Original Article

Bioinformatic analysis and *in vitro* validation of a five-microRNA signature as a prognostic biomarker of hepatocellular carcinoma

Wang Li[1#], Xiangshuo Kong[2#], Tao Huang[1,3,4], Lujun Shen[1,3,4], Peihong Wu[1], Qi-Feng Chen[1,3,4]

[1]Department of Medical Imaging and Interventional Radiology, Sun Yat-sen University Cancer Center, Guangzhou, China; [2]Department of Oncology, Yantai Yuhuangding Hospital, Yantai, China; [3]State Key Laboratory of Oncology in South China, Guangzhou, China; [4]Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Contributions: (I) Conception and design: QF Chen, W Li, X Kong, P Wu; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Dr. Qi-Feng Chen; Prof. Peihong Wu. Department of Medical Imaging and Interventional Radiology, Sun Yat-sen University Cancer Center, 651 Dongfeng Road East, Guangzhou 510060, China. Email: chenqf25@sysucc.org.cn; wuph@sysucc.org.cn.

Background: Existing research has identified correlations between numerous microRNAs (miRNAs) and the prognosis of hepatocellular carcinoma (HCC). However, the role of a combination of miRNAs in predicting HCC survival requires further elucidation.

Methods: miRNA expression profiles and clinical data from HCC patients were downloaded from The Cancer Genome Atlas (TCGA). Differentially expressed (DE) miRNAs in tumor versus normal samples were identified. All HCC patients were randomly assigned to a training cohort or a validation cohort at a ratio of 1 to 1. A least absolute shrinkage and selection operator (LASSO) Cox regression model was subsequently employed to establish the miRNA signature. The constructed miRNA signature was then developed and validated.

Results: In total, 127 DE miRNAs were detected between HCC and paracancerous tissue using HCC RNA sequencing (RNA-Seq) data extracted from TCGA database. LASSO Cox regression generated a five-miRNA signature consisting of has-mir-105-2, has-mir-9-3, has-mir-137, has-mir-548f-1, and has-mir-561 in the training cohort. This risk model was significantly related to survival (P=5.682e-6). Log-rank tests and multivariate Cox regression analyses revealed the five-miRNA signature as an independent prognostic indicator [HR =3.285, 95% confidence interval (CI): 1.737–6.213], with the area under curve (AUC) of the miRNA signature being 0.728. The effects of the miRNA signature were further confirmed in the validation cohort and in the OncomiR Cancer Database and Gene Expression Omnibus (GEO) dataset. Functional enrichment analysis revealed the potential effects of the five-miRNA signature in tumor-related biological pathways and processes. Cell Counting Kit-8, Transwell, and wound healing assays, were used to evaluate the role of has-mir-137 in HCC cell proliferation and migration *in vitro*.

Conclusions: We established a novel five-miRNA signature which reliably predicted prognosis in HCC patients and which could be used to assist in both strategic counseling and personalized management in HCC.

Keywords: Hepatocellular carcinoma (HCC); microRNAs (miRNAs); prognostic signature; least absolute shrinkage and selection operator (LASSO)

Submitted Mar 14, 2020. Accepted for publication Aug 28, 2020.
doi: 10.21037/atm-20-2509
View this article at: http://dx.doi.org/10.21037/atm-20-2509

Introduction

Hepatocellular carcinoma (HCC) ranks fifth in terms of cancer mortality in the world, and h morbidity has increased each year (1). Dysregulated gene expression has been extensively demonstrated to exert a vital role in tumorigenesis and cancer progression. Growing evidence has uncovered diagnostic, prognostic, and even therapeutic targets for different types of cancers through the different gene expression profiles. Moreover, some mutant genes have recently been suggested to participate in the carcinogenesis and progression of HCC (2). However, HCC is a highly heterogeneous disease, with large variations in HCC progression and significantly complicated prognostic prediction. Thus, the urgent exploration of effective diagnostic and prognostic biomarkers is required.

MicroRNAs (miRNAs), a group of small noncoding RNA molecules, control about one-third of protein-coding genes via post-transcriptional gene regulation, thereby exerting direct or indirect effects on the majority of cellular pathways (3). The critical biological roles of miRNAs have been validated in previous research, demonstrating miRNAs influence in carcinogenesis, tumor progression, metastasis, and patient survival (4,5). A number of miRNAs have been suggested as possible diagnostic and prognostic biomarkers and promising therapeutic targets (6). Additionally, several miRNA signatures have been constructed for prognostic estimation in HCC patients (7-12). However, the common drawbacks of these studies are that they lack validation, ongoing enrollment patients, discovery of novel miRNAs, and uniformity in processing and analyzing the data. Thus, the current prognostic miRNA signatures still need to be optimized, and the identification of practical and potent miRNA signatures may play a critical role for prognostic prediction in clinical practice.

The present study was designed to establish and verify a miRNA-based classifier by using the least absolute shrinkage and selection operator (LASSO) Cox regression model. The robust miRNA expression-based signature was constructed to improve prognostic prediction of HCC by comprehensively analyzing genomic data. The model miRNA was further confirmed by *in vitro* analysis. We present the study in accordance with the MDAR reporting checklist (available at http://dx.doi.org/10.21037/atm-20-2509).

Methods

Patient datasets and processing

Publicly accessible raw counts of miRNA expression profiles, along with the corresponding clinical data of HCC patients, were downloaded from The Cancer Genome Atlas (TCGA) (https://tcga-data.nci.nih.gov/tcga/, accessed July 2019). The data of 373 HCC patients, comprising 375 HCC cancer samples and 50 adjacent normal samples, were downloaded. The data from one patient were excluded due to a lack of follow-up information. In total, data for 372 HCC patients were randomly categorized into either a training group or a validation group at a ratio of 1 to 1 for comprehensive integrated analysis using the "caret" package. The flow chart is shown in *Figure 1*. This study used open-access data and so did not require additional approval from an ethics committee. Data processing was performed strictly according to the National Institutes of Health (NIH) TCGA Human Subject Protection and Data Access Policies. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Data processing and screening of differentially expressed (DE) miRNAs

The raw count of miRNA from each sample was used to form the expression matrix, and was followed by the normalization of the expression matrix using "edgeR" package derived from R language. |log2 foldchange (FC)| >2.0 along with adjusted P value <0.05 was employed for the identification of DE miRNAs. Heat map and volcano plots were further performed on DE miRNAs for better differentiation between normal and HCC tumor samples.

Construction of the miRNA signature

The correlation expression of each miRNA with patient survival was calculated by univariate Cox model, where a P value <0.05 indicated statistical significance for a specific miRNA. "Survminer" package in R project was used to determine the optimal cutoff miRNA values to investigate the relationship between the miRNAs signature and the prognosis of patients. R software was further used for random number generation to equally categorize patients into the training and validation cohorts. In the training cohort LASSO was conducted using R package "glmnet" for selection of the most valuable predictive miRNAs, and was simultaneously employed to obtain shrinkage and variable selection; meanwhile, 10-time cross-validations were used to determine the optimal values of penalty parameter λ. Moreover, the risk score of the miRNA signature for each patient was calculated according to the expression of every

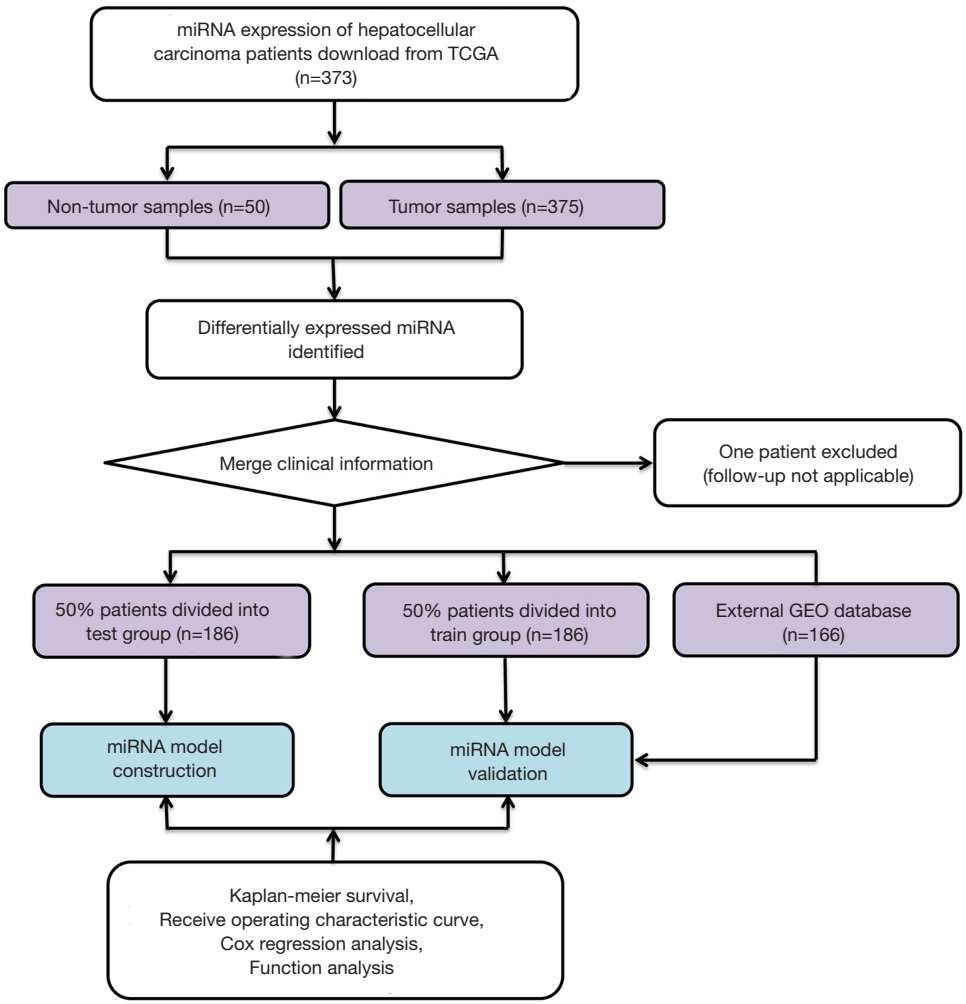

Figure 1 Overall design of this study.

prognostic miRNA along with its relevant coefficient, which is shown as follows: $\sum_{i=1}^{n}(coefi \times Expi)$; where Exp*i* is the miRNA signature expression of patient i, and coef*i* is the LASSO coefficient of miRNA i.

Confirmation of the miRNA signature

Patients were distributed based on their risk score and survival information. HCC patients were further categorized into high- and low-risk groups with the median risk score being used as the cutoff value; this was followed by the generation of Kaplan-Meier (KM) survival curves and Cox proportional hazards regression in the two groups. The prognostic performance of miRNA risk scores was assessed by a time-dependent receive operating characteristic (ROC) curve via the comparison of sensitivity and specificity. Multivariate Cox regression analysis was subsequently conducted to confirm whether the predictive performance of miRNA risk score was independent of other clinicopathological parameters. Additionally, the predictive performance of the miRNA model was verified in the validation and entire groups. A two-sided P value <0.05 suggested statistical significance, and R language (version 3.6.0; R Foundation) was used for all analyses.

miRNA signature validation using the OncomiR database and the GEO dataset

To further test the proposed miRNA-derived prognostic value, we used the OncomiR online database (http://www.oncomir.org/) to cross-validate the reliability of the model. Using OncomiR, we analyzed our miRNA signature for

HCC survival outcome. The user-defined signature in predicting overall survival (OS) was demonstrated through KM curves and log-rank tests. The Gene Expression Omnibus (GEO) GSE31384 dataset was used as an external validation cohort (n=166). Patients were also classified according to the median risk score threshold as high- or low-risk, and KM survival curves were plotted.

Functional analysis

StarBase (http://starbase.sysu.edu.cn/) was employed in predicting mRNAs specific to miRNAs. Using StarBase, we searched the interactions of miRNA-targets using multiple target-predicting programs (PITA, RNA22, miRmap, DIANA-microT, miRanda, PicTar, and TargetScan). Then, protein-coding genes with significance were selected to analyze the functional enrichment using Metascape (http://metascape.org) (13). Gene set enrichment analysis (GSEA), using JAVA program (http://software.broadinstitute.org/gsea/index.jsp), was employed for assessing the possible mechanisms). The number of random sample permutations was set at 1,000, and a P value <0.05 was set as the significance threshold.

In vitro analysis

Cells from one human normal hepatocyte cell line, MIHA (RRID:CVCL_SA11), and two hepatocyte cell lines, HepG2 (RRID:CVCL_0027) and HCCLM3 (RRID:CVCL_6832) were provided by the Shanghai Cell Bank of the Chinese Academy of Sciences, and cultivated within Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Invitrogen). Cells were then cultivated in the absence of antibiotics at 37 ℃ in an atmosphere of 5% CO_2 and 99% relative humidity. They were then transfected with miR-137 mimic (miR-137) or the negative control scramble miR (miR-NC) according to the manufacturer's instructions (purchased from Guangzhou RIBOBIO, China). The total cellular or tissue RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) in accordance with manufacture protocols. The $2^{-\Delta\Delta Ct}$ approach was utilized to normalize target RNA expression. HepG2 (1×10^3 cells per well) and HCCLM3 (1×10^3 cells per well) cells were seeded in 96-well plates. Cell viability was measured with the Cell Counting Kit-8 (CCK-8, Dojindo). Transwell assay was carried out to determine cell migratory capacity. In the migration assay, transduced cells were inoculated into the 24-well plates. After 24 h, the cells were serum-starved overnight and followed by trypsin digestion. Later, 200 μL serum-free medium containing 5×10^4 cells was placed into the upper chamber, while 700 μL culture medium supplemented with 10% FBS was placed into the lower chamber to grow for 24 h at 37 ℃. The subsequent migratory cells were subjected to 4% paraformaldehyde fixation and 0.5% crystal violet staining, and the number of cells stained was calculated under the microscope. In addition, HCC cell lines were cultivated within the culture dish, scratched with a 10-μL pipette tip, and cultured in FBS-free DMEM. Finally, the inverted microscope (OLYMPUS IX73) was used to observe the migratory cells at 24 h post injury.

Statistical analysis

LASSO regression was conducted to identify potential miRNAs. Survival outcomes were compared according to the KM method using the survival package in R. Time-dependent ROC curves were constructed using the R software package "survival ROC". Cox regression was conducted to verify the relationship of miRNA risk score prediction with other clinical parameters. Comparisons between groups were performed with Student's *t*-test and one-way ANOVA. A two-sided P value <0.05 was defined as statistically significant. All statistical analyses were performed using R (https://www.r-project.org) software.

Results

Clinical features

The baseline characteristics of all HCC patients are shown in *Table 1*. All samples were randomly categorized into two groups: a training cohort (n=186) and a testing cohort (n=186). The clinical features of patients in the training and testing cohorts were required to be similar. The expression profiles and clinical data from the training and testing cohorts are shown in files (https://cdn.amegroups.cn/static/application/15aac4fbcd5b505fe70f39214ebb9de2/ATM-20-2509-1.xlsx and https://cdn.amegroups.cn/static/application/e178bf8ebc5c5590f86bad217abc42ee/ATM-20-2509-2.xlsx), respectively.

Identification of DE miRNAs

In total, 127 DE miRNAs were identified between tumor samples and non-tumor samples (https://cdn.amegroups.cn/static/application/008ed35cca4ff295542

Table 1 Baseline characteristics of all hepatocellular carcinoma patients

Characteristic	Entire cohort, n (%)	Training cohort, n (%)	Test cohort, n (%)	P value
Total	372 (100.0)	186 (50.0)	186 (50.0)	–
Median follow-up days	556 (0–3,675)	555 (0–3,675)	563 (0–3,478)	0.99
Age	59.4±13.0	60.2±13.6	58.6±13.3	0.33
Gender				0.99
Male	253 (68.0)	126 (33.9)	127 (34.1)	
Female	119 (32.0)	60 (16.1)	59 (15.9)	
Tumor grade				0.88
I	55 (14.8)	25 (6.7)	30 (8.1)	
II	176 (47.3)	92 (24.7)	84 (22.6)	
III	124 (33.3)	62 (16.7)	62 (16.7)	
IV	13 (3.5)	4 (1.1)	9 (2.4)	
Unknown	4 (1.1)	3 (0.8)	1 (0.3)	
Stage				0.89
I	172 (46.2)	80 (21.5)	92 (24.7)	
II	86 (23.1)	47 (12.6)	39 (10.5)	
III	85 (22.8)	42 (11.3)	43 (11.6)	
IV	5 (1.3)	4 (1.1)	1 (0.3)	
Unknown	24 (6.5)	13 (3.5)	11 (3.0)	
T stage				0.52
I	182 (48.9)	82 (22.0)	100 (26.9)	
II	94 (25.3)	51 (13.7)	43 (11.6)	
III	79 (21.2)	43 (11.6)	36 (9.7)	
IV	13 (3.5)	6 (1.6)	7 (1.9)	
Unknown	4 (1.1)	4 (1.1)	0 (0.0)	
N stage				0.94
Without metastasis	254 (68.3)	123 (33.1)	131 (35.2)	
With metastasis	4 (1.1)	2 (0.5)	2 (0.5)	
Unknown	114 (30.6)	61 (16.4)	53 (14.2)	
M stage				0.37
Without metastasis	269 (72.3)	131 (35.2)	138 (37.1)	
With metastasis	4 (1.1)	4 (1.1)	0 (0.0)	
Unknown	99 (26.6)	51 (13.7)	48 (12.9)	
Tumor status during follow-up				0.97
With tumor	112 (30.1)	58 (15.6)	54 (14.5)	
Tumor free	233 (62.6)	116 (31.2)	117 (31.5)	
Unknown	27 (7.3)	12 (3.2)	15 (4.0)	
Vascular invasion				0.79
Micro	94 (25.3)	48 (12.9)	46 (12.4)	
Macro	17 (4.6)	11 (3.0)	6 (1.6)	
No	261 (70.2)	127 (34.1)	134 (36.0)	
Family history of cancer				0.96
No	211 (56.7)	103 (27.7)	108 (29.0)	
Yes	110 (29.6)	55 (14.8)	55 (14.8)	
Unknown	51 (13.7)	28 (7.5)	23 (6.2)	

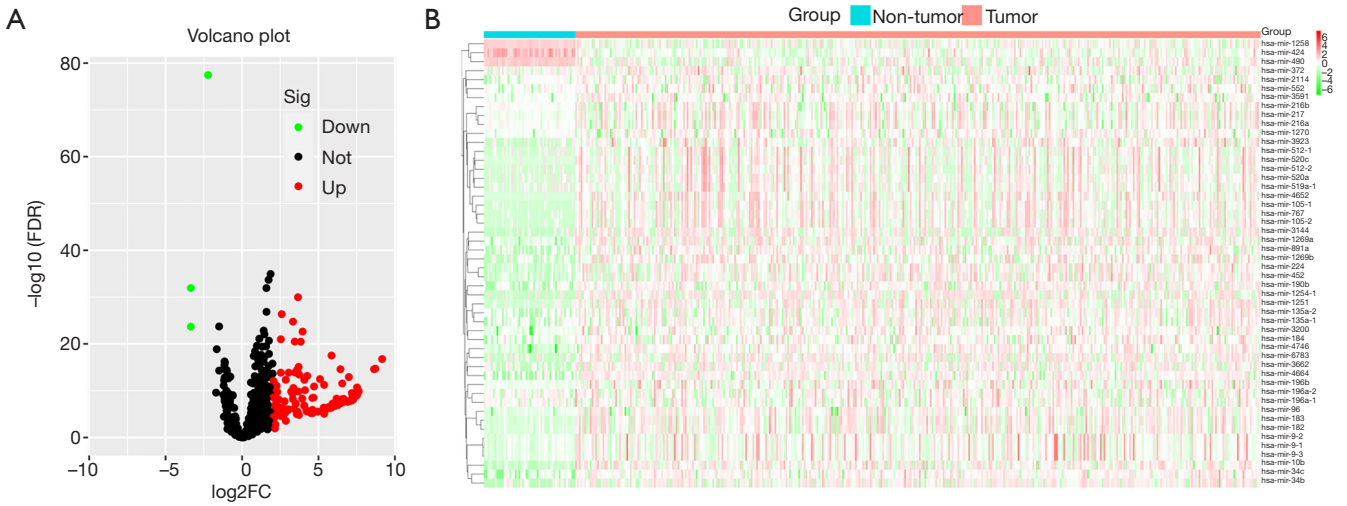

Figure 2 Heatmap and volcano plot. (A) The heatmap shows the differential expression of miRNAs; the X axis indicates the sample category, and the Y axis suggests DE miRNAs. Green represents the downregulated genes, while red represents the upregulated genes. (B) Volcano plot displaying DE miRNAs; the X axis represents the false discovery rate (FDR) values following log transformation, and the Y axis indicates the difference of mean miRNA expression. A red dot indicates upregulated miRNAs in tumor, a green dot indicates downregulated miRNAs in tumor, and a black dot indicates insignificant DE miRNAs. miRNA, microRNA; DE, differentially expressed.

c5299b6e9d29d/ATM-20-2509-3.xlsx), including 124 upregulated miRNAs and 3 downregulated miRNAs. A heat map was used to display the top DE miRNAs (*Figure 2A*), and a volcano plot was used to show miRNAs with significant differences between cancer and normal samples (*Figure 2B*).

Construction of the miRNA signature

Univariate Cox regression analysis revealed a significant association between 16 DE miRNAs and survival. Using "survminer" package, we determined the optimal cutoff miRNAs values, and the 16 corresponding survival curves are shown in *Figure 3*. LASSO regression was further used for validation of variables in the training cohort (*Figure 4A,B*), giving rise to five miRNAs, including has-mir137, hs-mir-105-2, has-mir-548f-1, has-mir-9-3, and has-mir-561. The forest plot of the correlation between each miRNA and survival is shown in *Figure 4C*. The prognostic risk score was imputed as follows: (0.6078 × has-mir-105-2) + (0.8071 × has-mir-9-3) + (0.8529 × has-mir-137) + (0.6551 × has-mir-548f-1) + (0.7866 × has-mir-561).

Confirmation of the miRNA signature

A risk score was calculated for each patient, and the patients were divided into high- or low-risk groups based on the median cutoff value. Risk score and survival time distributions in the training group are shown in *Figure 5A,B*, respectively. KM curves for high- and low-risk groups are displayed in *Figure 5C*, and show poorer OS in patients with high risk scores (P=5.682e-6). The ROC curve was further plotted for assessing the prognostic capacity of miRNAs biomarkers. The area under the curve (AUC) of the miRNA biomarker prognostic model was 0.726 (*Figure 5D*). The AUC of miRNA model was higher than that of every single miRNA or TNM stage (Figure S1). In the univariate Cox proportional hazards regression model, the hazard ratio (HR) of risk score was 2.979 [95% CI (confidence interval): 1.661–5.341] (*Figure 5E*), which was consistent with the outcomes from the multivariate Cox proportional hazards regression model (HR =3.285, 95% CI: 1.737–6.213) after adjustment of the clinical covariate (*Figure 5F*).

miRNA model in the validation cohort, OncomiR, and GEO

To determine if the proposed miRNA model exerted similar prognostic significance in other groups, we applied the same formula in the validation cohort. The miRNA risk score and survival time distributions in the validation group are shown in *Figure 6A,B*, respectively. KM curves in

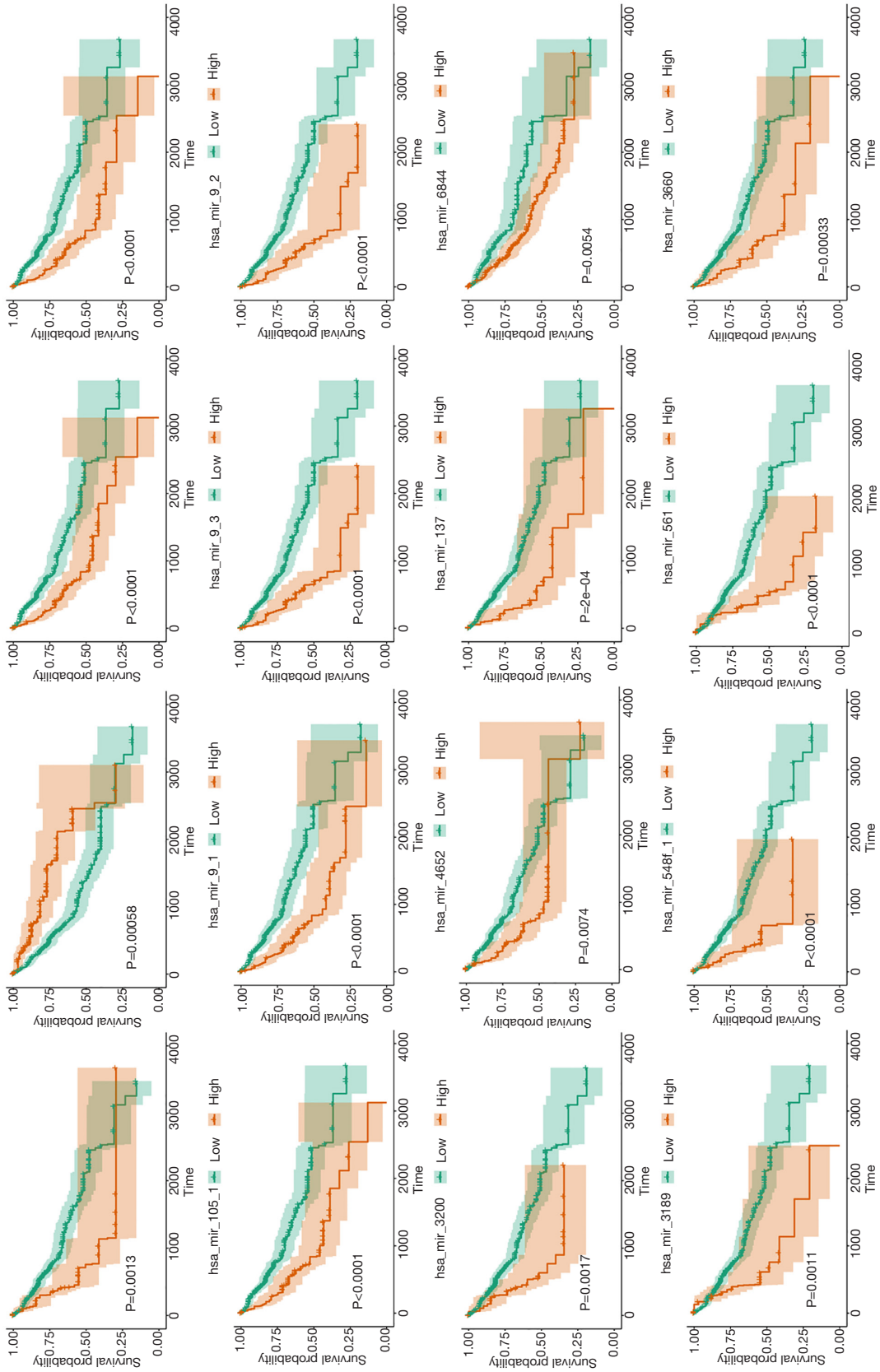

Figure 3 Survival analysis of 16 miRNAs by stratifying patients with an optimal cutoff value.

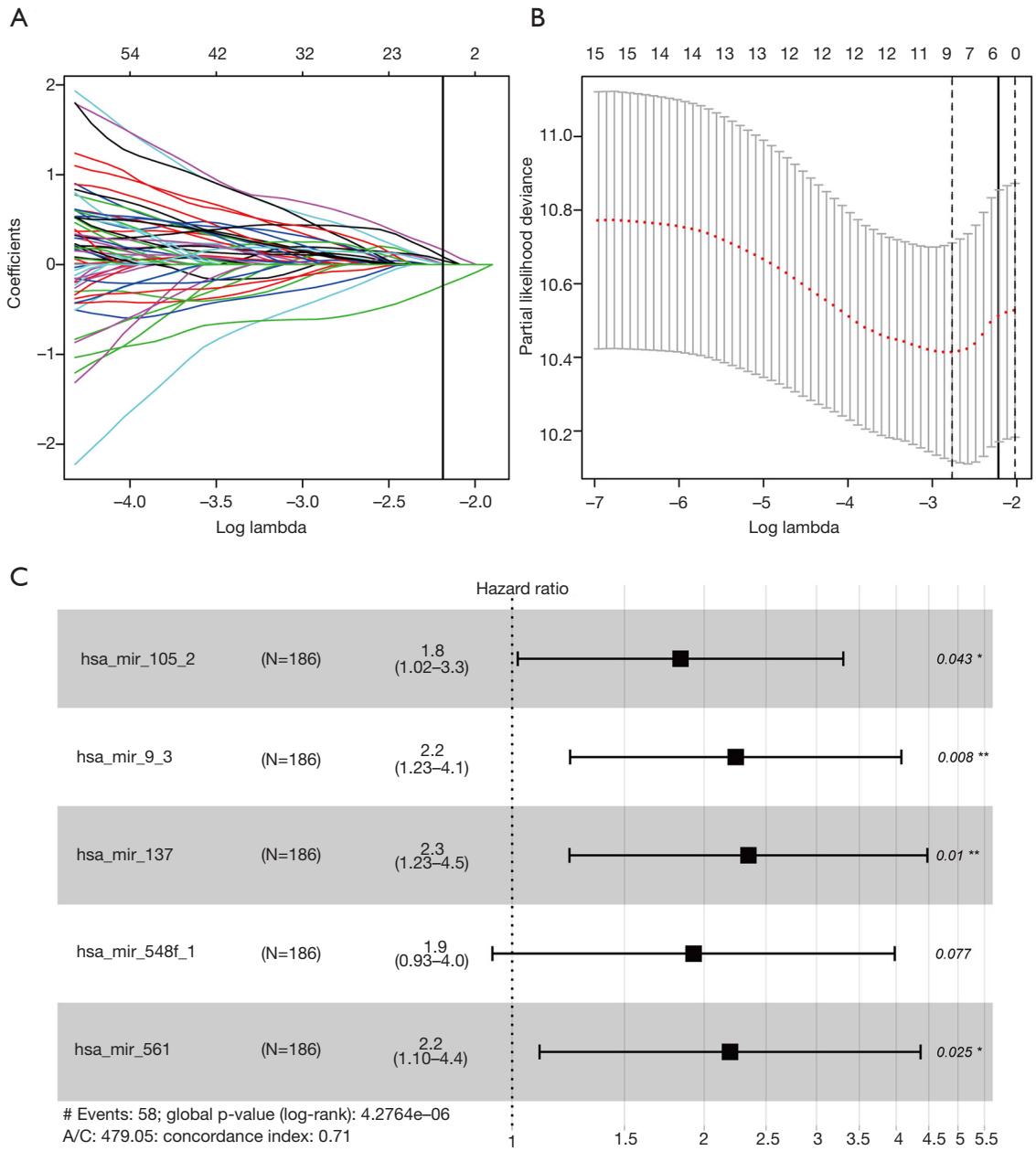

Figure 4 A plots of regression coefficient diagram using the LASSO regression approach. (A) LASSO coefficient profiles of the fractions of significant DE miRNAs. Coefficient profiles decreased when the lambda value was enlarged. LASSO coefficient profiles of five miRNAs; a vertical solid line was plotted based on 10-fold cross-validation. (B) Cross-validation to tune parameter selection in the LASSO model. The vertical lines were plotted at optimal values by minimum criteria and one-standard error criteria. A vertical solid line represents the final five identified miRNAs. (C) Forest plots displaying the correlation of different miRNA subsets with OS in the training cohort. Unadjusted HRs are presented with 95% CI. *, P<0.05; **, P<0.01. HR, hazard ratio; CI, confidence interval; LASSO, least absolute shrinkage and selection operator.

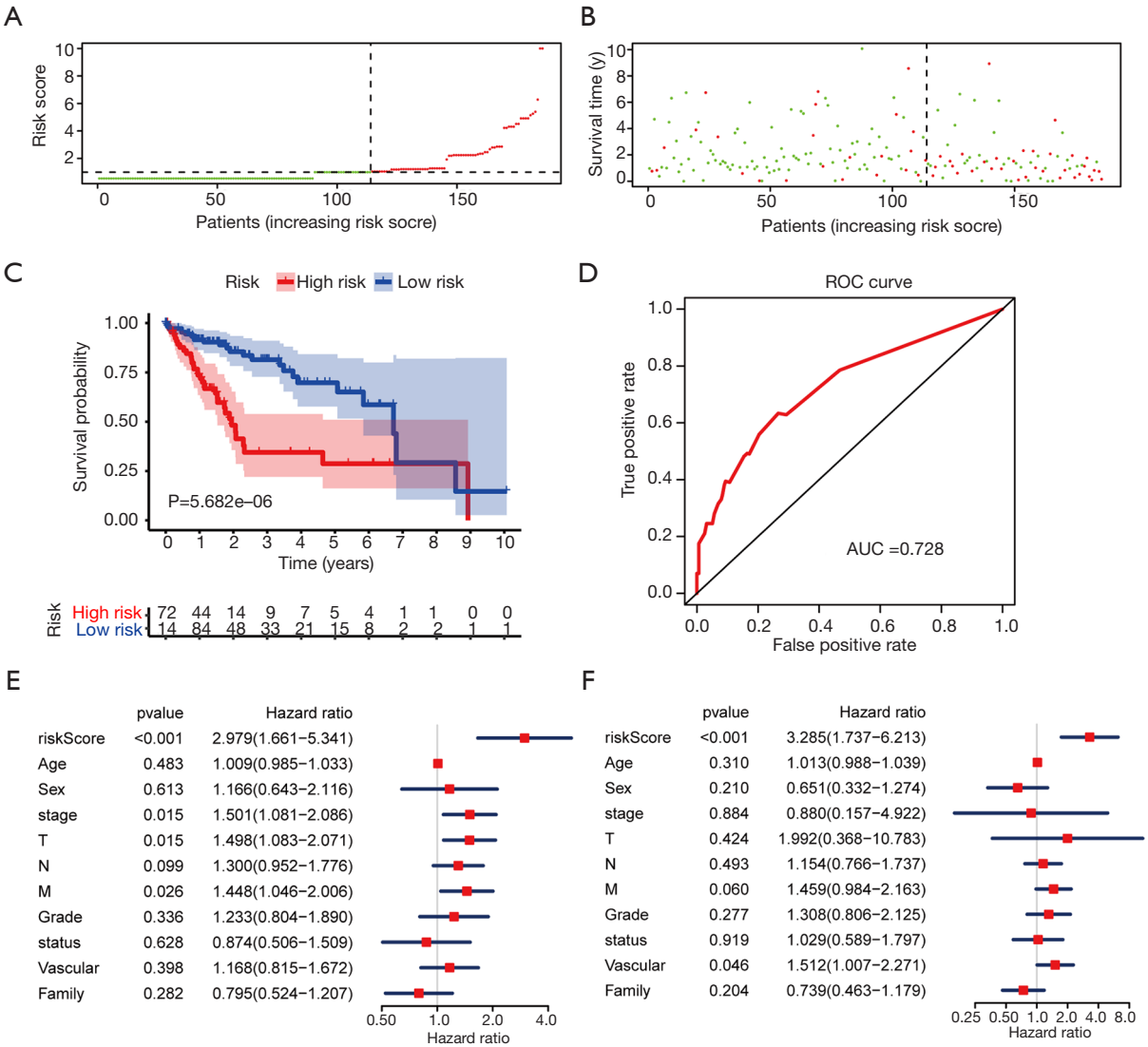

Figure 5 Establishment of a prognostic miRNA signature for HCC in the training cohort. (A) miRNA risk score distribution. Red indicates a high miRNA risk score; green indicates a low miRNA risk score. (B) The vital status, survival time, and risk score distribution. (C) KM survival curve between the high-risk and low-risk groups divided by median risk score. (D) Receive operating characteristic (ROC) curve of survival discriminated by the miRNA signature. (E) Univariate Cox regression analysis for the survival-related clinicopathological features confirmed the miRNA model as an independent factor. (F) Multivariate Cox analysis confirmed the miRNA model as an independent risk factor. Stage, pathological stage; T, T stage; N, N stage; M, M stage; grade, histological grade; status, tumor status during follow-up (with tumor/tumor free); vascular, vascular invasion; family, family history of cancer. HCC, hepatocellular carcinoma.

the high- and low-risk groups are displayed in *Figure 6C*. Patients with a low-risk score had a survival benefit in the validation group (P=7.865e-6). The AUC for the validation group was 0.721 (*Figure 6D*). The miRNA model AUC was also higher than every single miRNA or TNM stage AUC (Figure S1). Similar to the training cohort, univariate and multivariate analyses of the validation cohort demonstrated the presently established miRNA model as an independent prognostic indicator (*Figure 6E,F*). A similar pattern was seen in OncomiR. Patients with a high risk score had poor OS (P=5.035e-5, *Figure 7A*). To confirm the external validity, the model was applied in the external GEO data. *Figure 7B* shows the high-risk group had a significantly shorter survival than the low-risk group in the GSE31384

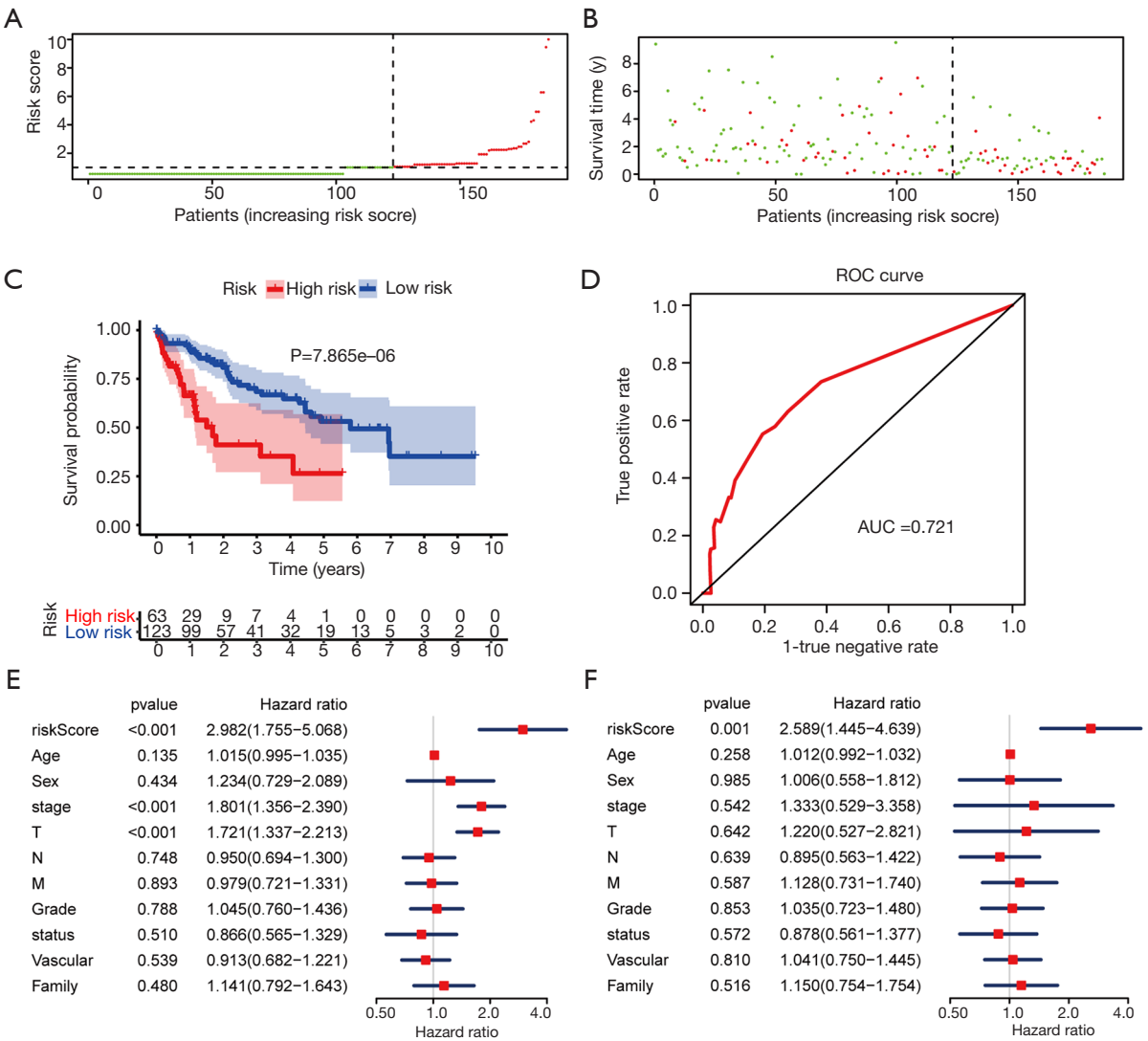

Figure 6 Further verification on the prognostic miRNA signature in HCC in the validation group. (A) miRNA risk score distribution. (B) The vital status, survival time, and risk score distribution. (C) KM survival. (D) Receive operating characteristic (ROC) curve of miRNA signature. (E) Univariate Cox regression. (F) Multivariate Cox analysis. Stage, pathological stage; T, T stage; N, N stage; M, M stage; grade, histological grade; status, tumor status during follow-up (with tumor/tumor free); vascular, vascular invasion; family, family history of cancer. HCC, hepatocellular carcinoma.

cohorts (P=0.020). ROC curve analysis showed that the risk signature prognosis prediction could attain an AUC value of 0.716 (*Figure 7C*).

Functional analysis of the prognostic miRNA model

For in-depth investigation of the possible biological effects of these identified miRNAs, miRNA-related mRNAs were filtered out. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were conducted on these selected mRNAs using Metascape. Specifically, terms having the P value of <0.01 and ≥3 enriched genes were regarded as significant. The top 20 pathways screened from GO and KEGG analyses are shown in *Figure 8*. GSEA was further conducted to illustrate the biological roles of the miRNA model, and indicated

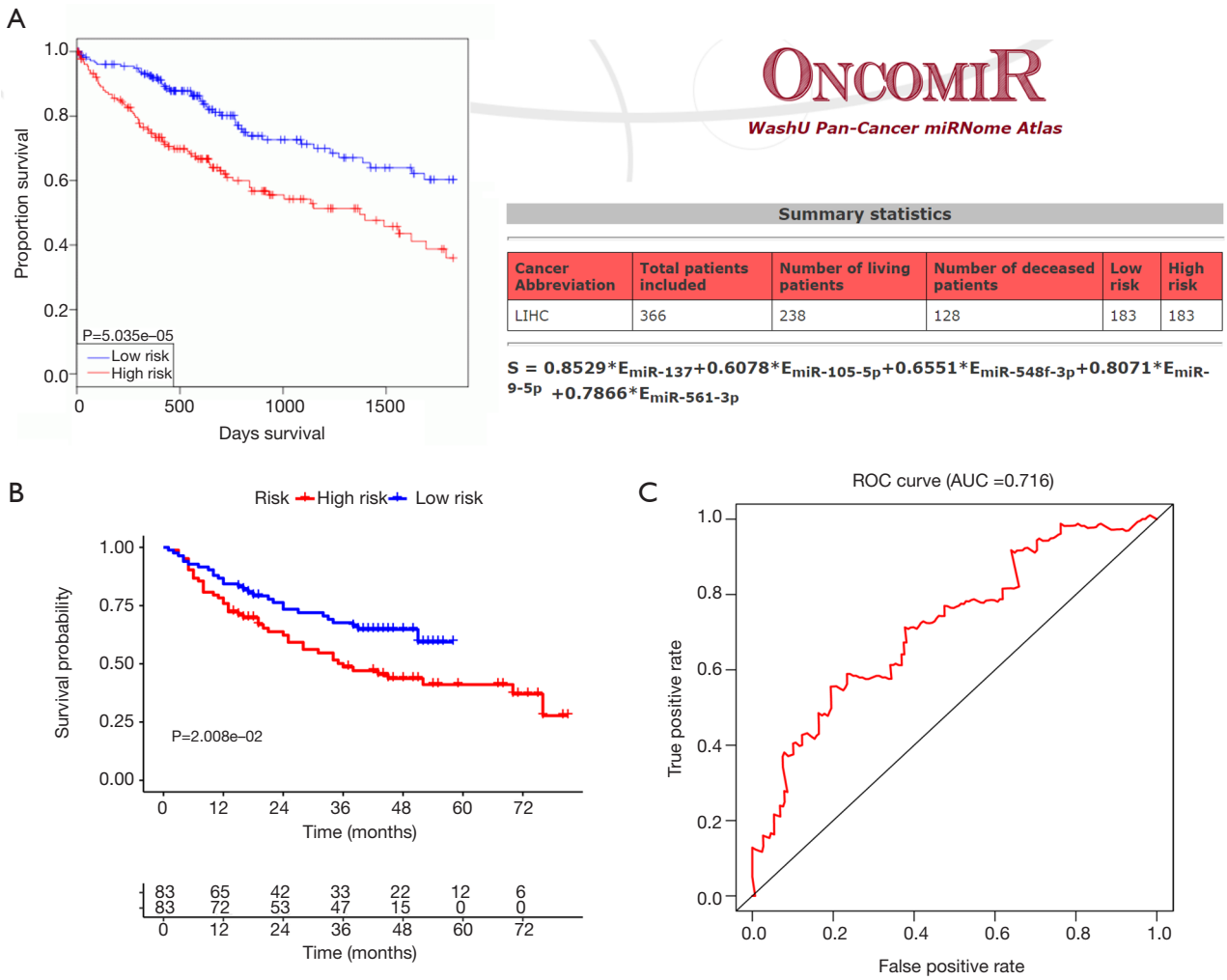

Figure 7 Further verification of the prognostic miRNA signature in HCC. (A) Survival curves in OncomiR. (B) Survival curves in GSE31384. (C) Receive operating characteristic (ROC) curve in GSE31384. HCC, hepatocellular carcinoma.

that the high-scoring miRNA model showed significant enrichment in Wnt signaling pathways, mTOR signaling pathways, and several cancer pathways (*Figure 9*).

miR-137 contributed to HCC progression and its target genes prediction

As miR−137 expression was the most upregulated of the five prognostic miRNAs in HCC, we performed *in vitro* experiments to further investigate the role of miR-137. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) showed that the miR-137 expression was high in HCC cell lines, and significantly increased following transfection (*Figure 10A*). We then performed the CCK-8 assay to detect the effect of miR-137 overexpression on cell proliferation. After miR-137 overexpression, HepG2 and HCCLM3 cell proliferation significantly increased compared with the miR-NC groups (*Figure 10B*, P<0.001). Transwell was conducted in order to determine the miR-137 role in the migratory capacity of HCC cells. Relative to controls, ectopic miR-137 expression within HepG2 (*Figure 10C*) and HCCLM3 (*Figure 10D*) cell lines resulted in the markedly upregulated migratory capacities of cells. A wound healing assay revealed that miR-137 overexpression significantly accelerated wound healing in HepG2 and HCCLM3 cells. A high miR-137 level correlated with poorer survival according to KM Plotter online tools (*Figure 10E*). The target genes of miR-137 were predicted,

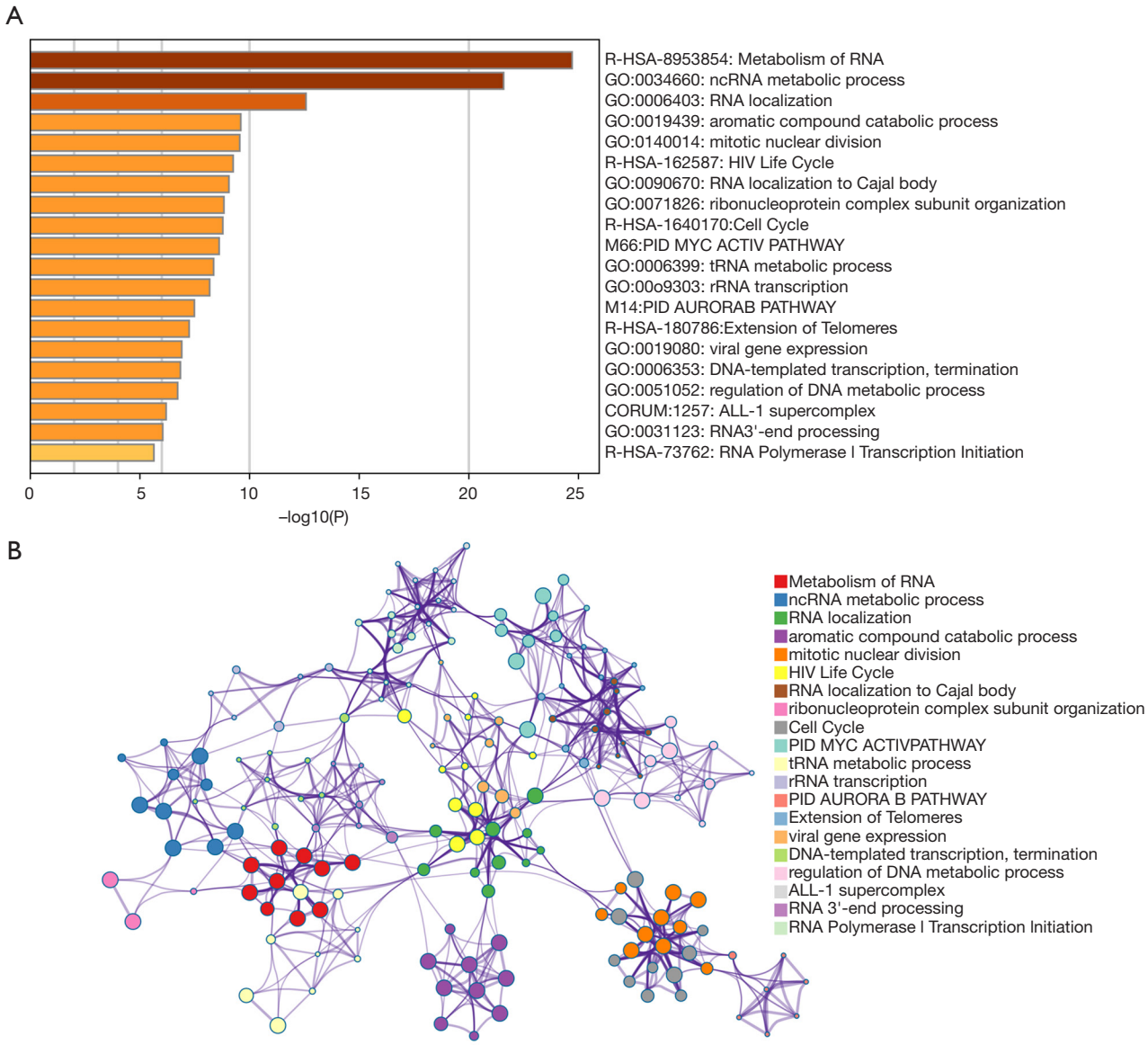

Figure 8 Enrichment analyses of the counterpart protein-coding genes. (A) The top 20 enrichment results are visualized by a bar graph. (B) The top 20 enrichment results are visualized by enrichment networks colored cluster IDs.

enrichment-analyzed, and are visualized in *Figure 10F*. Catenin-beta-interacting protein 1 (CTNNBIP1), a negative regulator of the Wnt signaling pathway, was predicted to be the target gene of miR-137 (https://cdn.amegroups.cn/static/application/aeb3ecf6ed3e8488be77e16af0f66be8/ATM-20-2509-4.xlsx). The above results suggested that miR-137 promotes HCC progression.

Discussion

To date, only a small number of miRNA-based models on HCC prognosis have been established, with common drawbacks including small study populations, inconsistent approaches to processing and analyzing data, and a lack of validation. To overcome these limitations, we conducted the current research in HCC patients to predict their survival using miRNA signatures. DE miRNAs that significantly correlated with survival were comprehensively screened by applying biostatistical methods and univariate Cox analysis. Five miRNAs were subsequently filtered out by

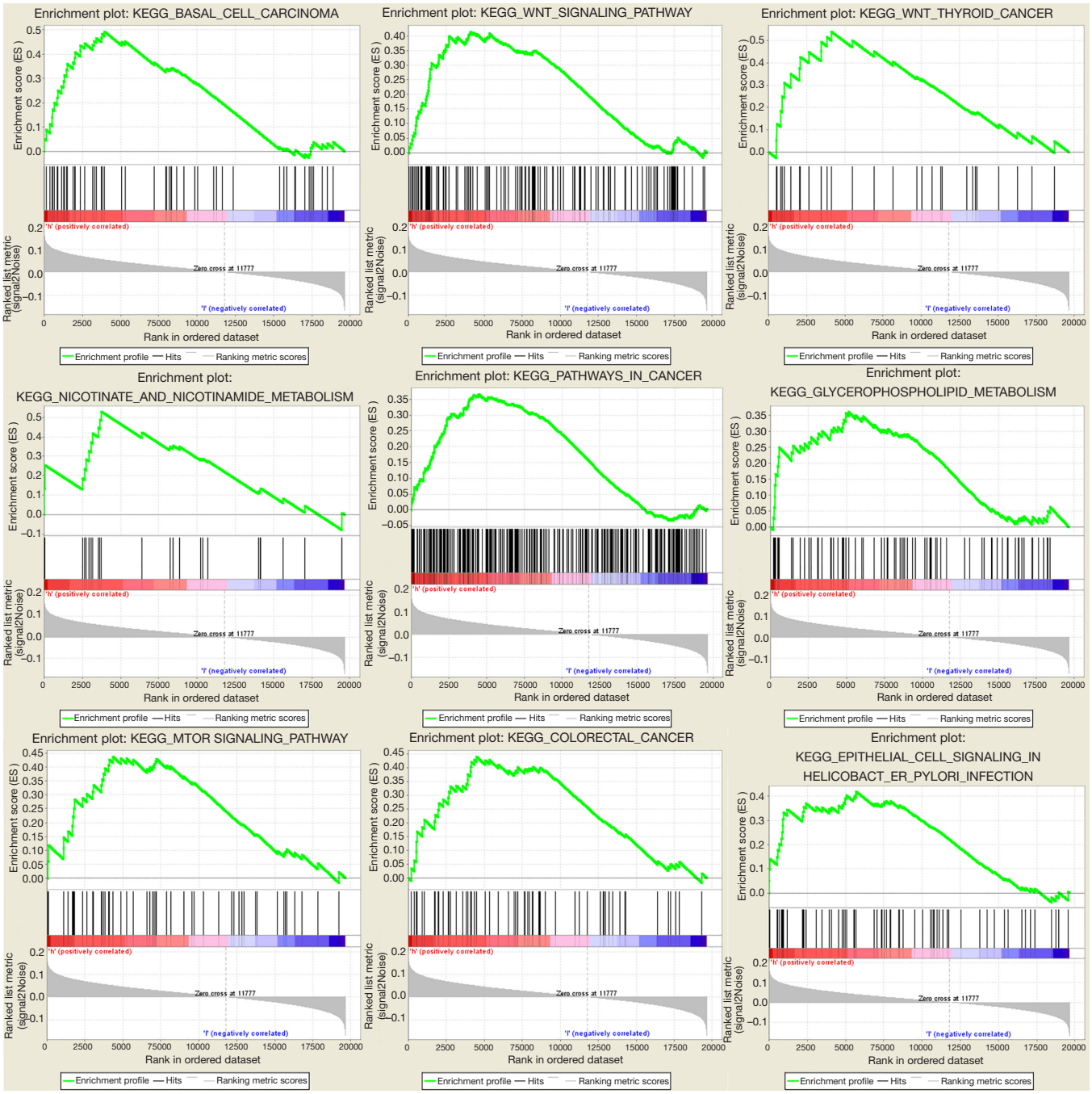

Figure 9 Gene set enrichment analysis indicating the biological pathways related to the miRNA model.

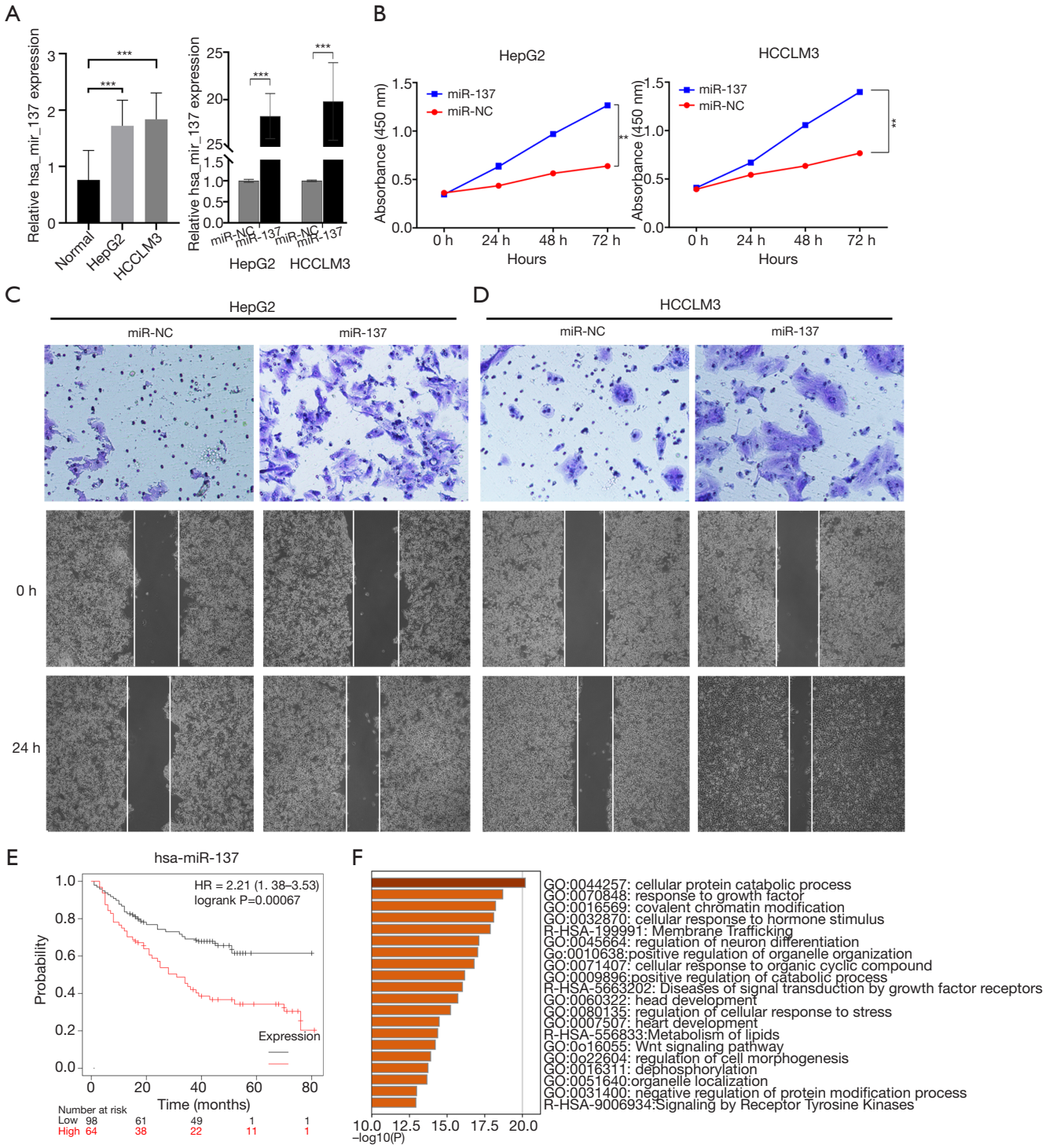

Figure 10 miR-137 contributed to HCC progression. (A) RT-qPCR confirmed the upregulated miR-137 expression in HCC cell lines and the miR-137 overexpression in the HepG2 and HCCLM3 cells transfected with miR-137 mimic compared to miR-NC. ***, P<0.001 *vs.* miR-NC. MTT (B), Transwell with 0.5% crystal violet staining ×20 (C), and wound healing assay ×4 (D) were conducted to examine the cell proliferation and migration of HepG2 (C) and HCCLM3 (D) cells. **, P<0.01; ***, P<0.001 *vs.* miR-NC. (E) Survival curves of miR-137. (F) Enrichment analysis of miR-137 predicted target genes. RT-qPCR, reverse transcription-quantitative polymerase chain reaction; miR, microRNA; miR-NC, negative control scramble miR. HCC, hepatocellular carcinoma.

using LASSO regression on the miRNA data in TCGA. Significant miRNAs were further incorporated for the establishment of prognostic models. KM curves, time-dependent ROC curves, and Cox regression analysis, confirmed the five-miRNA signature to be an independent prognostic predictor in HCC. Further validation was carried out in the validation group with OncomiR and GEO being used to confirm the reliability of our results. Functional analyses including GO, KEGG, and GSEA were conducted to demonstrate the biological role of the miRNA model in HCC progression. Finally, we conducted *in vitro* assays to evaluate the promotive role of has-mir-137 in HCC cell proliferation and migration. Shi *et al.* constructed an 11-miRNA expression signature in HCC using datasets published prior to 2013 from the GEO, Web of Science, and ArrayExpress databases (8). However, as the data were published many years ago an and data homogeneity was as an issue, the data should be interpreted cautiously. Wang *et al.* identified a miRNA profile in hepatitis B virus-induced HCC (14), while Fu *et al.* performed an analysis specifically for resectable HCC (11). Following this, Liao *et al.* (10) and Bai *et al.* (12) identified potential miRNA biomarkers for prognostic prediction in HCC, while Lin *et al.* (7) and Han *et al.* (9) proposed miRNA signatures for vascular invasion to predict survival in HCC. However, the abovementioned studies only used conventional Cox regression rather than LASSO regression for model construction and thus lacked further validation.

Our study design was superior to other research concerning the prognostic significance of miRNAs in HCC in a number of ways. To begin, all HCC patients in the TCGA group were enrolled for analysis, with a considerable sample size. Second, unlike the conventional stepwise regression in previous research, the accuracy of bioinformatics analysis was enhanced by LASSO penalized regression, which is capable of simultaneously analyzing all independent parameters and identifying the most influential ones (15). After introduction of a penalty following a regularization path, the coefficients of variables with less influence become zero (16). Thus, the LASSO approach provides significantly higher accuracy compared with the conventional multivariate Cox stepwise regression model, particularly in cases of extremely large datasets, like genomics (17). Third, to ensure the reliability of the results, we produced the miRNA signature in both the training group and the validated model. Finally, we conducted an *in vitro* study to clarify the functional role of miR-137.

Although a number of miRNAs have been reported to be associated with tumorigenesis and tumor progression, the functions of most miRNAs remain largely undefined. Anwar *et al.* (18) found that aberrant hypermethylation and concomitantly decreased expression of intergenic miRNA genes were commonly detected in HCC; one of these genes, hsa-mir-9-3 (50%), was also identified by Potapova to be a novel target of abnormal DNA methylation in human HCC (19). Additionally, miR-137 might function as a prognostic biomarker in HCC population, which is also a possible target for eliminating liver cancer stem cells (20-22). Previous studies have also reported that miR-137 has either cancer cell-promoting or -suppressing capabilities (23). miR-137 was reported to be upregulated and to promote invasion in breast, bladder, non-small cell lung cancer (24-26), and more controversially, in liver cancer. Some research has found miR-137 to be downregulated in liver cancer, inhibiting tumor progression (22,27,28). However, other recent studies found miR-137 to have the opposite effect. Sakabe *et al.* found that patients with high miR-137 expression had worse survival prognosis, more frequent invasion of vessels and the bile duct, and high alpha-fetoprotein levels (20). Furthermore, Wei *et al.* showed that miR-137 upregulation in HCC contributed to poor survival, and could bind to Afamin, promoting invasion and metastasis of HCC cell lines (29). Our data also implicated miR-137 in promoting invasive growth of HCC cell lines, while its elevated expression was associated with worse survival of HCC patients. Another study performed miR-137-predicted target gene analysis and found that Wnt/β-catenin inhibitory protein, CTNNBIP1, could inhibit liver tumor progression and invasion (30). The combinatory effect of miR-137 and target gene, CTNNBIP1, on the regulation of liver cancer remains to be explored. In addition to hsa-mir-9-3 and miR-137, our results, for the first time, indicate that three other miRNAs (has-mir-105-2, has-mir-548f-1, and hsa-mir-561) may be possible prognostic predictors for HCC; further in-depth studies are warranted to confirm the present findings and to examine their molecular features.

Pathway enrichment analysis showed that these miRNAs are likely to impact the carcinogenesis and progression of HCC by pathways in a fashion according to their respective biological functions in HCC. Moreover, a portion of the enriched pathways reported in literature are involved in the carcinogenesis and progression of HCC. Although the biological roles of the miRNAs contributing to HCC

remain unclear, pathway enrichment analysis showed that these miRNAs are likely to impact the carcinogenesis and progression of HCC via the mTOR (31,32) or Wnt signaling pathway (33-37).

A few limitations in this study should also be addressed. First, the prognostic effects of miRNA signature in HCCs were not validated by further functional assays, but were deduced from online datasets using bioinformatics approaches. Second, while the exploration of circulatory plasmid/serum miRNAs in the circulating biofluids as potential biomarkers for cancer diagnosis and prognosis has attracted increasing attention, whether or not the proposed miRNAs can be cancer-specific and reliably detected in the blood remains unclear. Thus, further verification of the outcomes of this research is still required. Third, *in vitro* results need further verification, such as miRNA inhibitor experiment.

In summary, using comprehensive bioinformatics analysis and the genetic profiles and clinical features of a selected cohort, we demonstrated that a novel five-miRNA signature could be an independent prognostic biomarker for HCC. Further studies are warranted to validate the possible functions and mechanisms of these miRNAs in HCC progression.

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (No. 81801804).

Footnote

Reporting Checklist: The authors present the study in accordance with the MDAR reporting checklist. Available at http://dx.doi.org/10.21037/atm-20-2509

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form(available at: http://dx.doi.org/10.21037/atm-20-2509). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Chen QF, Huang T, Shen L, et al. Predictive value of a nomogram for hepatocellular carcinoma with brain metastasis at initial diagnosis: A population-based study. PLoS One 2019;14:e0209293.
2. Llovet JM, Zucman-Rossi J, Pikarsky E, et al. Hepatocellular carcinoma. Nat Rev Dis Primers 2016;2:16018.
3. Song JL, Nigam P, Tektas SS, et al. microRNA regulation of Wnt signaling pathways in development and disease. Cell Signal 2015;27:1380-91.
4. Hayes J, Peruzzi PP, Lawler S. MicroRNAs in cancer: biomarkers, functions and therapy. Trends Mol Med 2014;20:460-9.
5. Di Leva G, Garofalo M, Croce CM. MicroRNAs in cancer. Annu Rev Pathol 2014;9:287-314.
6. Budhu A, Jia HL, Forgues M, et al. Identification of metastasis-related microRNAs in hepatocellular carcinoma. Hepatology 2008;47:897-907.
7. Lin Z, Cai YJ, Chen RC, et al. A microRNA expression profile for vascular invasion can predict overall survival in hepatocellular carcinoma. Clin Chim Acta 2017;469:171-9.
8. Shi KQ, Lin Z, Chen XJ, et al. Hepatocellular carcinoma associated microRNA expression signature: integrated bioinformatics analysis, experimental validation and clinical significance. Oncotarget 2015;6:25093-108.
9. Han B, Zheng Y, Wang L, et al. A novel microRNA signature predicts vascular invasion in hepatocellular carcinoma. J Cell Physiol 2019;234:20859-68.
10. Liao X, Zhu G, Huang R, et al. Identification of potential prognostic microRNA biomarkers for predicting survival in patients with hepatocellular carcinoma. Cancer Manag Res 2018;10:787-803.
11. Fu Q, Yang F, Xiang T, et al. A novel microRNA signature

predicts survival in liver hepatocellular carcinoma after hepatectomy. Sci Rep 2018;8:7933.
12. Bai F, Zhou H, Ma M, et al. A novel RNA sequencing-based miRNA signature predicts with recurrence and outcome of hepatocellular carcinoma. Mol Oncol 2018;12:1125-37.
13. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019;10:1523.
14. Wang G, Dong F, Xu Z, et al. MicroRNA profile in HBV-induced infection and hepatocellular carcinoma. BMC Cancer 2017;17:805.
15. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33:1-22.
16. Gao J, Kwan PW, Shi D. Sparse kernel learning with LASSO and Bayesian inference algorithm. Neural Netw 2010;23:257-64.
17. McNeish DM. Using Lasso for Predictor Selection and to Assuage Overfitting: A Method Long Overlooked in Behavioral Sciences. Multivariate Behav Res 2015;50:471-84.
18. Anwar SL, Albat C, Krech T, et al. Concordant hypermethylation of intergenic microRNA genes in human hepatocellular carcinoma as new diagnostic and prognostic marker. Int J Cancer 2013;133:660-70.
19. Potapova A, Albat C, Hasemeier B, et al. Systematic cross-validation of 454 sequencing and pyrosequencing for the exact quantification of DNA methylation patterns with single CpG resolution. BMC Biotechnol 2011;11:6.
20. Sakabe T, Azumi J, Umekita Y, et al. Prognostic relevance of miR-137 in patients with hepatocellular carcinoma. Liver Int 2017;37:271-9.
21. Gao M, Liu L, Li S, et al. Inhibition of cell proliferation and metastasis of human hepatocellular carcinoma by miR-137 is regulated by CDC42. Oncol Rep 2015;34:2523-32.
22. Liu LL, Lu SX, Li M, et al. FoxD3-regulated microRNA-137 suppresses tumour growth and metastasis in human hepatocellular carcinoma by targeting AKT2. Oncotarget 2014;5:5113-24.
23. Mahmoudi E, Cairns MJ. MiR-137: an important player in neural development and neoplastic transformation. Mol Psychiatry 2017;22:44-55.
24. Ying X, Sun Y, He P. MicroRNA-137 inhibits BMP7 to enhance the epithelial-mesenchymal transition of breast cancer cells. Oncotarget 2017;8:18348-58.
25. Xiu Y, Liu Z, Xia S, et al. MicroRNA-137 upregulation increases bladder cancer cell proliferation and invasion by targeting PAQR3. PLoS One 2014;9:e109734.
26. Chang TH, Tsai MF, Gow CH, et al. Upregulation of microRNA-137 expression by Slug promotes tumor invasion and metastasis of non-small cell lung cancer cells through suppression of TFAP2C. Cancer Lett 2017;402:190-202.
27. Zhu M, Li M, Wang T, et al. MicroRNA-137 represses FBI-1 to inhibit proliferation and in vitro invasion and migration of hepatocellular carcinoma cells. Tumour Biol 2016;37:13995-4008.
28. Wu DC, Zhang MF, Su SG, et al. HEY2, a target of miR-137, indicates poor outcomes and promotes cell proliferation and migration in hepatocellular carcinoma. Oncotarget 2016;7:38052-63.
29. Wei Q, Zhao L, Jiang L, et al. Prognostic relevance of miR-137 and its liver microenvironment regulatory target gene AFM in hepatocellular carcinoma. J Cell Physiol 2019;234:11888-99.
30. Fu X, Zhu X, Qin F, et al. Linc00210 drives Wnt/beta-catenin signaling activation and liver tumor progression through CTNNBIP1-dependent manner. Mol Cancer 2018;17:73.
31. Matter MS, Decaens T, Andersen JB, et al. Targeting the mTOR pathway in hepatocellular carcinoma: current state and future trends. J Hepatol 2014;60:855-65.
32. Malakar P, Shilo A, Mogilevsky A, et al. Long Noncoding RNA MALAT1 Promotes Hepatocellular Carcinoma Development by SRSF1 Upregulation and mTOR Activation. Cancer Res 2017;77:1155-67.
33. Dai B, Ma Y, Yang T, et al. Synergistic effect of berberine and HMQ1611 impairs cell proliferation and migration by regulating Wnt signaling pathway in hepatocellular carcinoma. Phytother Res 2019;33:745-55.
34. Li B, Cao Y, Meng G, et al. Targeting glutaminase 1 attenuates stemness properties in hepatocellular carcinoma by increasing reactive oxygen species and suppressing Wnt/beta-catenin pathway. EBioMedicine 2019;39:239-54.
35. Hu P, Ke C, Guo X, et al. Both glypican-3/Wnt/beta-catenin signaling pathway and autophagy contributed to the inhibitory effect of curcumin on hepatocellular carcinoma. Dig Liver Dis 2019;51:120-6.
36. Tan A, Li Q, Chen L. CircZFR promotes hepatocellular carcinoma progression through regulating miR-3619-5p/CTNNB1 axis and activating Wnt/beta-catenin pathway.

Arch Biochem Biophys 2019;661:196-202.
37. Huynh H, Ong R, Goh KY, et al. Sorafenib/MEK inhibitor combination inhibits tumor growth and the Wnt/betacatenin pathway in xenograft models of hepatocellular carcinoma. Int J Oncol 2019;54:1123-33.

(English Language Editors: J. Brown and J. Gray)

Cite this article as: Li W, Kong X, Huang T, Shen L, Wu P, Chen QF. Bioinformatic analysis and *in vitro* validation of a five-microRNA signature as a prognostic biomarker of hepatocellular carcinoma. Ann Transl Med 2020;8(21):1422. doi: 10.21037/atm-20-2509

