



## OPEN ACCESS

EDITED BY  
Jaqueline Bohrer Schuch,  
Federal University of Rio Grande do  
Sul, Brazil

REVIEWED BY  
Yanfang Liu,  
Jinan University, China  
Jiankun Zang,  
Jinan University, China

\*CORRESPONDENCE  
Jiayun Liu  
jiayun@fmmu.edu.cn  
Jingzhou Chen  
chendragon1976@aliyun.com

SPECIALTY SECTION  
This article was submitted to  
Neurogenomics,  
a section of the journal  
Frontiers in Neuroscience

RECEIVED 25 July 2022  
ACCEPTED 14 November 2022  
PUBLISHED 29 November 2022

CITATION  
Bai C, Hao X, Zhou L, Sun Y, Song L,  
Wang F, Yang L, Liu J and Chen J  
(2022) Machine learning-based  
identification of the novel circRNAs  
circERBB2 and circCHST12 as  
potential biomarkers of intracerebral  
hemorrhage.  
*Front. Neurosci.* 16:1002590.  
doi: 10.3389/fnins.2022.1002590

COPYRIGHT  
© 2022 Bai, Hao, Zhou, Sun, Song,  
Wang, Yang, Liu and Chen. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which  
does not comply with these terms.

# Machine learning-based identification of the novel circRNAs circERBB2 and circCHST12 as potential biomarkers of intracerebral hemorrhage

Congxia Bai<sup>1</sup>, Xiaoyan Hao<sup>1</sup>, Lei Zhou<sup>1</sup>, Yingying Sun<sup>2</sup>,  
Li Song<sup>2</sup>, Fengjuan Wang<sup>1</sup>, Liu Yang<sup>1</sup>, Jiayun Liu<sup>1\*</sup> and  
Jingzhou Chen<sup>2,3\*</sup>

<sup>1</sup>Department of Clinical Laboratory Medicine, Xijing Hospital, Fourth Military Medical University, Xi'an, China, <sup>2</sup>State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, <sup>3</sup>National Health Commission Key Laboratory of Cardiovascular Regenerative Medicine, Fuwai Central-China Hospital, Central-China Branch of National Center for Cardiovascular Diseases, Zhengzhou, China

**Background:** The roles and potential diagnostic value of circRNAs in intracerebral hemorrhage (ICH) remain elusive.

**Methods:** This study aims to investigate the expression profiles of circRNAs by RNA sequencing and RT-PCR in a discovery cohort and an independent validation cohort. Bioinformatics analysis was performed to identify the potential functions of circRNA host genes. Machine learning classification models were used to assess circRNAs as potential biomarkers of ICH.

**Results:** A total of 125 and 284 differentially expressed circRNAs (fold change > 1.5 and FDR < 0.05) were found between ICH patients and healthy controls in the discovery and validation cohorts, respectively. Nine circRNAs were consistently altered in ICH patients compared to healthy controls. The combination of the novel circERBB2 and circCHST12 in ICH patients and healthy controls showed an area under the curve of 0.917 (95% CI: 0.869–0.965), with a sensitivity of 87.5% and a specificity of 82%. In combination with ICH risk factors, circRNAs improved the performance in discriminating ICH patients from healthy controls. Together with hsa\_circ\_0005505, two novel circRNAs for differentiating between patients with ICH and healthy controls showed an AUC of 0.946 (95% CI: 0.910–0.982), with a sensitivity of 89.1% and a specificity of 86%.

**Conclusion:** We provided a transcriptome-wide overview of aberrantly expressed circRNAs in ICH patients and identified hsa\_circ\_0005505 and novel circERBB2 and circCHST12 as potential biomarkers for diagnosing ICH.

#### KEYWORDS

intracerebral hemorrhage, RNA sequencing, circular RNA, biomarkers, machine learning algorithms

## Introduction

Stroke causes high levels of mortality and disability globally. Intracerebral hemorrhage (ICH) is a deadly stroke subtype with an estimated annual incidence of 16 per 100,000 persons worldwide (Wilkinson et al., 2018). ICH accounts for approximately 23.8% of stroke cases in China, compared with Western countries, where it accounts for 10–15% of stroke cases, causing a median fatality ratio of 40.4% per month (Qureshi et al., 2009; Benjamin et al., 2017). The diagnosis of stroke is often made with computed tomography (CT) or magnetic resonance imaging (MRI), and although most patients are hospitalized with typical neurological symptoms, it is difficult to distinguish ICH from ischemic stroke (IS) in the super acute period (Hankey, 2017). Thus, identifying potential biomarkers for the early prediction and diagnosis of ICH is important.

Non-coding RNAs (ncRNAs) have been extensively studied in the pathophysiology of cerebrovascular diseases (Weng et al., 2022). Changes in RNA levels during stroke have the potential to aid stroke diagnosis and provide insight into stroke diagnosis and management (Montaner et al., 2020). Emerging evidence has revealed that ncRNA expression profiles are altered in the peripheral blood of patients with ICH (Kim et al., 2019; Li et al., 2019; Cheng et al., 2020). CircRNAs are a novel class of ncRNAs that are produced in eukaryotic cells during posttranscriptional processes; these covalently closed RNAs lack a free 3' or 5' end and are resistant to exonuclease digestion (Kristensen et al., 2019). Thus, circRNAs are promising diagnostic and prognostic biomarkers for many human diseases because of their stability, specificity and abundance in human blood (Jeck and Sharpless, 2014; Zhang et al., 2018). Growing evidence has demonstrated that circRNAs are implicated in a variety of pathological conditions, including coronary artery disease (Cardona-Monzonis et al., 2020), acute ischemic stroke (Liu Y. et al., 2022) and cancers (Kristensen et al., 2022). Moreover, the expression of circRNAs was found to be significantly altered in IS (Tiedt et al., 2017; Dong et al., 2020; Li et al., 2020; Lu et al., 2020; Ostolaza et al., 2020; Zuo et al., 2020), and these studies implied that aberrantly expressed circRNAs may be novel biomarkers for IS diagnosis and prognosis. Our previous study revealed that circRNA profiles were significantly altered in hypertensive ICH patients compared to hypertensive subjects without

ICH and found that hsa\_circ\_0001240, hsa\_circ\_0001947 and hsa\_circ\_0001386 were potential biomarkers for predicting and diagnosing hypertensive ICH (Bai et al., 2021). In addition, circRNA expression is significantly altered in rat brain tissue after ICH (Dou et al., 2020; Zhong et al., 2020), indicating that circRNAs are novel clinical biomarkers for ICH. However, comprehensive circRNA expression profiles and their potential diagnostic value in the peripheral blood of ICH patients remain elusive.

Artificial intelligence techniques such as machine learning tools have been increasingly used in precision diagnosis (Chang et al., 2021). Machine learning algorithms are artificial intelligence techniques used to select the best model from a set of alternatives to fit a set of observations (Li, 2018). Machine learning has remained a fundamental and indispensable tool due to its efficacy and efficiency in both feature extraction of relevant biomarkers and the classification of samples as validation of the discovered biomarkers (Ledesma et al., 2021).

In this study, we investigated the expression profile of circRNAs in peripheral blood cells from patients with ICH, patients with IS and healthy controls by RNA sequencing in the discovery and validation cohorts. The significantly altered circRNA host genes were examined with Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses to characterize the potential functions. We further validated the altered circRNAs by quantitative reverse transcription-PCR (RT-PCR) analysis of all samples. Logistic regression models were performed to identify whether circRNAs were independent factors for ICH. Additionally, we performed Spearman's correlation analysis to investigate the correlation between ICH risk factors and candidate circRNAs. Furthermore, machine learning classification algorithms and ROC curves were used to assess circRNAs as potential biomarkers of ICH.

## Materials and methods

### Study design and sample collection

We recruited 64 patients with ICH, 59 patients with IS and 50 sex- and age-matched healthy controls between 2014 and 2019 from two individual cohorts for RNA sequencing. In the discovery cohort, 44 patients with ICH, 43 patients

with IS and 31 healthy controls were enrolled from Cangzhou Central Hospital between 2014 and 2017. In the validation cohort, 20 patients with ICH were enrolled from the Affiliated Hospital of Hebei University, 16 patients with IS were enrolled from General Hospital of Ningxia Medical University, and 19 healthy control subjects were enrolled from the Tsinghua University Hospital between 2017 and 2019. Patients with ICH were diagnosed by professional neurologists based on their histories and examinations, and ICH was confirmed by CT or MRI. Healthy controls without a history of stroke or cardiovascular events were selected. The demographic and clinical characteristics of the study population were obtained through a face-to-face survey and by checking hospital records or medical examination records. The exclusion criteria included autoimmune diseases, cardiac disease, liver diseases, renal diseases, cancer or a history of stroke and cerebral infarction with hemorrhagic transformation. This study was reviewed and approved by the Human Ethics Committee, Fuwai Hospital (Approval No. 2016-732), and conducted in accordance with the principles of Good Clinical Practice and the Declaration of Helsinki. Written informed consent was obtained from all participants or their legal proxies.

## RNA isolation and sequencing

RNA was isolated from human peripheral blood and used to perform RNA sequencing by Annoroad

Gene Technology Company Ltd. (Beijing, China), as previously described (Bai et al., 2021). Total RNA from all samples was isolated with an RNeasy Mini kit (QIAGEN). An Agilent 2100 RNA Nano 6000 Assay Kit (Agilent Technologies, CA, USA) was used to measure RNA integrity. The libraries were constructed using an RNA integrity number  $\geq 7.5$  and a 28S:18S rRNA ratio  $\geq 1.8$ . Ribo-Zero™ Gold Kits (Illumina, San Diego, CA, USA) were utilized to eliminate all ribosomal RNAs from total RNA. RNase R (Epicenter, Madison, WI, USA) digestion was used to eliminate linear RNAs. The purified circRNAs were subjected to the NEB Next Ultra Directional RNA Library Prep Kit for Illumina (NEB, Ipswich, USA) according to the manufacturer's instructions. The obtained libraries were subjected to paired-end sequencing with 150 bp reads performed on the Illumina PE150 platform. The sequence depth was approximately 15G. The raw sequencing data were analyzed using Q30 statistics from FastQC, and clean reads were obtained by removing adaptor-polluted and low-quality reads. The RNA-seq data have been deposited into the Genome Sequence Archive (Chen T. et al., 2021) in the National Genomics Data Center (CNCB-NGDC Members and Partners, 2022), China National Center for Bioinformatics/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: HRA001807), which are publicly accessible at <https://ngdc.cncb.ac.cn/gsa-human>.

TABLE 1 Demographics and characteristics of the discovery and validation cohorts.

	Discovery cohort			P-value	Validation cohort			P-value
	Control (n = 31)	ICH (n = 44)	IS (n = 43)		Control (n = 19)	ICH (n = 20)	IS (n = 16)	
Age, y	58.9 ± 5.3	55.9 ± 7.2	57.4 ± 5.5	0.09	57.2 ± 7.0	56.7 ± 7.1	57.2 ± 7.7	0.86
Men,%	17 (54.8)	24 (54.5)	21 (48.8)	0.83	10 (52.6)	10 (50)	8 (50)	0.98
BMI, kg/m <sup>2</sup>	24.8 ± 2.9	26.1 ± 6.6	27.6 ± 6.9	0.09	24.9 ± 2.4	25.8 ± 6.8	25.0 ± 2.6	0.90
SBP, mmHg	125.7 ± 10.1	137.4 ± 17.6	138.6 ± 13.6	< 0.001	120.3 ± 9.7	171.2 ± 25.7	150.6 ± 19.4	< 0.001
DBP, mmHg	79.2 ± 4.3	87.9 ± 10.7	91.8 ± 16.6	< 0.001	77.6 ± 9.1	103.7 ± 13.3	89.3 ± 13.9	< 0.001
HDL-C, mmol/L	1.4 ± 0.3	1.1 ± 0.3	1.1 ± 0.2	< 0.001	1.3 ± 0.3	0.9 ± 0.5	1.0 ± 0.3	0.007
LDL-C, mmol/L	2.9 ± 0.7	2.4 ± 0.8	2.3 ± 0.8	< 0.001	2.9 ± 0.9	2.8 ± 0.8	2.7 ± 0.9	0.82
TC, mmol/L	5.5 ± 1.0	4.5 ± 1.0	4.5 ± 1.0	< 0.001	4.5 ± 1.0	4.3 ± 0.9	4.9 ± 1.3	0.23
TG, mmol/L	1.4 ± 0.8	1.5 ± 0.9	1.6 ± 0.6	0.41	1.2 ± 0.5	1.4 ± 0.6	2.3 ± 1.5	0.004
GLU, mmol/L	6.0 ± 1.8	6.3 ± 1.6	5.9 ± 1.3	0.17	5.3 ± 0.6	5.5 ± 1.7	6.0 ± 1.1	0.21
Smoking,%				0.95				0.92
Never	19 (61.3)	28 (63.7)	26 (60.5)		13 (68.4)	14 (70)	11 (68.7)	
Former	4 (12.9)	5 (13.6)	8 (18.6)		3 (15.8)	2 (10)	2 (12.5)	
Current	8 (25.8)	11 (22.7)	9 (20.9)		3 (15.8)	4 (20)	3 (18.8)	
Drinking,%				0.98				0.96
Non-drinker	20 (64.5)	28 (63.6)	27 (62.8)		11 (57.9)	12 (60)	10 (62.5)	
Drinker	11 (35.5)	16 (36.4)	16 (37.2)		8 (42.1)	8 (40)	6 (37.5)	

Data are expressed as the mean ± standard deviation or n (%).

BMI, Body mass index; SBP, Systolic blood pressure; DBP, Diastolic blood pressure; TC, Total cholesterol; TG, Triacylglycerol; HDL-C, High-density lipoprotein cholesterol; LDL-C, Low-density lipoprotein cholesterol; GLU, Glucose; ICH, Intracerebral hemorrhage; IS, ischemic stroke.

Statistical comparisons for percentages were performed using the chi-square test. Comparisons between means or medians were performed using one-way ANOVA.

## Differential expression analysis

The differential expression circRNA analysis was performed as previously described (Bai et al., 2021). Briefly, CIRI2 (Gao et al., 2018) was used to detect paired chiasmic clipping signals according to the mapping of reads. The reads were mapped to the reference genome<sup>1</sup> using the BWA-MEM method. Back-spliced junction reads were integrated and measured by spliced reads per billion mapping to quantify circRNA. Differential expression analysis was performed using the DESeq2 R package (Wang et al., 2010) and edgeR (Robinson et al., 2010). Fold differences of each circRNA were calculated to identify differentially expressed circRNAs between ICH patients and healthy controls (or IS patients) by Student's *t*-test. A *P* value was assigned to each circRNA and adjusted by multiple testing using the Benjamini–Hochberg method for controlling the false discovery rate (FDR). The differentially expressed circRNAs were defined as those with a fold change  $\geq 1.5$  and FDR  $< 0.05$ .

## Bioinformatics analysis

Volcano plots and hierarchical clustering using heatmaps were generated based on the normalized values of differentially expressed genes using the R package. Venn diagrams were used to present the consistently differentially expressed genes in the discovery and validation cohorts. GO enrichment and KEGG analyses were performed to determine the biological functions and pathways of differentially expressed circRNA host genes. *P* values were calculated using Fisher's exact test with the hypergeometric algorithm.

## Quantitative real-time polymerase chain reaction validation

To validate the expression levels of differentially expressed circRNAs identified by RNA-seq, the candidate circRNAs were selected for further validation of expression levels by quantitative RT-PCR. Total RNA was incubated with RNase R or RNase-free water as a control at 37°C for 30 min to purify the circRNAs. After incubation, cDNA synthesis was completed using 1  $\mu$ g of total RNA and a Transcriptor First Stand cDNA Synthesis Kit (Takara, Dalian, China), and Taq premix (Takara, Dalian, China) was added to start PCR according to the manufacturer's protocol. The products were used for Sanger sequencing. Quantitative RT-PCR was performed using SYBR Master Mix (Yeasen, Shanghai, China) on the ViiA 7 Real-time PCR System (Applied Biosystems) according to

the manufacturer's instructions. The circRNA primers were designed to overlap the back-spliced junction using the NCBI Primer-BLAST website.<sup>2</sup> The primers used in this study are listed in **Supplementary Table 7**. The relative expression of the corresponding genes was quantified and normalized to that of GAPDH.

## Performance evaluation of candidate biomarkers with classification algorithms

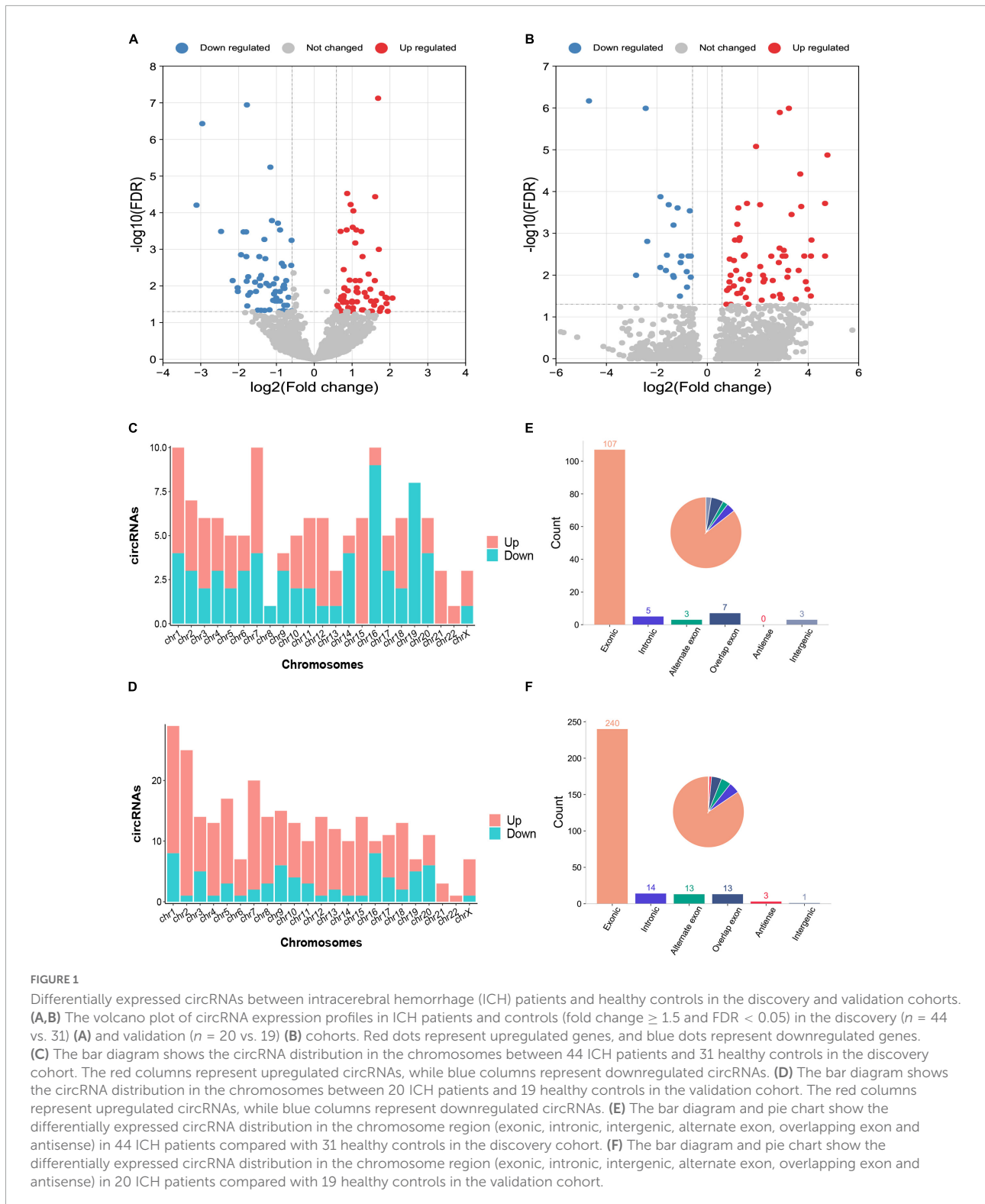
To evaluate the applicable biomarkers for ICH, we used mutual information (MI) (Blokh and Stambler, 2017) and random forest (RF) algorithms (Ambale-Venkatesh et al., 2017; Kawakami et al., 2019) to screen circRNA biomarker signatures according to the expression levels in all samples. To assess the diagnostic values of the specific circRNAs, we used six machine learning classification algorithms (Chang et al., 2021; Chen Y. et al., 2021; Liu D. et al., 2022), support vector machine (SVM), RF, K-nearest neighbor (KNN), logistic regression (LR), decision tree (DT) and Gaussian naive Bayes (GNB), to discriminate ICH patients from healthy controls or IS patients according to the expression levels of circRNAs by Python packages. To ensure the stability and accuracy of the classifiers, we used 10-fold cross-validation; 90% of the data were used for the training set, and 10% were used for the test set. We calculated five measurements, including sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) (Shu et al., 2020). The ROC curve was illustrated based on sensitivity and 1-specificity scores. For each area under the curve (AUC) value, the 95% CI was computed with 1000 stratified bootstrap replicates.

## Statistical analysis

Statistical analysis was performed using SPSS 21.0 (IBM Corp., NY, USA). The sample distribution was determined using the Kolmogorov–Smirnov normality test. For parametric data, the two-tailed unpaired Student's *t*-test was used to determine differences between two groups. The data are represented as the means  $\pm$  standard deviations or medians (interquartile range). Statistical comparisons for percentages were performed using chi-square statistical analysis. In the RNA sequencing analysis, differentially expressed RNAs were selected if there were significant differences (fold change  $> 1.5$  and FDR  $< 0.05$ ) between the ICH patients and healthy controls (or IS patients) using Student's *t*-test. Logistic regression models were used to evaluate

<sup>1</sup> <http://www.ensembl.org/index.html>

<sup>2</sup> <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>



whether circRNAs were independent predictive factors for ICH. Spearman’s correlation analysis was performed to investigate the correlation between ICH risk factors and circRNAs. The net reclassification index (NRI) and integrated

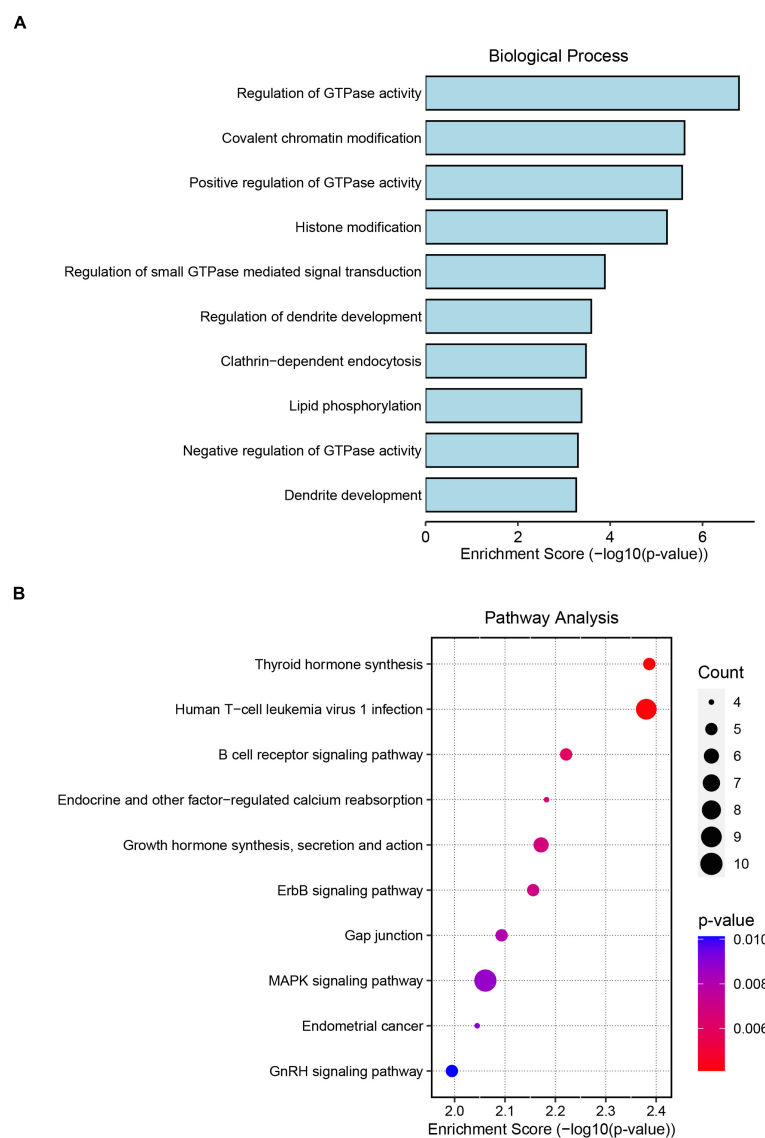
discrimination improvement (IDI) were calculated to evaluate the effect of the candidate biomarkers as previously described (Wu et al., 2020).  $P < 0.05$  was considered indicative of statistical significance.

## Results

### CircRNA expression profiles were significantly altered in intracerebral hemorrhage patients

The characteristics and demographics of the cohorts of ICH patients, IS patients and healthy controls are shown in **Table 1**. In RNA sequencing, the significantly differentially expressed circRNAs were determined by a fold change > 1.5 and FDR < 0.05 by DESeq2 methods. In total, 125 circRNAs were significantly altered between

patients with ICH and controls, including 63 upregulated circRNAs and 62 downregulated circRNAs in the discovery cohort (**Figure 1A** and **Supplementary Table 1**), and 284 circRNAs were significantly altered between patients with ICH and healthy controls in the validation cohort, including 218 upregulated circRNAs and 66 downregulated circRNAs (**Figure 1B** and **Supplementary Table 2**). Additionally, the circRNAs were distributed across all chromosomes in both cohorts (**Figures 1C,D**). There were 107 circRNAs produced by classic exon back-splicing, 3 alternate exons, 5 introns, 7 overlapping exons, and 3 intergenic circRNAs detected between ICH patients and controls in the discovery cohort



**FIGURE 2** Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses of significantly altered circRNA host genes. **(A)** The top 10 biological process terms from GO enrichment analysis of differentially expressed circRNA host genes. **(B)** The top 10 KEGG pathway analyses of differentially expressed circRNA host genes.

(Figure 1E), and 240 circRNAs produced by classic exon back-splicing, 13 alternate exons, 14 introns, 13 overlapping exons, 3 antisense and 1 intergenic circRNA were detected

between ICH patients and controls in the validation cohort (Figure 1F). Moreover, we observed that 302 and 395 circRNAs were significantly altered between ICH and IS patients in the

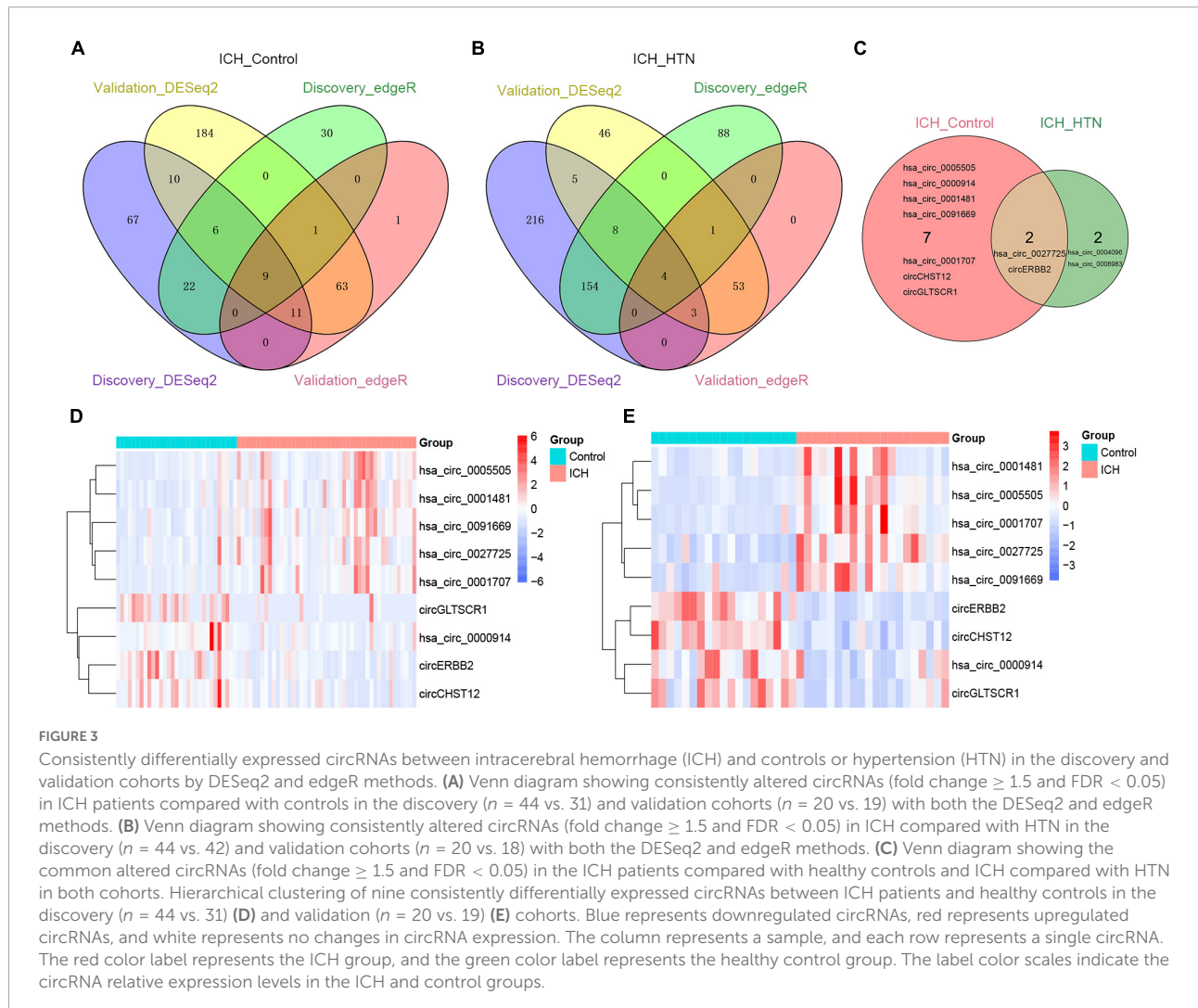


TABLE 2 The consistently altered circRNAs in intracerebral hemorrhage (ICH) patients compared with controls.

Location	circRNA ID	Discovery		Validation		Up/Down	Host gene	Type
		FC	FDR	FC	FDR			
chr7:48541721-48542148: +	hsa_circ_0001707	3.038	2.4E-04	3.817	2.1E-09	up	Null	Intronic
chrX:147733519-147744289: +	hsa_circ_0091669	1.827	0.005	2.138	1.4E-06	up	AFF2	Exonic
chr12:66597490-66622150: +	hsa_circ_0005505	2.903	0.007	7.329	2.8E-10	up	IRAK3	Exonic
chr5:49694940-49707217: -	hsa_circ_0001481	1.850	0.012	2.803	3.9E-06	up	EMB	Exonic
chr12:94562928-94580249: +	hsa_circ_0027725	2.179	0.037	2.281	4.9E-07	up	PLXNC1	Exonic
chr7:2477438-2483381: +	circCHST12	0.334	0.007	0.395	5.5E-07	down	CHST12	Alternate exon
chr19:18648410-18649246: -	hsa_circ_0000914	0.510	0.010	0.595	4.6E-06	down	FKBP8	Exonic
chr17:37866065-37872192: +	circERBB2	0.278	0.020	0.184	1.7E-10	down	ERBB2	Exonic
chr19:48185232-48198731: +	circGLTSCR1	0.183	0.037	0.275	1.3E-05	down	GLTSCR1	Exonic

FC, fold change; FDR, false discovery rate.

discovery and validation cohorts, respectively (Supplementary Figures 1A,B).

## Gene ontology enrichment and kyoto encyclopedia of genes and genomes pathway analyses of circRNA host genes

To assess the potential regulatory mechanism of differentially expressed circRNAs in host gene transcription after ICH, we performed GO and KEGG pathway analyses of the host genes of the altered circRNAs in the two cohorts. The top GO terms in the biological process category indicated that the host genes were involved in the regulation of GTPase activity, covalent chromatin modification, histone modification, regulation of dendrite development and lipid phosphorylation (Figure 2A). KEGG pathway analysis showed that the host genes were mainly involved in the MAPK signaling network, B-cell receptor signaling, ERBB receptor signaling network, thyroid hormone synthesis and lysine degradation (Figure 2B).

## Consistently altered circRNAs in the discovery and validation cohorts

To elucidate the underlying mechanism by which the circRNAs affected ICH more specifically, we screened the common circRNAs in the two cohorts by both DESeq2 and edgeR methods (Supplementary Tables 1–4) and found that 9 circRNAs overlapped between the ICH patients and controls (Figure 3A). Similarly, there were 4 consistent circRNAs between ICH and hypertension (HTN) in our previous study (Figure 3B) (Bai et al., 2021); 2 of them were consistently altered in the two comparison groups, including hsa\_circ\_0027725 and a novel circRNA (host gene *ERBB2*) we named circERBB2 (Figure 3C).

The nine consistently altered circRNAs included five upregulated circRNAs and four downregulated circRNAs. The five upregulated circRNAs in ICH were hsa\_circ\_0001707, hsa\_circ\_0091669, hsa\_circ\_0005505, hsa\_circ\_0001481 and hsa\_circ\_0027725; the 4 downregulated circRNAs in ICH were hsa\_circ\_0000914 and three novel circRNAs that we named according to their host genes, circCHST12 (host gene *CHST12*), circERBB2 and circGLTSCR1 (host gene *GLTSCR1*) (Table 2). The 9 circRNA expression variants are shown with hierarchical clustering heatmaps in the discovery and validation cohorts (Figures 3D,E), which indicated that the circRNA expression profiles in ICH patients were distinct from those in healthy control groups.

Likewise, we detected 20 consistent circRNAs between ICH and IS patients in the two cohorts by both DESeq2 and edgeR methods (Supplementary

Figure 1C). Notably, 3 circRNAs were in the intersection between ICH versus controls (9 consistent circRNAs) and ICH versus IS (20 consistent circRNAs), including circERBB2, circCHST12 and hsa\_circ\_0005505 (Supplementary Figure 1D).

## Investigation of the nine circRNAs as independent predictors of intracerebral hemorrhage

To further explore the potential value of candidate circRNAs as ICH biomarkers, logistic regression models were performed to identify whether nine circRNAs could be predictors of ICH occurrence. As shown in Table 3, after adjusting for age, sex, body mass index (BMI), systolic blood pressure (SBP), diastolic blood pressure (DBP), total cholesterol (TC), triacylglycerol (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), smoking and alcohol consumption, per unit of increase in hsa\_circ\_0001707, hsa\_circ\_0091669, hsa\_circ\_0005505, hsa\_circ\_0001481 and hsa\_circ\_0027725, the odds ratios for ICH occurrence were 2.23 (95% CI: 1.294–3.842;  $P = 0.004$ ), 3.372 (95% CI: 1.665–6.867;  $P = 0.001$ ), 2.216 (95% CI: 1.363–3.316;  $P = 0.001$ ), 4.750 (95% CI: 2.054–10.985;  $P < 0.001$ ) and 2.156 (95% CI: 1.170–3.974;  $P = 0.014$ ), respectively. In addition, the adjusted ORs were 0.009 (95% CI: 0.001–0.097;  $P < 0.001$ ), 0.160 (95% CI: 0.051–0.507;  $P = 0.002$ ), 0.019 (95% CI: 0.002–0.157;  $P < 0.001$ ) and 0.122 (95% CI: 0.037–0.410;  $P = 0.001$ ) per unit increase in circCHST12, hsa\_circ\_0000914, circERBB2 and circGLTSCR1, respectively.

TABLE 3 Logistic regression analysis to identify circRNAs as independent predictive factors of intracerebral hemorrhage (ICH).

circRNA ID	Adjusted risk factors			Up/Down	Host gene
	OR	95% CI	P-value		
hsa_circ_0001707	2.230	1.294–3.842	0.004	up	Null
hsa_circ_0091669	3.372	1.655–6.867	0.001	up	AFF2
hsa_circ_0005505	2.216	1.363–3.316	0.001	up	IRAK3
hsa_circ_0001481	4.750	2.054–10.985	< 0.001	up	EMB
hsa_circ_0027725	2.156	1.170–3.974	0.014	up	PLXNC1
circCHST12	0.009	0.001–0.097	< 0.001	down	CHST12
hsa_circ_0000914	0.160	0.051–0.507	0.002	down	FKBP8
circERBB2	0.019	0.002–0.157	< 0.001	down	ERBB2
circGLTSCR1	0.122	0.037–0.410	0.001	down	GLTSCR1

Risk factors included SBP, systolic blood pressure; DBP, diastolic blood pressure; TG, triacylglycerol; TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol, smoking and alcohol consumption; ICH, intracerebral hemorrhage; OR, odds ratio; CI, confidence interval.

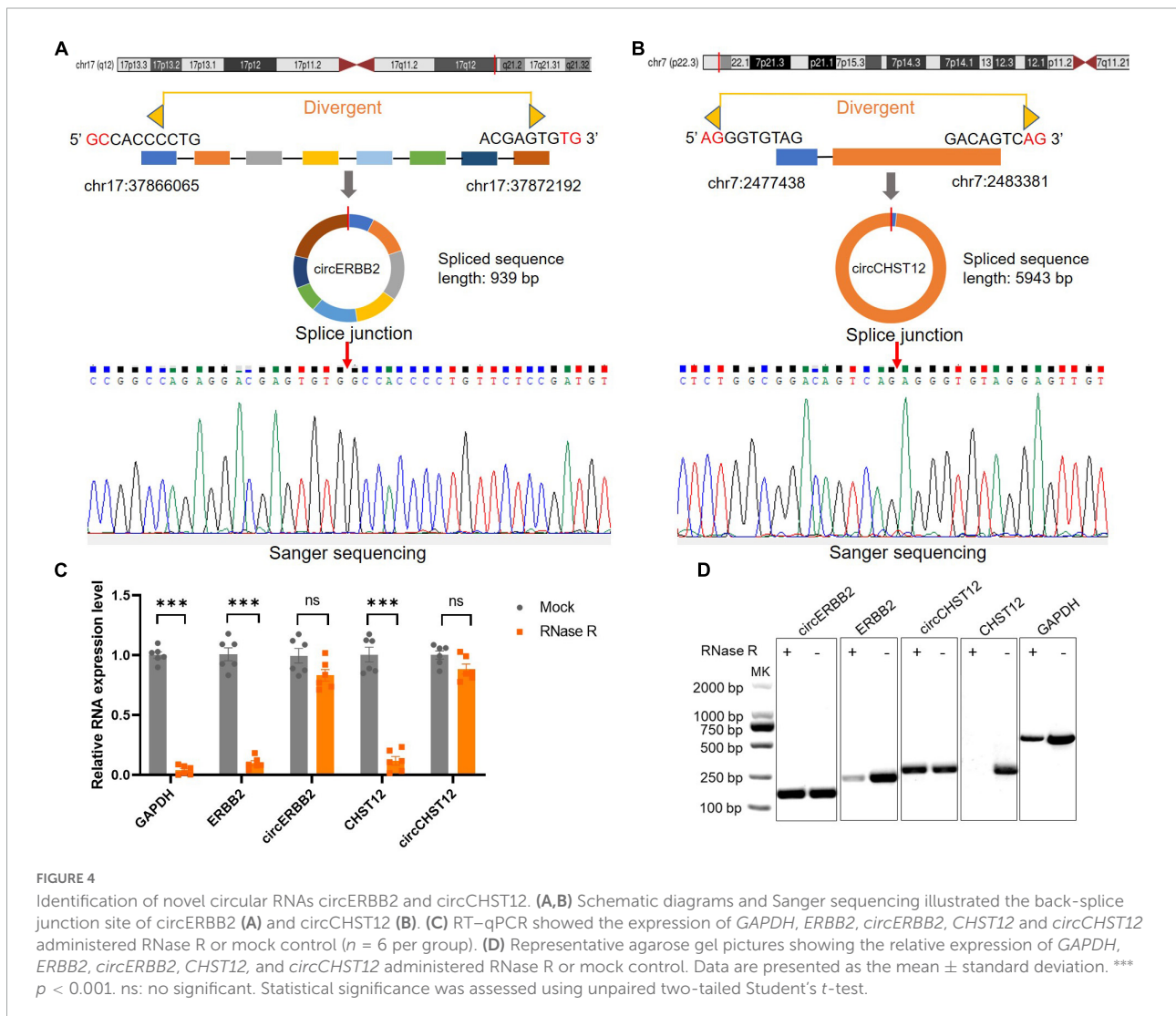


## Validation of the differentially expressed circRNAs by quantitative real-time polymerase chain reaction

To verify the novel circRNAs circERBB2 and circCHST12 are really circular form, we first blasted the sequences and confirmed the back-splice junction sites and assayed them by RT-PCR with divergent primers. Next, Sanger sequencing was performed to illustrate the junction site. The results showed that circERBB2, located at chr17:37866065-37872192 (genomic length: 6127 bp, spliced sequence length: 939 bp), was derived from exons 9–16 of the *ERBB2* gene (Figure 4A). circCHST12, located at chr7:2477438-2483381 (genomic length: 5943 bp, spliced sequence length: 5943 bp), was derived from exon 1 and partial exon 2 of the *CHST12* gene (Figure 4B). RT-qPCR analysis of total RNA after RNase R or control treatment indicated that circERBB2 and circCHST12 were resistant, while

*ERBB2*, *CHST12* and *GAPDH* mRNA transcripts were degraded (Figures 4C,D). These data established that circERBB2 and circCHST12 are two *bona fide* circRNAs.

Next, to confirm the expression of circRNAs in the high-throughput results, we selected three upregulated circRNAs (hsa\_circ\_0001707, hsa\_circ\_0005505 and hsa\_circ\_0027725) and three downregulated circRNAs (hsa\_circ\_0000914, circERBB2 and circCHST12) of the above consistently altered circRNAs for further validation by RT-qPCR in all samples. The expression levels of these circRNAs were consistent with the RNA sequencing results, including three upregulated circRNAs and three downregulated circRNAs that were significantly altered in patients with ICH compared with control subjects (Figures 5A–F). Moreover, the expression levels of circERBB2, circCHST12 and hsa\_circ\_0005505 were also significantly altered between ICH and IS patients (Figures 5G–I). These results were consistent with the levels obtained by RNA sequencing, supporting the accuracy and reliability of the data.



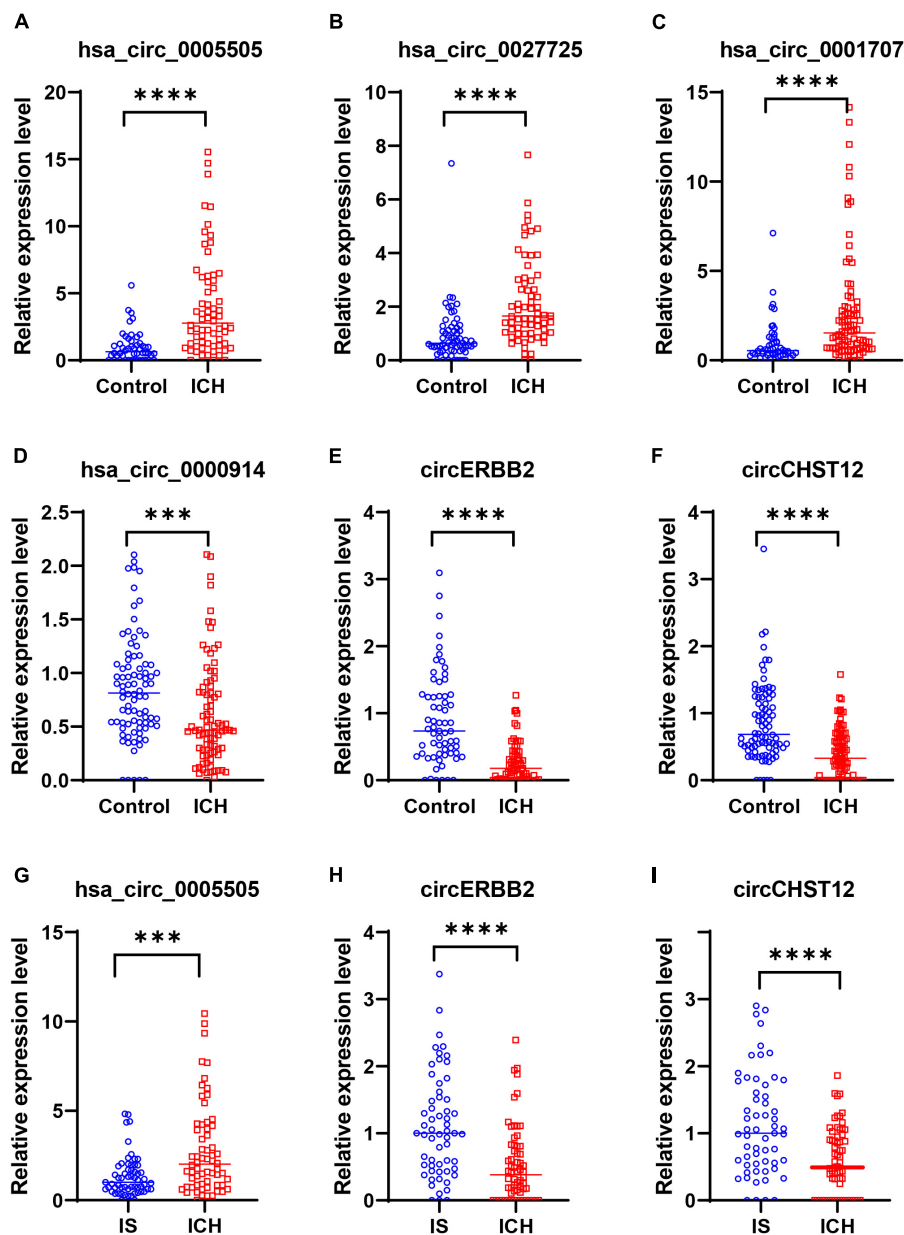


FIGURE 5

Validation of circRNA expression levels by quantitative real-time polymerase chain reaction (RT-qPCR). (A–F) RT-qPCR results validated the expression levels of candidate circRNAs in all samples between 64 intracerebral hemorrhage (ICH) patients and 50 healthy controls. (A) hsa\_circ\_0005505, (B) hsa\_circ\_0027725, (C) hsa\_circ\_0001707, (D) hsa\_circ\_0000914, (E) circERBB2 and (F) circCHST12. (G–I) RT-qPCR results validated the expression levels of hsa\_circ\_0005505 (G), circERBB2 (H) and circCHST12 (I) between 64 ICH patients and 59 ischemic stroke (IS) patients. The data are presented as the median (interquartile range). \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . Statistical significance was assessed using the Mann–Whitney U test.

## Performance evaluation of the candidate circRNAs with classification algorithms

To evaluate applicable biomarkers for ICH, we used mutual information (MI) and random forest (RF) algorithms to screen circRNA marker signatures according to the expression levels in

all samples. We obtained the signature of the top 10 circRNAs in the two algorithms and found 4 circRNAs [hsa\_circ\_0005806, circERBB2, circCHST12, circFBRS (host gene *FBRS*)] in the intersection (Supplementary Table 5). However, there was no significant difference in hsa\_circ\_0005806 or circFBRS expression levels between the ICH patients and controls in the validation cohort (Supplementary Figure 2). Finally, we

focused on evaluating the diagnostic value of circERBB2 and circCHST12 as potential ICH biomarkers in further statistical analysis.

Furthermore, six different classifier algorithms were executed to assess the validity of the candidate circRNAs. By using 10-fold cross-validation, the average performance measurement values of the candidate circRNAs in ICH were computed and are summarized in Table 4. The six machine learning classifiers based on test accuracies and AUCs in the training set and validation set are presented in Figure 6. The RF provides greater accuracy values of 0.995 and 0.910 than the other five classifiers in the training and test sets between ICH and controls, respectively (Figures 6A,B). We also evaluated the performance of the circERBB2 and circCHST12 signatures for discriminating ICH from IS patients and observed that the RF had the highest value of 0.989 in the training set and the SVM had the highest value of 0.779 in the test set (Figures 6C,D and Supplementary Table 6). These results indicate that the combination of the circERBB2 and circCHST12 signatures is capable of identifying ICH with high accuracy according to expression levels.

## Correlation of the circERBB2 and circCHST12 expression levels with clinical characteristics

Additionally, we performed Spearman's correlation analysis to test the correlation of the expression levels of circCHST12

and circERBB2 with ICH patient clinical characteristics. The results showed that the circERBB2 expression levels positively correlated with HDL-C and negatively correlated with SBP, DBP and alcohol consumption in ICH patients ( $P < 0.05$ ); the circCHST12 expression levels positively correlated with LDL-C and negatively correlated with SBP, DBP, glucose, white blood cells and alcohol consumption ( $P < 0.05$ ) (Table 5). These results indicated that circERBB2 and circCHST12 may be involved in the pathogenesis of ICH.

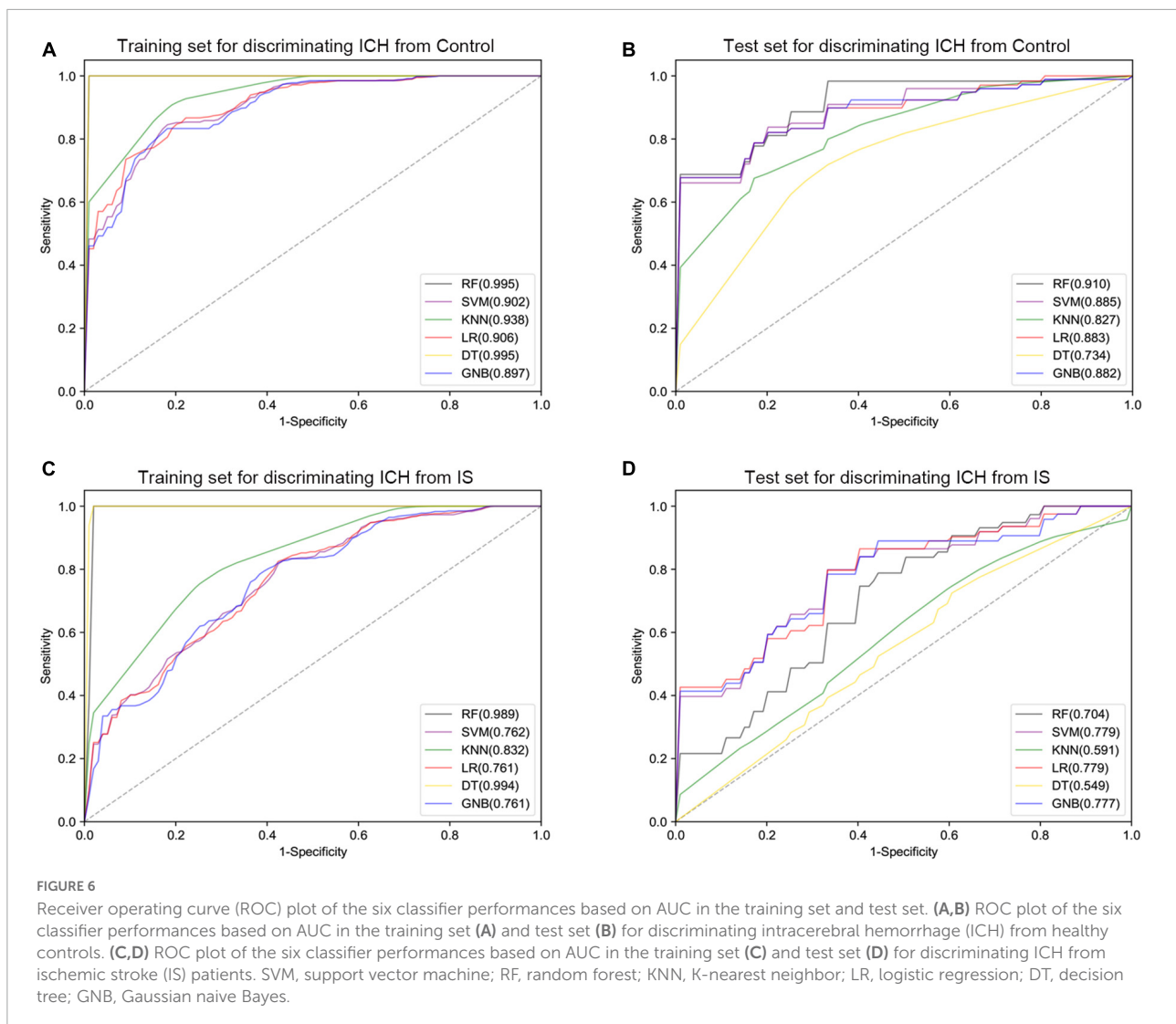
## Evaluation of the diagnostic value of circERBB2 and circCHST12 in intracerebral hemorrhage patients

Receiver operating curve (ROC) analysis was performed to explore the potential diagnostic value of circERBB2 and circCHST12. The signatures of circERBB2 for differentiating between patients with ICH and healthy control subjects showed an AUC of 0.883 (95% CI: 0.811–0.937) with a sensitivity of 68.2% and a specificity of 92%; the signatures of circCHST12 showed an AUC of 0.838 (95% CI: 0.769–0.908) with a sensitivity of 93% and a specificity of 71.6% (Figure 7A). The combination of circERBB2 and circCHST12 for differentiating between patients with ICH and healthy controls showed an AUC of 0.917 (95% CI: 0.869–0.965), with a sensitivity of 87.5% and a specificity of 82% (Figure 7A). We next performed a multifactor risk logistic regression model, the combination of circERBB2 and circCHST12 together with the risk factors (age, sex, BMI,

TABLE 4 Classification performance for the two-circRNA signatures between intracerebral hemorrhage (ICH) patients.

	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)	AUC
<b>RF</b>						
Training set	100	100	100	100	100	0.995(0.983–1)
Test set	80.83	74.81	78.48	80.10	77.48	0.910(0.857–0.963)
<b>KNN</b>						
Training set	91.71	81.32	87.26	86.66	88.33	0.938(0.894–0.982)
Test set	80.56	69.63	76.89	78.08	72.62	0.827(0.753–0.901)
<b>DT</b>						
Training set	100	100	100	100	100	0.995(0.983–1)
Test set	77.03	69.81	73.26	78.21	66.48	0.734(0.644–0.824)
<b>LR</b>						
Training set	86.35	76.82	82.28	83.16	80.99	0.906(0.852–0.960)
Test set	84.03	72.64	80.38	80.62	76.33	0.883(0.822–0.944)
<b>GNB</b>						
Training set	90.90	65.11	79.79	77.48	84.41	0.897(0.840–0.954)
Test set	90.14	63.98	79.62	77.46	82.17	0.882(0.821–0.943)
<b>SVM</b>						
Training set	93.25	64.25	80.75	77.51	87.90	0.902(0.846–0.957)
Test set	88.89	63.98	79.62	76.78	85.83	0.885(0.825–0.945)

ICH, intracerebral hemorrhage; RF, random forest; KNN, K-nearest neighbor; LR, logistic regression; DT, decision tree; GNB, Gaussian naive Bayes; SVM, support vector machine; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve.



SBP, DBP, TC, TG, HDL-C, LDL-C, smoking and alcohol consumption) showed that the AUC was increased to 0.980 (95% CI: 0.959–1), the sensitivity was 93.8%, and the specificity was 96% (Figure 7B). The addition of circERBB2 and circCHST12 to the previously known risk factors improved the predictive ability, with an NRI of 20.3% and IDI of 23.7% ( $P < 0.001$ ). The AUC of circERBB2 and circCHST12 for differentiating between ICH and IS patients was 0.765 (95% CI: 0.682–0.847); the sensitivity was 57.6%, and the specificity was 85.9% (Figure 7C).

hsa\_circ\_0005505 was upregulated in both ICH compared with controls and ICH compared IS patients. Furthermore, we evaluated the diagnostic values of the two novel circRNA combinations of hsa\_circ\_0005505 for identifying ICH. The combination of hsa\_circ\_0005505, circERBB2 and circCHST12 for differentiating between patients with ICH and healthy controls showed an AUC of 0.946 (95% CI: 0.910–0.982), with a sensitivity of 89.1% and a specificity of 86% (Figure 7A); the AUC was 0.799 (95% CI: 0.722–0.875), with a sensitivity

of 59.3% and a specificity of 89.5% for differentiating between patients with ICH and IS patients (Figure 7D). These results indicate that hsa\_circ\_0005505, novel circERBB2 and circCHST12, individually or combined, serve as potential diagnostic biomarkers for identifying ICH (Figure 8).

## Discussion

In the present study, we first investigated the circRNA profiles in the peripheral blood of ICH patients and healthy controls by using RNA sequencing in two independent cohorts. Functional analysis indicated that the differentially expressed circRNAs are involved in many pathophysiologic processes of ICH. By using two independent analysis strategies, we obtained nine circRNAs that were consistently altered in both cohorts, including five upregulated circRNAs and four downregulated circRNAs. Furthermore, based on machine

TABLE 5 Correlation between baseline characteristic and circRNA levels in intracerebral hemorrhage (ICH) patients.

Parameters	circERBB2		circCHST12	
	Coefficient	P-value	Coefficient	P-value
Age, y	0.143	0.128	-0.018	0.850
Sex (male)	0.017	0.895	0.022	0.814
BMI, kg/m <sup>2</sup>	0.044	0.646	0.103	0.274
SBP, mmHg	-0.373	< 0.001*	-0.240	0.010*
DBP, mmHg	-0.418	< 0.001*	-0.309	0.001*
HDL-C, mmol/L	0.190	0.043*	0.153	0.104
LDL-C, mmol/L	0.157	0.096	0.224	0.016*
TC, mmol/L	0.165	0.079	0.016	0.419
TG, mmol/L	-0.085	0.367	-0.164	0.182
GLU, mmol/L	-0.06	0.525	-0.273	0.003*
UA, $\mu$ mol/L	0.193	0.097	0.218	0.060
TBIL, $\mu$ mol/L	-0.023	0.846	0.001	0.992
BUN, mmol/L	-0.027	0.817	0.094	0.442
WBC, 10 <sup>9</sup> /L	-0.283	0.014*	-0.366	0.001*
Smoking	-0.063	0.504	-0.153	0.104
Alcohol consumption	-0.215	0.022*	-0.307	0.001*

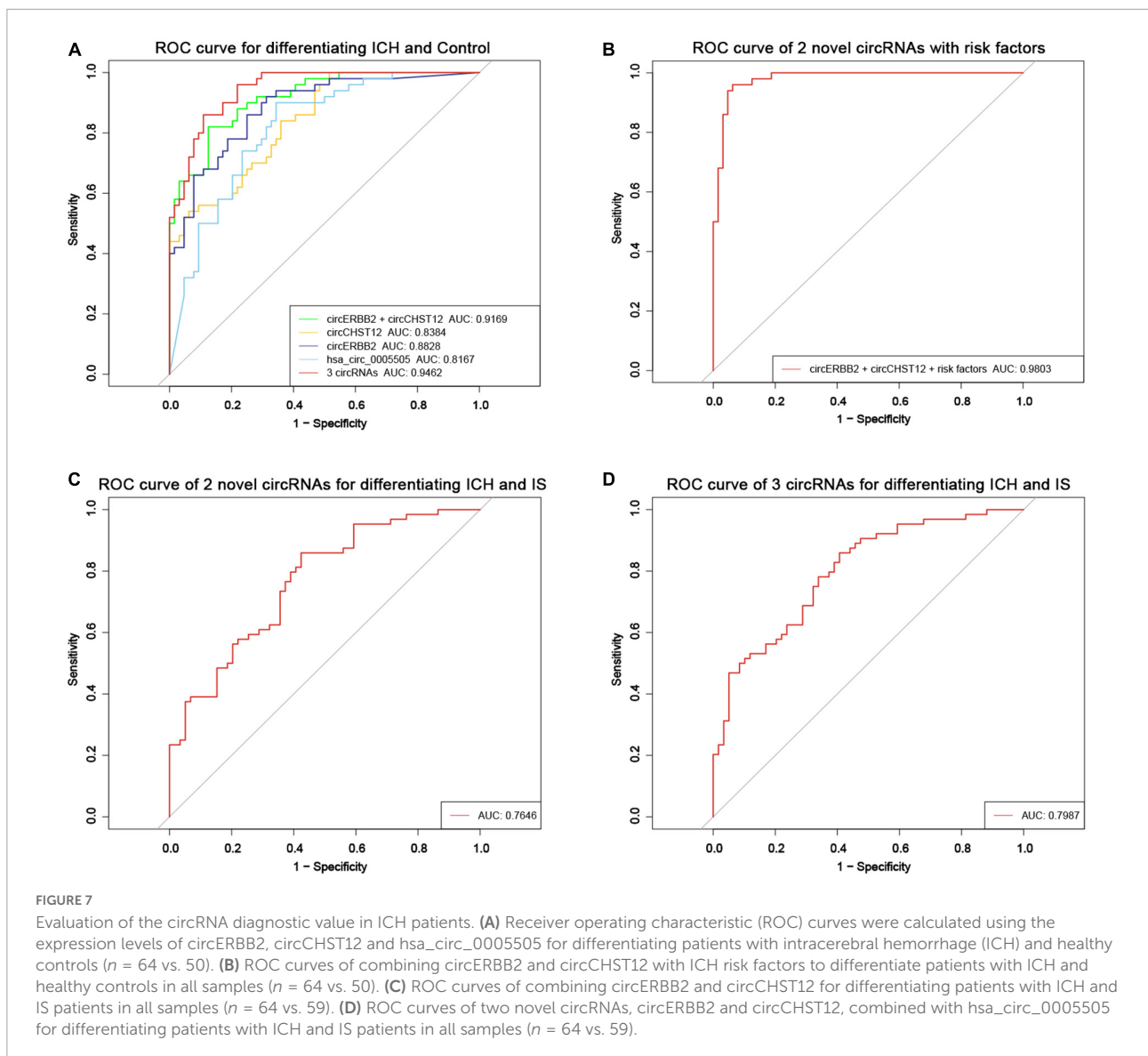
ICH, Intracerebral hemorrhage; BMI, Body mass index; SBP, Systolic blood pressure; DBP, Diastolic blood pressure; TC, Total cholesterol; TG, Triacylglycerol; HDL-C, High-density lipoprotein cholesterol; LDL-C, Low-density lipoprotein cholesterol; GLU, Glucose; UA, Uric acid; TBIL, Total bilirubin; BUN, Blood urea nitrogen; WBC, White blood cell. \* $p < 0.05$ .

learning classification, we screened two candidates, circERBB2 and circCHST12, to explore their diagnostic value as potential biomarkers in ICH patients. The AUC was 0.917 (95% CI: 0.869–0.965), with a sensitivity of 87.5% and a specificity of 82% for distinguishing between ICH patients and healthy controls. In combination with ICH risk factors, the AUC was 0.980 (95% CI: 0.959–1), sensitivity was 93.8% and specificity was 96% in ICH diagnosis. Moreover, logistic regression analysis and Spearman's correlation test demonstrated that downregulation of circERBB2 and circCHST12 may be independent risk factors for ICH. Additionally, the expression level of circERBB2 correlated with SBP and HDL-C; circCHST12 expression levels correlated with LDL-C, SBP, DBP and white blood cells, indicating that circERBB2 and circCHST12 might be heavily involved in the pathology of ICH. Our data show that circERBB2 and circCHST12 may be novel biomarkers for ICH diagnosis. Together with hsa\_circ\_0005505, circERBB2 and circCHST12 showed high accuracy for identifying ICH. A previous study revealed that hsa\_circ\_0005505 was upregulated in ruptured intracranial aneurysm tissues, promoted proliferation and migration and suppressed apoptosis of vascular smooth muscle cells *in vitro* (Chen X. et al., 2021), indicating that hsa\_circ\_0005505 may be associated with the pathological process of cerebrovascular diseases.

Intracerebral hemorrhage (ICH) is a multifactorial disease with high incidence and mortality that imposes a large

socioeconomic burden. Identifying novel potential biomarkers for the early diagnosis of ICH would be part of risk prediction. CircRNAs are produced by host gene back-splicing, and closed RNAs without a free 3' or 5' end are resistant to exonuclease digestion (Jeck and Sharpless, 2014), which makes them more stable and better biomarkers of human disease. Furthermore, circRNAs are highly expressed in many tissues, particularly the human brain, and in blood (Patop et al., 2019). There is growing evidence that the circRNA expression profile is altered in IS (Dong et al., 2020; Ostolaza et al., 2020; Zuo et al., 2020; Liu Y. et al., 2022), indicating that circRNAs have the potential to serve as biomarkers and therapeutic targets in IS. Moreover, the circRNA expression profiles were altered in rat brain tissues after ICH (Zhong et al., 2020; Bai et al., 2021). However, the changes in circRNA expression in the peripheral blood of ICH patients remain unclear. Our previous study demonstrated that hsa\_circ\_0001240, hsa\_circ\_0001947 and hsa\_circ\_0001386 were promising biomarkers for predicting and diagnosing hypertensive ICH (Bai et al., 2021). In this study, we first investigated whether circRNA profiles were significantly altered between ICH patients and healthy controls, which provides new insights into understanding the epigenomic mechanisms of ICH.

In this study, we found that circERBB2 may serve as a novel biomarker in ICH diagnosis. Previous studies have identified blood biomarkers, such as glial fibrillary acid protein (GFAP), retinol binding protein 4 and N-terminal pro B-type natriuretic peptide, that distinguish IS from ICH with moderate accuracy (Bustamante et al., 2021) and metabolic biomarkers for ICH diagnosis (Zhang et al., 2021). The AUCs of S100 and IL6 were 0.65 and 0.59 (Bhatia et al., 2020), respectively, and GFAP had a sensitivity of 78% and a specificity of 95% between ICH and IS (Kumar et al., 2020). ncRNAs have been identified as critical novel regulators of cardiovascular risk factors and cell functions and are thus important candidates to improve diagnostics and prognosis assessment (Poller et al., 2018). In the present study, we identified that the AUC of circERBB2 was 0.883 for distinguishing between ICH patients and healthy controls, with a sensitivity and specificity of 68.2% and 92%, respectively. The signatures of circCHST12 showed an AUC of 0.838 with a sensitivity of 93% and a specificity of 71.6%. The combination of circERBB2 and circCHST12 with ICH risk factors increased the predictive value for the identification of ICH. These findings were better than the diagnostic value of three previously identified circRNAs [hsa\_circ\_0001240 (AUC = 0.808), hsa\_circ\_0001947 (AUC = 0.798) and hsa\_circ\_0001386 (AUC = 0.806)] in ICH (Bai et al., 2021). Additionally, we observed that downregulation of circERBB2 was positively associated with HDL-C and negatively correlated with SBP and DBP. Lowering blood lipids was associated with an increased risk of ICH (Sun et al., 2019), and high blood pressure was found to be the most prevalent stroke risk factor (Feigin et al., 2016; Wang et al., 2017). Thus,

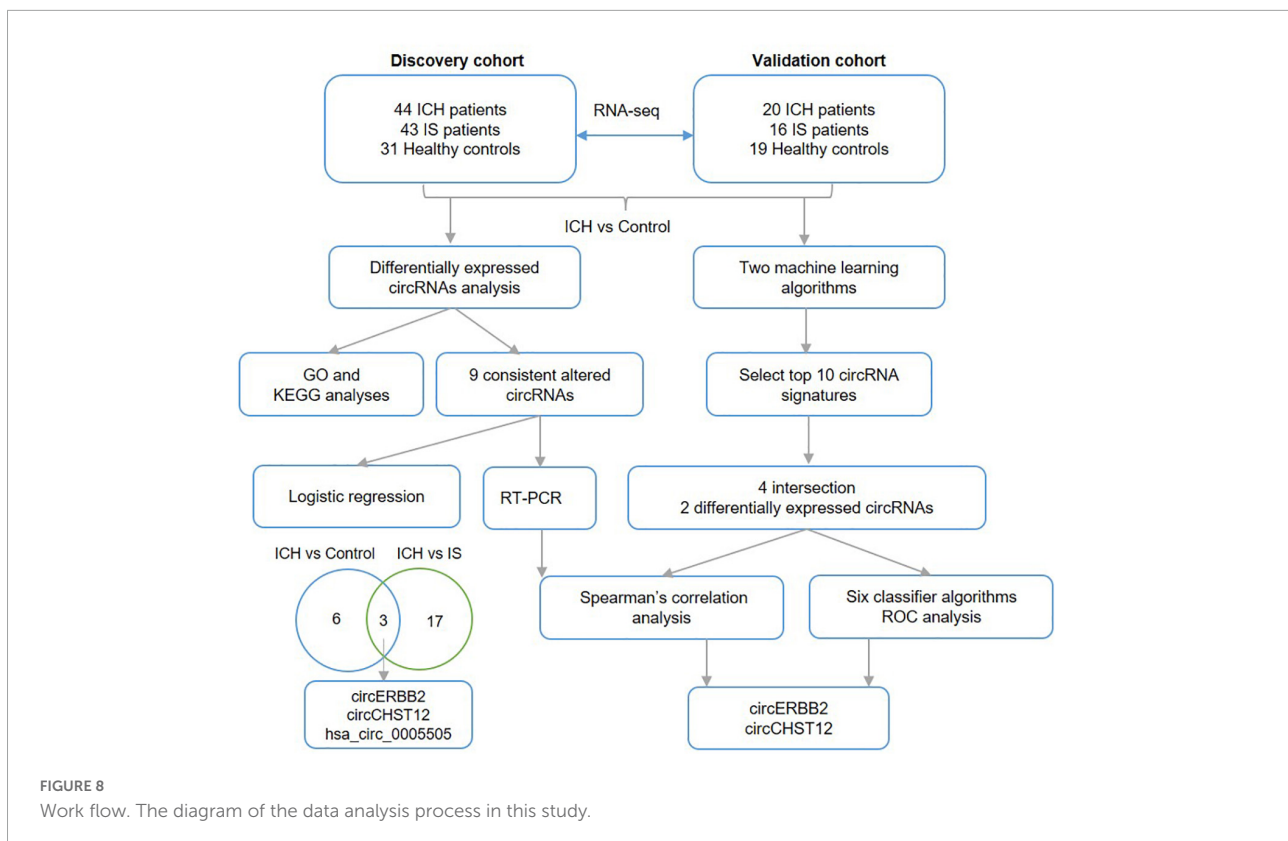


we speculate that a decrease in circERBB2 expression levels might correlate with an increased risk of ICH occurrence. These findings indicate that circERBB2 might play vital roles in the pathogenesis and pathology of ICH.

The protein ERBB2 is a member of a family of epidermal growth factor receptors that are involved in aberrant signaling and cell migration, growth, adhesion, and differentiation (Strickler et al., 2022). A previous study demonstrated that circERBB2 (chr17: 39,708,320–39,710,481; length: 676 bp) serves as an important regulator of cancer cell proliferation and has the potential to be a new therapeutic target for gallbladder cancer (Huang et al., 2019) and breast cancer (Huang Y. et al., 2021). Our study identified circERBB2 (chr17: 37,866,065–37,872,192; genomic length: 6127 bp, spliced sequence length: 939 bp), which is a novel back-splicing circRNA that has never been reported thus far, at a different chromosomal

position. Carbohydrate sulfotransferases (CHSTs) are a class of key enzymes that contribute to tissue remodeling. CHST12 is a significant member of the CHST family, and a previous study demonstrated that CHST12 may be a novel biomarker for glioblastoma; it regulates cell proliferation and mobility via the WNT/ $\beta$ -catenin pathway (Wang et al., 2021). One study reported that hsa\_circ\_0134005 (chr7:2472197–2477555; genomic length: 5358 bp, spliced sequence length: 5358 bp) is derived from the CHST12 gene (Rybak-Wolf et al., 2015). This study identified circCHST12 (chr7:2477438–2483381; genomic length: 5943 bp, spliced sequence length: 5943 bp) derived from exon 1 and partial exon 2 of the CHST12 gene, which is a novel back-splicing circRNA that has never been reported thus far at a different chromosomal position.

CircRNAs are involved in the translational and transcriptional regulation of the pathological mechanisms



of many disorders (Shan et al., 2017; Aufiero et al., 2019). CircRNAs can act as miRNA sponges and are expected to influence downstream miRNA function, further regulating the expression levels of target mRNAs (Hansen et al., 2013). We performed GO and KEGG analyses to investigate the enrichment of differentially expressed circRNAs. Functional analysis demonstrated that the circRNA host genes were mainly involved in GTPase activity, covalent chromatin modification, histone modification, the MAPK signaling pathway and the ERBB signaling pathway. Activation of the MAPK signaling pathway is involved in the progression of injury following ICH (Ding et al., 2020; Guo et al., 2020). Recently, research identified that knockdown of circERBB2 suppressed the PDGF-BB-induced proliferation, migration, and inflammatory response of human airway smooth muscle cells via miR-98-5p/IGF1R signaling (Huang J. Q. et al., 2021). The phenotype of smooth muscle cells transforming from a contractile to a synthetic phenotype plays an essential role in the onset of brain vascular pathological progression (Bennett et al., 2016; Rho et al., 2017). In this study, we speculated that the downregulation of the novel circERBB2 in ICH patients might contribute to the pathogenesis of ICH via the phenotype of smooth muscle cell transformation.

Notably, there are some limitations of this study. First, we should perform a larger multicenter study with more participants to externally validate the candidate biomarkers.

Second, further studies should be performed to explore how hsa\_circ\_0005505, circERBB2 and circCHST12 contribute to the pathogenesis and development of ICH with cell- or animal-based experiments. Additionally, our study lacked follow-up information for ICH patients, and the prognostic value of these candidate circRNAs should be assessed in subsequent studies. We expect that hsa\_circ\_0005505, circERBB2 and circCHST12 will provide new insights for a better understanding of the pathogenesis of ICH and help to improve the diagnosis and prognostic assessment of ICH in clinical practice.

## Conclusion

In this study, we provided a transcriptome-wide overview of aberrantly expressed circRNAs in the peripheral blood of ICH patients and identified hsa\_circ\_0005505 and novel circERBB2 and circCHST12 as promising biomarkers for diagnosing ICH based on machine learning algorithms.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

and accession number(s) can be found in the article/[Supplementary material](#).

## Ethics statement

The studies involving human participants were reviewed and approved by Human Ethics Committee, Fuwai Hospital (Approval No. 2016-732). The patients/participants provided their written informed consent to participate in this study.

## Author contributions

CB, YS, and LS: design and experiment. XH and FW: data analyses. CB and LZ: manuscript preparation. JL, LY, and JC: manuscript review. All authors contributed to the article and approved the submitted version.

## Funding

This study was supported by the National Natural Science Foundation of China (91539113 and 82130013 to JC), the National Basic Research Program of China (2014CB541601 to JC), and the CAMS Innovation Fund for

## References

- Ambale-Venkatesh, B., Yang, X., Wu, C. O., Liu, K., Hundley, W. G., McClelland, R., et al. (2017). Cardiovascular event prediction by machine learning: The multi-ethnic study of atherosclerosis. *Circ. Res.* 121, 1092–1101. doi: 10.1161/CIRCRESAHA.117.311312
- Aufiero, S., Reckman, Y. J., Pinto, Y. M., and Creemers, E. E. (2019). Circular RNAs open a new chapter in cardiovascular biology. *Nat. Rev. Cardiol.* 16, 503–514. doi: 10.1038/s41569-019-0185-2
- Bai, C., Liu, T., Sun, Y., Li, H., Xiao, N., Zhang, M., et al. (2021). Identification of circular RNA expression profiles and potential biomarkers for intracerebral hemorrhage. *Epigenomics* 13, 379–395. doi: 10.2217/epi-2020-0432
- Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., et al. (2017). Heart disease and stroke statistics-2017 update: A report from the American heart association. *Circulation* 135, e146–e603. doi: 10.1161/CIR.0000000000000485
- Bennett, M. R., Sinha, S., and Owens, G. K. (2016). Vascular smooth muscle cells in atherosclerosis. *Circ. Res.* 118, 692–702. doi: 10.1161/CIRCRESAHA.115.306361
- Bhatia, R., Warrier, A. R., Sreenivas, V., Bali, P., Sisodia, P., Gupta, A., et al. (2020). Role of blood biomarkers in differentiating ischemic stroke and intracerebral hemorrhage. *Neurol. India* 68, 824–829. doi: 10.4103/0028-3886.293467
- Blokh, D., and Stambler, I. (2017). The application of information theory for the research of aging and aging-related diseases. *Prog. Neurobiol.* 157, 158–173. doi: 10.1016/j.pneurobio.2016.03.005
- Bustamante, A., Penalba, A., Orset, C., Azurmendi, L., Llombart, V., Simats, A., et al. (2021). Blood biomarkers to differentiate ischemic and hemorrhagic strokes. *Neurology* 96, e1928–e1939. doi: 10.1212/WNL.00000000000011742
- Cardona-Monzonis, A., Garcia-Gimenez, J. L., Mena-Molla, S., Pareja-Galeano, H., de la Guia-Galipienso, F., and Pallardo, F. V. (2020). Non-coding RNAs and

Medical Sciences (2021-CXGC02-3CAMS-I2 M and 2021-1-I2 M-007 to JC).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2022.1002590/full#supplementary-material>

coronary artery disease. *Adv. Exp. Med. Biol.* 1229, 273–285. doi: 10.1007/978-981-15-1671-9\_16

Chang, C. H., Lin, C. H., and Lane, H. Y. (2021). Machine learning and novel biomarkers for the diagnosis of Alzheimer's disease. *Int. J. Mol. Sci.* 22:2761. doi: 10.3390/ijms22052761

Chen, T., Chen, X., Zhang, S., Zhu, J., Tang, B., Wang, A., et al. (2021). The genome sequence archive family: Toward explosive data growth and diverse data types. *Genom. Proteom. Bioinform.* 19, 578–583.

Chen, X., Yang, S., Yang, J., Liu, Q., Li, M., Wu, J., et al. (2021). The potential role of hsa\_circ\_0005505 in the rupture of human intracranial aneurysm. *Front. Mol. Biosci.* 8:670691. doi: 10.3389/fmolb.2021.670691

Chen, Y., Chen, B., Song, X., Kang, Q., Ye, X., and Zhang, B. (2021). A data-driven binary-classification framework for oil fingerprinting analysis. *Environ. Res.* 201:111454. doi: 10.1016/j.envres.2021.111454

Cheng, X., Ander, B. P., Jickling, G. C., Zhan, X., Hull, H., Sharp, F. R., et al. (2020). MicroRNA and their target mRNAs change expression in whole blood of patients after intracerebral hemorrhage. *J. Cereb. Blood Flow Metab.* 40, 775–786. doi: 10.1177/0271678X19839501

CNCB-NGDC Members and Partners (2022). Database resources of the national genomics data center, china national center for bioinformatics in 2022. *Nucleic Acids Res.* 50, D27–D38.

Ding, Y., Flores, J., Klebe, D., Li, P., McBride, D. W., Tang, J., et al. (2020). Annexin A1 attenuates neuroinflammation through FPR2/p38/COX-2 pathway after intracerebral hemorrhage in male mice. *J. Neurosci. Res.* 98, 168–178. doi: 10.1002/jnr.24478

Dong, Z., Deng, L., Peng, Q., Pan, J., and Wang, Y. (2020). CircRNA expression profiles and function prediction in peripheral blood mononuclear cells of patients with acute ischemic stroke. *J. Cell. Physiol.* 235, 2609–2618. doi: 10.1002/jcp.29165



- Dou, Z., Yu, Q., Wang, G., Wu, S., Reis, C., Ruan, W., et al. (2020). Circular RNA expression profiles alter significantly after intracerebral hemorrhage in rats. *Brain Res.* 1726:146490. doi: 10.1016/j.brainres.2019.146490
- Feigin, V. L., Roth, G. A., Naghavi, M., Parmar, P., Krishnamurthi, R., Chugh, S., et al. (2016). Global burden of stroke and risk factors in 188 countries, during 1990–2013: A systematic analysis for the global burden of disease study 2013. *Lancet Neurol.* 15, 913–924. doi: 10.1016/S1474-4422(16)30073-4
- Gao, Y., Zhang, J., and Zhao, F. (2018). Circular RNA identification based on multiple seed matching. *Brief. Bioinform.* 19, 803–810. doi: 10.1093/bib/bb1014
- Guo, F., Xu, D., Lin, Y., Wang, G., Wang, F., Gao, Q., et al. (2020). Chemokine CCL2 contributes to BBB disruption via the p38 MAPK signaling pathway following acute intracerebral hemorrhage. *FASEB J.* 34, 1872–1884. doi: 10.1096/fj.201902203RR
- Hankey, G. J. (2017). Stroke. *Lancet* 389, 641–654. doi: 10.1016/S0140-6736(16)30962-X
- Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388. doi: 10.1038/nature11993
- Huang, J. Q., Wang, F., Wang, L. T., Li, Y. M., Lu, J. L., and Chen, J. Y. (2021). Circular RNA ERBB2 contributes to proliferation and migration of airway smooth muscle cells via miR-98-5p/IGF1R signaling in asthma. *J. Asthma Allergy* 14, 1197–1207. doi: 10.2147/JAA.S326058
- Huang, X., He, M., Huang, S., Lin, R., Zhan, M., Yang, D., et al. (2019). Circular RNA circERBB2 promotes gallbladder cancer progression by regulating PA2G4-dependent rDNA transcription. *Mol. Cancer* 18:166. doi: 10.1186/s12943-019-1098-8
- Huang, Y., Zheng, S., Lin, Y., and Ke, L. (2021). Circular RNA circ-ERBB2 elevates the warburg effect and facilitates triple-negative breast cancer growth by the MicroRNA 136-5p/pyruvate dehydrogenase kinase 4 axis. *Mol. Cell. Biol.* 41:e0060920. doi: 10.1128/MCB.00609-20
- Jeck, W. R., and Sharpless, N. E. (2014). Detecting and characterizing circular RNAs. *Nat. Biotechnol.* 32, 453–461. doi: 10.1038/nbt.2890
- Kawakami, E., Tabata, J., Yanaihara, N., Ishikawa, T., Koseki, K., Iida, Y., et al. (2019). Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers. *Clin. Cancer Res.* 25, 3006–3015. doi: 10.1158/1078-0432.CCR-18-3378
- Kim, J. M., Moon, J., Yu, J. S., Park, D. K., Lee, S. T., Jung, K. H., et al. (2019). Altered long noncoding RNA profile after intracerebral hemorrhage. *Ann. Clin. Transl. Neurol.* 6, 2014–2025. doi: 10.1002/acn3.50894
- Kristensen, L. S., Andersen, M. S., Stagsted, L. V. W., Ebbesen, K. K., Hansen, T. B., and Kjems, J. (2019). The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* 20, 675–691. doi: 10.1038/s41576-019-0158-7
- Kristensen, L. S., Jakobsen, T., Hager, H., and Kjems, J. (2022). The emerging roles of circRNAs in cancer and oncology. *Nat. Rev. Clin. Oncol.* 19, 188–206. doi: 10.1038/s41571-021-00585-y
- Kumar, A., Misra, S., Yadav, A. K., Sagar, R., Verma, B., Grover, A., et al. (2020). Role of glial fibrillary acidic protein as a biomarker in differentiating intracerebral haemorrhage from ischaemic stroke and stroke mimics: A meta-analysis. *Biomarkers* 25, 1–8. doi: 10.1080/1354750X.2019.1691657
- Ledesma, D., Symes, S., and Richards, S. (2021). Advancements within modern machine learning methodology: Impacts and prospects in biomarker discovery. *Curr. Med. Chem.* 28, 6512–6531. doi: 10.2174/0929867328666210208111821
- Li, L., Wang, P., Zhao, H., and Luo, Y. (2019). Noncoding RNAs and intracerebral hemorrhage. *CNS Neurol. Disord. Drug Targets* 18, 205–211. doi: 10.2174/1871527318666190204102604
- Li, R. (2018). data mining and machine learning methods for dementia research. *Methods Mol. Biol.* 1750, 363–370. doi: 10.1007/978-1-4939-7704-8\_25
- Li, S., Chen, L., Xu, C., Qu, X., Qin, Z., Gao, J., et al. (2020). Expression profile and bioinformatics analysis of circular RNAs in acute ischemic stroke in a South Chinese han population. *Sci. Rep.* 10:10138. doi: 10.1038/s41598-020-66990-y
- Liu, D., Zhao, L., Jiang, Y., Li, L., Guo, M., Mu, Y., et al. (2022). Integrated analysis of plasma and urine reveals unique metabolomic profiles in idiopathic inflammatory myopathies subtypes. *J. Cachexia Sarcopenia Muscle* 13, 2456–2472. doi: 10.1002/jcsm.13045
- Liu, Y., Li, Y., Zang, J., Zhang, T., Li, Y., Tan, Z., et al. (2022). CircOGDH Is a penumbra biomarker and therapeutic target in acute ischemic stroke. *Circ. Res.* 130, 907–924. doi: 10.1161/CIRCRESAHA.121.319412
- Lu, D., Ho, E. S., Mai, H., Zang, J., Liu, Y., Li, Y., et al. (2020). Identification of blood circular RNAs as potential biomarkers for acute ischemic stroke. *Front. Neurosci.* 14:81. doi: 10.3389/fnins.2020.00081
- Montaner, J., Ramiro, L., Simats, A., Tiedt, S., Makris, K., Jickling, G. C., et al. (2020). Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke. *Nat. Rev. Neurol.* 16, 247–264. doi: 10.1038/s41582-020-0350-6
- Ostolaza, A., Blanco-Luquin, I., Urdanoz-Casado, A., Rubio, I., Labarga, A., Zandio, B., et al. (2020). Circular RNA expression profile in blood according to ischemic stroke etiology. *Cell Biosci.* 10:34. doi: 10.1186/s13578-020-00394-3
- Patop, I. L., Wust, S., and Kadener, S. (2019). Past, present, and future of circRNAs. *EMBO J.* 38:e100836. doi: 10.15252/embj.2018100836
- Poller, W., Dimmeler, S., Heymans, S., Zeller, T., Haas, J., Karakas, M., et al. (2018). Non-coding RNAs in cardiovascular diseases: Diagnostic and therapeutic perspectives. *Eur. Heart J.* 39, 2704–2716. doi: 10.1093/eurheartj/ehx165
- Qureshi, A. I., Mendelow, A. D., and Hanley, D. F. (2009). Intracerebral haemorrhage. *Lancet* 373, 1632–1644. doi: 10.1016/S0140-6736(09)60371-8
- Rho, S. S., Ando, K., and Fukuhara, S. (2017). Dynamic regulation of vascular permeability by vascular endothelial cadherin-mediated endothelial cell-cell junctions. *J. Nippon Med. Sch.* 84, 148–159. doi: 10.1272/jnms.84.148
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp166
- Rybak-Wolf, A., Stottmeister, C., Glazar, P., Jens, M., Pino, N., Giusti, S., et al. (2015). Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol. Cell* 58, 870–885. doi: 10.1016/j.molcel.2015.03.027
- Shan, K., Liu, C., Liu, B. H., Chen, X., Dong, R., Liu, X., et al. (2017). Circular noncoding RNA HIPK3 mediates retinal vascular dysfunction in diabetes mellitus. *Circulation* 136, 1629–1642. doi: 10.1161/CIRCULATIONAHA.117.029004
- Shu, T., Ning, W., Wu, D., Xu, J., Han, Q., Huang, M., et al. (2020). Plasma proteomics identify biomarkers and pathogenesis of COVID-19. *Immunity* 53, 1108–1122e5. doi: 10.1016/j.immuni.2020.10.008
- Strickler, J. H., Yoshino, T., Graham, R. P., Siena, S., and Bekaii-Saab, T. (2022). Diagnosis and treatment of ERBB2-positive metastatic colorectal cancer: A review. *JAMA Oncol.* 8, 760–769. doi: 10.1001/jamaoncol.2021.8196
- Sun, L., Clarke, R., Bennett, D., Guo, Y., Walters, R. G., Hill, M., et al. (2019). Causal associations of blood lipids with risk of ischemic stroke and intracerebral hemorrhage in Chinese adults. *Nat. Med.* 25, 569–574. doi: 10.1038/s41591-019-0366-x
- Tiedt, S., Prestel, M., Malik, R., Schieferdecker, N., Duering, M., Kautzky, V., et al. (2017). RNA-seq identifies circulating miR-125a-5p, miR-125b-5p, and miR-143-3p as potential biomarkers for acute ischemic stroke. *Circ. Res.* 121, 970–980. doi: 10.1161/CIRCRESAHA.117.311572
- Wang, J., Xia, X., Tao, X., Zhao, P., and Deng, C. (2021). Knockdown of carbohydrate sulfotransferase 12 decreases the proliferation and mobility of glioblastoma cells via the WNT/beta-catenin pathway. *Bioengineered* 12, 3934–3946. doi: 10.1080/21655979.2021.1944455
- Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: An R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138. doi: 10.1093/bioinformatics/btp612
- Wang, W., Jiang, B., Sun, H., Ru, X., Sun, D., Wang, L., et al. (2017). Prevalence, incidence, and mortality of stroke in china: Results from a nationwide population-based survey of 480 687 adults. *Circulation* 135, 759–771. doi: 10.1161/CIRCULATIONAHA.116.025250
- Weng, R., Jiang, Z., and Gu, Y. (2022). Noncoding RNA as diagnostic and prognostic biomarkers in cerebrovascular disease. *Oxid. Med. Cell. Longev.* 2022:8149701. doi: 10.1155/2022/8149701
- Wilkinson, D. A., Pandey, A. S., Thompson, B. G., Keep, R. F., Hua, Y., and Xi, G. (2018). Injury mechanisms in acute intracerebral hemorrhage. *Neuropharmacology* 134(Pt B), 240–248. doi: 10.1016/j.neuropharm.2017.09.033
- Wu, J., Zhang, H., Li, L., Hu, M., Chen, L., Xu, B., et al. (2020). A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: A population-based analysis. *Cancer Commun.* 40, 301–312. doi: 10.1002/cac2.12067

Zhang, J., Su, X., Qi, A., Liu, L., Zhang, L., Zhong, Y., et al. (2021). Metabolomic profiling of fatty acid biomarkers for intracerebral hemorrhage stroke. *Talanta* 222:121679. doi: 10.1016/j.talanta.2020.121679

Zhang, Z., Yang, T., and Xiao, J. (2018). Circular RNAs: Promising biomarkers for human diseases. *EBioMedicine* 34, 267–274. doi: 10.1016/j.ebiom.2018.07.036

Zhong, Y., Li, X., Li, C., Li, Y., He, Y., Li, F., et al. (2020). Intracerebral hemorrhage alters circular RNA expression profiles in the rat brain. *Am. J. Transl. Res.* 12, 4160–4174.

Zuo, L., Zhang, L., Zu, J., Wang, Z., Han, B., Chen, B., et al. (2020). Circulating circular RNAs as biomarkers for the diagnosis and prediction of outcomes in acute ischemic stroke. *Stroke* 51, 319–323. doi: 10.1161/STROKEAHA.119.027348