

Comprehensive analysis of miRNA sequencing profiles identifies novel deregulated and prognostic biomarkers in hepatocellular carcinoma

Hui Zhang

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital <https://orcid.org/0000-0002-5172-4039>

Senmiao Ni

Nanjing Medical University

Changxian Li

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Haoming Zhou

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Jianling Bai

Nanjing Medical University

Xiongiong Pan

Jiangsu Province Hospital and Nanjing Medical University First Affiliated Hospital

Zhengshan Wu (✉ wwzzss355@163.com)

Research

Keywords: miRNA sequencing profiles, HCC, differential expression, prognostic signatures, functional annotation

Posted Date: April 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-20213/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Liver cancer is the fourth most common cause of cancer-related death and rank sixth in terms of incident cases. We aim to identify a set of miRNAs and a miRNA-based signature related to tumorigenesis and prognosis in patients with hepatocellular carcinoma (HCC).

Methods: We analyzed the miRNA sequencing profiles of 373 HCC patients downloaded from The Cancer Genome Atlas LIHC program. The isoform quantification profiles were transformed into 5p and 3p mature miRNA names. Differentially expressed (DE) miRNAs between tumor and adjacent normal tissues were identified by Wald test based on the negative binomial distribution. Prognostic miRNAs associated with overall survival were confirmed by multivariate Cox proportional hazards models. The miRNA-based signatures were obtained from the linear predictors of cox regression, and the prognostic performance was compared by Harrel's C-index and revealed by the restricted mean survival (RMS) curve.

Results: The selected twelve DE miRNAs showed a good performance to classify tumor tissues from normal tissues. Meanwhile, a miRNA-based prognostic signature of eight mature miRNAs was constructed, which significantly stratified patients into high- vs low-risk groups in terms of overall survival (hazard ratio, 4.11; 95% CI, 2.71-6.24; $P < 0.001$). When integrated with clinical information, the composite miRNA-clinical signature showed improved prognostic accuracy relative to the eight-miRNA signature alone. As we set the follow-up time at 5 years, the estimated RMST difference between low- and high-risk group stratified by miRNA index was 1.39 (95% CI: 0.95-1.83) months, which is lesser than the difference between miRNA-clinical risk groups (1.63, 95%CI: 1.20-2.06). Functional enrichment analysis indicated that the target mRNAs of selected miRNAs were mainly enriched in cancer-related pathways and vital cell biological processes.

Conclusions: The proposed DE miRNAs and miRNA-clinical signature are promising biomarkers for discrimination and predicting overall survival respectively in HCC patients. These biomarkers may have significant relevance for development of new drug research and targeting therapies for HCC patients.

Introduction

Liver cancer is predicted to be the sixth most commonly diagnosed cancer and the fourth leading cause of cancer death worldwide in 2018, with about 841,000 new cases and 782,000 deaths annually [1]. With a 5-year survival of 18%, liver cancer is the second most lethal tumor, after pancreatic cancer [2]. Despite improvements in therapeutic strategies of liver cancer over recent years, such as liver resection, transplantation, ablation, embolization and chemotherapy [3], the long-time survival of patients with HCC is not improved due to tumor recurrence and metastasis after liver resection [4]. Exploring the molecular mechanisms related to tumorigenesis and progression of HCC survival is helpful for developing new drug research and targeting therapies in early stage to prolong patients' survival time.

MicroRNAs (miRNAs) are produced through two steps via primary miRNAs (pri-miRNAs) and precursor miRNAs (pre-miRNAs) with cleavage by Drosha and Dicer.[5] The pre-miRNA, approximately 70 nucleotides in length, is composed of a 5p arm, a 3p arm and a terminal loop. Mature sequences of pre-miRNAs are named as miR-#-5p and miR-#-3p, nearly 22 nucleotides in length [6, 7]. They play essential roles across various

biological processes in post-transcriptional regulation, which selectively bind to 3' untranslated region of message RNAs (mRNAs) thus inducing to mRNA degradation or protein translation repression [7]. Several length and/or sequence variants of the same miRNA, which were called isoforms of miRNA (isomiR), are observed by using high-throughput sequencing technology [8]. Various miRNAs have elucidated to be related with kinds of cancers, such as breast cancer [9], non-small cell lung cancer [10], ovarian cancer [11] and liver cancer [12, 13]. Accumulating evidence showed that miRNAs played critical roles in pathological processes of HCC, such as cell proliferation, metastasis, migration, DNA methylation and so on [14].

The available public database, such as The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO), are helpful for data researchers to conduct an in-depth analysis of identifying potential biomarkers across various cancer types [15, 16]. Recent study showed that a six-miRNA signature as an independent factor for predicting recurrence and outcome of patients with HCC was identified based on precursor miRNA profiles [17]. Furthermore, a miRNA combination was established as a promising tool for HCC diagnosis [18]. However, an isomiR-based signature related to tumorigenesis and patients' prognosis of HCC has not been published.

In this study, we aim to identify a class of mature miRNAs that might serve as clinical biomarkers of HCC discrimination and survival outcome from a perspective of isomiR profiling. Firstly, we identified the differentially expressed (DE) miRNAs between tumor and normal tissues with the miRNA sequencing data of HCC from TCGA. Then we constructed a miRNA-based signature that effectively distinguish patients into a high- or low- risk. This signature might be helpful in guiding the development of new drug research and more effective individualized therapies, and improving survival outcome of patients with HCC.

Materials And Methods

Data processing and normalization

Forty hundred and twenty-five miRNA expression profiles and their clinical information on 373 cases with liver cancer were downloaded from the Cancer Genome Atlas (TCGA) data portal (<https://tcga-data.nci.nih.gov/>). The miRNA sequencing (miRNAseq) data were generated by Canada's Michael Smith Genome Sciences Centre (GSC) at the British Columbia (BC) Cancer Agency between 2010 and 2015. After checking the annotations, 13 profiles of 10 patients were deleted as they are not hepatocellular cases. One profile of non-matched normal tissue and three profiles of recurrent tumor tissues were also excluded. Raw read counts of mature miRNAs annotated MIMAT accession IDs were derived from isoforms quantification profiles with self-defined Python codes. Then we translated 2166 MIMATs into 5p and 3p mature strand names and removed 14 dead miRNAs using miRBase version 22.1 (<http://www.mirbase.org/>). We generated a $\log_2(\text{RPM} + 1)$ normalized data, first normalizing read counts of each mature miRNAs to total reads of per million mapped reads, then scaling to $\log_2(\text{RPM} + 1)$.

Differential Expression Analysis

For the comparison of primary tumors and adjacent normal tissues, the differentially expressed (DE) miRNAs were identified with DESeq2 (release 3.9) package, which was based on the assumption that the read counts of deep sequencing data follow the negative binomial distribution [19]. Also, paired t-test was used to the

$\log_2(\text{RPM} + 1)$ normalized data and p-values were adjusted with false discovery rate (FDR) [20]. The significantly DE miRNAs were defined as $|\text{mean difference between paired tissues}| \geq 2$ and FDR adjusted p-value < 0.05 . Considering the abundance of miRNAs varied greatly in tumor tissues, we select those DE miRNAs with a median tumor expression (362 samples) greater than or equal to 20 RPM for drawing a heatmap. DE miRNAs (rows) and paired tissues (columns) were clustered using unsupervised hierarchical clustering with Euclidean distance as distance measure and Ward's hierarchical clustering as clustering method [21].

Construction of a prognostic miRNA signature and a miRNA-clinical signature

Prognostic miRNAs were selected using Cox univariate analyses to assess the association between each mature miRNAs and patients' overall survival. To ensure the robustness of results for great varies in tumor expression, prognostic miRNAs with a p-value less than 0.05 and a median tumor expression greater than or equal to 20 RPM were candidates to build the miRNA risk index. To future reduce the prognostic miRNAs, we applied the Cox proportional hazards regression model combined with the least absolute shrinkage and selection operator (LASSO, glmnet package, version 3.0–2) [22]. The penalty parameter was estimated by 10-fold cross-validation set at the minimum partial likelihood deviance [23]. Final model was constructed with only those statistically significant miRNAs using multivariate Cox stepwise regression method.

The miRNA risk index was achieved from the linear predictors of cox regression with multiple prognostic miRNAs. To separate patients into low- and high-risk groups, the optimal cutoff was determined by a time-dependent receiver operating characteristic (ROC) curve (survivalROC package, version 1.0.3) at 5 years [24]. We used the nearest neighbor estimation method to estimate the survival ROC curve [25]. Here we chose zero instead of the optimal value as a cutoff value because the optimal value (0.13) was close to zero and it would infinitely approach to zero as the sample size increased.

The prognostic miRNA signature was assessed in patients with different groups of clinical factors, including gender, age group, race, treatment type and tumor stage. Then we integrated miRNA risk index with the clinical features to a composite miRNA-clinical signature in a multivariate cox model. The cutoff value for miRNA-clinical signature was set as zero other than the optimal value (-0.034) to separate patients into low- and high-risk. The prognostic performance of the miRNA-clinical signature was compared with that of the miRNA signature in terms of C-index and revealed by the restricted mean survival (RMS) curve. RMS represents the life expectancy at 10 years for patients with different risk. A higher RMS time ratio corresponds to a larger prognostic difference.

Functional Annotation And Enrichment Analysis

Potential gene target of DE miRNAs and prognostic miRNAs were obtained from two bioinformatics algorithm prediction database (targetScan v7.2 [26] and miRDB v6.0 [27]) and one experimentally validated miRNA-target interactions database (miRTarBase v7.0 [28]). Gene2ensembl database was used to perform gene id conversion. At first, the potential targets of miRNAs were separately obtained from three databases. Those targets predicted by all the three databases were kept, which ensured each mRNA target with experimental evidence. The biological processes of gene ontology and KEGG pathway enrichment analyses

were performed with DAVID (Database for Annotation, Visualization and Integrated Discovery) Bioinformatics Tool (<https://david.ncifcrf.gov/>, version 6.8) [29]. Both biological processes of gene ontology and KEGG pathway with $P < 0.10$ were considered as significantly enriched function annotation.

Statistical analysis

All statistical analyses were performed using R (<http://www.r-project.org/>, version 3.6.1). Differential expression analysis was performed with DESeq2 (version 3.9) packages and paired t-test. Univariate analyses and multivariate analysis of the association of miRNAs and other clinical factors with overall survival were performed with the Cox proportional hazards regression model. The Harrel's C-index was calculated and compared with compare (version 1.3.1). The RMS curve and RMS time ratio were estimated with survival (version 3.0) [30] and survRM2 (version 1.0–2) [31] packages. Statistical significance was defined as $P < 0.05$ unless specified otherwise.

Results

Differentially expressed miRNAs between normal and tumor tissues

To systematically identify mature miRNAs related to tumorigenesis and prognosis, approximately 2.0 billion reads for 408 samples from 362 HCC patients were sequenced using miRNA sequencing (miRNAseq). The flowchart of this study was shown in Fig. 1. After translating MIMAT accession IDs into 5p and 3p mature strand names using miRBase v22.1, 2152 mature miRNAs were available. For the 46 paired tumor and adjacent normal tissues, 12 miRNAs were differentially expressed (DE), including four upregulated and eight downregulated. The results of DESeq2 for raw read counts and paired t-test for $\log_2(\text{RPM} + 1)$ transformed data were very similar to each other. The expression abundance of the selected 12 miRNAs (miR-199a-5p, miR-199a-3p, miR-139-5p, miR-10b-5p, miR-182-5p, miR-183-5p, miR-130a-3p, miR-424-5p, miR-451a, miR-199b-3p, miR-1269a, miR-675-3p) in normal and tumor tissues were listed in Table 1 and Figure S1. The discriminant abilities of the DE miRNAs were represented as the area under the ROC curve (AUC), ranging from 0.77 to 0.97, which showed a good performance to identify tumor tissues from normal tissues. Unsupervised hierarchical clustering of 12 DE miRNAs was shown in Fig. 2. There were only three of 92 samples misclassified, one normal tissue and two tumor tissues.

Table 1

The differentially expressed miRNAs in paired tumor and adjacent normal tissues of 46 patients with hepatocellular carcinoma.

Mature miRNAs	Tumor tissues (n = 46)		Paired normal tissues (n = 46)		$\log_2(\text{FC})$	P_{adj}	AUC (95% CI)
	Read counts	Normalized	Read counts	Normalized			
	Median (Range)	Mean \pm SD	Median (Range)	Mean \pm SD			
hsa-miR-1269a	825.5(2-21610)	6.46 \pm 4.04	14.5(0-870)	2.67 \pm 2.06	3.79	1.72E-29	0.77 (0.67, 0.87)
hsa-miR-10b-5p	98561.5(2042-920864)	14.14 \pm 1.64	7641(1790-104797)	10.58 \pm 1.12	3.56	2.45E-52	0.95 (0.91, 1.00)
hsa-miR-183-5p	6746.5(108-264988)	10.64 \pm 2.29	1333(398-6773)	8.12 \pm 0.81	2.52	5.42E-38	0.86 (0.77, 0.95)
hsa-miR-182-5p	17583(461-483585)	12.05 \pm 2.23	4809(1802-22784)	10.05 \pm 0.76	2.00	3.83E-30	0.81 (0.71, 0.91)
hsa-miR-451a	1409.5(233-126308)	8.49 \pm 1.44	8277(1400-94069)	10.56 \pm 0.99	-2.07	1.18E-15	0.91 (0.84, 0.97)
hsa-miR-130a-3p	123(15-727)	4.70 \pm 1.02	606.5(261-1636)	6.86 \pm 0.47	-2.16	6.64E-29	0.96 (0.91, 1.00)
hsa-miR-424-5p	493(79-7092)	7.00 \pm 1.03	3248(674-12509)	9.19 \pm 0.77	-2.19	2.06E-18	0.95 (0.90, 1.00)
hsa-miR-139-5p	289(48-1803)	6.16 \pm 1.14	1879.5(798-5058)	8.49 \pm 0.46	-2.33	4.11E-19	0.97 (0.93, 1.00)
hsa-miR-199b-3p	1845(104-40470)	8.82 \pm 1.89	12028.5(4942-47804)	11.31 \pm 0.52	-2.49	6.26E-07	0.89 (0.82, 0.96)

*Those differentially expressed miRNAs with a median tumor expression (n = 362) greater than or equal to 20 RPM were listed above. Normalized expression data means $\log_2(\text{RPM} + 1)$; RPM was a data normalization as counts normalized to reads per million mapped reads. SD stands for standard deviation; FC stands for fold change between tumor and adjacent normal tissues, and here $\log_2(\text{FC})$ is calculated as the mean difference of normalized data between paired tissues. P-values were derived from Wald test based on the negative binomial distribution (Bioconductor package DESeq2, release 3.9). AUC (95%CI) were obtained from modeling univariate logistic regression, showing the discriminate accuracy of each miRNA.

Mature miRNAs	Tumor tissues (n = 46)		Paired normal tissues (n = 46)		$\log_2(\text{FC})$	P_{adj}	AUC (95% CI)
	Read counts	Normalized	Read counts	Normalized			
	Median (Range)	Mean \pm SD	Median (Range)	Mean \pm SD			
hsa-miR-199a-3p	1852(105-40540)	8.82 \pm 1.89	12053.5(4947-47891)	11.31 \pm 0.52	-2.49	6.22E-07	0.89 (0.82, 0.96)
hsa-miR-199a-5p	769.5(34-12695)	7.53 \pm 1.95	5604.5(1408-30839)	10.09 \pm 0.72	-2.56	1.68E-07	0.87 (0.80, 0.95)
hsa-miR-675-3p	105.5(5-25274)	5.27 \pm 2.45	1728(446-9085)	8.39 \pm 0.97	-3.12	8.69E-05	0.87 (0.78, 0.95)
<p>*Those differentially expressed miRNAs with a median tumor expression (n = 362) greater than or equal to 20 RPM were listed above. Normalized expression data means $\log_2(\text{RPM} + 1)$; RPM was a data normalization as counts normalized to reads per million mapped reads. SD stands for standard deviation; FC stands for fold change between tumor and adjacent normal tissues, and here $\log_2(\text{FC})$ is calculated as the mean difference of normalized data between paired tissues. P-values were derived from Wald test based on the negative binomial distribution (Bioconductor package DESeq2, release 3.9). AUC (95%CI) were obtained from modeling univariate logistic regression, showing the discriminate accuracy of each miRNA.</p>							

Prognostic miRNAs Associated With Overall Survival

A total of 185 mature miRNAs were selected as candidate miRNAs to assess the association between each miRNA and patients' overall survival, removing those miRNAs' RPM less than 20 in tumor expression. Also, the observations were deleted if their overall survival (OS) time were missing (n = 7) or less than 30 days (n = 20). Therefore, 335 patients with 118 meeting death events were included in the survival analysis. With the $\log_2(\text{RPM} + 1)$ transformed data, univariate cox regression identified 65 miRNAs with statistical significance. As shown in Table 2 and Figure S2, 16 miRNAs were selected using LASSO method with the seed number as 10101 and lambda as the minimum partial likelihood deviance (0.04221). The miRNA risk index was calculated from multivariate Cox regression model with 8 miRNAs (let-7a-3p, miR-148a-3p, miR-151a-3p, miR-210-3p, miR-215-5p, miR-361-5p, miR-582-3p, miR-9-5p) at a significant level. Zero was used as a cutoff value to separate the patients into high- and low-risk groups (Figure S3).

Table 2

The prognostic miRNAs associated with HCC long-term survival selected by LASSO method and multivariate Cox regression. (N = 335, death events = 118)

miRNA	Univariate analysis			Multivariate analysis		
	Coef.	HR (95%CI)	P	Coef.	HR (95%CI)	P
hsa-let-7a-3p	0.3661	1.44 (1.10,1.88)	7.21E-03	0.3792	1.46 (1.12,1.90)	4.96E-03
hsa-miR-10b-5p	0.1101	1.12 (1.01,1.23)	2.91E-02			
hsa-miR-122-5p	-0.1489	0.86 (0.78,0.95)	2.29E-03			
hsa-miR-1269a	0.0502	1.05 (1.01,1.10)	2.93E-02			
hsa-miR-139-5p	-0.3361	0.71 (0.62,0.82)	3.76E-06			
hsa-miR-148a-3p	-0.3358	0.71 (0.60,0.85)	2.09E-04	-0.2999	0.74 (0.60,0.91)	4.45E-03
hsa-miR-151a-3p	0.4551	1.58 (1.24,2.01)	2.23E-04	0.3689	1.45 (1.11,1.89)	7.01E-03
hsa-miR-15b-3p	0.5188	1.68 (1.38,2.04)	1.62E-07			
hsa-miR-186-5p	0.5597	1.75 (1.33,2.31)	7.74E-05			
hsa-miR-210-3p	0.2573	1.29 (1.17,1.43)	6.33E-07	0.1805	1.20 (1.08,1.33)	9.90E-04
hsa-miR-215-5p	0.1272	1.14 (1.01,1.27)	2.79E-02	0.1555	1.17 (1.03,1.33)	1.81E-02
hsa-miR-22-3p	-0.5706	0.57 (0.44,0.73)	1.68E-05			
hsa-miR-328-3p	0.2227	1.25 (1.01,1.54)	3.67E-02			
hsa-miR-361-5p	0.5102	1.67 (1.24,2.24)	6.92E-04	0.4731	1.60 (1.11,2.33)	1.24E-02
hsa-miR-582-3p	0.2796	1.32 (1.16,1.50)	2.18E-05	0.1358	1.15 (1.00,1.31)	4.61E-02

*With the LASSO method used, 16 miRNAs were selected when lambda set as the minimum partial likelihood deviance 0.042 and seed number was 10101. In final multivariate models, 8 statistically significant miRNAs were selected using Cox stepwise regression. Concordance = 0.741 (se = 0.025). The eight-miRNA risk index score was derived from the linear predictors of multivariate Cox model based on eight-miRNA.

	Univariate analysis			Multivariate analysis		
hsa-miR-9-5p	0.1674	1.18 (1.10,1.27)	1.37E-05	0.1312	1.14 (1.05,1.24)	1.82E-03
*With the LASSO method used, 16 miRNAs were selected when lambda set as the minimum partial likelihood deviance 0.042 and seed number was 10101. In final multivariate models, 8 statistically significant miRNAs were selected using Cox stepwise regression. Concordance = 0.741 (se = 0.025). The eight-miRNA risk index score was derived from the linear predictors of multivariate Cox model based on eight-miRNA.						

The hazard ratio of the eight-miRNA signature between high-risk and low-risk group was 4.11(95% CI: 2.71–6.24) with statistically significance ($P < 0.001$). The relation between eight-miRNA risk group and other clinical factors was assessed using univariate cox regression and the results were shown in Fig. 3. In early and mid-term (stage I and stage II), the Kaplan-Meier curves were significantly separated from each other; otherwise, the curves for stage III were somewhat crossed although the p -value < 0.05 (Figure S4).

Integrated prognostic miRNA-clinical risk index by combining the miRNA risk index with clinical factors

In multivariate analysis, clinical features and eight-miRNA risk index were independent factors. To further improve accuracy, we derived a miRNA-clinical risk index by combining important clinical features (gender, age group, race and tumor stage) and eight-miRNA risk index (Table S1). If we set the follow-up time at 5 years, the estimated RMST difference between low- and high-risk group stratified by eight-miRNA risk index was 1.39 (95% CI: 0.95–1.83) months, which is lesser than the difference between miRNA-clinical risk groups (1.63, 95%CI: 1.20–2.06) (Table S2). Similar results were observed in the Kaplan-Meier curves for the two risk groups derived from the prognostic miRNA risk index and miRNA-clinical risk index (Fig. 4). We also plotted the restricted mean survival curve for the two signatures, the estimation of overall survival was significantly improved (mean C-index, 0.74 vs 0.77; $P = 0.0067$) (Fig. 5). Additionally, the accuracy of prediction of five-year survival was slightly increased while the clinical factors were combined with the eight-miRNA risk index (AUC of predicting the five-year survival, 0.808 vs 0.823) (Figure S5).

Functional Annotation For Predicted Target mRNAs Of Selected miRNAs

A total of 13409 potential target mRNAs of 12 DE mature miRNAs were firstly identified from three miRNA-target interactions databases. Then 279 target mRNAs predicted by all the three databases were selected. A functional enrichment analysis of these target mRNAs showed 181 GO biological processes and 61 KEGG pathways with significant enrichment. The top 15 enriched GO biological processes terms and KEGG pathways for DE miRNAs targeted mRNAs are shown in Fig. 6A-B. Our results indicated that the twelve miRNAs played an important role in cancer-related pathways and cell biological functions, such as transcription from RNA polymerase II promoter, MAPK signaling pathway, focal adhesion.

Similarly, a total of 12736 potential targets of eight prognostic miRNAs were identified and 266 target mRNAs were finally selected. A functional enrichment analysis of these target mRNAs showed 161 significant enriched GO biological processes and 17 significant enriched KEGG pathways. The top 15 enriched GO biological processes terms and KEGG pathways for HCC prognosis-related miRNAs targeted genes are shown

in Fig. 6C-D. These results demonstrated that the eight prognostic miRNAs might be associated with cell proliferation, apoptotic process and TGF- β signaling pathway [28]. Our results support use of this eight-miRNA signature as an independent factor for predicting the overall survival of patients with HCC.

Discussion

Patients with HCC are at substantially high risk for recurrence and death, even after surgical resection. A prognostic signature can help oncologists to proceed systematic therapy on patients at a high risk as early as possible. In this study, based on the isomiR profiles, we identified twelve differentially expressed miRNAs related to tumorigenesis through 46 paired tissues, then a miRNA-clinical signature, which was derived from multivariate Cox regression by combining eight-miRNA risk index with several clinical factors, significantly stratified the high-risk to the low-risk patients with HCC. These results prompted that the miRNA screened in this study have potential value for the discrimination and prognosis of hepatocellular carcinoma, and might be useful for the development of novel drug research and individualized treatment of HCC.

It is widely accepted that a series of molecular pathways and biological processes are involved in the pathogenesis and progression of HCC. The miR-199 family members (miR-199a-3p, miR-199a-5p and miR-199b-3p in our study) have been reported to under-expression in HCC cells, promoting HCC cell proliferation, increasing invasion and inhibiting apoptosis with different signaling pathway [32, 33]. The upregulation of miR-10b-5p promotes cell invasion and metastasis in HCC by suppressing tumor suppressor genes, such as cell adhesion molecule 1 (CADM1) and matrix metalloproteinases (MMPs) [34, 35]. Qiu et al. indicated that miR-139-5p functions as a suppressor for epithelial-mesenchymal transition (EPT) and metastasis of HCC by targeting ZEB1 and ZEB2 genes [36]. Recent study shows that miR-182-5p directly inhibited the expression of FOXO3a, thus activates AKT/FOXO3a pathway to promote HCC progression [37]. However, miR-361-5p inhibits HCC cell proliferation, migration and invasion by targeting CXCR6 directly, which is inconsistent with ours [38].

Besides, the mechanism of one DE miRNA (miR-675-3p) and three prognostic miRNAs (let-7a-3p, miR-148a-3p, and miR-151a-3p) in our results is rarely reported. The DE analysis demonstrated that miR-675-3p expression is inhibited significantly in tumor tissues, suggesting as a suppressor in tumorigenesis. With the Cox model used, we found that let-7a-3p and miR-151a-3p may be the risk factors for patients' overall survival ($HR > 1$, $P < 0.05$) while miR-148a-3p may be a protective factor ($HR < 1$, $P < 0.05$). Unlike to Bai et al. [17], our data mining were based on the isoform quantification profiles other than the pre-miRNA expression data. The analysis strategy of Bai et al. is first identifying the prognostic miRNA, then using two-sided Student's t-test to determine significance differences between tumor and nontumor tissues. In fact, the aim of differential expression analysis is to select those miRNAs that might be helpful for cancer diagnosis, and the miRNA-based signature is related to cancer prognosis, which are two separate things. In our study, there might be some limitations. Another independent database should be used to validate our proposed miRNA-based signatures. Functional experiments are required to provide a more convincing explanation for the biological implications of these DE miRNAs and prognostic miRNAs in HCC. Furthermore, prospective studies are needed to further validate its predictive accuracy for estimating prognosis and to test its clinical utility in individualized management of HCC.

In summary, our proposed differentially expressed miRNAs and miRNA-clinical signature are promising discriminate and prognostic biomarkers in HCC patients, especially in the early- and mid-term of tumor progression. These signatures may have significant relevance for development of new drug research and targeting therapies as well as prognostic prediction for patients with HCC.

Declarations

Acknowledgement

This study was supported by National Science and Technology Major Project of China (2017ZX10203207-004-004), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant No. 17KJB330003).

Disclosure Statement

The authors who have taken part in this study declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

Author Contributions

Z.H. and N.S.M. conceived and designed the study. Z.H. and N.S.M did data preparing, statistical analysis and drafted the manuscript. L.C.X, Z.H.M. and B.J.L provided the critical revision of the manuscript for important intellectual; P.X.X and W.Z.S supervised the study. All authors read and approved the final version of the manuscript.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **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. Villanueva A: **Hepatocellular Carcinoma.** *N Engl J Med* 2019, **380**(15):1450-1462.
3. Mokdad AA, Singal AG, Yopp AC: **Treatment of liver cancer.** *Jama* 2016, **315**(1):100-100.
4. Fan ST, Mau Lo C, Poon RTP, Yeung C, Leung Liu C, Yuen WK, Ming Lam C, Ng KKC, Ching Chan S: **Continuous Improvement of Survival Outcomes of Resection of Hepatocellular Carcinoma.** *Annals of Surgery* 2011, **253**(4):745-758.
5. Lee Y, Jeon K, Lee JT, Kim S, Kim VN: **MicroRNA maturation: stepwise processing and subcellular localization.** *EMBO J* 2002, **21**(17):4663-4670.
6. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function.** *cell* 2004, **116**(2):281-297.
7. Bartel DP: **MicroRNAs: target recognition and regulatory functions.** *cell* 2009, **136**(2):215-233.
8. Neilsen CT, Goodall GJ, Bracken CP: **IsomiRs—the overlooked repertoire in the dynamic microRNAome.** *Trends Genet* 2012, **28**(11):544-549.

9. Shimono Y, Zabala M, Cho RW, Lobo N, Dalerba P, Qian D, Diehn M, Liu H, Panula SP, Chiao E *et al*: **Downregulation of miRNA-200c Links Breast Cancer Stem Cells with Normal Stem Cells.** *Cell* 2009, **138**(3):592-603.
10. Hu Z, Chen J, Tian T, Zhou X, Gu H, Xu L, Zeng Y, Miao R, Jin G, Ma HJTJoci: **Genetic variants of miRNA sequences and non–small cell lung cancer survival.** *The Journal of clinical investigation* 2008, **118**(7):2600-2608.
11. Dahiya N, Sherman-Baust CA, Wang T-L, Davidson B, Shih I-M, Zhang Y, Wood III W, Becker KG, Morin PJJPo: **MicroRNA expression and identification of putative miRNA targets in ovarian cancer.** *PloS one* 2008, **3**(6).
12. Zhang S, Chen L, Wang Y, Tang W, Chen Y, Liu L: **Investigation of the Association of miRNA-499, miRNA-146a, miRNA-196a2 Loci with Hepatocellular Carcinoma Risk: A Case–Control Study Involving 1507 Subjects.** *DNA and Cell Biology* 2020.
13. Ji J, Shi J, Budhu A, Yu Z, Forgues M, Roessler S, Ambs S, Chen Y, Meltzer PS, Croce CM *et al*: **MicroRNA Expression, Survival, and Response to Interferon in Liver Cancer.** *N Engl J Med* 2009, **361**(15):1437-1447.
14. Ambros V: **The functions of animal microRNAs.** *Nature* 2004, **431**(7006):350-355.
15. Shen S, Bai J, Wei Y, Wang G, Li Q, Zhang R, Duan W, Yang S, Du M, Zhao Y *et al*: **A seven-gene prognostic signature for rapid determination of head and neck squamous cell carcinoma survival.** *Oncol Rep* 2017, **38**(6):3403-3411.
16. Li B, Cui Y, Diehn M, Li R: **Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer.** *JAMA Oncol* 2017, **3**(11):1529-1537.
17. Bai F, Zhou H, Ma M, Guan C, Lyu J, Meng QH: **A novel RNA sequencing-based mi RNA signature predicts with recurrence and outcome of hepatocellular carcinoma.** *Molecular oncology* 2018, **12**(7):1125-1137.
18. Ji J, Chen H, Liu X-P, Wang Y-H, Luo C-L, Zhang W-W, Xie W, Wang F-B: **A miRNA combination as promising biomarker for hepatocellular carcinoma diagnosis: a study based on bioinformatics analysis.** *Journal of Cancer* 2018, **9**(19):3435.
19. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome biology* 2014, **15**(12):550.
20. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *Journal of the Royal statistical society: series B* 1995, **57**(1):289-300.
21. Murtagh F, Legendre P: **Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?** *Journal of classification* 2014, **31**(3):274-295.
22. Friedman J, Hastie T, Tibshirani RJRp: **glmnet: Lasso and elastic-net regularized generalized linear models.** *R package version* 2009, **1**(4).
23. Simon N, Friedman J, Hastie T, Tibshirani RJJoss: **Regularization paths for Cox's proportional hazards model via coordinate descent.** *Journal of statistical software* 2011, **39**(5):1.
24. Heagerty P, Saha PJB: **SurvivalROC: time-dependent ROC curve estimation from censored survival data.** *Biometrics* 2000, **56**(2):337-344.
25. Heagerty PJ, Lumley T, Pepe MSJB: **Time-dependent ROC curves for censored survival data and a diagnostic marker.** *Biometrics* 2000, **56**(2):337-344.

26. Agarwal V, Bell GW, Nam J-W, Bartel DP: **Predicting effective microRNA target sites in mammalian mRNAs.** *elife* 2015, **4**:e05005.
27. Chen Y, Wang X: **miRDB: an online database for prediction of functional microRNA targets.** *Nucleic acids research* 2020, **48**(D1):D127-D131.
28. Chou C-H, Shrestha S, Yang C-D, Chang N-W, Lin Y-L, Liao K-W, Huang W-C, Sun T-H, Tu S-J, Lee W-HJNar: **miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions.** *Nucleic acids research* 2018, **46**(D1):D296-D302.
29. Huang DW, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic acids research* 2009, **37**(1):1-13.
30. Therneau TM, Lumley TJSaPoC: **Package 'survival'.** *Survival analysis Published on CRAN* 2014, **2**:3.
31. Uno H, Tian L, Cronin A, Battiou C, Horiguchi M: **survRM2: comparing restricted mean survival time. R package version 1.0-2. Published 2017.** In.; 2018.
32. Shen Q, Cicinnati VR, Zhang X, Iacob S, Weber F, Sotiropoulos GC, Radtke A, Lu M, Paul A, Gerken G: **Role of microRNA-199a-5p and discoidin domain receptor 1 in human hepatocellular carcinoma invasion.** *Molecular cancer* 2010, **9**(1):227.
33. Ren K, Li T, Zhang W, Ren J, Li Z, Wu G: **miR-199a-3p inhibits cell proliferation and induces apoptosis by targeting YAP1, suppressing Jagged1-Notch signaling in human hepatocellular carcinoma.** *Journal of biomedical science* 2016, **23**(1):79.
34. Li Q-j, Zhou L, Yang F, Wang G-x, Zheng H, Wang D-s, He Y, Dou K-f: **MicroRNA-10b promotes migration and invasion through CADM1 in human hepatocellular carcinoma cells.** *Tumor Biology* 2012, **33**(5):1455-1465.
35. Liao C-g, Kong L-m, Zhou P, Yang X-l, Huang J-g, Zhang H-l, Lu N: **miR-10b is overexpressed in hepatocellular carcinoma and promotes cell proliferation, migration and invasion through RhoC, uPAR and MMPs.** *Journal of translational medicine* 2014, **12**(1):234.
36. Qiu G, Lin Y, Zhang H, Wu D: **miR-139-5p inhibits epithelial–mesenchymal transition, migration and invasion of hepatocellular carcinoma cells by targeting ZEB1 and ZEB2.** *Biochemical and biophysical research communications* 2015, **463**(3):315-321.
37. Cao M-Q, You A-B, Zhu X-D, Zhang W, Zhang Y-Y, Zhang S-Z, Zhang K-w, Cai H, Shi W-K, Li X-L: **miR-182-5p promotes hepatocellular carcinoma progression by repressing FOXO3a.** *Journal of hematology & oncology* 2018, **11**(1):12.
38. Sun J-J, Chen G-Y, Xie Z-T: **MicroRNA-361-5p inhibits cancer cell growth by targeting CXCR6 in hepatocellular carcinoma.** *Cellular Physiology and Biochemistry* 2016, **38**(2):777-785.

Figures

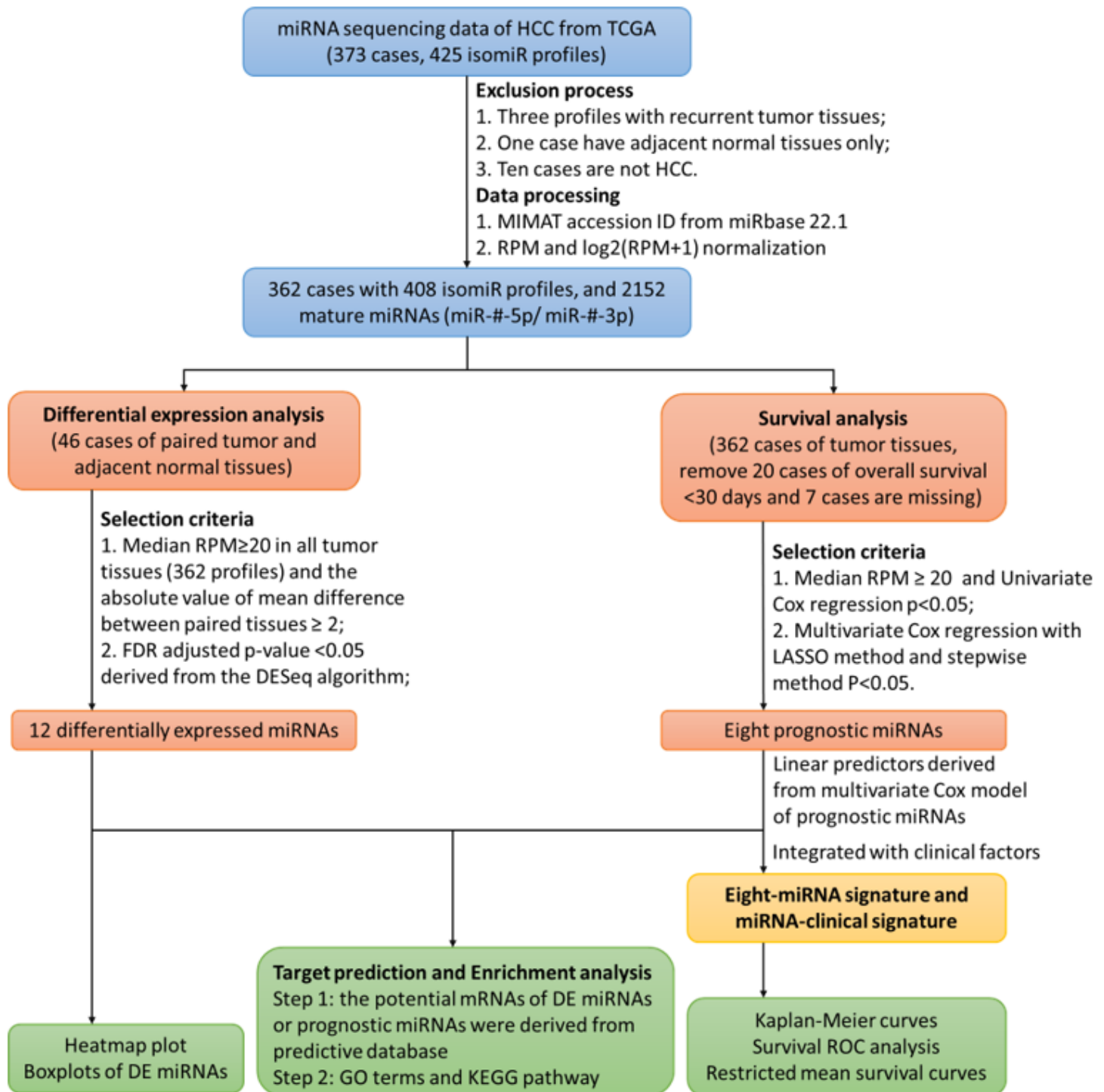


Figure 1

Flowchart of study design.

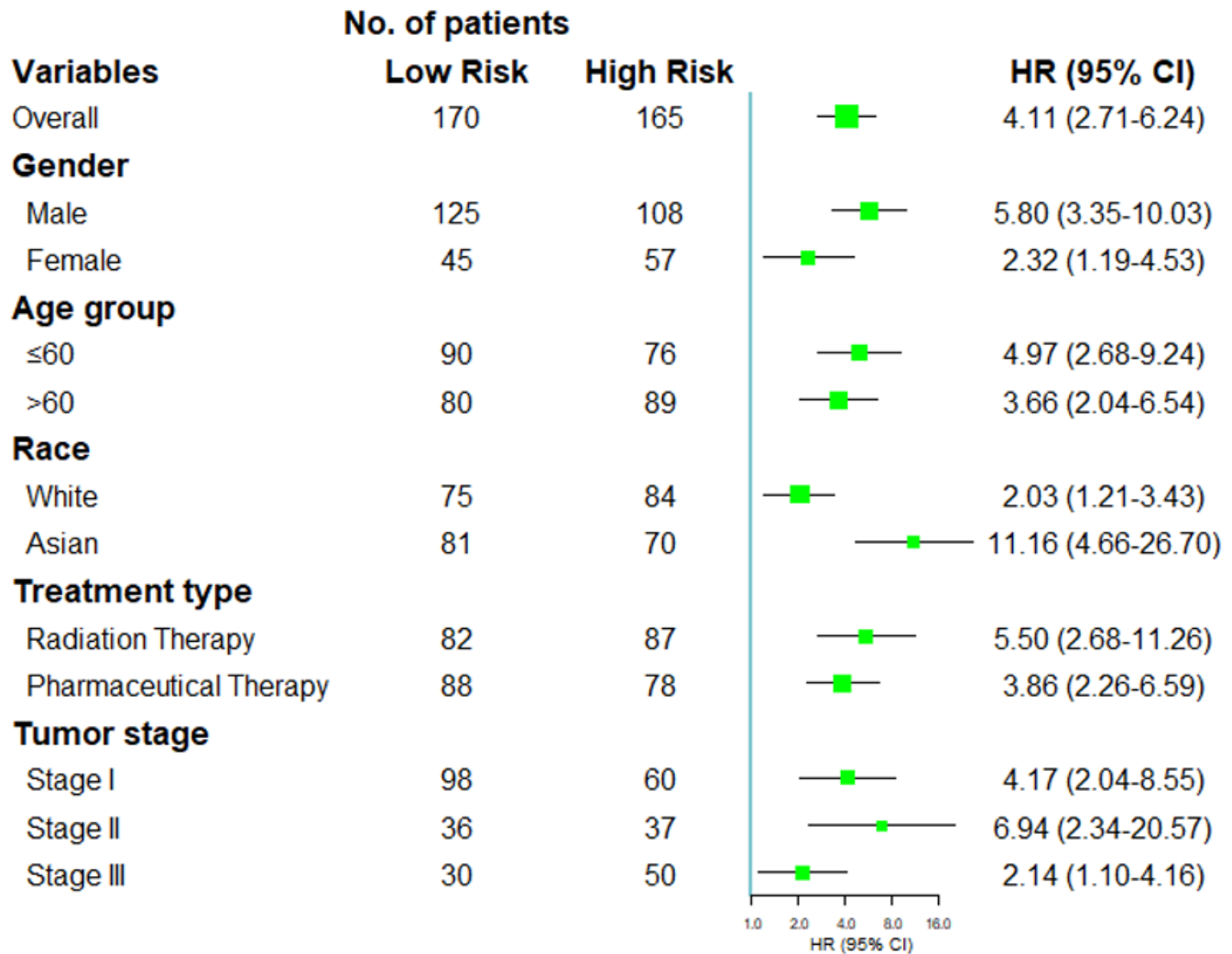


Figure 5

Associations between the prognostic eight-miRNA signature and clinical factors for overall survival of HCC patients.

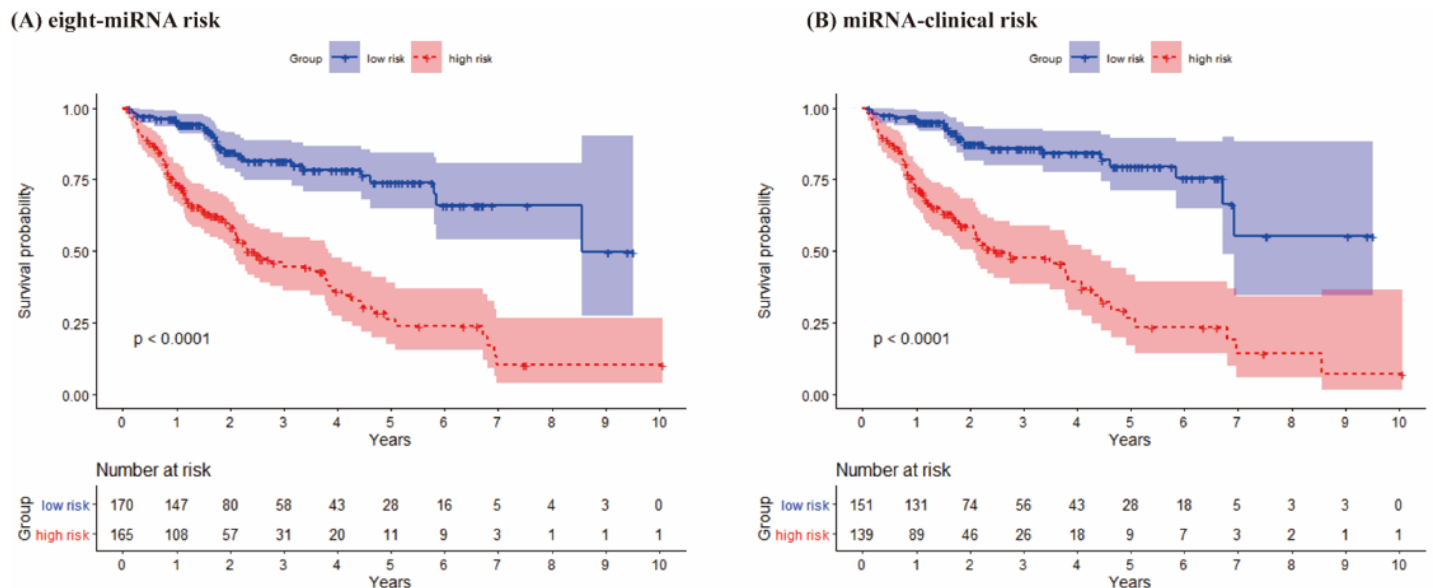


Figure 7

Kaplan-Meier curves of HCC patients' overall survival for the prognostic eight-miRNA signature or miRNA-clinical signature.

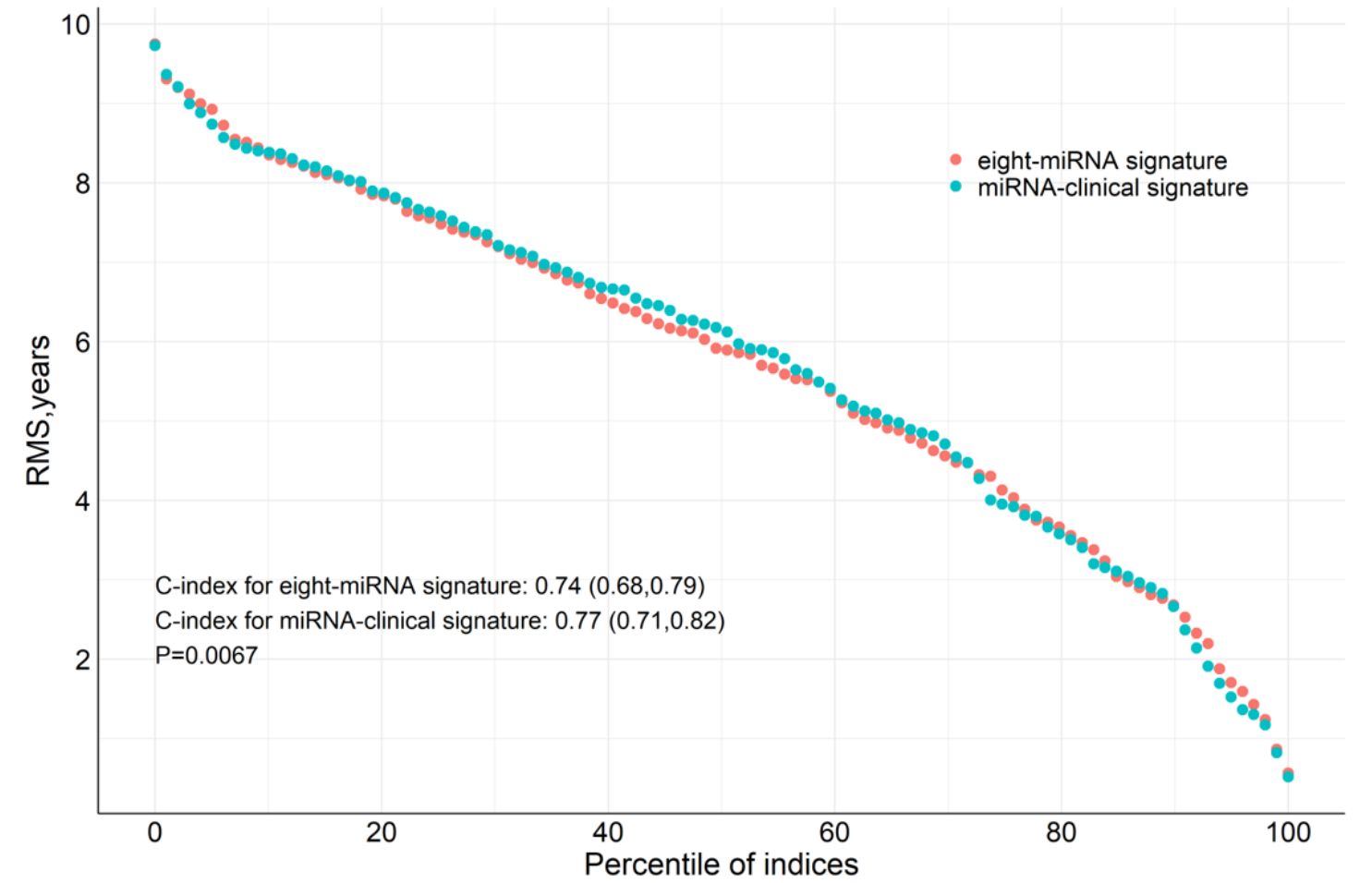


Figure 9

The RMS curves for the prognostic eight-miRNA signature and miRNA-clinical signature.

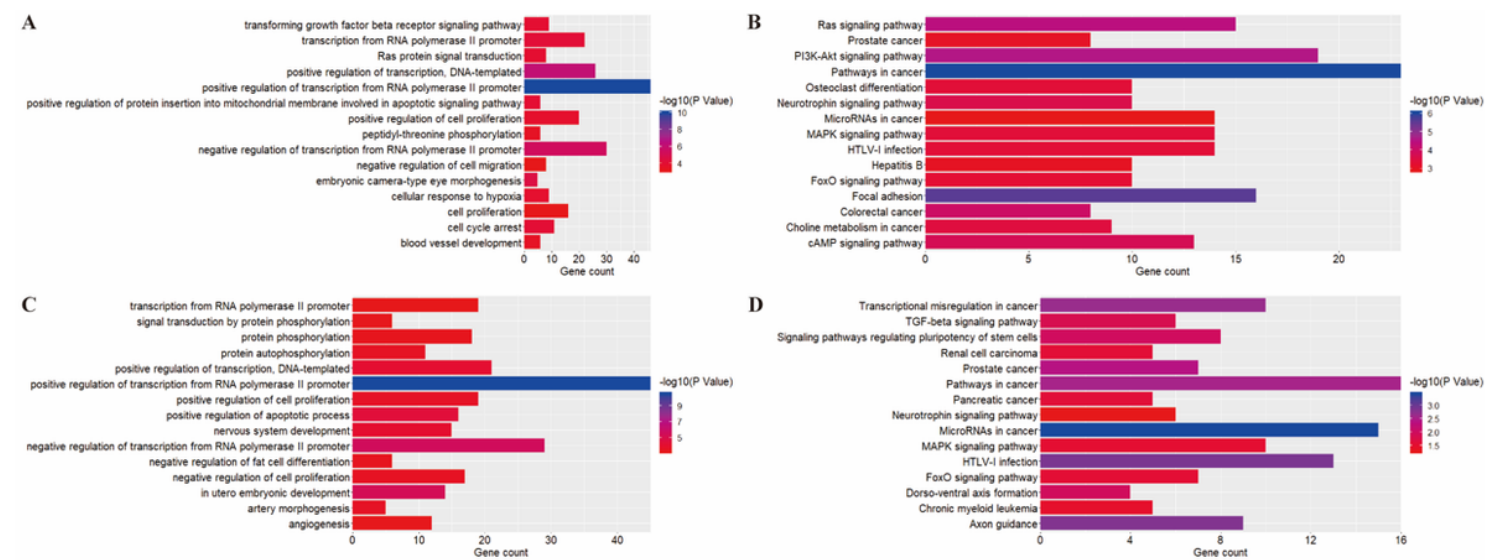


Figure 11

Functional annotation for predicted target mRNAs of DE or prognostic miRNAs. (A) Biological process of GO terms for target mRNAs of 12 DE miRNAs. (B) KEGG pathways for target mRNAs of 12 DE miRNAs. (C) Biological process of GO terms for target mRNAs of 8 prognostic miRNAs. (D) KEGG pathways for target mRNAs of 8 prognostic miRNAs. The x-axis represents the number of genes, and the y-axis represents the GO terms of biological process and KEGG pathway names. The color indicates the P-value.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplemental.docx](#)
- [Supplemental.docx](#)