.267 | | | |
| hypertension(1) | 0.143 | 3.867(0.634–23.585) | – | – |
| hypertension(2) | 0.772 | 1.208(0.336–4.344) | – | – |
| hypertension(3) | 0.115 | 2.32(0.816–6.599) | – | – |

Cds: Existence of cardiogenic stroke etiology.

delineation of the model. In addition, an accompanying online prediction tool (https://www.origingenetic.com/CardiogenicStroke-FCER1G)for distinguishing CE from non-CE by analyzing *FCER1G* expression combined with clinical pathological features was constructed, which can be used as an auxiliary tool to assist physicians in making a quick, accurate diagnosis.

## 4. Discussion

Stroke is a serious cerebrovascular disease, and the current guidelines for ischemic stroke treatment are based on the TOAST classification system [27]. CE refers to a type of stroke that occurs when a cardiogenic embolism is dislodged and travels to the brain, causing ischemic stroke in a cerebral artery [28]. Although the classification system of stroke etiology is constantly being improved,

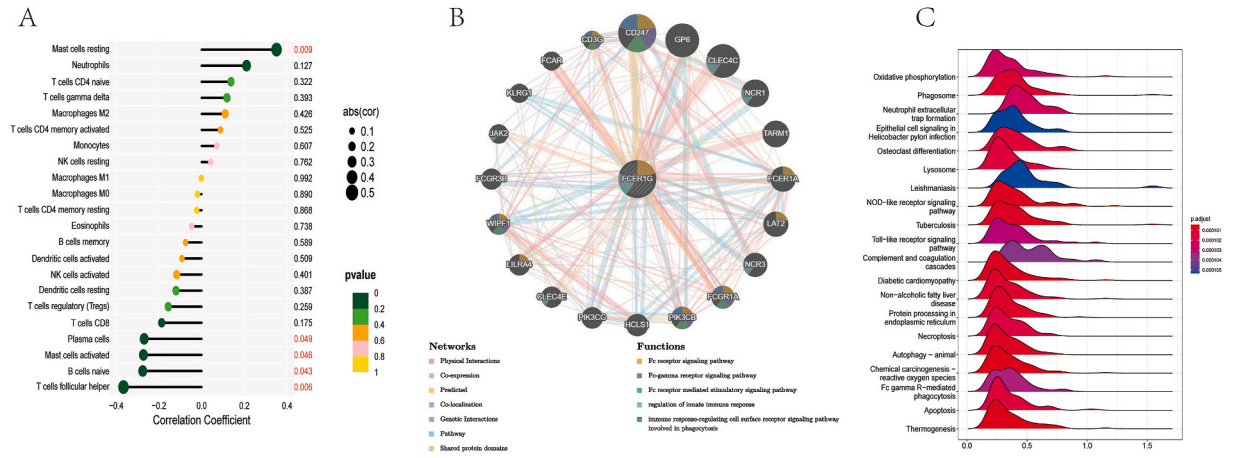


**Fig. 9.** Correlation analysis of the *FCER1G*.
(A) Immune correlation lollipop chart. The association between *FCER1G* and different immune cell infiltration in Cardiogenic Stroke. (A) Gene network analysis of *FCER1G* using the GeneMANIA database. (B) Single-gene GSEA enrichment analysis of *FCER1G*.

there may be variances among different evaluators owing to personal experiences, professional knowledge, and individual knowledge levels, which may lead to diagnostic dilemmas. These limitations underscore the inadequacies of the causative stroke classification system. As AF promotes blood clot formation, patients with AF are at an increased risk of rethrombosis and generally experience higher rates of morbidity and mortality than individuals with other types of stroke [29,30]. To prevent the recurrence of CE, it is recommended to perform a thorough evaluation to identify potential causes [31]. However, approximately 30 % of AF patients cannot be diagnosed through routine electrocardiogram testing, and there is no difference in the risk of stroke between paroxysmal AF and persistent AF. Clinically, many undetected paroxysmal AF-induced CE cases are classified as cryptogenic stroke, which leads to the failure to administer reasonable secondary preventive medication treatment to these patients [32]. AF can remain occult because of its paroxysmal nature and asymptomatic nature even after an extensive workup [33]. AF can be prevented by anticoagulation; however, indiscriminate anticoagulation for suspected but unproven cardioembolism has not been shown to be effective and increases the risk of bleeding [32]. Therefore, accurately identifying the etiology and classification of stroke patients has important clinical significance for reducing the recurrence rate and mortality rate of stroke patients.

In this study, according to the GO analysis, the DEGs were found to be enriched in certain biological functions or pathways, such as neutrophil chemotaxis and activation, as well as cellular chemotaxis. There is a strong association between these biological processes and the progression of ischemic stroke [34]. Limiting brain damage progression during or after stroke is considered an important role for immunity in acute stroke pathogenesis [35]. In the case of ischemic stroke, infiltration by natural killer (NK) cells into the brain occurs. This infiltration of NK cells is a response to stroke and is believed to be part of the immune-related mechanisms that occur during the condition [36]. In addition, some studies have shown that immune regulation can effectively delay the progression of ischemic stroke, restore neurological function, and improve patient prognosis [37]. These studies further emphasize the importance of maintaining immune microenvironment balance in protecting the central nervous system [38]. The results of KEGG enrichment analysis indicated that DEGs were enriched in pathways related to immunity, such as the IL-17 signaling pathway and natural killer cell-mediated cytotoxicity. Moreover, DEGs were predominantly enriched in neutrophil degranulation, positive regulation of apoptotic processes, neutrophil activation, peptide ligand-binding receptors, apoptotic signaling pathway, and negative regulation of cytokine production, which play a crucial role in the progression and pathogenesis of stroke. Interestingly, DO enrichment analysis showed that DEGs were mainly associated with arteriosclerotic cardiovascular disease, atherosclerosis, psoriatic arthritis, lymph node disease, and lymphadenitis, while *FCER1G* was not enriched in these diseases.

The number of biomarkers for stroke is continuously growing. Zheng et al. showed that 4 key immune-related genes (*ADM, ANXA3, SLC22A4* and *VIM*) were reliable serum markers for the diagnosis of CIS [39]. Wang et al. found that lncRNA zinc finger antisense 1 was sensitive for identifying LAA stroke patients (0.89), but the specificity was only 0.48 [40]. A comparative diagnostic meta-analysis reported that BNP showed a superior sensitivity (0.65) and NT-proBNP showed a superior specificity (0.93) in distinguishing CE from non-CE in CIS [41]. Jickling et al. reported that a 40-gene panel could distinguish CE from LAA within the first 24 h of stroke onset [42]. A single 37-gene panel was able to differentiate between AF and non-AF embolic stroke with a sensitivity and specificity >90 % [42]. In stroke, a substantial number of candidate blood biomarkers have shown promise for translation into clinical practice. However, the potential diagnostic utility and clinical feasibility of the gene panel need further verification. Through the application of bioinformatics analysis, it becomes feasible to detect a multitude of genes that exhibit differential expression in samples of cardiogenic stroke patients when compared to samples of noncardiogenic stroke patients. For example, *Li* et al. identified three hub proteins (*C3, APOA4* and *S100A9*) that found significant immune-related functions in a middle carotid artery occlusion mouse model based on proteome and transcriptome analysis exploring stroke progression [43]. The same conclusion was reached in another study of lipid cells in stroke mice [44]. *FCER1G* is a member of the Fc family (including *FCAR, FCGR1A, FCGR3B*, and *FCER1A*) [29]. *FCGR3B* has been recognized as a prospective gene of interest in the progression of AF [45](Table 4). In this study, PPI network analysis showed that

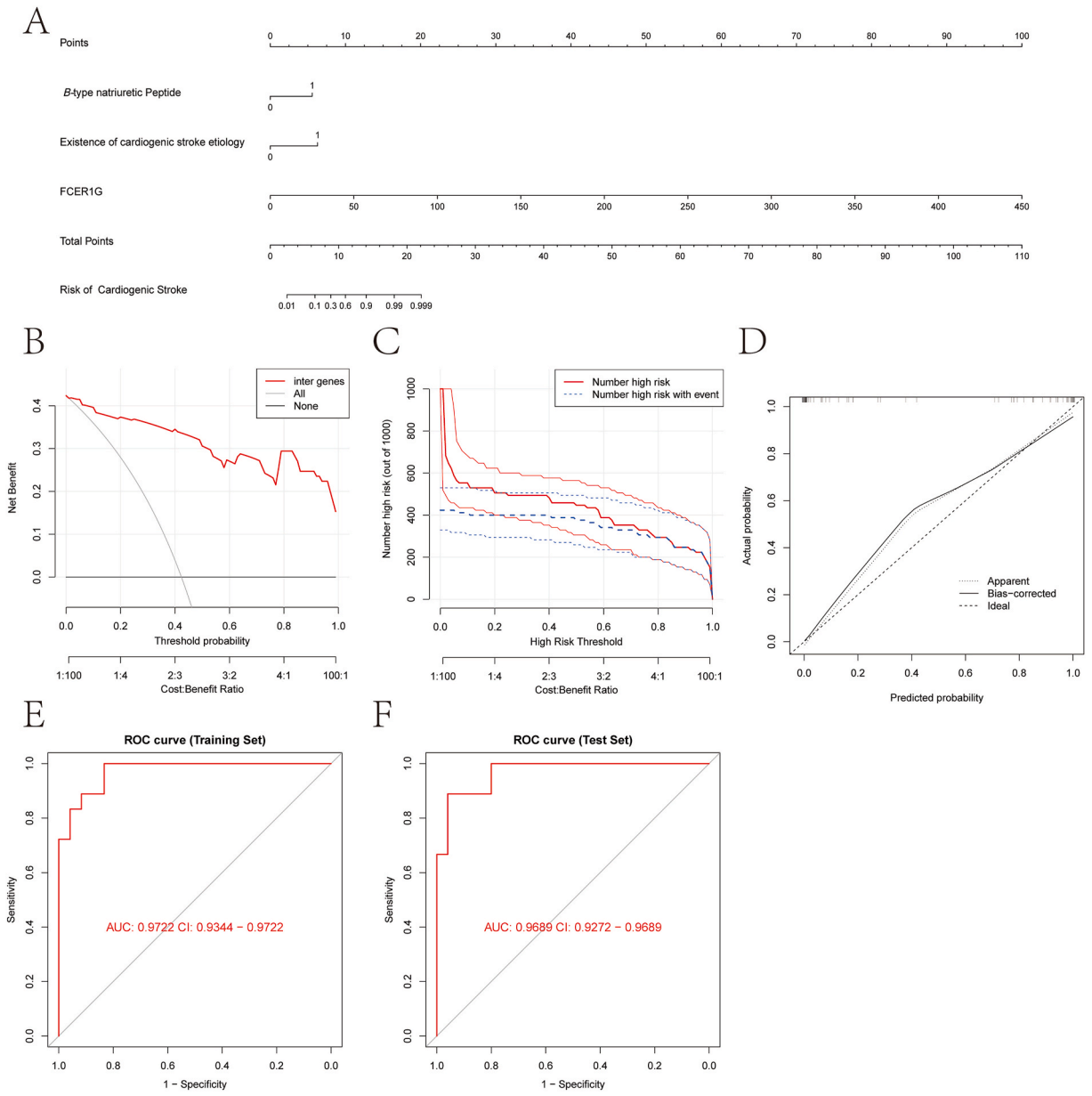


**Fig. 10.** Construction and verification of the nomogram for prediction of CE risk for stroke patients.
(A) Nomogram; (B) decision curve; (C) clinical impact curve; (D) calibration curve; and ROC curves of the (E) training set and (F) validation set.

*FCER1G* is a *FCGR3B* interacting protein. In the present study, after conducting a thorough examination of the extracted data, we evaluated whether *FCER1G* expression was elevated in both publicly available clinical datasets and clinical specimens of CE. Through machine learning algorithms and WGCNA, *FCER1G* was identified as a novel blood biomarker in patients with CE. In addition, we observed no variation in *FCER1G* expression in the blood samples obtained from noncardiogenic stroke datasets, including the datasets for ischemic stroke induced by atherosclerosis and coronary artery disease. We verified *FCER1G* expression in blood samples of CE patients in both publicly available clinical datasets and clinical specimens, and the AUCs were above 0.9 for both the publicly dataset and clinical specimens. Additionally, according to the experimental data, the expression of *FCER1G* was notably elevated in the peripheral blood of CE patients in comparison to that of patients with LAA stroke. In particular, the CE groups exhibited a significant rise in relative *FCER1G* expression when compared to both the non-CE and healthy control groups, which is expected to have implications for differentiating between CE and non-CE when detected in the peripheral blood of patients with CIS.

The analysis of enrichment and infiltration of the immune system suggests a strong association between the developmental progression of CE and immune-related functions. The potential role of *FCER1G* in autoimmunity has been investigated [46], but its role in CE remains unclear. Previously, it has been reported that *FCER1G* is mainly expressed in monocyte/macrophages and plays an

**Table 4**
Biomarkers used for the differential diagnosis of stroke.

| Clinical purpose or question | Type | Biomarker | Sample size | Sample source | Region | Ref. |
|---|---|---|---|---|---|---|
| Serum markers for the diagnosis of CIS | mRNA | ADM, ANXA3, SLC22A4 and VIM | 132 samples from 89 strokes and 43 controls | peripheral blood | China | Peng-Fei Zheng et al., 2022 |
| Identifying LAA stroke patients | lncRNA | ZFAS1 | 176 strokes, 111 controls | peripheral blood | China | Wang et al., 2018 |
| Distinguishing CE from non-CE in CIS | Protein | NT-proBNP | 1840 IS (meta-analysis) | peripheral blood | N/A | Bai et al., 2018 |
| Distinguish CE from LAA | 40-gene panel | ADAMTSL4, AP3S2, ARHGEF12,ARHGEF5, BANK1, C16orf68, C19orf28,CD46, CHURC1, CLEC18A, COL13A1, EBF1, ENPP2, EXT2, FCRL1, FLJ40125, GRM5, GSTK1, HLA-DOA, IRF6, LHFP, LHFP, LOC284751, LRRC37A3, OOEP, P2RX5, PIK3C2B, PTPN20A, TFDP1, TMEM19, TSKS, ZNF185, ZNF254 | 76 cryptogenic strokes (194 samples) | peripheral blood | USA | Jickling et al. 2010 |
| Differentiate between AF and non-AF embolic stroke | A separate 37 gene profile | BRUNOL6, C6orf164, CMTM1, COL13A1, DUSP16, GPR176, GRLF1, HIPK2, LOC100129034, LRRC43, MAP3K7IP1, MIF/// SLC2A11, PER3, PPIE, SDC4, SMC1A, SNORA68, TTC7A | 76 cryptogenic strokes (194 samples) | peripheral blood | USA | Jickling et al., 2010 |
| Stroke versus no stroke | Protein | C3, APOA4 and S100A9 | 18 MACO mouse Samples (23 Samples) | brain | N/A | Li et al., 2020 |

important role in cancer immune infiltration and tumour microenvironment [47]. In addition, *FCER1G* gene hypomethylation was associated the activity of rheumatoid arthritis, which is a chronic autoimmune disease [48]. Recently, there are also some literatures that reported *FCER1G* been implicated in human cardiovascular disease processes. For instance, two bioinformatics studies reported that *FCER1G* is a predictor of acute myocardial infarction [49,50]. Our analysis revealed that *FCER1G* exerts a suppressive effect on plasma cells, mast-activated cells, naive B cells, and follicular helper T cells and exhibits a positive association with resting quiescent mast cells. This results suggest that the *FCER1G* may be a immune-associated biomarker for CE. However, future validations are warranted *in vitro* and *in vivo*.

Most importantly we developed a computational model to forecast the likelihood of CE in individuals who have suffered from CIS. The calibration curve of this model showed high accuracy, indicating good discriminatory power and clinical relevance. On the basis of this model, we developed an online tool for convenient use. The prediction tool indicates disease risk for patients as a percentage, with values closer to 100 % suggesting a higher probability of CE. The development of online prediction tools has greatly simplified the use of nomograms in clinical settings, thereby enhancing the ability of clinicians to assess and choose treatment options for their patients.

## 5. Conclusions

Our study suggests that *FCER1G* may serve as a promising biomarker for CE when detected in the peripheral blood of patients with CIS. Additionally, the online predictive website is expected to have implications for differentiating between CE and non-CE and further provide more optimization of treatment decision supports. The primary constraint of this research was the limited number of participants. The principal limitation of this study was the restricted participant pool. Therefore, it is imperative that future multicenter studies with a substantial sample size be conducted to authenticate the findings of this research and refine the model under optimal conditions. Further exploration of the role and mechanisms of *FCER1G* in CE necessitates additional in *vivo* and in *vitro* investigations for validation.

## Availability of data and materials

The datasets (GSE66724、GSE41177、GSE58294、GSE20129、GSE146882、GSE20680、GSE20681、GSE42148) analyzed during this study are available in the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The dataset and clinical information gathered from the hospital, which were analyzed in the present study, are available from the corresponding author on reasonable request.

## Ethics approval and consent to participate

The studies involving human participants were reviewed and approved by ethics committee of the Nanyang Central Hospital (Nanyang, Henan, P.R. China). Each participant gave written informed consent prior to the tissue sampling.

## CRediT authorship contribution statement

**Yuanzheng Hu:** Data curation, Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. **Xiangxin Li:** Formal analysis, Investigation, Validation. **Kaiqi Hou:** Data curation, Software. **Shoudu Zhang:** Writing – review & editing. **Siyi Zhong:** Writing – original draft. **Qian Ding:** Writing – original draft. **Wuyang Xi:** Writing – original draft. **Zongqing Wang:** Writing – original draft. **Juan Xing:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization. **Fanghui Bai:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization. **Qian Xu:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization.

## Declaration of competing interest

The authors declare that they have no competing interests.

## Acknowledgement

The present study was supported by the Natural Science Foundation of China (Grant No. #81701185), the Specialized Science and Technology Key Project of Henan Province (Grant No. #232102310042), the Special Research and Development Project of Henan Academy of Sciences (Grant No. #220914011) and the key Research and Development Project in Henan Province (Grant No. #231111310900)

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e33846.

## References

[1] W. Shang, J. Liu, Stroke subtype classification: a comparative study of ASCO and modified TOAST, J. Neurol. Sci. 314 (2012) 66–70.
[2] H.P. Adams Jr., B.H. Bendixen, L.J. Kappelle, J. Biller, B.B. Love, D.L. Gordon, et al., Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment, Stroke 24 (1993) 35–41.
[3] Y.J. Wang, Z.X. Li, H.Q. Gu, Y. Zhai, Y. Jiang, X.Q. Zhao, et al., China stroke statistics 2019: a report from the national center for healthcare quality management in neurological diseases, China national clinical research center for neurological diseases, the Chinese stroke association, national center for chronic and non-communicable disease control and prevention, Chinese center for disease control and prevention and institute for global neuroscience and stroke collaborations, Stroke Vasc Neurol 5 (2020) 211–239.
[4] D. Yang, M.S.V. Elkind, Current perspectives on the clinical management of cryptogenic stroke, Expert Rev. Neurother. 23 (2023) 213–226.
[5] Y.N.V. Reddy, B.A. Borlaug, B.J. Gersh, Management of atrial fibrillation across the spectrum of heart failure with preserved and reduced ejection fraction, Circulation 146 (2022) 339–357.
[6] S.N. Azahar, S. Sulong, W.A. Wan Zaidi, N. Muhammad, Y. Kamisah, N. Masbah, Direct medical cost of stroke and the cost-effectiveness of direct oral anticoagulants in atrial fibrillation-related stroke: a cross-sectional study, Int J Environ Res Public Health 19 (2022).
[7] Y. Bai, Y.L. Wang, A. Shantsila, G.Y.H. Lip, The global burden of atrial fibrillation and stroke: a systematic review of the clinical epidemiology of atrial fibrillation in Asia, Chest 152 (2017) 810–820.
[8] V. Thijs, Atrial fibrillation detection: fishing for an irregular heartbeat before and after stroke, Stroke 48 (2017) 2671–2677.
[9] L. Bie, The status and research progress on vitamin D deficiency and atrial fibrillation, Braz. J. Cardiovasc. Surg. 34 (2019) 605–609.
[10] M. Marnane, C.A. Duggan, O.C. Sheehan, A. Merwick, N. Hannon, D. Curtin, et al., Stroke subtype classification to mechanism-specific and undetermined categories by TOAST, A-S-C-O, and causative classification system: direct comparison in the North Dublin population stroke study, Stroke 41 (2010) 1579–1586.
[11] F. Meric-Bernstam, J. Larkin, J. Tabernero, C. Bonini, Enhancing anti-tumour efficacy with immunotherapy combinations, Lancet 397 (2021) 1010–1022.
[12] S. Misra, J. Montaner, L. Ramiro, R. Arora, P. Talwar, M. Nath, et al., Blood biomarkers for the diagnosis and differentiation of stroke: a systematic review and meta-analysis, Int. J. Stroke 15 (2020) 704–721.
[13] J. Kamtchum-Tatuene, G.C. Jickling, Blood biomarkers for stroke diagnosis and management, NeuroMolecular Med. 21 (2019) 344–368.
[14] Y. Tian, Y. Lu, Y. Cao, C. Dang, N. Wang, K. Tian, et al., Identification of diagnostic signatures associated with immune infiltration in Alzheimer's disease by integrating bioinformatic analysis and machine-learning strategies, Front. Aging Neurosci. 14 (2022) 919614.
[15] E. Zhao, H. Xie, Y. Zhang, Predicting diagnostic gene biomarkers associated with immune infiltration in patients with acute myocardial infarction, Front Cardiovasc Med 7 (2020) 586871.
[16] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.
[17] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, et al., NCBI GEO: archive for functional genomics data sets–update, Nucleic Acids Res. 41 (2013) D991–D995.
[18] M. Allende, E. Molina, E. Guruceaga, I. Tamayo, J.R. Gonzalez-Porras, T.J. Gonzalez-Lopez, et al., Hsp70 protects from stroke in atrial fibrillation patients by preventing thrombosis without increased bleeding risk, Cardiovasc. Res. 110 (2016) 309–318.
[19] Y.H. Yeh, C.T. Kuo, Y.S. Lee, Y.M. Lin, S. Nattel, F.C. Tsai, et al., Region-specific gene expression profiles in the left atria of patients with valvular atrial fibrillation, Heart Rhythm 10 (2013) 383–391.
[20] B. Stamova, G.C. Jickling, B.P. Ander, X. Zhan, D. Liu, R. Turner, et al., Gene expression in peripheral immune cells following cardioembolic stroke is sexually dimorphic, PLoS One 9 (2014) e102550.
[21] C.C. Huang, D.M. Lloyd-Jones, X. Guo, N.M. Rajamannan, S. Lin, P. Du, et al., Gene expression variation between African Americans and whites is associated with coronary artery calcification: the multiethnic study of atherosclerosis, Physiol. Genom. 43 (2011) 836–843.
[22] M.R. Elashoff, J.A. Wingrove, P. Beineke, S.E. Daniels, W.G. Tingley, S. Rosenberg, et al., Development of a blood-based gene expression algorithm for assessment of obstructive coronary artery disease in non-diabetic patients, BMC Med. Genom. 4 (2011) 26.
[23] Y. Zhou, B. Zhou, L. Pache, M. Chang, A.H. Khodabakhshi, O. Tanaseichuk, et al., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets, Nat. Commun. 10 (2019) 1523.

[24] L.M. Schriml, J.B. Munro, M. Schor, D. Olley, C. McCracken, V. Felix, et al., The human disease ontology 2022 update, Nucleic Acids Res. 50 (2022) D1255–D1261.
[25] C.H. Chin, S.H. Chen, H.H. Wu, C.W. Ho, M.T. Ko, C.Y. Lin, cytoHubba: identifying hub objects and sub-networks from complex interactome, BMC Syst. Biol. 8 (Suppl 4) (2014) S11.
[26] M. Franz, H. Rodriguez, C. Lopes, K. Zuberi, J. Montojo, G.D. Bader, et al., GeneMANIA update 2018, Nucleic Acids Res. 46 (2018) W60–W64.
[27] H.C. Diener, J.D. Easton, R.G. Hart, S. Kasner, H. Kamel, G. Ntaios, Review and update of the concept of embolic stroke of undetermined source, Nat. Rev. Neurol. 18 (2022) 455–465.
[28] A.S. Kim, Evaluation and prevention of cardioembolic stroke, Continuum 20 (2014) 309–322.
[29] L. Moore, Z. Pan, M. Brotto, RNAseq of osteoarthritic synovial tissues: systematic literary review, Front. Aging 3 (2022) 836791.
[30] I. Migdady, A. Russman, A.B. Buletko, Atrial fibrillation and ischemic stroke: a clinical review, Semin. Neurol. 41 (2021) 348–364.
[31] C. Flach, W. Muruet, C.D.A. Wolfe, A. Bhalla, A. Douiri, Risk and secondary prevention of stroke recurrence: a population-base cohort study, Stroke 51 (2020) 2435–2444.
[32] S.H. Chang, I.J. Chou, Y.H. Yeh, M.J. Chiou, M.S. Wen, C.T. Kuo, et al., Association between use of non-vitamin K oral anticoagulants with and without concurrent medications and risk of major bleeding in nonvalvular atrial fibrillation, JAMA 318 (2017) 1250–1259.
[33] E. Vizzardi, A. Curnis, M.G. Latini, F. Salghetti, E. Rocco, L. Lupi, et al., Risk factors for atrial fibrillation recurrence: a literature review, J. Cardiovasc. Med. 15 (2014) 235–253.
[34] B. Zhu, Y. Pan, J. Jing, X. Meng, X. Zhao, L. Liu, et al., Neutrophil counts, neutrophil ratio, and new stroke in minor ischemic stroke or TIA, Neurology 90 (2018) e1870–e1878.
[35] N.P. Torres-Aguila, C. Carrera, A.K. Giese, N. Cullell, E. Muino, J. Carcel-Marquez, et al., Genome-wide association study of white blood cell counts in patients with ischemic stroke, Stroke 50 (2019) 3618–3621.
[36] Y. Gan, Q. Liu, W. Wu, J.X. Yin, X.F. Bai, R. Shen, et al., Ischemic neurons recruit natural killer cells that accelerate brain infarction, Proc Natl Acad Sci U S A. 111 (2014) 2704–2709.
[37] M.S.V. Elkind, R. Veltkamp, J. Montaner, S.C. Johnston, A.B. Singhal, K. Becker, et al., Natalizumab in acute ischemic stroke (ACTION II): a randomized, placebo-controlled trial, Neurology 95 (2020) e1091–e1104.
[38] P.J. Winklewski, M. Radkowski, U. Demkow, Cross-talk between the inflammatory response, sympathetic activation and pulmonary infection in the ischemic stroke, J. Neuroinflammation 11 (2014) 213.
[39] P.F. Zheng, L.Z. Chen, P. Liu, H.W. Pan, W.J. Fan, Z.Y. Liu, Identification of immune-related key genes in the peripheral blood of ischaemic stroke patients using a weighted gene coexpression network analysis and machine learning, J. Transl. Med. 20 (2022) 361.
[40] J. Wang, J. Ruan, M. Zhu, J. Yang, S. Du, P. Xu, et al., Predictive value of long noncoding RNA ZFAS1 in patients with ischemic stroke, Clin. Exp. Hypertens. 41 (2019) 615–621.
[41] J. Bai, H. Sun, L. Xie, Y. Zhu, Y. Feng, Detection of cardioembolic stroke with B-type natriuretic peptide or N-terminal pro-BNP: a comparative diagnostic meta-analysis, Int. J. Neurosci. 128 (2018) 1100–1108.
[42] G.C. Jickling, H. Xu, B. Stamova, B.P. Ander, X. Zhan, Y. Tian, et al., Signatures of cardioembolic and large-vessel ischemic stroke, Ann. Neurol. 68 (2010) 681–692.
[43] L. Li, L. Dong, Z. Xiao, W. He, J. Zhao, H. Pan, et al., Integrated analysis of the proteome and transcriptome in a MCAO mouse model revealed the molecular landscape during stroke progression, J. Adv. Res. 24 (2020) 13–27.
[44] M. Arbaizar-Rovirosa, J. Pedragosa, J.J. Lozano, C. Casal, A. Pol, M. Gallizioli, et al., Aged lipid-laden microglia display impaired responses to stroke, EMBO Mol. Med. 15 (2023) e17175.
[45] L. Liu, Y. Yu, L.L. Hu, Q.B. Dong, F. Hu, L.J. Zhu, et al., Potential target genes in the development of atrial fibrillation: a comprehensive bioinformatics analysis, Med Sci Monit 27 (2021) e928366.
[46] C. Chou, X. Zhang, C. Krishna, B.G. Nixon, S. Dadi, K.J. Capistrano, et al., Programme of self-reactive innate-like T cell-mediated cancer immunity, Nature 605 (2022) 139–145.
[47] R. Yang, Z. Chen, L. Liang, S. Ao, J. Zhang, Z. Chang, et al., Fc Fragment of IgE Receptor Ig (FCER1G) acts as a key gene involved in cancer immune infiltration and tumour microenvironment, Immunology 168 (2023) 302–319.
[48] D. Podgorska, M. Ciesla, B. Kolarz, FCER1G gene hypomethylation in patients with rheumatoid arthritis, J. Clin. Med. 11 (2022).
[49] S. Guo, J. Wu, W. Zhou, X. Liu, Y. Liu, J. Zhang, et al., Identification and analysis of key genes associated with acute myocardial infarction by integrated, bioinformatics methods. Medicine (Baltimore) 100 (2021) e25553.
[50] S. Xiao, Y. Zhou, Q. Wu, Q. Liu, M. Chen, T. Zhang, et al., FCER1G and PTGS2 serve as potential diagnostic biomarkers of acute myocardial infarction based on integrated bioinformatics analyses, DNA Cell Biol. 40 (2021) 1064–1075.