PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2020, 47(4): 344~360
www.pibb.ac.cn

The Identification and Analysis of a miRNA Risk Score Model for Hepatocellular Carcinoma Prognosis*

MEN Jing-Rui, TAN Jian-Jun**, SUN Hong-Liang
(*College of Life Science and Bioengineering, Beijing University of Technology, Beijing* 100124, China)

Abstract Hepatocellular carcinoma (HCC) is one of the malignancies with high morbidity and mortality in the world. The purpose of our study was to search for HCC related miRNA prognostic biomarkers to predict the risk degree and survival time of HCC patients and provide effective prognostic information for HCC patients. Four methods were used to identify differentially expressed miRNAs (DEMs) from The Cancer Genome Atlas(TCGA). Kaplan-Meier survival curve, univariate and multivariable Cox regression analysis were used to identify prognostic miRNAs of HCC from DEMs. Four prognostic miRNAs biomarkers (hsa-miR-132-3p, hsa-miR-139-5p, hsa-miR-3677-3p, hsa-miR-500a-3p) of HCC were identified at last, and combined into a risk score model. There was no experimental evidence that hsa-mir-3677-3p is related to HCC, and it was a newly discovered miRNA in this study. The evaluation results of various bioinformatics methods, including survival curve, ROC curve, *chi-square* test, *et al.*. All indicated that the risk score calculated by the model can effectively predict the risk degree of patients(P<0.000, hazard ratio=2.551, 95% confidence interval=1.751−3.717). 1−5 year survival rates of HCC patients in the low risk group had 20%−30% higher than in the high risk group. Through the clinical data analysis, it was found that the combined biomarkers have a better prognostic effect than other clinical indicators, and can also be used as an independent prognostic factor. Target genes of four miRNAs were predicted, including AGO2, FOXO1, ROCK2, RAP1B, CYLD, *et al.*, and enriched in biological processes such as cell proliferation, migration, apoptosis and immune response.

Key words miRNA, hepatocellular carcinoma, prognosis, biomarker, TCGA
DOI: 10.16476/j.pibb.2019.0286

Hepatocellular carcinoma (HCC) is one of the malignant tumors in the world with the characteristic of easy metastasis and poor prognosis. In 2018, there were about 850 000 new HCC cases in the world (48.4% in Asia) and about 780 000 death cases(57.3% in Asia) [1]. HCC was the sixth cancer with high incidence after lung, breast, colorectal, prostate and stomach cancer and the fourth mortal cancer. HCC has a high incidence and mortality rate in Asia, accounting for 48.4% of new cases and 57.3% of deaths globally. Most patients are diagnosed with HCC in the middle or late stages because the early symptoms are not obvious. In addition, most HCC patients have depression, tension, anxiety or other negative mental. This phenomenon will aggravate the disease. However, the study on the prognosis can predict the disease progression and provide prognosis information to HCC patients, relieve anxiety, and give patients motivation for survival. Therefore, prognostic research has great significance for the treatment and future development of HCC patients.

Nowadays, the main diagnostic methods of HCC are ultrasonic examination and serum alpha-fetoprotein(AFP) content detection[2]. However,

* This work was supported by grants from Beijing Natural Science Foundation (2202002) and The National Natural Science Foundation of China (21173014).
** Corresponding author.
Tel: 86-10-67392001, E-mail: tanjianjun@bjut.edu.cn
Received: November 26, 2019; Accepted: March 13, 2020

ultrasonic examination can only detect HCC and cannot predict the prognosis of patients. AFP is the most commonly used biomarker for HCC detection, but it is only reliable in the 3/4 stage of HCC. The novel and reliable biomarkers are needed to obtain accurate prognostic information. With the development of second-generation gene sequencing technology, some novel markers obtained from gene chip expression profile data have attracted wide attention, especially microRNAs(miRNAs). miRNA and its target genes form a regulatory network together. A single miRNA can regulate multiple genes, and multiple miRNA can also regulate the same gene. In tumor tissues, the dysregulation of miRNA expression may result in oncogenes over expression or tumor suppressor genes low expression[3-4]. Therefore, the abnormal expression miRNAs may be an effective biomarker for tumor diagnosis and prognosis.

The miRNAs are small non-coding RNAs with about 22 nt length, which play an important role in many cancer pathways in HCC such as JAK/STAT, TP53, WNT/β-catenin and PI3K/MAPK[5-7]. The JAK/STAT pathway is considered to be one of the major molecular pathways in HCC development. JAK participates in cell proliferation, apoptosis, and glucose metabolism by recruiting STAT3 protein to mediate gene transcription. SOCS2 is an inhibitor of the cytokine signaling (SOCS) family and a key negative regulator of the JAK/STAT signaling pathway, which inhibits the JAK/STAT pathway in HCC[8]. At the same time, SOCS2 is a functional target of miR-196a and miR-196b in HCC cells. miR-196a or miR-196b regulates the JAK/STAT pathway by targeting SOCS2 to promote the progression of HCC. That means the over expression of miR-196a and miR-196b in liver tissue indicates a poor prognosis for HCC patients. Therefore, miRNAs play an important role in cancer pathways. In addition, many other studies have also shown that there are a variety of abnormal miRNA expressions in liver cancer tissues. The detection and analysis of miRNAs expression in tumor tissues and adjacent tissues of HCC patients can provide the main basis for the HCC prognosis.

Gradually developed *vivo* detection technology is an important way to obtain liver cancer tissue cells of patients. Northern blot, polymerase chain reaction (PCR), microarray chip and other methods can obtain the expression of miRNAs after getting sample by *vivo* detection[9]. The prognosis of HCC can be determined and analyzed after obtaining the expression of biomarker miRNAs. *In vivo* detection technology has increased the feasibility and provided technical support for the theoretical study of prognosis using miRNAs. Meanwhile it also adds commercial value for the prognosis study of miRNA.

However, most research only focus on the diagnosis of HCC, but ignore the importance of prognosis. And at the same time, they have also focused on a single miRNA, but we believe that multiple miRNA combinations can provide more information than a single miRNA. Moreover, a lot of studies used miRNA precursors data for research, while in this study we used the mature miRNA data. Since miRNA regulation target genes play a role through mature miRNA, the mature miRNA data is more valuable than the precursor data. Therefore, we believe that using mature miRNA data and studying with multiple miRNA combinations can provide effective prognostic information for HCC patients.

The purpose of this study is to identify the prognostic-related miRNAs of HCC through the expression of mature miRNAs and clinical data from The Cancer Genome Atlas(TCGA, https://cancergenome. nih. gov/) database. The identified miRNAs were used to construct a model which can predict the prognosis of HCC patients. The model was used to predict the survival time and risk degree of HCC patients, and various bioinformatics methods were used to prove the relationship between the risk score calculated by the model and the survival time of HCC patients. The accuracy and practicability of the model were compared with other clinical indicators. And the effect of miRNAs *in vivo* were verified by target gene enrichment analysis.

1 Materials and methods

1.1 Data collection and pre−processing

HCC related data, including Level 3 mature miRNA expression data and clinical data were downloaded from TCGA(accessed August 2, 2018). The former data files consisted of 370 HCC tumor tissues and 50 adjacent normal liver tissues. The later data files consist of 9 indicators: age, gender, AFP, grade, pathologic stage, child-pugh, liver fibrosis, vascular invasion and race. The raw counts of mature

miRNA expression data needed to be preprocessed after downloading. The same patient's HCC tumor tissues needed to be paired with adjacent normal live tissues. Meanwhile, the miRNAs were deleted which sample data loss more than 20% and mean expression <1. The multiple interpolation method of mice in *R* software was used to predict the missing sample data[10]. In this method, each missing sample data was calculated by Gibbs sampling from other complete sample data for five times and a mean value of five predictions was filled in the missing location.

1.2 Differentially expressed miRNA （DEM） screening

DESeq, edgeR and Limma packages in *R* software were used to screen miRNA expression data[11-12]. The logarithm of fold change with base 2 and the false positive rate(FDR) were used as screening criteria to screen miRNA with $|\log_2$ fold change $|>0.8$ and FDR<0.001. Paired student's test (*t*-test) was performed between the HCC tumor group and the adjacent normal liver group of miRNAs, and miRNAs with *P* value less than 0.001 were retained. The overlapping miRNAs of the four methods were identified as final DEMs.

1.3 DEM−based prognostic biomarkers screening

The data of DEMs with all HCC tissue samples were transformed into $\log_2$(reads per million+1). Patients with a survival time less than 30 days would be deleted to avoid the influence of other factors (For example, death from non-liver cancer causes). Kaplan-Meier(K-M) and univariate Cox proportional hazards regression analysis were performed with a survival package of *R* software to screen HCC prognostic biomarkers and evaluate the prognostic value of DEMs. The screening criteria of the two methods were $P<0.01$ and $P<0.01$, respectively. The overlap of DEM-based biomarkers screening was regarded as prognostic biomarkers which significantly associated with the overall survival (OS) of HCC patients.

1.4 Construction of risk score model

In order to determine the survival status of HCC patients accurately, the hazard ratio(HR) value, *P* value, and other factors were comprehensively considered for finding the optimal prognostic DEM combination. In this process, multivariate COX proportional hazards regression analysis and "step" function were used for further screening to find a risk score model, which contained a small amount of miRNA but more characteristics and accurately predicted the prognosis, as shown in formula (1). This method is called stepwise feature addition.

$$\text{Risk score} = \beta_{\text{miRNA1}} \times \text{Exp}_{\text{miRNA1}} + \beta_{\text{miRNA2}} \times \text{Exp}_{\text{miRNA2}} + \cdots + \beta_{\text{miRNA}n} \times \text{Exp}_{\text{miRNA}n} \quad (1)$$

The multivariate Cox regression coefficient(β) was used as the weight for building the model. Exp is the expression value of miRNA.

According to the formula, calculate the risk score value of each sample and sort them according to the value. And the HCC patients were divided into two groups by the median: the high expression group, and the low expression group. The 1,3 and 5-year receiver-operating characteristic(ROC) curves were drawn by the survivalROC package to evaluate the reliability of the risk score model. The *Chi-square* test was used to determine the relationship between risk scores and HCC patient survival times.

1.5 Comprehensive analysis of risk score model and clinical indicators

Nine different clinical indicators were included for further discussion of the impact of the risk score model and clinical indicators on the HCC patient OS. After the univariate Cox proportional hazards regression analysis, clinical indicators with a *P* value less than 0.05 were selected for further multivariate Cox proportional hazards regression analysis. According to the results, the former analysis could find the key clinical indicators affecting the OS rate, and the latter analysis could determine whether the risk score model could predict the prognosis independently. K-M survival curves and box plots were drawn to study the value of clinical indicators and the potential application of the risk score model in clinical practice.

1.6 Target gene prediction and functional enrichment analysis

miRTarBase 6.0(http://mirtarbase. mbc. nctu. edu. tw/php/index. php, accessed March 28, 2019) was used to predict target genes of prognostic DEMs[13]. All target genes in miRTarBase 6.0 database have experimental evidence to support the relationship between miRNAs and target genes. Therefore, the selected target genes needed to be proved by strong experiments(such as Western blot, qRT-PCR, Luciferase reporter assay and Immunoprecipitation). The interaction networks between target genes and miRNAs were drawn by Cytoscape 3.6.1(http://www.

cytoscape. org/, accessed September 29, 2018) [14]. Though the enrichment analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) by ClusterProfiler package of *R* software to find the potential functions of target genes.

1.7 Statistical analysis

The *P* values calculated by edgeR, DESeq, Limma software package were adjusted by Benjamini-hochberg method to obtain the FDR values. K-M analysis was used to draw the survival curve, and the log-rank test was adopted to evaluate the difference between miRNAs and OS, as well as clinical indicators and OS. The miRNAs and clinical indicators were selected according to the *P* value for further multivariate Cox proportional hazards regression analysis. The purpose of it was to explore its potential association with the OS of HCC patients. All statistical analyses were performed with SPSS 21.0(IBM Corporation, Armonk, NY, USA) and R 3.5.1.

2 Results

2.1 DEM screening

Among the 50 adjacent normal liver tissue samples, 49 pairs were successfully matched with the HCC tumor tissue samples, and only one sample was not matched with the tumor samples (Table 1). A total of 2 165 mature miRNAs were extracted from the downloaded file, and 453 miRNAs were remained after pretreatment in which deleting miRNAs with a lot of data missing and very low expression. Four methods of edgeR, DESeq, Limma and *t*-test were used to screen 453 miRNAs according to the above criteria. The number of remaining miRNA was 114, 119, 123 and 155, respectively. The number of screened overlapping miRNA was 70, which can be regarded as DEMs. The overlapping graph and heatmap of DEMs were respectively drawn by UpSetR package(Figure 1a) and pheatmap package (Figure 1b).

Table 1 Matched comparison table of HCC tumor tissue samples and normal adjacent liver tissue samples

NO.	Cancer	Normal	NO.	Cancer	Normal
1	BC-A10Q-01A	BC-A10Q-11A	*26*	DD-A1EH-01A	DD-A1EH-11A
2	BC-A10R-01A	BC-A10R-11A	*27*	DD-A1EI-01A	DD-A1EI-11A
3	BC-A10T-01A	BC-A10T-11A	*28*	DD-A1EJ-01A	DD-A1EJ-11A
4	BC-A10U-01A	BC-A10U-11A	*29*	DD-A1EL-01A	DD-A1EL-11A
5	BC-A10W-01A	BC-A10W-11A	*30*	DD-A39V-01A	DD-A39V-11A
6	BC-A10X-01A	BC-A10X-11A	*31*	DD-A39W-01A	DD-A39W-11A
7	BC-A10Y-01A	BC-A10Y-11A	*32*	DD-A39X-01A	DD-A39X-11A
8	BC-A10Z-01A	BC-A10Z-11A	*33*	DD-A39Z-01A	DD-A39Z-11A
9	BC-A110-01A	BC-A110-11A	*34*	DD-A3A1-01A	DD-A3A1-11A
10	BC-A216-01A	BC-A216-11A	*35*	DD-A3A2-01A	DD-A3A2-11A
11	BD-A2L6-01A	BD-A2L6-11A	*36*	DD-A3A3-01A	DD-A3A3-11A
12	BD-A3EP-01A	BD-A3EP-11A	*37*	DD-A3A4-01A	DD-A3A4-11A
13	DD-A113-01A	DD-A113-11A	*38*	DD-A3A5-01A	DD-A3A5-11A
14	DD-A114-01A	DD-A114-11A	*39*	DD-A3A6-01A	DD-A3A6-11A
15	DD-A116-01A	DD-A116-11A	*40*	DD-A3A8-01A	DD-A3A8-11A
16	DD-A118-01A	DD-A118-11A	*41*	EP-A12J-01A	EP-A12J-11A
17	DD-A119-01A	DD-A119-11A	*42*	EP-A26S-01A	EP-A26S-11A
18	DD-A11A-01A	DD-A11A-11A	*43*	EP-A3RK-01A	EP-A3RK-11A
19	DD-A11B-01A	DD-A11B-11A	*44*	ES-A2HT-01A	ES-A2HT-11A
20	DD-A11C-01A	DD-A11C-11A	*45*	FV-A23B-01A	FV-A23B-11A
21	DD-A11D-01A	DD-A11D-11A	*46*	FV-A2QR-01A	FV-A2QR-11A
22	DD-A1EB-01A	DD-A1EB-11A	*47*	FV-A3I0-01A	FV-A3I0-11A
23	DD-A1EC-01A	DD-A1EC-11A	*48*	FV-A3I1-01A	FV-A3I1-11A
24	DD-A1EE-01A	DD-A1EE-11A	*49*	G3-A3CH-01A	G3-A3CH-11A
25	DD-A1EG-01A	DD-A1EG-11A	-	-	-

*01A represents the HCC tumor tissue sample and 11A represents the adjacent normal liver tissue sample.

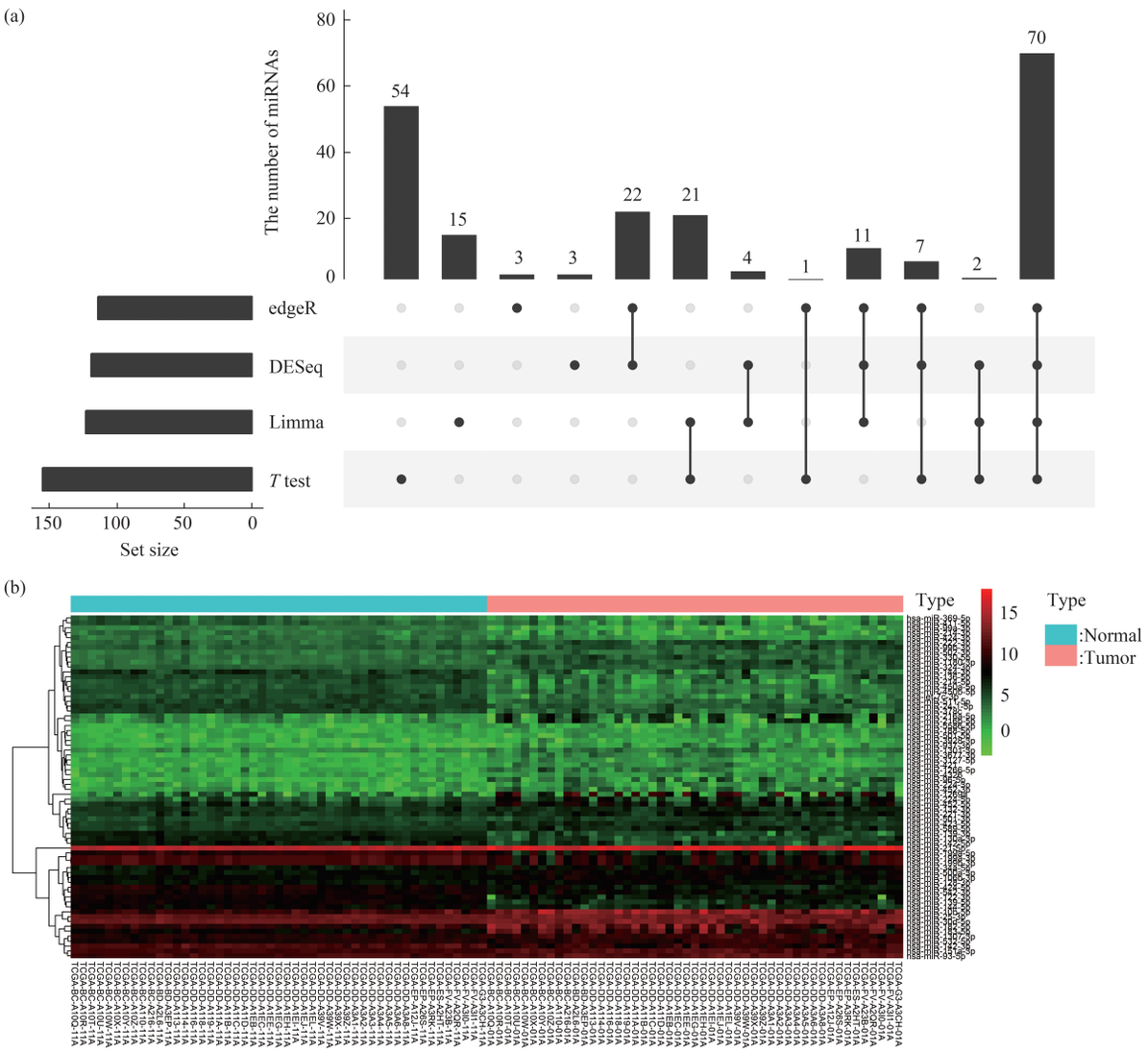

Fig. 1　The overlapping plot and heatmap of DEMs in HCC

（a）The overlapping plot of four methods screening the DEMs in HCC.（b）Heat map of 70 DEMs in HCC.

2.2　DEM−based prognostic biomarkers screening

HCC patient samples with survival time less than 30 days were deleted, leaving 363 samples. The results of survival curve and univariate Cox proportional risk regression analysis showed that there were 10 DEMs with *P* values less than 0.01(Table 2). The 10 DEMs were regarded as candidate prognostic biomarkers for multivariate Cox proportional hazards regression analysis and risk model construction.

Table 2　*P* values of univariate COX and Survival for 70 DEMs and identification of candidate biomarkers

NO.	miRNA	HR	Cox *P* value	significant	Survival *P* value	significant	Candidate
1	hsa-let-7c-3p	0.894	0.184	No	0.828	No	No
2	hsa-miR-106b-3p	1.468	0.001	Yes	0.091	No	No
3	hsa-miR-10b-5p	1.118	0.027	No	0.340	No	No
4	hsa-miR-1180-3p	1.318	0.000	Yes	0.010	No	No

Continued to Table 2

NO.	miRNA	HR	Cox *P* value	significant	Survival *P* value	significant	Candidate
5	hsa-miR-122-3p	0.854	0.005	Yes	0.065	No	No
6	hsa-miR-1266-5p	1.364	0.000	Yes	0.073	No	No
7	hsa-miR-1269a	1.040	0.090	No	0.129	No	No
8	hsa-miR-128-3p	1.083	0.503	No	0.667	No	No
9	hsa-miR-1301-3p	1.319	0.002	Yes	0.523	No	No
10	hsa-miR-1307-3p	1.349	0.005	Yes	0.318	No	No
11	hsa-miR-130a-3p	1.010	0.913	No	0.830	No	No
12	**hsa-miR-132-3p**	**1.605**	**0.001**	**Yes**	**0.004**	**Yes**	**Yes**
13	hsa-miR-136-5p	0.996	0.934	No	0.912	No	No
14	**hsa-miR-139-3p**	**0.802**	**0.005**	**Yes**	**0.001**	**Yes**	**Yes**
15	**hsa-miR-139-5p**	**0.734**	**0.000**	**Yes**	**0.000**	**Yes**	**Yes**
16	hsa-miR-142-3p	1.056	0.504	No	0.728	No	No
17	hsa-miR-142-5p	0.976	0.749	No	0.748	No	No
18	hsa-miR-144-3p	0.981	0.746	No	0.590	No	No
19	hsa-miR-144-5p	0.929	0.210	No	0.158	No	No
20	hsa-miR-151a-3p	1.583	0.000	Yes	0.014	No	No
21	hsa-miR-182-5p	1.088	0.045	No	0.130	No	No
22	hsa-miR-183-5p	1.101	0.014	No	0.070	No	No
23	hsa-miR-188-5p	1.392	0.003	Yes	0.010	No	No
24	hsa-miR-199a-3p	0.997	0.959	No	0.514	No	No
25	hsa-miR-199a-5p	0.955	0.313	No	0.153	No	No
26	hsa-miR-199b-3p	0.997	0.948	No	0.514	No	No
27	hsa-miR-214-3p	1.071	0.424	No	0.304	No	No
28	hsa-miR-214-5p	1.007	0.913	No	0.840	No	No
29	hsa-miR-21-5p	1.304	0.036	No	0.009	Yes	No
30	hsa-miR-216a-5p	1.009	0.791	No	0.775	No	No
31	hsa-miR-216b-5p	1.008	0.805	No	0.864	No	No
32	hsa-miR-221-3p	1.162	0.092	No	0.090	No	No
33	hsa-miR-222-3p	1.358	0.001	Yes	0.019	No	No
34	hsa-miR-224-3p	1.152	0.084	No	0.218	No	No
35	hsa-miR-224-5p	1.088	0.089	No	0.049	No	No
36	hsa-miR-24-1-5p	0.947	0.626	No	0.878	No	No
37	hsa-miR-25-3p	1.529	0.000	Yes	0.026	No	No
38	hsa-miR-30d-5p	0.956	0.661	No	0.456	No	No
39	**hsa-miR-3127-5p**	**1.752**	**0.000**	**Yes**	**0.001**	**Yes**	**Yes**
40	**hsa-miR-324-3p**	**1.467**	**0.000**	**Yes**	**0.000**	**Yes**	**Yes**
41	hsa-miR-330-5p	1.354	0.001	Yes	0.016	No	No
42	hsa-miR-34a-5p	0.808	0.045	No	0.074	No	No
43	**hsa-miR-3677-3p**	**1.617**	**0.000**	**Yes**	**0.000**	**Yes**	**Yes**
44	hsa-miR-369-5p	1.110	0.119	No	0.101	No	No
45	hsa-miR-378c	0.933	0.364	No	0.505	No	No
46	hsa-miR-3928-3p	1.537	0.003	Yes	0.616	No	No
47	hsa-miR-411-5p	1.057	0.343	No	0.422	No	No
48	**hsa-miR-421**	**1.562**	**0.000**	**Yes**	**0.003**	**Yes**	**Yes**
49	hsa-miR-423-5p	1.337	0.038	No	0.166	No	No
50	hsa-miR-424-3p	0.999	0.992	No	0.376	No	No
51	hsa-miR-424-5p	1.191	0.043	No	0.098	No	No

Continued to Table 2

NO.	miRNA	HR	Cox *P* value	significant	Survival *P* value	significant	Candidate
52	hsa-miR-4326	1.211	0.041	No	0.333	No	No
53	hsa-miR-450a-5p	1.048	0.645	No	0.302	No	No
54	hsa-miR-450b-5p	0.989	0.913	No	0.928	No	No
55	hsa-miR-452-3p	1.192	0.009	Yes	0.065	No	No
56	hsa-miR-452-5p	1.145	0.025	No	0.046	No	No
57	**hsa-miR-500a-3p**	**1.333**	**0.003**	**Yes**	**0.005**	**Yes**	**Yes**
58	hsa-miR-501-3p	1.311	0.003	Yes	0.028	No	No
59	**hsa-miR-501-5p**	**1.361**	**0.003**	**Yes**	**0.008**	**Yes**	**Yes**
60	hsa-miR-502-3p	1.202	0.129	No	0.140	No	No
61	hsa-miR-511-5p	1.153	0.123	No	0.320	No	No
62	hsa-miR-532-5p	1.027	0.805	No	0.283	No	No
63	hsa-miR-542-3p	0.963	0.727	No	0.891	No	No
64	hsa-miR-5586-5p	1.310	0.006	Yes	0.018	No	No
65	hsa-miR-589-5p	1.278	0.031	No	0.179	No	No
66	hsa-miR-93-5p	1.190	0.075	No	0.191	No	No
67	hsa-miR-937-3p	1.336	0.002	Yes	0.536	No	No
68	hsa-miR-96-5p	1.130	0.020	No	0.043	No	No
69	hsa-miR-99a-3p	0.832	0.050	No	0.686	No	No
70	**hsa-miR-99b-3p**	**1.400**	**0.000**	**Yes**	**0.002**	**Yes**	**Yes**

2.3 Construction of risk score model

Using 10 DEMs for multivariate Cox proportional hazards regression analysis, the number of miRNAs were gradually reduced to construct a risk score model, which has a strong predictive ability. Considering the factors mentioned above, it is found that better prognostic predictive results can be obtained using models constructed with hsa-miR-132-3p, hsa-miR-139-5p, hsa-miR-3677-3p and hsa-miR-500a-3p. The K-M survival curve and ROC curve of four miRNAs were shown in Figure 2a−d and Figure 2f−i. ROC curves were plotted using the adjacent normal liver group as a control to obtain *P* values. The risk score formula was shown in formula (2).

$$\text{Risk score} = (0.348 \times Exp_{\text{hsa - miR - 132 - 3p}}) - (0.178 \times Exp_{\text{hsa - miR - 139 - 5p}}) + (0.250 \times Exp_{\text{hsa - miR - 3677 - 3p}}) + (0.237 \times Exp_{\text{hsa - miR - 500a - 3p}}) \quad (2)$$

The expression of hsa-miR-132-3p marked as $Exp_{hsa-miR-132-3p}$ and it is same to others.

The HR values of hsa-miR-132-3p, hsa-miR-139-5p, hsa-miR-3677-3p and hsa-miR-500a-3p were shown as following: 1.416(95% confidence interval (CI) =1.069−1.882, *P*=0.016); 0.836(95%CI=0.713−0.999,*P*=0.036);1.283(95%CI=1.021−1.622,*P*=0.033); 1.267(95%CI=1.049−1.533,*P*=0.012)(Figure 3a). The median risk score for 363 HCC patients was 3.49. Patients with a risk score higher than 3.49 were in the high risk group and lower than 3.49 were in the low risk group. Meanwhile, hsa-miR-139-5p was a low-risk miRNA, which HR value less than 1, and the risk of death would decrease with the increase of expression level. On the contrary, hsa-miR-132-3p, hsa-miR-3677-3p and hsa-miR-500a-3p were high risk miRNAs, with a HR value over 1, and the risk of death would increase with the increase of expression level. With the increase of risk score calculated by the formula(2), the survival time of HCC patients became shorter and the number of deaths started increasing (HR=2.551, 95%CI=1.751−3.717, *P*<0.000; Figure 4a and b). The expression distribution of these four miRNAs between adjacent normal tissues and HCC tissues was shown in Figure 3b, as well as HCC tissues of low risk and high risk groups were shown in Figure 3c. When the expressions of hsa-miR-132-3p, hsa-miR-3677-3p, hsa-miR-500a-3p were up-regulated and the expression of hsa-miR-139-5p was down-regulated, the prognosis was poor. The results of the *Chi-square* test also demonstrated a significant correlation between risk scores and HCC patient survival times (Table 3). The 1, 3 and 5-year survival rates of HCC patients in low-risk group and high-risk group were 91.8%, 70.2%, 56.9% and 71.2%, 40.0%,

24.5%, respectively (Table 4). Overall, 1−5 year survival rates of HCC patients in the low-risk group were 20%−30% higher than high-risk group. The difference in survival rates between the two groups increased with time. And the Area Under Curve(AUC) corresponding was 0.775, 0.721 and 0.704, respectively, as shown in Figure 4c. Time-related ROC curve analysis also showed the model performed well in OS prediction of HCC prognosis. These all proved that HCC patients in the high-risk group and the low-risk group had significant differences in survival time and risk degree. It also proves that this grouping is reliable. Since the grouping was based on risk score, it also indicated the reliability of the four miRNAs and the model.

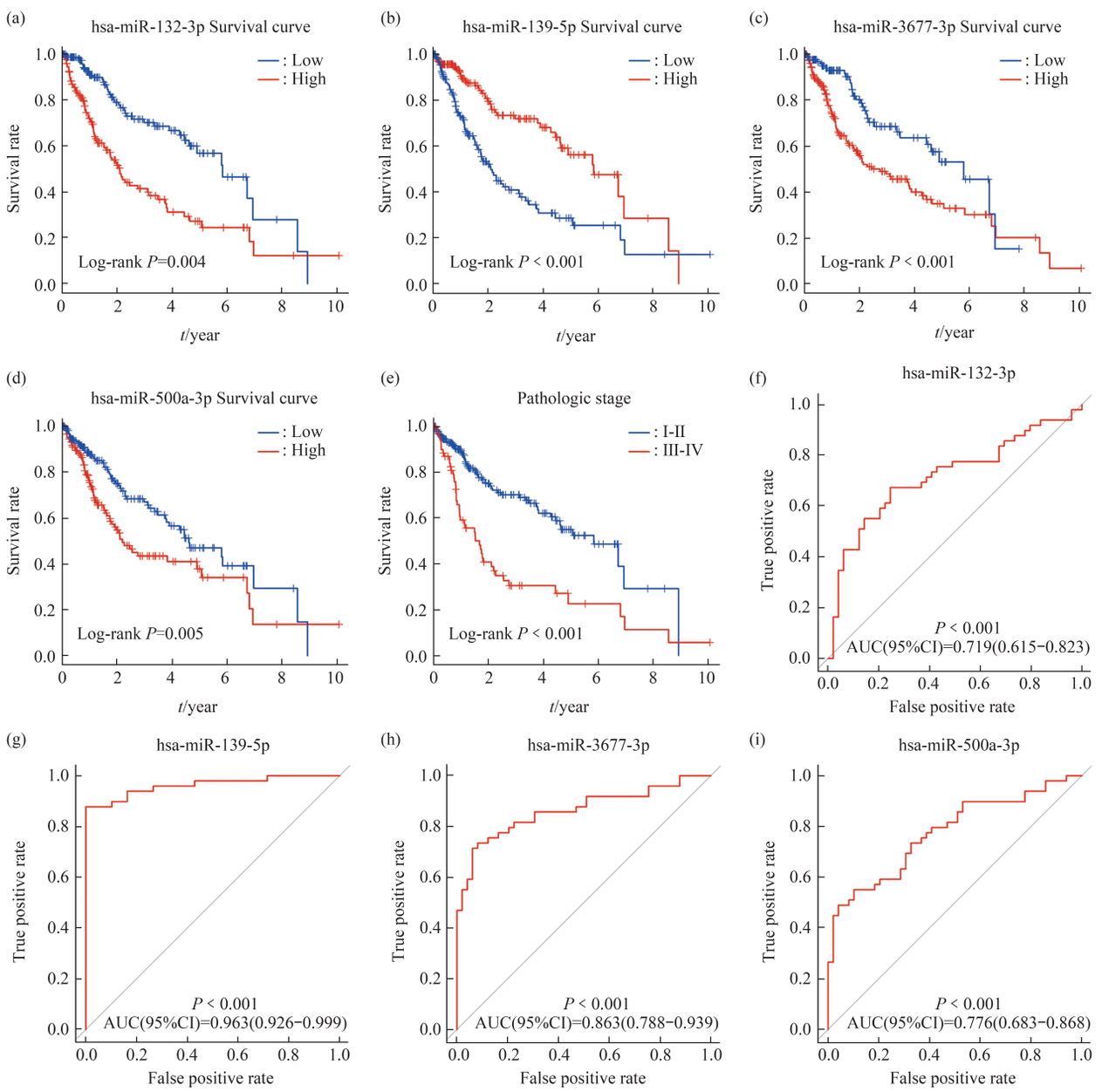

Fig. 2　The Kaplan−Meier and ROC curves of four prognostic DEMs and clinical indicator in HCC

The order of Kaplan-Meier curves were as follows：（a）hsa-miR-132-3p；（b）hsa-miR-139-5p；（c）hsa-miR-3677-3p；（d）hsa-miR-500a-3p；（e）Pathologic stage. The order of ROC curves of four prognostic DEMs were as follow：（f）hsa-miR-132-3p；（g）hsa-miR-139-5p；（h）hsa-miR-3677-3p；（i）hsa-miR-500a-3p.

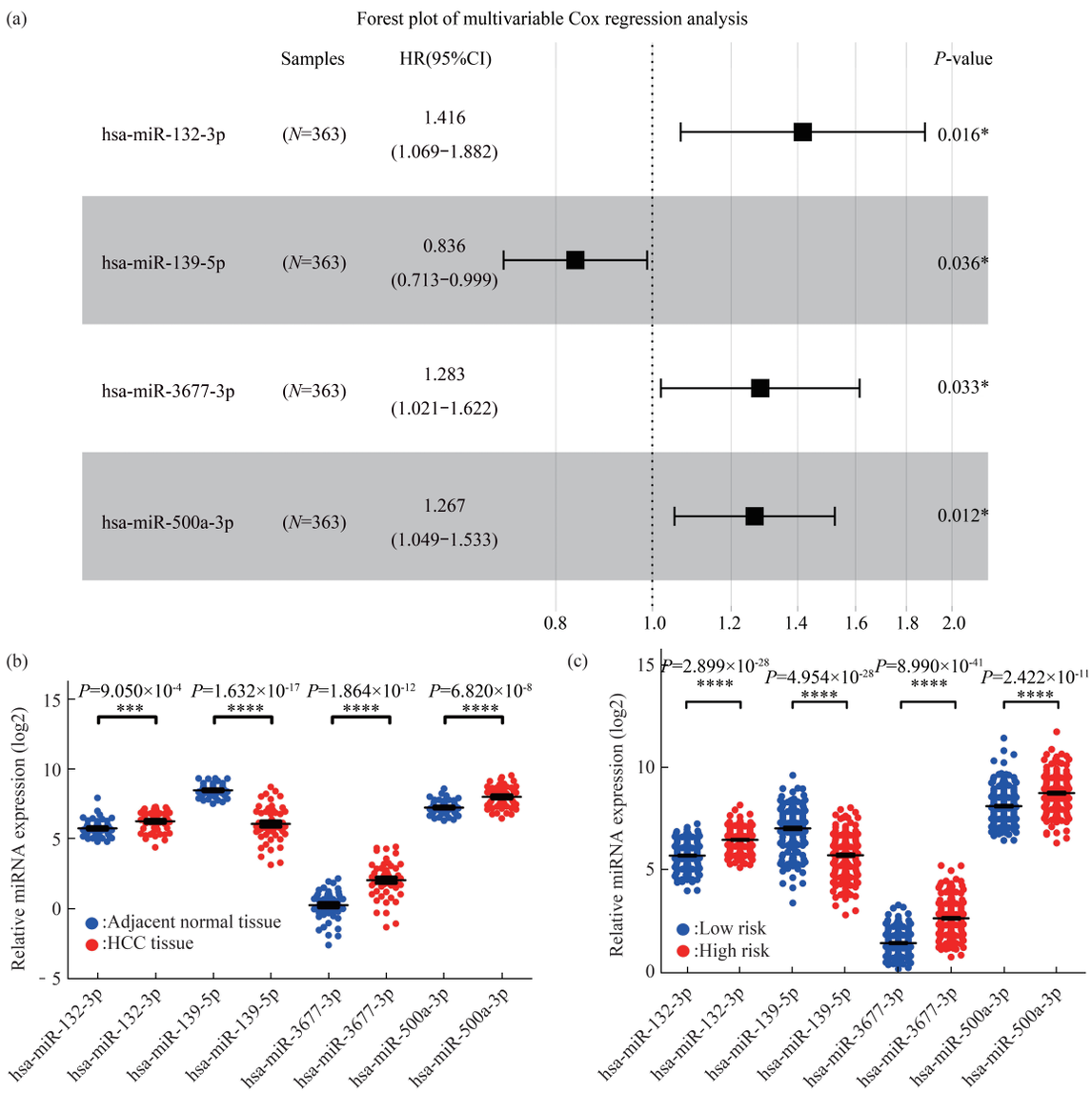

Fig. 3　Forest plot of multivariable Cox regression analysis and expression level of four prognostic DEMs between HCC tumor and adjacent normal liver tissues，and low− and high−risk groups

（a）Multivariable Cox proportional hazards regression analyses of four prognostic DEMs in HCC patients. The black squares represent the HR and the short transverse. lines represent 95% CI. **P*<0.05；（b）Scatter plot of four prognostic DEMs' expression level between HCC tumor and adjacent normal liver tissues；（c）Scatter plot of four prognostic DEMs' expression level between low-and high-risk groups. ****P*<0.001；*****P*< 0.0001. DEM，differentially expressed miRNA；HR，hazard ratio；CI，confidence interval；HCC，hepatocellular carcinoma.

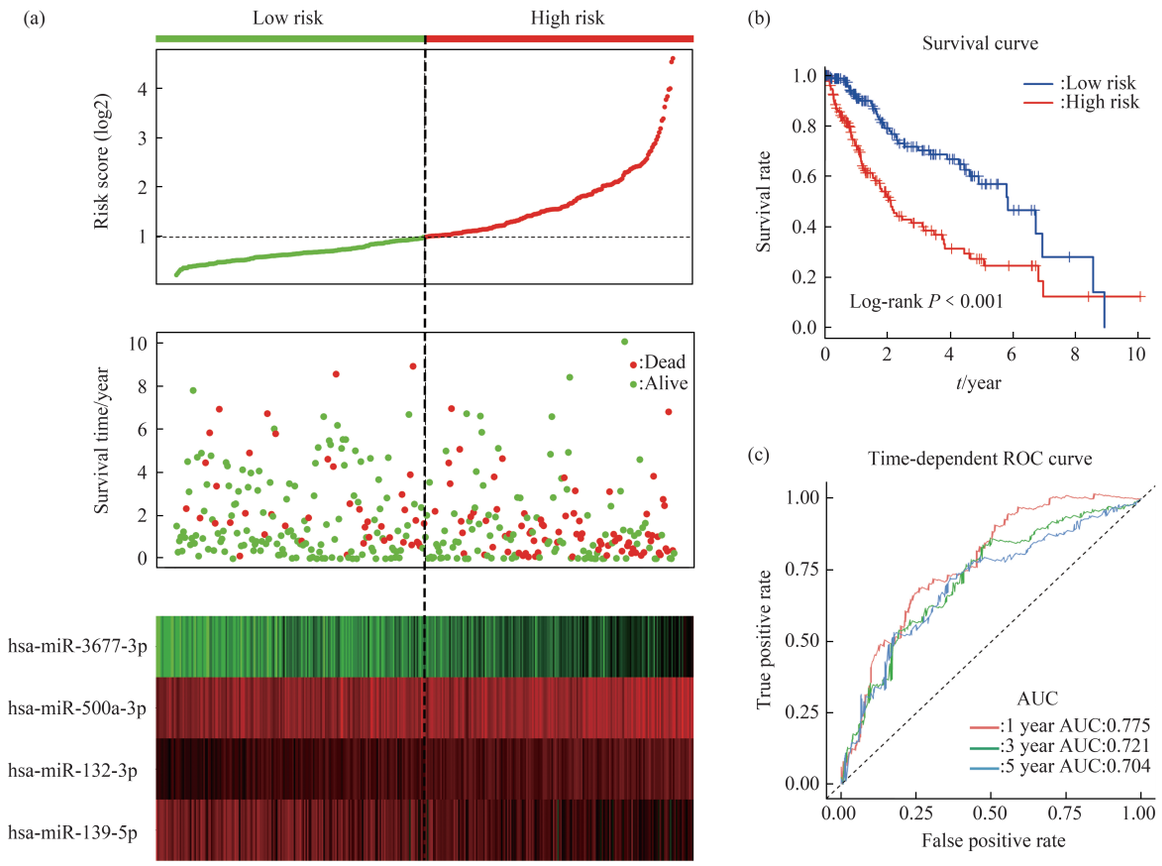

Fig. 4　Prognostic risk score model analysis of four prognostic DEMs in HCC patients

（a）From top to bottom are the risk score，patients' survival status distribution，and four prognostic DEMs' expression heat map for low- and high-risk groups；（b）K-M curves for low- and high-risk groups；（c）ROC curve for predicting survival in HCC patients by the risk score. ROC，receiver-operating characteristic；AUC，the area under the curve.

Table 3　*Chi-Square* tests between risk scores and HCC patient survival times

	Value	df	Asymptotic Significance（2-sided）
Pearson *Chi-Square*	8.951[1)]	1	0.003
Likelihood Ratio	8.989	1	0.003
N of Valid Cases	363	1	-

[1)] 0 cells（0.0%） have expected count less than 5. The minimum expected count is 90.25.

Table 4　Survival rates in the low-risk group and high-risk group

Survival time	Low risk group	High risk group
	Survival rate（95%CI）	Survival rate（95%CI）
<1 year	91.8%（87.2%-96.6%）	71.2%（63.9%-79.4%）
1-2 years	79.1%（71.6%-87.3%）	51.6%（43.2%-61.8%）
2-3 years	70.2%（61.5%-80.2%）	40.0%（31.2%-51.1%）
3-4 years	66.8%（57.5%-77.6%）	31.2%（22.5%-43.4%）
4-5 years	56.9%（45.6%-70.9%）	24.5%（15.9%-37.8%）
>5 years	<56.9%	<24.5%

2.4 Comprehensive analysis of risk score model and clinical indicators

The risk score model and 9 clinical indicators were also analyzed by univariate Cox proportional hazards regression. It was found that there were significant differences in Risk Score (HR=1.826, *P*< 0.000, 95% CI=1.516−2.189) and pathologic stage(HR =2.639, *P*<0.000, 95% CI=1.785−3.899) (Table 5). The K-M survival curves of pathologic stage were shown in Figure 2e. The two indicators were further analyzed by multivariate Cox proportional hazards regression analysis. It was found that Risk Score(HR= 1.644, *P*<0.000, 95% CI=1.342−2.014) and pathologic stage(HR=2.061, *P*=0.001, 95% CI=1.362− 3.119) still existed significant differences, as shown in Table 6. The results showed that there were two indicators that could affect OS of HCC patients. At the same time, risk score and pathologic stage were not affected by other indicators. It could predict the survival status of HCC patients independently and be belonged to independent prognostic factors. The box plots showed that the risk score of HCC patients was significantly correlated with pathologic stage, grade, AFP and vascular invasion(Figure 5). Therefore, those four clinical indicators could be predicted by calculating the patient's risk score. Through comprehensive analysis, we found that the combination of these four prognostic DEM performed well in not only OS prediction, but also clinical analysis.

Table 5 Univariable Cox proportional hazards regression analyses of risk score model and clinical indicators

Variables	Patients（*n*=363）	Univariate analysis	
		HR（95%CI）	*P*-value
Age（years）（<60/≥60）	169/190	1.167（0.811−1.679）	0.405
Gender（Male/Female）	247/116	1.154（0.797−1.671）	0.448
AFP（<400 μg/L/≥400 μg/L）	210/64	1.131（0.682−1.875）	0.633
Grade（G1−2/G3−4）	228/131	1.020（0.698−1.492）	0.917
Pathologic stage（I−II/II−IV）	**251/88**	**2.639（1.785−3.899）**	**<0.000**
Child-Pugh（A/B-C）	213/22	2.002（0.982−4.084）	0.056
Liver Fibrosis（No/Yes）	74/136	0.821（0.482−1.399）	0.468
Vascular Invasion（No/Yes）	200/93	1.373（0.857−2.199）	0.187
Race（Asian/Others）	159/192	1.217（0.827−1.792）	0.318
Risk Score（low/high）	**182/181**	**1.826（1.516−2.198）**	**<0.000**

Table 6 Multivariable Cox proportional hazards regression analyses of risk score model and pathologic stage

Variables	Patients（*n*=363）	Multivariate analysis	
		HR（95%CI）	*P*-value
Pathologic stage（I−II/III−IV）	251/88	2.061（1.362−3.119）	0.001
Risk Score（low/high）	182/181	1.644（1.342−2.014）	<0.000

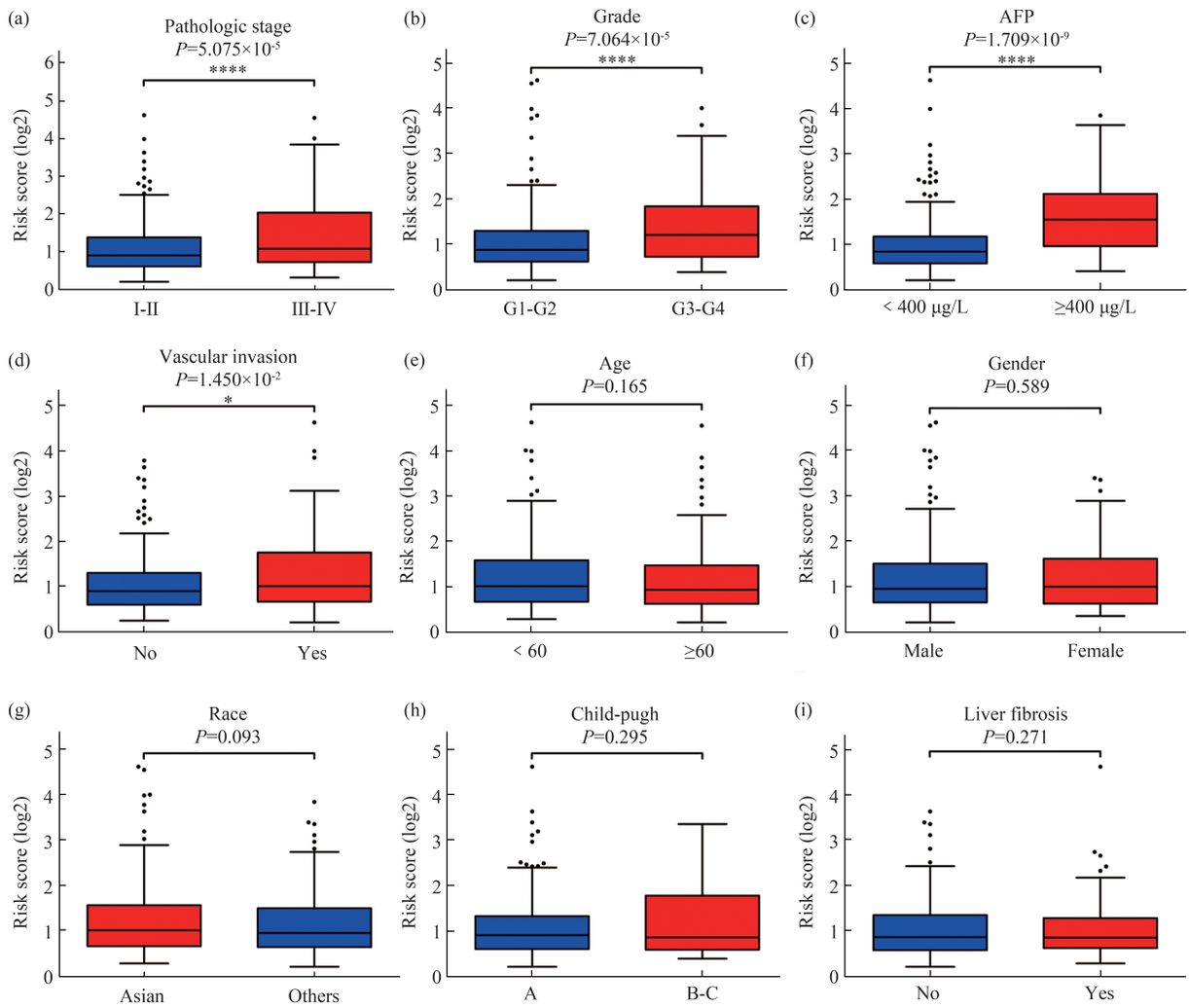

Fig. 5　The box plots of risk score differences in nine clinical indicators

（a）Pathologic stage；（b）Grade；（c）AFP；（d）Vascular invasion；（e）Age；（f）Gender；（g）Race；（h）Child-Pugh；（i）Liver fibrosis. *P<0.05；****P<0.0001. AFP，alpha-fetoprotein.

2.5　Target prediction and enrichment analysis

A total of 60 target genes confirmed by strong experiments were predicted by miRTarBase 6.0. The number of target genes of hsa-miR-132-3p, hsa-miR-139-5p and hsa-miR-500a-3p was 32, 25 and 3, respectively. However, hsa-mir-3677-3p didn't have target genes in miRTarBase verified by strong experiment, only have the target genes verified by weak experiment. These 60 target genes were constructed the interaction network between target genes and miRNAs (Figure 6a). The enrichment results of target genes include GO and KEGG enrichment. The GO enrichment also includes three parts: molecular function, biological process and cellular component. The GO enrichment results in Figure 6b showed the 60 target genes were major enriched in biological processes, such as immune response signaling pathway, regulation of cell mitotic cycle, modification of protein phosphorylation and negative regulation of the metabolic process. The KEGG enrichment results in Figure 6c showed the target genes of the four miRNAs were enriched in

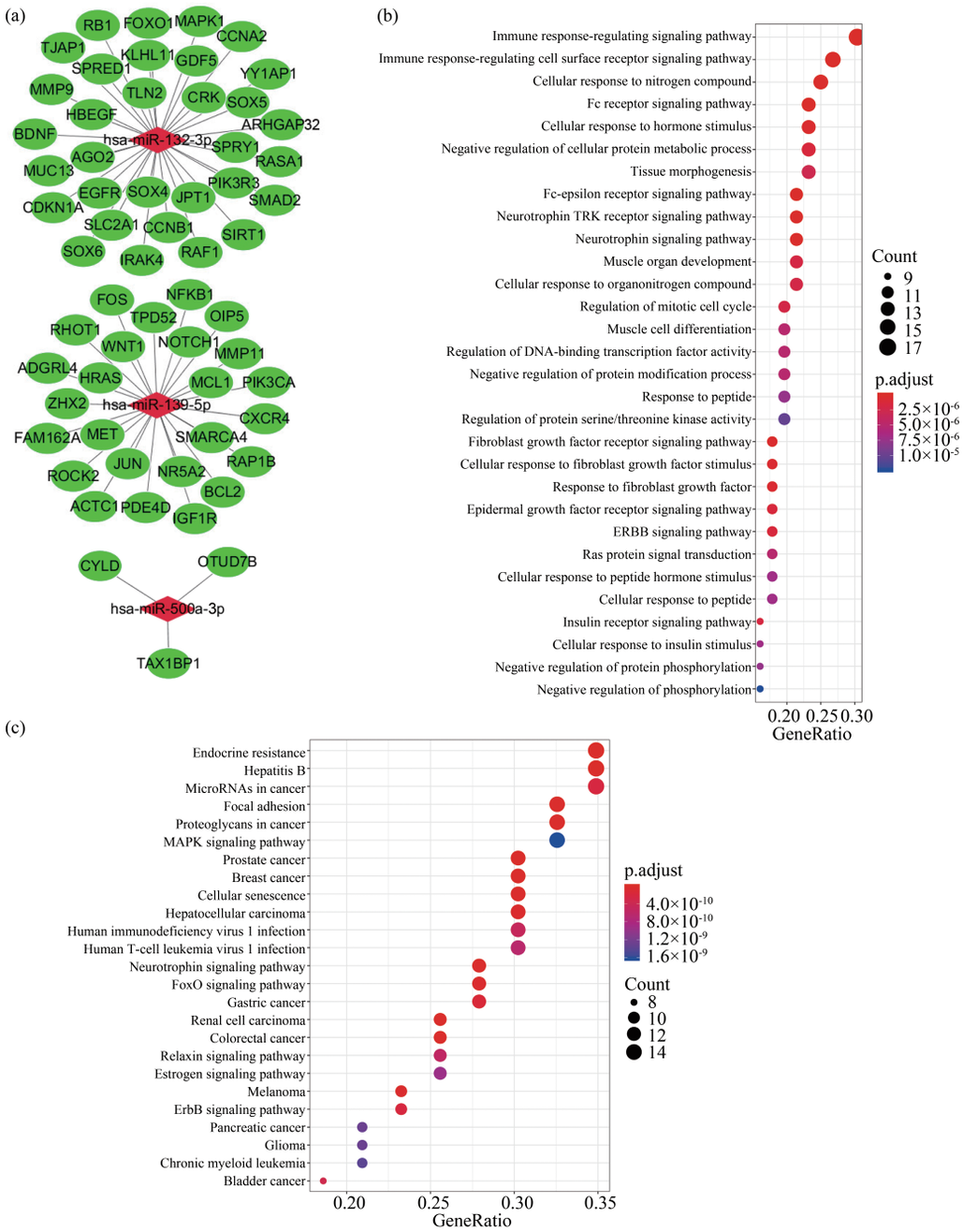

Fig. 6　The prognostic miRNAs target genes interactions network and their enrichment analysis results

（a）The prognostic miRNAs-target genes interactions network；（b）GO term enrichment results；（c）KEGG enrichment results.

MAPK signaling pathway, ErbB signaling pathway, FoxO signaling pathway, cell senescence, adhesion spots, and other aspects. The enrichment of GO and KEGG demonstrated that the predicted target genes could affect the immune response, protein modification and metabolism, cell proliferation, migration, and apoptosis.

3 Discussion

Currently, miRNAs have been used in many kinds of research as biomarkers for diagnosis and prognosis analysis of breast, colorectal, prostate, stomach, lung, liver and other cancers[15-17]. The prognosis of HCC often is poor, so it is necessary to find reliable miRNAs as prognostic biomarkers to provide effective prognostic information for HCC patients. The advantages of this study include the use of mature miRNA data and pre-processing of the raw count's data to avoid the impact of a large number of absences and low expression miRNAs on the screening. Meanwhile, four methods were also used to screen DEM simultaneously to ensure the reliability of screening. The optimal model for HCC prognosis was built by the function of step in R software. A variety of bioinformatics methods have demonstrated that the risk score calculated by the model can accurately predict the OS of HCC patients, especially in the chi-square test. In addition, the study explored the potential prognostic application of the model through box plots and analysis between the model and clinical indicators.

Many research have suggested that dysregulated miRNAs can be used as biomarkers for diagnosis or prognosis of HCC[18]. Four miRNAs(hsa-miR-132-3p, hsa-miR-139-5p, hsa-miR-500a-3p and hsa-miR-3677-3p) with dysregulated were found in our study. The target genes of has-miR-132-3p include *AGO2*, *FOXO1*, *MAPK1*, *SOX4*, *etc.* AGO2 is an essential protein in miRNA processing and gene silencing[19-20]. The overexpression of hsa-miR-132 affects the downstream pathway by inhibiting AGO2 expression. In addition, FOXO1 has been shown to be a tumor suppressor protein by blocking the G1/S transition in the cell cycle and inducing apoptosis[21]. Activation of PI3K/AKT signaling pathway promotes cell proliferation, which leads to poor prognosis. hsa-miR-132 reduces FOXO1 expression and promotes cancer progression by directly targeting FOXO1 or activating the PI3K/AKT pathway. hsa-miR-132-3p is a recognized biomarker for early diagnosis of HCC, indicating that its expression level changes have high research value[22].

The target genes of has-miR-139-5p include *ROCK2*, *RAP1B*, *NFKB1*, *WNT1*, *etc*. ROCK2 overexpression is associated with tumor invasion. RAP1B, a member of the Ras superfamily G-protein, has the properties of promoting tumor cell proliferation and mitosis. In summary, ROCK2 and RAP1B promote the proliferation and migration of cancer cells. However, these genes play a carcinogenic role not only through its expression products, but also through its mRNAs, which can interact to affect the biochemical properties of cells. For example, ROCK2 and RAP1B compete with each other in binding with miR-139-5p to execute carcinogenic functions that regulate proliferation and migration[23]. The low expression of miR-139-5p failed to inhibit the expression of ROCK2 and RAP1B, which led to its function as an oncogene, promoting the proliferation and metastasis of cancer cells. Therefore, this process may be an important reason for the poor prognosis of HCC. Similar to the result of our study, Li *et al.* [24] also found that hsa-miR-139-5p was significantly down-regulated in both cancer tissues and plasma of HCC patients, and the low expression of hsa-miR-139-5p significantly reduced the OS of HCC patients. This demonstrates that hsa-miR-139-5p is one of the most important prognostic factors for HCC.

Only three target genes, *CYLD*, *OTUD7B* and *TAX1BP1*, were enriched in hsa-miR-500a-3p. *CYLD* is a potential miRNA target gene that encodes deubiquitinating enzymes. *CYLD* is considered a tumor suppressor because of its low expression in many human tumor types[25]. It has been reported that the deletion of *CYLD* gene in hepatocytes leads to the formation of HCC[26]. Therefore, we speculated that the high expression of miR-500a-3p may target *CYLD* and promote cells to proliferate, resulting in a poor prognosis. Jiang *et al.* [27] showed that miR-500a-3p was significantly increased in HCC tissues, cells and blood, and its high expression reduced OS of HCC patients. According to the analysis of clinical data, the expression of hsa-miR-500a-3p is positively correlated with tumor size, distant metastasis and clinical stage of HCC patients. In our study of HCC, the report on hsa-miR-3677-3p is still in the theoretical stage, and there is no experimental evidence to prove the relationship between hsa-miR-3677-3p and HCC. But, it is known that hsa-miR-3677-5p in HCC cells is considered as a novel miRNA that affects the expression of STAT3 and JAK1 proteins[28]. Therefore, we speculated that hsa-miR-3677-3p may also be related to the occurrence

and development of HCC. This study revealed that the expression of hsa-miR-3677-3p had a certain impact on the prognosis of HCC patients. Further biological experiments are needed to explore the specific mechanism of adverse prognosis of HCC patients due to the high expression of hsa-miR-3677-3p.

The functions enrichment of target genes of hsa-miR-132-3p, hsa-miR-139-5p and hsa-miR-500a-3p show that these target genes will affect the immune response regulation, protein phosphorylation modification and degradation metabolism, as well as cell proliferation and migration in HCC patients. It has been reported that hsa-miR-132-3p activates PI3K signals in T cells during the activation of T cells to establish precise immune response[29]. hsa-miR-139-5p and hsa-miR-500a-3p affect cell proliferation, migration and invasion by regulating multiple pathways. Due to the expression of hsa-miR-139-5p has an important influence on the development of HCC cells, it can also be used as an HCC inhibitor. Furthermore, the abnormal expression of hsa-miR-139-5p inhibits Akt phosphorylation[30]. These researches show that these three miRNAs not only can serve as biomarkers for the HCC prognosis, but may also serve as therapeutic strategies for HCC.

Four miRNAs were identified and combined into a risk score model in our study, which confirmed their important role in the prognosis of HCC through a variety of bioinformatics methods. At the same time, a large number of related articles have been retrieved to prove the validity of the model. After clinical data analysis, the model can also be used as an independent prognostic factor. All the above results indicate that the risk score model can be used as a prognostic combined biomarker for HCC, and provides effective prognostic information and clinical application value for HCC patients. In commercial application, *in vivo* detection technology can be used to obtain HCC cell tissues and detect the expression of these four miRNAs. The expression of 4 miRNAs was substituted into the formula to calculate the risk score, and the survival time and risk degree of patients were predicted accordingly.

5 Conclusion

Four prognostic miRNA biomarkers of HCC were identified and combined into a risk score model. The relationship between risk score and survival time of HCC patients was verified. And the risk score can be calculated by using the expression of the four miRNAs to evaluate any HCC patients, and providing useful prognostic information.

References

[1] Freddie B, Jacques F, Isabelle S, *et al*. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 2018, **68**(6): 394-424

[2] Nobuhiro T, Sawada Y, Endo I, *et al*. Biomarkers for the early diagnosis of hepatocellular carcinoma. World J Gastroenterol, 2015, **21**(37): 10573-10583

[3] Chen X, Zhang Q, Ma W J, *et al*. The abnormal expression of microRNA-542-3p in hepatocellular carcinoma and its clinical significance. Dis Markers, 2018, **2018**: 1-9

[4] Shao P, Sun D, Wang L, *et al*. Deep sequencing and comprehensive expression analysis identifies several molecules potentially related to human poorly differentiated hepatocellular carcinoma. FEBS Open Bio, 2017, **7**(11): 1696-1706

[5] Huang F Y, Wong K H, Tsui W M, *et al*. Targeted genomic profiling identifies frequent deleterious mutations in FAT4, and TP53, genes in HBV-associated hepatocellular carcinoma. BMC Cancer, 2019, **19**(1): 789-799

[6] Vilchez V. Targeting Wnt/β -catenin pathway in hepatocellular carcinoma treatment. World J Gastroenterol, 2016, **22**(2): 823-832

[7] Gedaly R, Angulo P, Hundley J, *et al*. PKI-587 and sorafenib targeting PI3K/AKT/mTOR and Ras/Raf/MAPK pathways synergistically inhibit HCC cell proliferation. J Surg Res, 2012, **176**(2): 542-548

[8] Ren W, Wu S, Wu Y, *et al*. MicroRNA-196a/-196b regulate the progression of hepatocellular carcinoma through modulating the JAK/STAT pathway *via* targeting SOCS2. Cell Death Dis, 2019, **10**(5): 333-345

[9] Bellingham S A, Shambrook M, Hill A F. Quantitative analysis of exosomal miRNA *via* qPCR and digital PCR. Methods Mole Biol, 2017, **1545**: 55-70

[10] Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. Ann Transl Med, 2016, **4**(2): 30-35

[11] Varet H, Brillet-Guéguen L, Coppée, J Y, *et al*. SARTools: a DESeq2- and edgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. Plos One, 2016, **11**(6): e0157022

[12] Wang L, Feng Z, Wang X, *et al*. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinf, 2010, **26**(1): 136-138

[13] Chou C H, Chang N W, Shrestha S. MiRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res, 2015, **44**: 239-247

[14] Kohl M, Wiese S, Warscheid B. Cytoscape: software for

visualization and analysis of biological networks. Methods Mole Boil, 2011, **696**: 291-303

[15] Chen W, Cai F, Zhang B, *et al*. The level of circulating miRNA-10b and miRNA-373 in detecting lymph node metastasis of breast cancer: potential biomarkers. Tumor Biol, 2013, **34**(1): 455-462

[16] Slattery M L, Herrick J S, Pellatt D F, *et al*. Site-specific associations between miRNA expression and survival in colorectal cancer cases. Oncotarget, 2014, **7**(37): 60193-60205

[17] Li X, Zhang Y, Zhang H, *et al*. miRNA-223 promotes gastric cancer invasion and metastasis by targeting tumor suppressor EPB41L3. Mol Cancer Res, 2011, **9**(7): 824-833

[18] Fiorino S, Bacchi-Reggiani M L, Visani M, *et al*. microRNAs as possible biomarkers for diagnosis and prognosis of hepatitis B- and C-related hepatocellular carcinoma. World J Gastroenterol, 2016, **22**(15): 3907-3936

[19] Wang G, Dong F, Xu Z, *et al*. microRNA profile in HBV-induced infection and hepatocellular carcinoma. BMC Cancer, 2017, **17**(1): 805-815

[20] Leonov G, Shah K, Yee D, *et al*. Suppression of AGO2 by miR-132 as a determinant of miRNA-mediated silencing in human primary endothelial cells. Int J Biochem Cell Biol, 2015, **69**: 75-84

[21] Lian R, Lu B, Jiao L, *et al*. miR-132 plays an oncogenic role in laryngeal squamous cell carcinoma by targeting FOXO1 and activating the PI3K/AKT pathway. Eur J Pharmacol, 2016, **792**: 1-6

[22] Wen Y, Han J, Chen J, *et al*. Plasma miRNAs as early biomarkers for detecting hepatocellular carcinoma. Int J Cancer, 2015, **137**(7): 1679-1690

[23] Shen K, Mao R, Ma L, *et al*. Post-transcriptional regulation of the tumor suppressor miR-139-5p and a network of miR-139-5p-mediated mRNA interactions in colorectal cancer. FEBS J, 2014, **281**(16): 3609-3624

[24] Li T, Yin J, Yuan L, *et al*. Downregulation of microRNA-139 is associated with hepatocellular carcinoma risk and short-term survival. Oncol Rep, 2014, **31**(4): 1699-1706

[25] Vecchio F D, Gallo F, Marco A D, *et al*. Bioinformatics approach to predict target genes for dysregulated microRNAs in hepatocellular carcinoma: study on a chemically-induced HCC mouse mode. BMC Bioinf, 2015, **16**: 408-418

[26] Ni F, Zhao H, Cui H, *et al*. microRNA-362-5p promotes tumor growth and metastasis by targeting CYLD in hepatocellular carcinoma. Cancer Lett, 2014, **356**(2): 809-818

[27] Jiang C, Long J, Liu B, *et al*. miR-500a-3p promotes cancer stem cells properties *via* STAT3 pathway in human hepatocellular carcinoma. J Exp Clin Canc Res, 2017, **36**(1): 99-112

[28] Servais F A, Kirchmeyer M, Hamdorf M, *et al*. Modulation of the IL-6-signaling pathway in liver cells by miRNAs targeting gp130, JAK1, and/or STAT3. Mol Ther Nucl Acids, 2019, **16**: 419-433

[29] Gutiérrez-Vázquez C, Rodríguez-Galán A, Fernández-Alfara M, *et al*. miRNA profiling during antigen-dependent T cell activation: a role for miR-132-3p. Sci. Rep, 2017, **7**(1): 3508-3516

[30] Wang Z, Ding Q, Li Y, *et al*. Reanalysis of microRNA expression profiles identifies novel biomarkers for hepatocellular carcinoma prognosis. Tumor Biol, 2016, **37**(11): 1-9

肝癌预后miRNA风险评分模型的鉴定和分析*

门婧睿　谭建军**　孙洪亮
（北京工业大学生命科学与生物工程学院，北京 100124）

摘要　肝细胞癌（hepatocellular carcinoma，HCC）是世界上高发病率和高死亡率的恶性肿瘤之一. 研究目的是寻找HCC相关的miRNA预后生物学标志物，预测HCC患者的风险程度和生存时间，为他们提供有效的预后信息. 使用4种方法从TCGA中识别差异表达的miRNAs （DEMs）. 并用Kaplan-Meier生存曲线、单因素和多因素Cox回归分析从DEMs中筛选肝癌预后相关的miRNA. 最终4个HCC的预后miRNA生物学标志物（hsa-miR-132-3p、hsa-miR-139-5p、hsa-miR-3677-3p、hsa-miR-500a-3p）被筛选出来组合成一个风险评分模型. 目前还没有实验证据表明组合中的hsa-mir-3677-3p与HCC相关，是本研究新发现的miRNA. 生存曲线、ROC曲线、卡方检验等多种生物信息学方法的评价结果均表明，该模型计算出的风险分值能有效预测患者的风险程度（P<0.000，风险比=2.551，95%置信区间=1.751−3.717）. 低风险组HCC患者1−5年生存率比高风险组高20%−30%. 通过与临床数据分析发现，组合的生物学标志物较其他临床指标相比具有更好的预后效果，也可以作为独立的预后因子. 最后，预测了4种miRNA的靶基因，包括*AGO2*、*FOXO1*、*ROCK2*、*RAP1B*、*CYLD*等，并在细胞增殖、迁移、凋亡、免疫应答等生物学过程中富集.

关键词　miRNA，肝细胞癌，预后，生物学标志物，TCGA
中图分类号　R735.7，R318.04　　**DOI**：10.16476/j.pibb.2019.0286

* 北京市自然科学基金（2202002）和国家自然科学基金（21173014）资助项目.
** 通讯联系人. Tel：010-67392001，E-mail：tanjianjun@bjut.edu.cn
收稿日期：2019-11-26，接受日期：2020-03-13