

Construction of an HCC recurrence model based on the investigation of immune-related lncRNAs and related mechanisms

Xiang-Xu Wang,¹ Li-Hong Wu,² Liping Ai,¹ Wei Pan,¹ Jing-Yi Ren,¹ Qiong Zhang,¹ and Hong-Mei Zhang¹

¹Department of Clinical Oncology, Xijing Hospital, Fourth Military Medical University, Xi'an, Shaanxi 710032, China; ²Xijing 986 Hospital Department, Fourth Military Medical University, Xi'an, Shaanxi 710032, China

Long noncoding RNAs (lncRNAs) are emerging as critical regulators of gene expression and play fundamental roles in immune regulation. Growing evidence suggests that immune-related genes and lncRNAs can serve as markers to predict the prognosis of patients with cancers, including hepatocellular carcinoma (HCC). This study aimed to construct an immune-related lncRNA (IR-lncRNA) signature for prospective assessment to predict early recurrence of HCC. A total of 319 HCC samples under radical resection were randomly divided into a training cohort (161 samples) and a testing cohort (158 samples). In the training dataset, univariate, lasso, and multivariate Cox regression analyses identified a 9-IR-lncRNA signature closely related to disease-free survival. Kaplan-Meier analysis, principal component analysis, gene set enrichment analysis, and nomogram were used to evaluate the risk model. The results were further confirmed in the testing cohort. Furthermore, we constructed a competitive endogenous RNA regulatory network. The results of the present study indicated that this 9-IR-lncRNA signature has important clinical implications for improving predictive outcomes and guiding individualized treatment in HCC patients. These IR-lncRNAs and regulated genes may be potential biomarkers associated with the prognosis of HCC.

INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the most common human malignancies and a leading cause of death worldwide.^{1,2} Although radical resection is the cornerstone of the therapeutic option for early-stage HCC patients,³ up to 70% of HCC patients present tumor recurrence within 5 years after resection.⁴ Early recurrence is one of the most prominent risk factors for HCC prognosis. The current strategies for assessing the risk of relapse include imaging (computed tomography/magnetic resonance imaging), the alpha-fetoprotein (AFP) serum biomarker, and several tumor characteristics, such as microvascular invasion (MVI) and portal vein tumor thrombus. Although these strategies provide some value for predicting the risk of recurrence, they still face great difficulty in providing an accurate evaluation of the risk of recurrence, owing to their limited sensitivity and specificity.⁵ Plasma biomarker detection and identification of patients with potential recurrence remain clinical dilemmas. Therefore,

it is highly desirable to identify novel molecular biological markers to accurately predict the risk of early recurrence in HCC patients.

Long noncoding RNAs (lncRNAs) are longer than 200 nucleotides in length and have no protein coding functions.⁶ lncRNAs play an essential role in cancer development and progression.^{7,8} Evidence suggests that lncRNAs are engaged in epigenetic regulation, gene expression, genetic imprinting, chromatin remodeling, transcription, and post-transcriptional processes.^{9,10} In addition, lncRNAs play key roles in many cellular processes, such as the cell cycle, cell differentiation, and DNA repair.^{11,12} Emerging evidence suggests that lncRNAs, as regulators, play an important role in cancer immunity, such as antigen release, immune cell migration, immune cell infiltration, antigen presentation, and immune activation.^{13,14} Recent studies have suggested that some lncRNAs may serve as potential prognostic biomarkers in HCC.¹⁵ Several studies have reported the potential value of an lncRNA signature for predicting prognosis in HCC.^{16–18} However, most studies have focused on overall survival (OS). Currently, studies are lacking regarding the immune-related lncRNA signature in the early recurrence of HCC patients after radical resection.

In the present study, we screened 319 HCC patients under radical resection from The Cancer Genome Atlas (TCGA) to obtain patient clinical symptoms and the expression of lncRNAs. A total of 319 samples were randomly divided into a training cohort (161 samples) and a validation testing cohort (158 samples). In the training cohort, we used Cox regression models and lasso regression models to identify nine immune-related lncRNAs (9-IR-lncRNAs) that are tightly associated with the recurrence of HCC. We developed a model to predict the risk of recurrence (risk score) in HCC patients based on the 9-IR-lncRNA signature. The results revealed that HCC patients with higher risk scores had worse disease-free survival (DFS) than patients with lower risk scores. The finding in the training cohort was validated in the testing cohort.

Received 13 July 2021; accepted 3 November 2021;
<https://doi.org/10.1016/j.omtn.2021.11.006>

Correspondence: Hong-Mei Zhang, Department of Clinical Oncology, Xijing Hospital, Fourth Military Medical University, Xi'an, Shaanxi 710032, China.
E-mail: zhm@fmmu.edu.cn



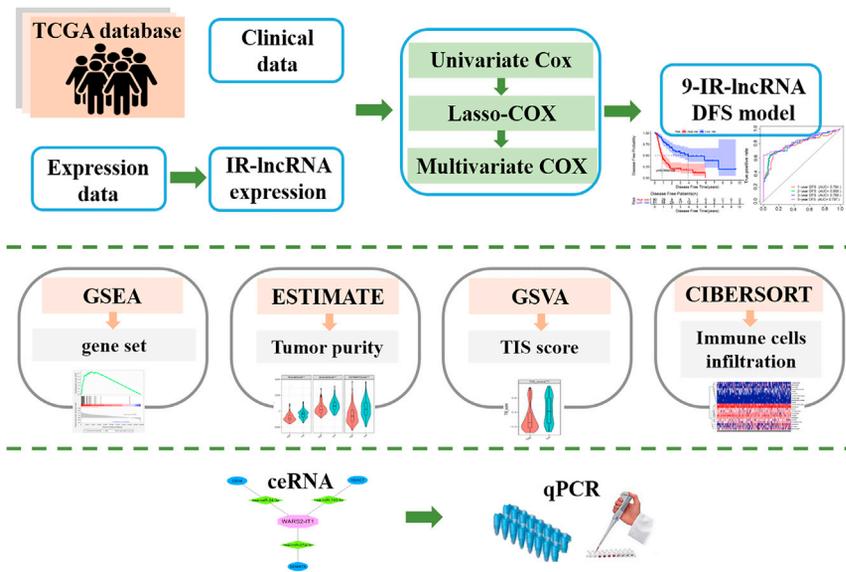


Figure 1. Analysis overview

Gene expression and clinical data of HCC under radical resection were collected from the TCGA database. To construct a relapse risk prediction model, nine DFS-related IR-lncRNAs were selected using univariate Cox, lasso-Cox, and multivariate Cox analyses. For each HCC sample, tumor purity, TIS, and the fraction of immune cell infiltration were estimated using the ESTIMATE and the gene set variation analysis (GSVA) function packages in R. In the GSVA function, samples were scored according to the level of expression of a comprehensive set of gene signatures related to the immune response. The resulting scores were used to cluster the metastatic samples based on their immune profile.

This 9-IR-lncRNA signature not only improves the ability to predict prognosis in HCC patients with early recurrence and promote better clinical treatment strategies, but also helps elucidate the underlying mechanisms. Finally, we further verified that these hub lncRNAs were significantly differentially expressed between HCC cell lines by quantitative real-time PCR (qRT-PCR).

RESULTS

Clinical characteristics and IR-lncRNA identification of 319 patients in the TCGA-HCC cohort

In the present study, 319 HCC samples with complete DFS information were included after screening 374 TCGA-HCC samples. The detailed analysis overview is shown in Figure 1. The "caret" package in R was used to randomly divide 319 patients into two groups as follows: 161 in the training cohort and 158 in the testing cohort. The clinical characteristics of HCC patients are shown in Table 1. There were no significant differences between the two groups in terms of age, gender, ethnic distribution, serum AFP level, vascular invasion, Child-Pugh grade, American Joint Committee on Cancer (AJCC) stage, recurrence status, and survival status. A total of 344 IR-lncRNAs were identified by expression correlation analysis of immune-related genes (IRGs) and lncRNAs ($P < 0.001$, $\text{cor} > 0.7$).

Construction of an HCC DFS prognostic model in the training cohort using nine IR-lncRNAs

In the training cohort, 100 IR-lncRNAs related to DFS were selected through univariate Cox analysis, 13 IR-lncRNAs (Figures S1A–S1C) were further screened using lasso-Cox regression, and nine IR-lncRNAs were included in the multivariate Cox regression. We used the multivariate Cox "stepwise regression method" to incorporate the nine core IR-lncRNAs related to DFS to establish a model (Figure S1D) and calculate the risk score using the following formula:

$$\text{risk score} = \sum_i^n \text{coef}_i * \exp(\text{lncRNA}_i), \text{ where } \text{coef}_i \text{ is the regression co-}$$

efficient, and $\exp(\text{lncRNA}_i)$ is the expression of nine IR-lncRNAs (Table 2). We further performed Kaplan-Meier curve survival analysis of the nine IR-lncRNAs. The results showed that high expression of AL031770, AC079328, AC011603, AL031985, AC087294, and LINC02193 significantly correlated with poor prognosis ($P < 0.05$, Figures S2A–S2G), while high expression of WARS2-IT1 and LINC002481 indicated a better prognosis ($P < 0.1$, Figures S2H and S2I).

HCC DFS prediction performance analysis of the 9-IR-lncRNA model

In the training cohort, with an increase in the risk score, the patient's recurrence or mortality gradually increased (Figures 2A and 2B), and the expression levels of WARS2-IT1 and LINC002481 decreased. In addition, the expression levels of AL008718, AL031770, AC079328, AC011603, AL031985, AC087294, and LINC02193 gradually increased with an increase in the risk score (Figure 2C). The 3D-principal components analysis diagram showed that the 9-IR-lncRNA signature distinguished the high-risk and low-risk groups (Figure 2D). Similar phenomena were observed in the verification cohort (Figures 2E–2H).

According to the median risk score of 1.02, the training cohort was divided into a high-risk group ($n = 80$) and a low-risk group ($n = 81$). The DFS of the high-risk group was significantly lower than that of the low-risk group (Figure 3A, $P < 0.001$). The median DFS time of the high-risk group was 0.8 years and that of the low-risk group was 3.8 years. Using the same threshold, the verification cohort was divided into a high-risk group ($n = 74$) and a low-risk group ($n = 84$). Similar to the training cohort, the DFS of the high-risk group in the verification cohort was significantly lower than that of the low-risk group (Figure 3B, $P < 0.001$). The median DFS time of the high-risk group was 1 year and that of the low-risk group was 3 years. The receiver operating characteristic (ROC) curve was used to test the model prediction performance. In the training cohort, the area under the curve (AUC) values of the risk score for 1-, 2-, 3-, and 5-year DFS were 0.784, 0.808, 0.768, and 0.797, respectively (Figure 3B). In the

Table 1. Clinical features of HCC patients with radical resection in each cohort

| Variables | Training cohort (n = 161)/(%) | Testing cohort (n = 158)/(%) | p |
|---------------------|----------------------------------|---------------------------------|-------|
| Age | | | 0.633 |
| <50 | 32 (20) | 33 (21) | |
| 50–59 | 51 (32) | 41 (26) | |
| 60–69 | 53 (33) | 53 (33) | |
| ≥70 | 25 (15) | 31 (20) | |
| Gender | | | 0.629 |
| Male | 113 (70) | 106 (67) | |
| Female | 48 (30) | 52 (33) | |
| AFP (U/LN) | | | 0.182 |
| <1 | 58 (36) | 47 (30) | |
| 1–50 | 42 (26) | 36 (23) | |
| ≥50 | 20 (12) | 30 (19) | |
| NA | 41 (26) | 45 (28) | |
| Vascular Invasion | | | 0.160 |
| None | 97 (60) | 80 (51) | |
| Yes | 42 (26) | 50 (32) | |
| NA | 22 (14) | 28 (18) | |
| Child-Pugh Grade | | | 0.469 |
| A | 102 (63) | 90 (57) | |
| B | 8 (5) | 10 (6) | |
| C | 1 (1) | 0 (0) | |
| NA | 50 (31) | 58 (37) | |
| AJCC Stage | | | 0.175 |
| Stage I | 85 (53) | 72 (46) | |
| Stage II | 36 (22) | 43 (27) | |
| Stage III | 33 (20) | 40 (25) | |
| Stage IV | 4 (23) | 0 (0) | |
| NA | 3 (2) | 3 (2) | |
| Disease-Free Status | | | 0.654 |
| Disease Free | 71 (44) | 74 (47) | |
| Recurred/Progressed | 90 (56) | 84 (53) | |
| Survival Status | | | 0.196 |
| Alive | 116 (72) | 124 (78) | |
| Death | 45 (28) | 34 (22) | |

Note: AFP, α -fetoprotein; AJCC, American Joint Committee on Cancer; HCC, hepatocellular carcinoma; NA, not available; ULN, upper limit of normal.

verification cohort, the AUC values of the risk score for 1-, 2-, 3-, and 5-year DFS were 0.773, 0.748, 0.678, and 0.679, respectively (Figure 3D). Furthermore, to test the advantages of our model, we performed ROC curve analysis and compared the 1-year, 2-year, 3-year, and 5-year AUCs of four published lncRNA recurrence models,^{21,22} which demonstrated that our DFS model had a better performance than previous models. In terms of 1-year recurrence prediction, the AUC value of our model was 0.780, while previous

Table 2. Nine immune-related long noncoding RNAs were identified for the construction of a prognostic model by a multivariate Cox regression analysis in training cohort

| Variables | Univariate Analysis | | Multivariate Analysis | | |
|------------|---------------------|--------|-----------------------|------------------|-------|
| | HR (95% CI) | p | coef | HR (95% CI) | p |
| LINC002481 | 0.81 (0.66–0.99) | 0.044 | −0.42 | 0.66 (0.52–0.83) | 0.001 |
| WARS2-IT1 | 0.81 (0.67–0.98) | 0.028 | −0.35 | 0.71 (0.58–0.87) | 0.001 |
| AC079328 | 1.38 (1.08–1.77) | 0.011 | 0.26 | 1.29 (0.96–1.74) | 0.092 |
| AC011603 | 1.67 (1.19–2.35) | 0.003 | 0.34 | 1.41 (0.92–2.16) | 0.115 |
| AL031770 | 1.42 (1.13–1.78) | 0.003 | 0.23 | 1.25 (0.96–1.64) | 0.101 |
| AC087294 | 1.45 (1.16–1.80) | 0.001 | 0.21 | 1.23 (0.97–1.56) | 0.095 |
| AL031985 | 1.55 (1.24–1.95) | <0.001 | 0.30 | 1.35 (1.02–1.79) | 0.038 |
| LINC02193 | 1.32 (1.12–1.56) | 0.001 | 0.20 | 1.22 (1.00–1.49) | 0.054 |
| AL008718 | 1.59 (1.25–2.02) | <0.001 | 0.28 | 1.33 (1.02–1.72) | 0.033 |

Note: CI, confidence interval; coef, coefficient; HR, hazard ratio.

models ranged from 0.53 to 0.610 (Gu JX et al.¹⁹ 3-lncRNA model, AUC = 0.563; Gu JX et al.²⁰ 6-lncRNA model, AUC = 0.610; Zhang Q et al.²¹ 15-lncRNA model, AUC = 0.662; and Zhang ZJ et al.²² 14-lncRNA model, AUC = 0.563). Similar results were found for the 2-year, 3-year, and 5-year AUC values (Figures 3E–3H). These results indicated that the 9-IR-lncRNA model has moderate prediction accuracy. Next, we analyzed the differential expression of the 9-IR-lncRNAs in cancer and normal tissues in the entire TCGA-HCC cohort. The results indicated that AL008718, AL031770, AC079328, AC011603, AL031985, AC087294, LINC02193, and LINC002481 showed significantly high expression in tumors, while WARS2-IT1 showed low expression in tumors (Figure S3A). Similar results were found in HCC cell lines (Figure S3B). Compared with normal tissues, LINC002481 was significantly upregulated in HCC, leading to a better prognosis. Interestingly, this phenomenon is slightly contradictory, as some genes related to immune regulation have similar phenomena. For instance, CXCL11 expression is significantly upregulated in colon cancer but is associated with better prognosis; high CXCL11 expression results in a higher fraction of antitumor immune cells and a lower fraction of protumor immune cells.²³ These findings suggested that WARS2-IT1 and LINC002481 may play an inhibitory role in tumorigenesis and the immune microenvironment.

Comparison of the 9-IR-lncRNA risk score to the combined clinical score and each individual clinical feature

Clinical characteristics, such as age, gender, serum AFP level, Child-Pugh grade, vascular invasion, and AJCC staging, were analyzed by univariate and multivariate analyses. In the training cohort, univariate analysis showed that the risk score (hazard ratio [HR] = 1.721, $P < 0.001$), vascular invasion (HR = 2.348, $P < 0.001$), and AJCC stage (HR = 1.940, $P < 0.001$) were significantly different (Figure 4A). Multivariate analysis showed that the risk score was an independent risk factor for the prediction of HCC recurrence (HR = 1.824, $P < 0.001$) (Figure 4B). In the testing cohort, univariate analysis showed that the risk score (HR = 1.177, $P < 0.001$), vascular invasion

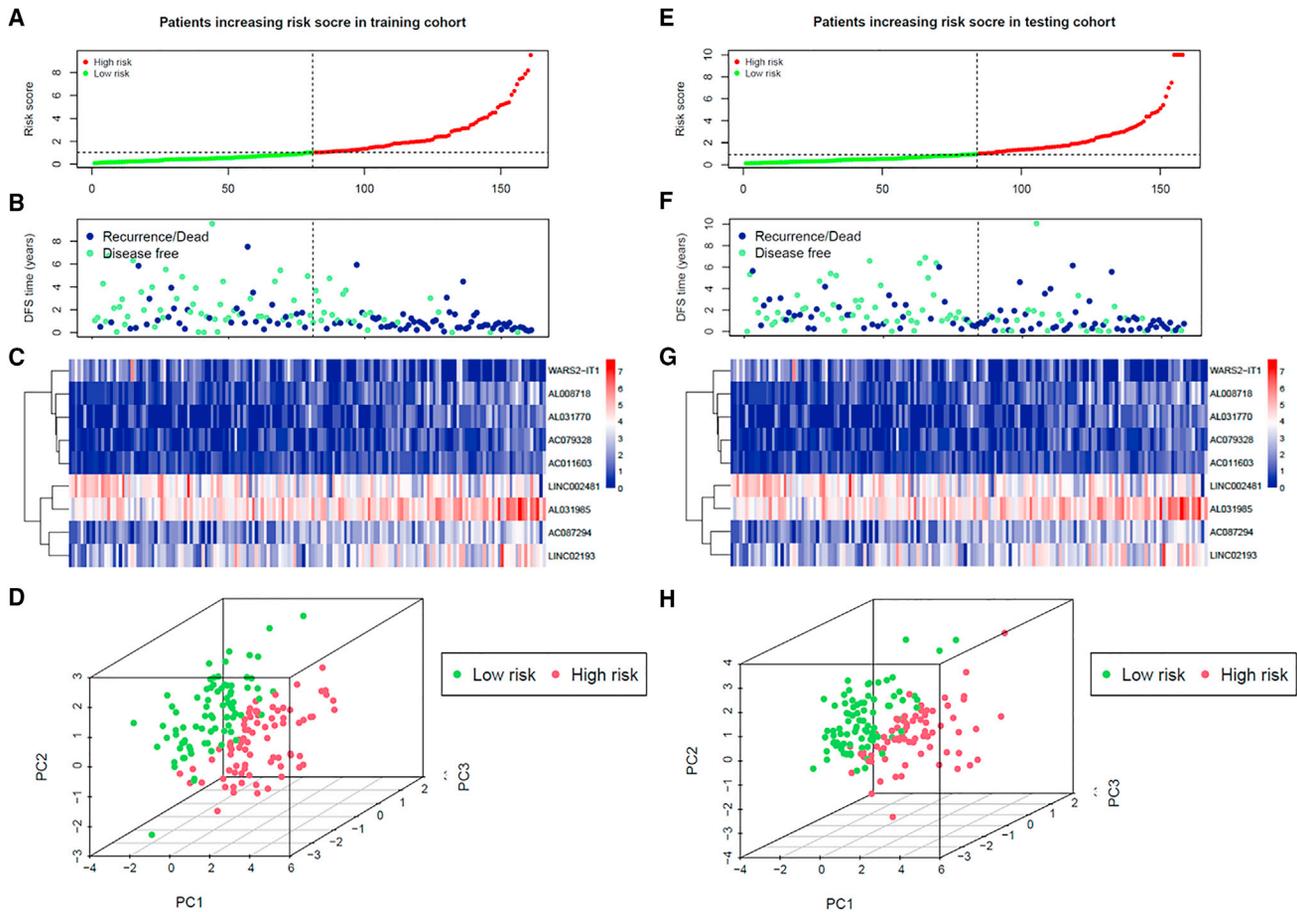


Figure 2. Relationship between the risk score and 9-IR-lncRNAs in the training cohort and testing cohort

(A) HCC patients were sorted by risk score. Red indicates high risk, and green indicates low risk. (B) DFS status of HCC patients. Dark blue indicates recurrence or dead, and light green indicates disease free. (C) Heatmap of the expression of 9-IR-lncRNAs and (D) PCA in the training cohort. (E–H) Risk score, DFS status, heatmap, and PCA in the testing cohort.

(HR = 1.711, $P = 0.037$), and AJCC stage (HR = 1.688, $P < 0.001$) were significantly different (Figure 4C). Multivariate analysis showed that risk score (HR = 1.743, $P < 0.001$) and AJCC stage (HR = 1.207, $P = 0.004$) were independent risk factors for the prediction of HCC recurrence (Figure 4D). The clinical risk score was a weighted sum of clinical characteristics, of which the weights were estimates from the multivariable Cox regression analysis. The AUC was used to evaluate the predictive performance of the 9-IR-lncRNA risk score, the clinical risk score, and each clinical feature. In the training cohort, the AUC value of the 9-IR-lncRNA risk score in the ROC curve of the 1-year DFS was the highest (AUC = 0.784) followed by clinical risk score (AUC = 0.719), AJCC stage (AUC = 0.698), AFP (AUC = 0.652), vascular invasion (AUC = 0.645), Child-Pugh grade (AUC = 0.532), gender (AUC = 0.507), and age (AUC = 0.437) (Figure 4E). The AUC values of the 2-year DFS ROC curve had the following order from high to low in the training cohort (Figure 4F): 9-IR-lncRNA risk score (AUC = 0.808) > clinical risk score (AUC = 0.702) > AJCC stage (AUC = 0.702) > vascular invasion (AUC = 0.621) > AFP (AUC = 0.607) > age (AUC = 0.542) > Child-Pugh grade (AUC = 0.525) >

gender (AUC = 0.487). Similar results were found in the verification cohort. For the 1-year DFS ROC curve in the validation cohort, the order of AUC values from high to low was as follows: 9-IR-lncRNA risk score (AUC = 0.773) > clinical risk score (AUC = 0.748) > AJCC stage (AUC = 0.711) > AFP (AUC = 0.616) > vascular invasion (AUC = 0.557) > Child-Pugh grade (AUC = 0.532) > sex (AUC = 0.445) > age (AUC = 0.431) (Figure 4G). The order of AUC values from high to low for the 2-year DFS ROC curve in the validation cohort was as follows: clinical risk score (AUC = 0.763) > 9-IR-lncRNA risk score (AUC = 0.748) > AJCC stage (AUC = 0.718) > AFP (AUC = 0.585) > vascular invasion (AUC = 0.537) > age (AUC = 0.505) > Child-Pugh grade (AUC = 0.481) > gender (AUC = 0.431). These results showed that the accuracy of early liver cancer recurrence prediction of the 9-IR-lncRNA model was better than that of the clinical risk score and the existing clinical indicators, such as AJCC stage, AFP, vascular invasion, and Child-Pugh grade.

To further explore the potential relationships between the risk score and multiple clinicopathologic factors, correlation analysis was

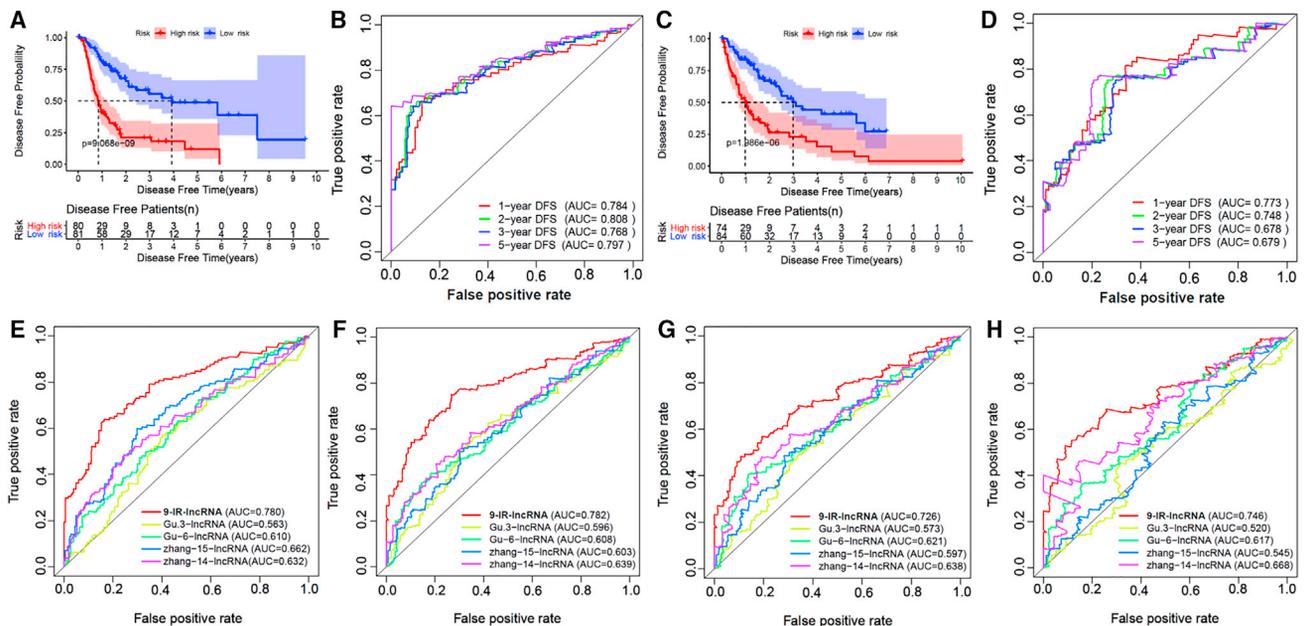


Figure 3. Kaplan-Meier curve survival analysis and time-ROC curve analysis of the IR-IncRNA signature

(A) Kaplan-Meier curve survival analysis between the high-risk and low-risk groups of the training cohort. The red line indicates the high-risk group, and the blue line indicates the low-risk group. (B) Time-ROC curve analysis of the 9-IR-IncRNA signature in the training cohort. The red, green, blue, and purple lines represent the 1-, 2-, 3-, and 5-year DFS, respectively. (C and D) Kaplan-Meier curve survival analysis and time-ROC curve analysis of the 9-IR-IncRNA signature in the testing cohort. (E–H) Time-ROC curve analysis of the 9-IR-IncRNA signature compared with other published models. (E) One-year DFS, (F) 2-year DFS, (G) 3-year DFS (H) and 5-year DFS. The red line represents our 9-IR-IncRNA model, and the yellow line represents the 3-IncRNA model by Gu et al.¹⁹ The green line represents the 6-IncRNA model by Gu et al.²⁰, and the blue line represents the 15-IncRNA model by Zhang et al.²¹ The purple line represents the 14-IncRNA model by Zhang et al.²²

conducted via independent t tests. In the training and testing cohorts, the risk scores of AJCC stage II and III/IV were higher than that of stage I ($P < 0.05$, Figure S4A), and the risk scores of T2 and T3/4 were higher than that of T1 ($P < 0.05$, Figure S4B). Moreover, the risk scores of N0, N1, and NX were not significantly different in the training and testing cohorts ($P > 0.05$, Figure S4C). The patients with macro- and microvascular invasion had higher risk scores than patients with no vascular invasion ($P < 0.05$, Figure S4D). A higher risk score was more commonly detected in patients with increased AFP ($P < 0.05$, Figure S4E), but the risk scores of Child-Pugh grades A and B were not significantly different ($P > 0.05$, Figure S4F).

Construction of HCC DFS prognostic nomogram

In the entire TCGA-HCC cohort, we selected three parameters, including risk score, AJCC stage, and vascular invasion, which exhibited significant differences ($P < 0.05$) in univariate Cox regression analysis (both training and testing cohorts), to construct the nomogram model for predicting the combined risk of HCC recurrence (Figure 5A). The calibration chart showed that the nomogram model predicted DFS at 1, 2, and 3 years, exhibiting high consistency with the actual observed DFS (Figures 5B–5D), and the C-index of the DFS nomogram model was 0.732 (95% confidence interval 0.686–0.778). The AUC value of the time ROC curve is used to evaluate the combined risk score, risk score, AJCC stage, and vascular invasion

of the nomogram model to predict the accuracy of DFS for 1, 2, and 3 years. The results showed that the highest AUC value of the 1-year DFS ROC curve was the combined risk score (AUC = 0.805) followed by the risk score (AUC = 0.782), AJCC stage (AUC = 0.658), and vascular invasion (AUC = 0.615) (Figure 5E). In the 2-year DFS ROC curve, the highest AUC value was also the combined risk score (AUC = 0.822) followed by the risk score (AUC = 0.777), AJCC stage (AUC = 0.696), and vascular invasion (AUC = 0.614) (Figure 5F). The same trend was also found in the 3-year DFS ROC curve with the combined risk score AUC value being the highest (AUC = 0.750) followed by the risk score (AUC = 0.718), AJCC stage (AUC = 0.645), and vascular invasion (AUC = 0.592) (Figure 5G). These results showed that the combined risk score of the nomogram model is generally more accurate than the risk score, AJCC stage, and vascular invasion in the prediction of HCC DFS.

Gene set enrichment analysis and immune-related characteristic analysis of the high- and low-risk groups of HCC

To further explore the potential molecular mechanisms of immune-related lncRNA signaling related to the risk of HCC recurrence, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) gene set enrichment analysis (GSEA) between the high-risk and low-risk groups. The results showed that the high-risk group was mainly enriched in nine KEGG pathways, including DNA replication, spliceosome, mismatch repair, homologous recombination, nucleotide

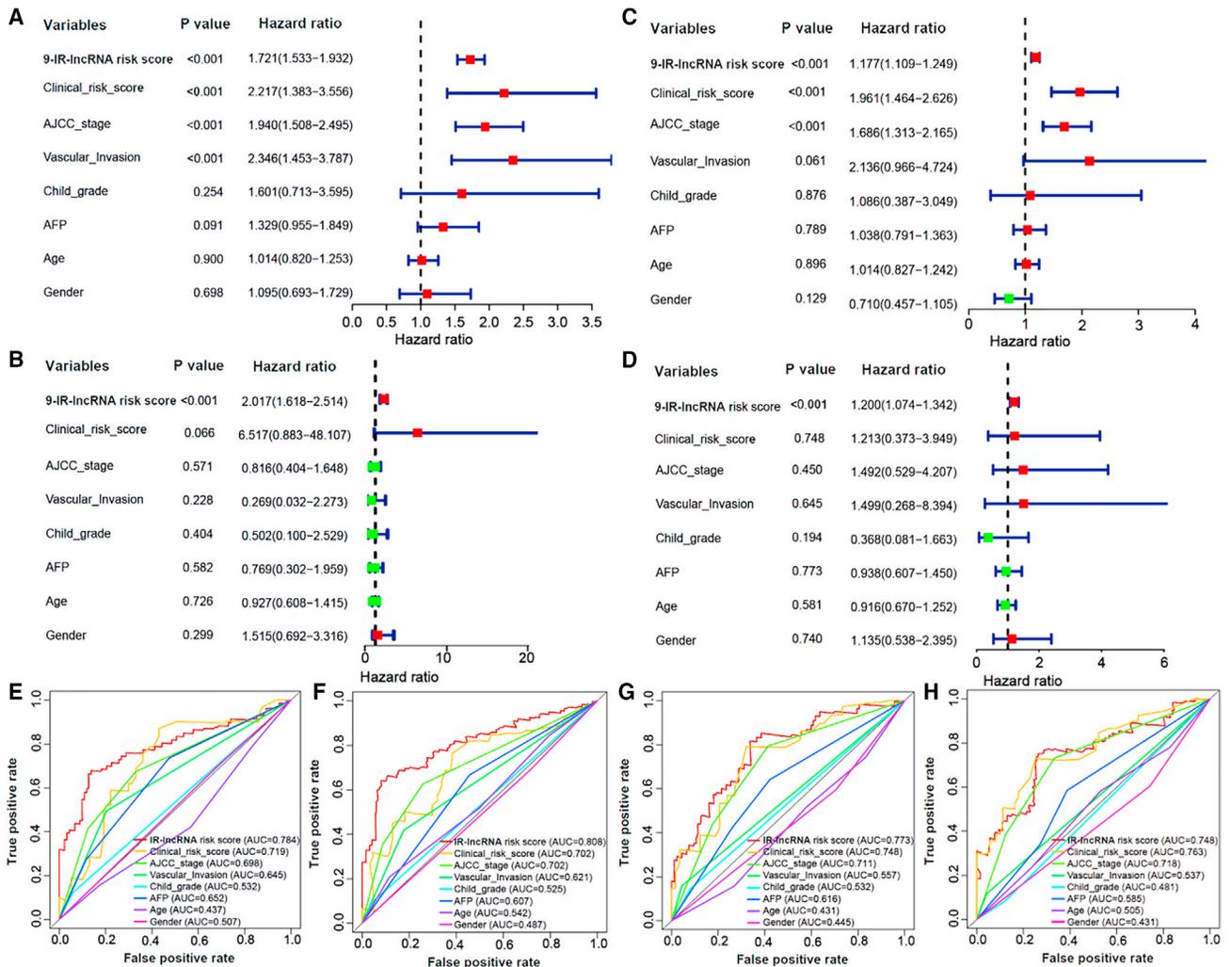


Figure 4. Univariate and multivariate Cox regression analyses of clinical characteristics and the 9-IR-IncRNA risk score in the training cohort and testing cohort

(A) Univariate and (B) multivariate Cox regression forest plots of clinical characteristics and the 9-IR-IncRNA risk score. (C) Univariate and (D) multivariate Cox regression forest plots of clinical characteristics and the 9-IR-IncRNA risk score. (E) One-year and (F) 2-year DFS ROC curve analysis of the 9-IR-IncRNA risk score and six clinical characteristics in the training cohort. The red, yellow, light green, green, light blue, blue, purple, and purplish red lines represent the 9-IR-IncRNA risk score, clinical risk score, AJCC stage, vascular invasion, Child-Pugh grade, AFP, age, and gender, respectively. (G) One-year and (H) 2-year DFS ROC curve analysis of the 9-IR-IncRNA risk score and six clinical characteristics in the testing cohort.

excision repair, base excision repair, proteasome, RNA degradation, and cell cycle (Figures S5A–S5I).

To explore immune-related characteristics, we compared the expression of immune checkpoint molecules, ESTIMATE ((Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) score, T cell inflammatory signature (TIS), and mRNA-based stem-like index (mRNAsi) between the high-risk and low-risk groups. Compared with the low-risk group, the high-risk group showed a trend of decreased expression of seven genes (CTLA-4, PD-L1, PD-L2, TIGIT, VISIR, LAG3, and TIM3), with two of these genes (PD-L2 and TIGIT) being significantly different

($p < 0.05$) (Figure 6A). Similarly, the stromal score, immune score, and ESTIMATE score in the high-risk group were significantly lower than those in the low-risk group (Figure 6B). The TIS, which is used to predict the response to anti-PD-L1 drugs,²⁴ was significantly lower in the high-risk group than that in the low-risk group (Figure 6C). However, the mRNAsi in the high-risk group was significantly higher than that in the low-risk group (Figure 6D).

IR-IncRNAs regulate the infiltration of a variety of immune cells. We used the CIBERSORT method to calculate the infiltration fraction of 22 immune cells between the high-risk and low-risk groups, and we analyzed the differences (Figures 7A and 7B). The fraction

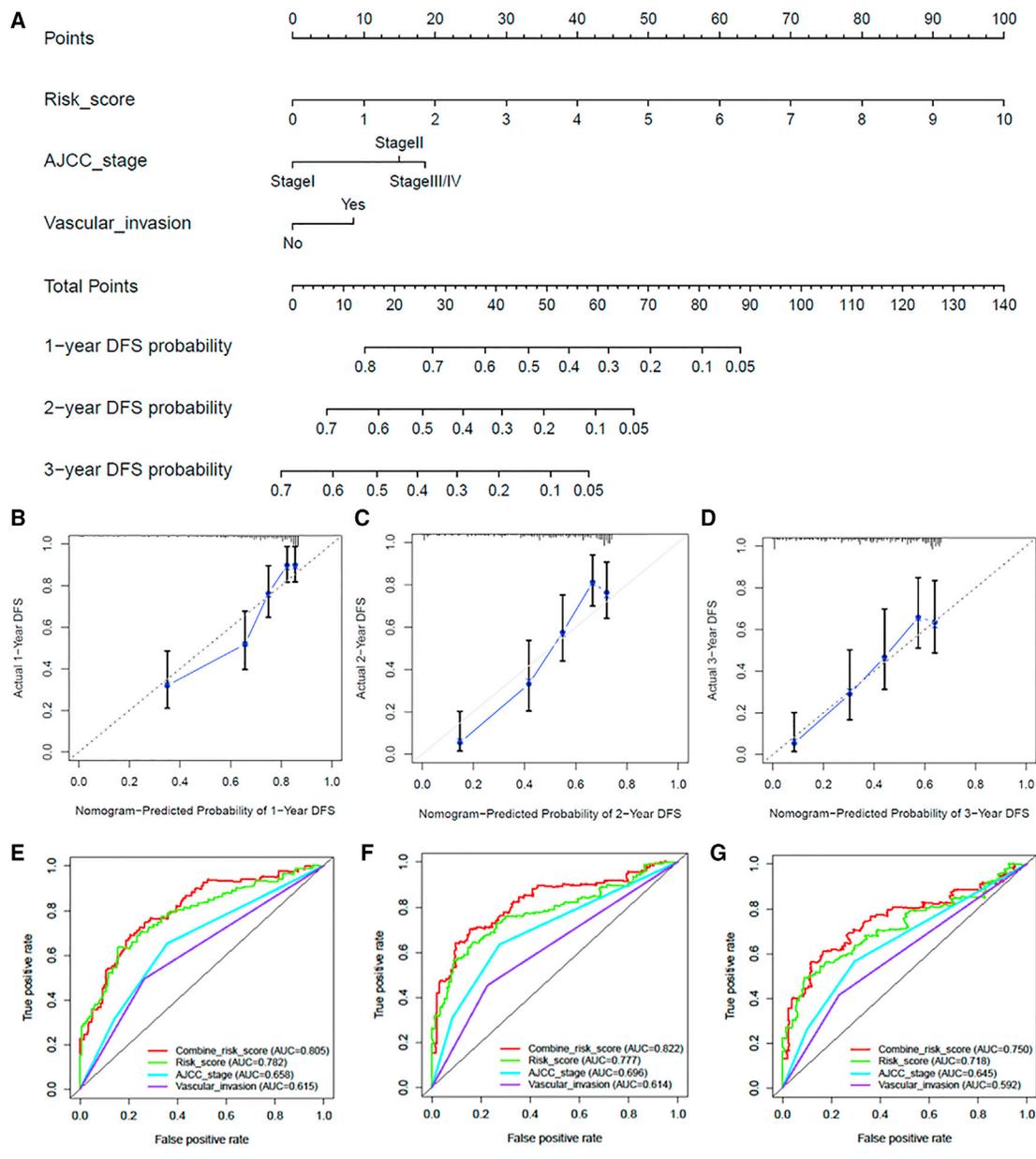


Figure 5. Construction of the nomogram based on clinical characteristics and IR-lncRNA risk score in the entire TCGA cohort

(A) Nomogram for predicting 1-year, 2-year, and 3-year DFS in the entire TCGA cohort. (B–D) Calibration curves of the nomogram for the consistency between predicted and observed 1-year, 2-year, and 3-year DFS in the entire TCGA cohort. The dashed line at 45° indicates a perfect prediction, and the actual performances of our nomogram are shown as blue lines. (E–G) One-year, 2-year, and 3-year DFS ROC curve analysis of the combined risk score, 9-IR-lncRNA risk score, AJCC stage, and vascular invasion in the entire TCGA cohort.

of M0 macrophages in the high-risk group was higher than that in the low-risk group ($P < 0.05$), and the fraction of M1 macrophages, CD8+ T cells, resting memory CD4+ T cells, activated memory CD4+ T cells, and naive CD4+ T cells was lower in the high-risk group than in the low-risk group ($P < 0.05$). These results suggested that the high-risk group is dominated by an increase in suppressor immune cells and M0 macrophages and that the low-risk group is

dominated by an increase in killer and antigen-presenting immune cells.

Construction of ceRNA regulatory network in HCC

To better understand the role of the nine-hub IR-lncRNAs screened out in HCC and to further clarify the interaction between IR-lncRNAs and miRNAs, we constructed a lncRNA-miRNA-mRNA related

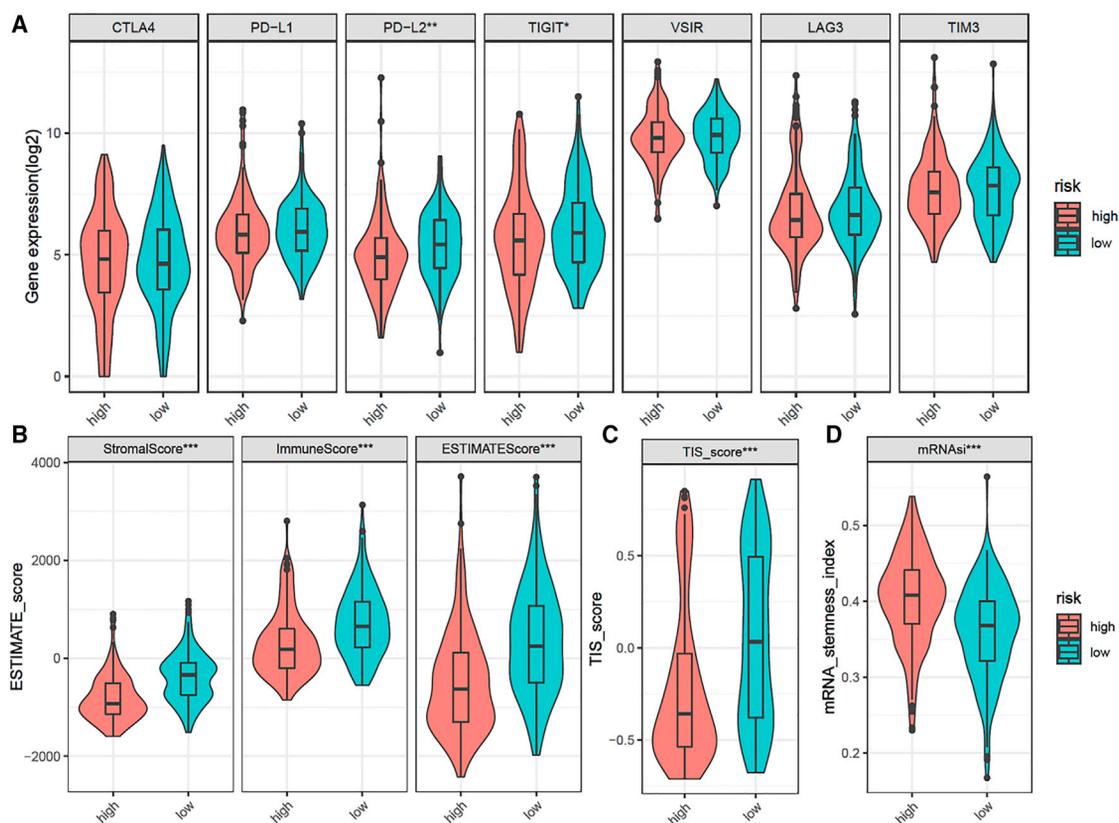


Figure 6. Comparison of immune checkpoint gene expression, ESTIMATE score, TIS, and mRNAi between the high- and low-risk groups

Violin plot showing the expression of immune checkpoint genes (A), ESTIMATE score (B), TIS (C), and mRNAi (D) between the high- and low-risk groups. The green bars represent the low-risk group, and the red bars represent the high-risk group. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

ceRNAs network. First, we retrieved 70 interacting lncRNA and miRNA pathways from two lncRNAs (WARS2-IT1 and AC087294) and 67 miRNAs from the miRcode database. First, we used TargetScan target gene data to identify a total of 163,558 pathways of 27 miRNAs targeting 17,781 mRNAs. Cytoscape software (version 3.7.3; <https://www.cytoscape.org/>) was used to construct the ceRNA regulatory network (Figure 8A), which consisted of two lncRNAs, 27 miRNAs, and 69 mRNAs. To ensure the reliability of the results, we then crossed the interaction pathways of the three databases (miRTarBase, miRDB, and TargetScan) and identified a total of 566 pathways of 17 miRNAs targeting 509 mRNAs (Figure 8B). Finally, Cytoscape software was used to construct a stricter ceRNA regulatory network (Figure 8C), which consisted of lncRNAs (WARS2-IT1), miRNAs (hsa-miR-24-3p, hsa-miR-27a-3p, and hsa-miR-140-5p), and target genes (SEMA7A, HDAC7, and CD34). These results suggested that SEMA7A, HDAC7, and CD34 may be core genes that affect the recurrence of liver cancer. We further analyzed the expression of the three target genes and three miRNA genes between cancer and normal samples. Compared with normal samples, the expression of hsa-miR-24-3p and hsa-miR-27a-3p was significantly lower in tumors, and the expression of hsa-miR-140-5p was significantly higher in tumors (Figure S6A). In addition, the expression of CD34, HDAC7, and SEMA7A

was significantly higher in tumors compared with normal samples (Figure S6B). We verified the expression of seven ceRNA regulatory network genes in human normal hepatocellular (MIHA) cells and two HCC cell lines (SNU-739 and SNU-368). Compared with normal hepatocellular cell lines, the expression of SEMA7A, HDAC7, and CD34 was higher in HCC cell lines, and the expression of hsa-miR-24-3p, hsa-miR-27a-3p, and hsa-miR-140-5p was lower in HCC cell lines (Figure 8D). Due to different origins and molecular backgrounds in cell lines, the expression of hsa-miR-24-3p was higher in SNU-368 cell lines than in MIHA normal hepatocellular cell lines and the SNU-739 HCC cell lines.

DFS and OS Kaplan-Meier curve survival analysis showed that the prognosis of patients was differentially correlated with the expression levels of the three hub miRNAs and three mRNAs. Patients with high expression of hsa-miR-24-3p had poor prognosis, as indicated by low DFS and OS (Figure S7A and S7D; DFS, $p = 0.005$; OS, $p = 0.046$). However, patients with high expression of hsa-miR-140-5p had low DFS and better OS (Figures S7B and S7E; DFS, $p = 0.184$; OS, $p < 0.001$). Patients with high expression of hsa-miR-27a-3p had poor prognosis as indicated by low DFS and OS (Figures S7C and S7F; DFS, $p = 0.017$; OS, $p = 0.016$). Patients with high CD34 expression

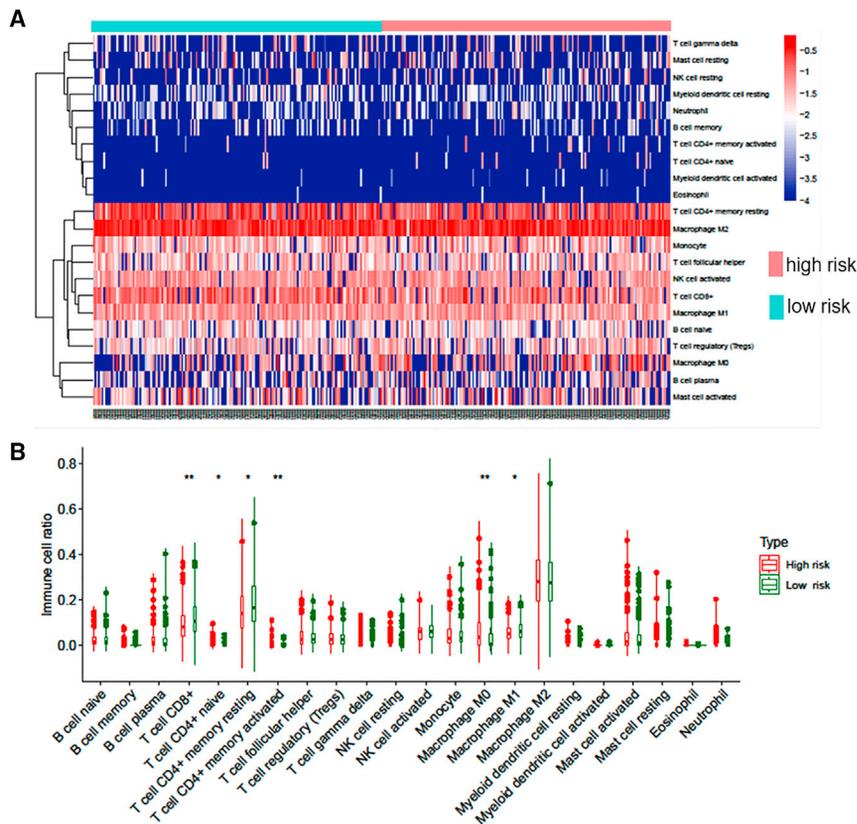


Figure 7. Immune cells infiltration landscape of HCC patients in high- and low-risk groups

(A) Heatmap of 22 immune cell ratios (log₁₀) between high- and low-risk patients. The red and light green bars represent the IR-lncRNA low- and high-risk groups, respectively. (B) Correlation analysis of 22 immune cell infiltrates between the high-risk and low-risk groups. The red and dark green bars represent the IR-lncRNA low- and high-risk groups, respectively.

Although many studies in the past few decades have been conducted to find reliable markers to predict the risk of recurrence, more research is required to construct a reliable system to assess the risk of recurrence in HCC. The majority of studies on predicting risk factors are focused on proteins encoding mRNAs or miRNAs.^{26,27} Currently, lncRNAs, especially IR-lncRNAs, which have not been previously studied, have generated broad interest. An increasing number of studies have revealed that the aberrant expression of lncRNAs is closely associated with the tumorigenesis, development, and metastasis of HCC and could become a potential biological marker for early diagnosis, immune therapy, and assessment of the risk of recurrence in HCC.²⁸ Therefore, to further determine the

value of IR-lncRNAs in the prediction of the recurrence of HCC, we employed univariate, multivariate Cox regression and lasso-Cox regression analyses to thoroughly screen the recurrence-related IR-lncRNAs in HCC patients with radical resection from TCGA database. Based on the screen, we constructed a 9-IR-lncRNA model to predict recurrence in HCC and further validated the model. The results indicated that the recurrence risk prediction model has good reliability in predicting recurrence in HCC and is a better predictor than currently used clinical markers, such as AJCC, MVI, and AFP. To further improve the prediction, we constructed a nomogram by combining the 9-IR-lncRNA risk score and clinical characteristics.

had a better prognosis as indicated by a better DFS and OS (Figures S8A and S8D; DFS, $p = 0.122$; OS, $p = 0.040$). Patients with high expression of HDAC7 had poor prognosis with low DFS and OS (Figures S8B and S8E; DFS, $p < 0.001$; OS, $p = 0.001$), and patients with high expression of SEMA7A had better prognosis with better DFS and OS (Figures S6C and S6F; DFS, $p = 0.001$; OS, $p = 0.026$).

Finally, we used the GSE76427 dataset to perform DFS and OS survival analysis verification of CD34, HDAC7, and SEMA7A (Figures S7A–S7F), and the ICGC-LIRI cohort was used to verify the OS analysis of CD34, HDAC7, and SEMA7A (Figures S7G–S7I). Patients with high CD34 expression had a better prognosis as indicated by a better DFS and OS (Figures S8A, S8D, and S8G; GSE76427S cohort DFS, $p = 0.122$; GSE76427S OS, $p = 0.040$; ICGC-LIRI cohort OS, $p = 0.004$). Patients with high expression of HDAC7 had poor prognosis with low DFS and OS (Figures S8B, S8E, and S8F; GSE76427S cohort DFS, $p = 0.014$; GSE76427S OS, $p = 0.932$; ICGC-LIRI cohort OS, $p = 0.288$). Patients with high expression of SEMA7A had better prognosis as indicated by better DFS and OS (Figures S8A, S8D, and S8G; GSE76427S cohort DFS, $p = 0.343$; GSE76427S OS, $p = 0.056$; ICGC-LIRI cohort OS, $p = 0.005$).

DISCUSSION

Recurrence of HCC is a critical cause of death worldwide in HCC patients due to the lack of a reliable method of prediction.²⁵

Our model differed from a previous lncRNA HCC prediction model.^{29–31} First, we used a gene expression correlation analysis to confirm IR-lncRNA. And followed by multiple step screening using univariate, lasso, and multivariate Cox regression analyses, which ensured the reliability of the lncRNAs that were selected for the model. In particular, the use of lasso regression allowed the variables to be simultaneously screened, excluded variables that contributed less to the model and selected the variables that had the most impact on the model.³² Therefore, our approach, which efficiently eliminated the collinearity of different variables and prevented overfitting of the model, is suitable for analyzing gene databases with large variables.³³

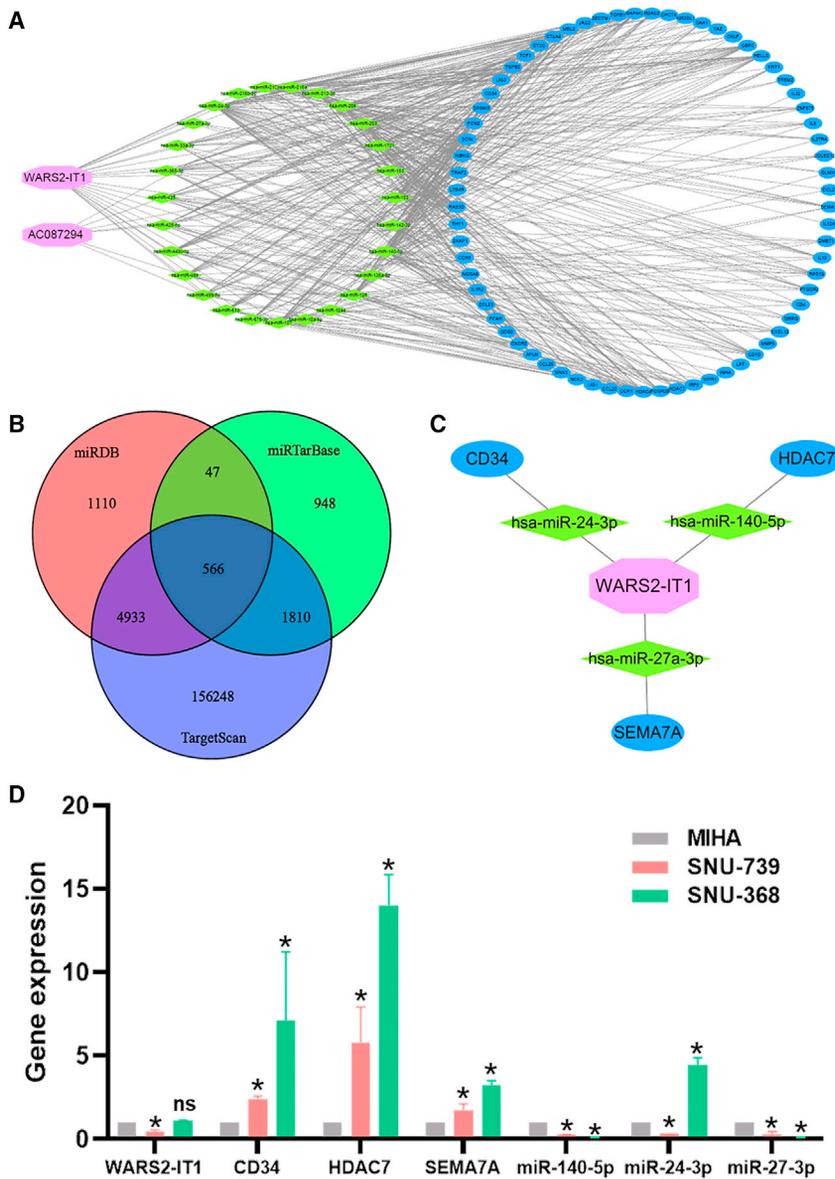


Figure 8. The ceRNAs regulatory network in hepatocellular carcinoma

(A) Plot of the ceRNA network based on the TargetScan database. The lncRNAs, miRNAs, and mRNAs are represented by nodes in pink, green, and blue. (B) Venn diagram of miRNA-mRNA pathways among the miRTarBase, miRDB, and TargetScan databases. (C) Plot of the hub ceRNA network based on the miRTarBase, miRDB, and TargetScan databases. (D) Bar plot of the expression of seven ceRNA network genes in a human normal hepatocellular cell line and HCC cell lines. Gray bars indicate the normal liver cell line (MIHA), the red and green bars indicate SNU-739 and SNU-368 HCC cell lines, respectively. Data are expressed as the mean \pm SEM.

and less potential benefit from immune therapy. Following immune infiltration analysis, we found that the high-risk group had significantly lower levels of CD8+ or CD4+ T cells. Third, while previous studies primarily focused on OS, we used DFS to evaluate the risk of recurrence. The use of DFS allowed us to evaluate the risk of recurrence in HCC patients and intervene promptly for high-risk patients after surgery, leading to improved survival post recurrence.

Finally, after predicting target genes using three databases (miRTarBase, miRDB, and TargetScan) and investigating the potential mechanisms of HCC recurrence with a ceRNA network, we constructed a ceRNA regulatory network consisting of lncRNAs (WARS2-IT1), miRNAs (hsa-miR-24-3p, hsa-miR-27a-3p, and hsa-miR-140-5p), and target genes (CD34, SEMA7A and HDAC7). Ye et al. and Yue et al. reported that lncRNA WARS2-IT1 affects the prediction of HCC recurrence.^{34,35} However, the mechanism behind lncRNA in HCC recurrence has not been investigated. Previous studies have indicated that the hsa-miR-24-3p miRNA is associated with many diseases, including atherosclerosis,³⁶ ischemic chest pain,³⁷ diabetes,³⁸ and Parkinson disease.³⁹ miR-24-3p plays a role in the development of multiple cancers. The expression of miR-24-3p in breast cancer has been found to be positively correlated with the expression of genes involved in hypoxia,⁴⁰ and miR-24-3p has been reported to inhibit the growth of nasopharyngeal carcinoma by directly regulating Jab1/CSN5.⁴¹ Chen et al. found that exosomes from cells with miR-24-3p deletion effectively block EMT.⁴² All the data suggest that miR-24-3p is associated with poor prognosis in multiple cancers, including HCC, which is consistent with our finding that patients with low expression of miR-24-3p had lower DFS and OS than those with high expression. CD34 is a downstream target of miR-24-3p. As a characteristic marker of blood vessel endothelium, elevated levels of CD34 in cancer generally indicate increased angiogenesis in tumor tissues.

Second, we used multiple approaches, including GSEA, ESTIMATE, TIS, mRNAsi, and immune cell infiltration, to evaluate the 9-IR-lncRNA risk score. We found that the high-risk group had a significant increase in nine molecular pathways, including DNA replication, spliceosome, mismatch repair, homologous recombination, nucleotide excision repair, base excision repair, proteasome, RNA degradation, and cell cycle. Subsequently, we compared the differences between the high- and low-risk groups regarding the relative expression of immune checkpoint genes, immune score, stromal score, TIS, and stem-like index. We found that the high-risk group had a lower level of expression of immune checkpoint genes, lower immune score, lower stromal score, lower TIS, and a higher stem-like index. These findings indicated that high-risk patients have a higher purity of tumor tissue with more stem-like features

Elevated angiogenesis may promote tumor growth and inhibit tumor growth by recruiting inflammatory cells to mediate inflammation and tumor immunity. Therefore, CD34 is a double-edged sword in cancer.

OIP5-AS1 regulates the HDAC7 epigenetic regulatory factor through miR-140-5p to promote metastasis of non-small cell lung cancer.⁴³ miR-140-5p plays a role in the recurrence of glioblastoma,⁴⁴ and the expression of miR-140-5p gradually decreases during the transformation of normal colon mucosa to tumor tissue with a further reduction in liver metastasis.⁴⁵ As such, miR-140-5p has been considered a critical regulator in the development and metastasis of colon cancer. Although miR-140-5p is associated with the poor prognosis of colon cancer, the relationship between miR-140-5p and the prognosis of HCC has not been reported. HDAC7 is a histone deacetylase that regulates tumorigenesis and tumor development in multiple cancers. In thoracic cancer, HDAC7 promotes tumorigenesis through downregulation of miR-4465, leading to upregulation of EphA2. HDAC7 has become a potential therapeutic target in thoracic cancer.⁴⁶ After analyzing TCGA data, Kim Freese et al. found that high expression levels of HDAC7 in HCC correlate with poor OS. However, there is no mechanistic study on the function of HDAC7 in HCC.⁴⁷

The expression of miR-27a-3p is elevated in multiple cancers, including triple-negative breast cancer,⁴⁸ glioblastoma,⁴⁹ and osteosarcoma,⁵⁰ and the expression level of miR-27a-3p is correlated with OS in these cancers. From the TCGA database, Xu et al. found that the expression of miR-27a-3p is elevated in HCC,⁵¹ but they did not further perform prognostic analysis or investigate the related mechanism. Semaphorin 7A (SEMA7A) is a highly effective immune regulator that is expressed in active lymphocytes and bone marrow cells. SEMA7A is significantly elevated in natural killer (NK) cells when stimulated by cytokines, specifically tagging active NK cells, leading to a strong capability of interferon- γ release.⁵² In lung adenocarcinoma clinical samples, high levels of SEMA7A predict poor response to EGFR-TKI therapy.⁵³ In breast cancer, SEMA7A is expressed in breast cells during gland degeneration, leading to changes in macrophage biology and the formation of lymphatic vessels, thereby promoting breast cancer metastasis. Overexpression of SEMA7A in estrogen receptor-positive (ER⁺) cells promotes cell growth.⁵⁴ Crump et al. reported that SEMA7A confers drug resistance to Fulvestrant in primary tumors and induces lung metastasis in ER⁺ breast cancer.²³ However, the study of SEMA7A in HCC has not been reported.

In conclusion, our research showed that the 9-IR-lncRNA model predicts the risk of HCC recurrence and that the lncRNA(WARS2-IT1)-miRNA(hsa-miR-24-3p, hsa-miR-27a-3p, and hsa-miR-140-5p)-mRNA (CD34, SEMA7A, and HDAC7) regulatory network is closely related to the prognosis of HCC, which is expected to become a potential target for HCC recurrence prediction and immunity therapy.

MATERIALS AND METHODS

Data acquisition and processing

RNA-sequencing (RNA-seq) data of HCC patients and clinical data of HCC samples were downloaded from the TCGA program

(<https://portal.gdc.cancer.gov>). The RNA-seq data were normalized using the trimmed mean of M-values (TMM) method in the “edgeR” package. The mean gene expression levels with a logarithmic fold change of >1 in all samples were retained. The normalized expression levels were further transformed using log₂(x+1) transformation. The initial dataset contains 374 HCC tissues and 50 adjacent normal tissues. Samples without radical resection treatment and complete survival data were removed, resulting in 319 radical resection HCC samples with complete clinical information for subsequent analysis in our study.

Identification of IR-lncRNAs

The IRGs were obtained from the Molecular Signatures Database v7.0 (Immune system process M13664, Immune response M19817).^{55,56} The IR-lncRNAs were identified by a Pearson correlation analysis between the IRGs and lncRNA expression levels in the HCC samples (Pearson correlation coefficient >0.7, p < 0.001). The “caret” package in R was used to randomly divide the samples into the training cohort (161 samples) and the testing cohort (158 samples). The expression levels of the IR-lncRNAs were generated from both the training cohort and testing cohort.

Construction of an IR-lncRNA signature and combined clinical risk score associated with DFS

The “survival,” “survminer,” and “forestplot” packages in R were used to perform univariate and multivariate Cox regression analyses. The “glmnet” package in R was used to perform a lasso regression analysis. IR-lncRNAs with a significant prognostic value (p < 0.05) were screened. In the multivariate Cox regression analysis, the IR-lncRNA signature was expressed as follows: $risk\ score = \sum_i^n coef_i * lncRNA_i$. The median risk score served as a cutoff value to classify the patients in the training cohort into the high- and low-risk groups. The testing cohort was also divided into high- and low-risk groups using the same cutoff value. We constructed a combined clinical characteristic risk score weighted sum of those clinical characteristics, of which the weights were estimates from the multivariable Cox regression analysis. The clinical risk score formula was as follows:

$Clinical_risk_score = 0.190 * AJCC_stage + 0.767 * Vascular_invasion + 0.500 * Child_grade + 0.332 * AFP + 0.128 * Age - 0.188 * Gender$. The “survminer” package in R was used to plot Kaplan-Meier survival curves. The “survivalROC” package in R was used to investigate the prognostic value of the IR-lncRNA risk score over time. A two-sided log rank p < 0.05 was considered significant in the survival analysis.

Construction of recurrence risk prediction nomogram based on clinical characteristics and IR-lncRNA risk score

A prognostic model was constructed by obtaining the independent prognostic factors and IR-lncRNA risk score. A time-dependent ROC curve was employed to compare the models. The most appropriate independent prognostic factors were selected for the prognostic model to construct a nomogram in the entire TCGA cohort. The

calibration plot and concordance index (C-index) were used to evaluate the capability of the nomogram to calibrate and discriminate against different factors (by a bootstrap method with 1,000 resamples).⁵⁷

Gene set enrichment analysis

To understand the KEGG pathways of the IR-lncRNA signature, a GSEA⁵⁸ was used to analyze the enrichment terms in the entire TCGA cohort. GSEA software version 4.1.0 (Cambridge, MA) was used to perform the analysis, and $p < 0.05$ was considered statistically significant.

ESTIMATE score, TIS, mRNAsi, and immune cell infiltration

ESTIMATE⁵⁹ is a tool to predict the tumor purity from gene signatures and calculate the following three scores: (1) stromal score, which predicts the presence of stromal cell types in tumor bulk; (2) immune score, which infers the infiltration of immune cells in tumor tissue; and (3) estimate score, which estimates the tumor purity. All patients were divided into high-risk and low-risk groups. Gene set variation analysis from the GSVA⁶⁰ package in R was performed to obtain the immune profile of the HCC samples. This function performs a nonparametric, unsupervised analysis for estimating variation of the given gene sets through the samples in the expression matrix, returning an enrichment score for each sample. The TIS scored with the GSVA method was used to predict the putative response to an anti-PD-L1 drug.⁶¹ This is a genetic profile reported as an acceptable predictor of clinical response to pembrolizumab across a wide variety of tumor types. In addition, the mRNAsi, which reflects gene expression, was adopted by means of one class logistic regression machine learning to quantify cancer stem-like indices.^{62,63} The fraction of 22 immune cell types infiltrated in tumor tissue was assessed using the “CIBERSORT” algorithm.⁶⁴

Construction of the ceRNA network

We used an miRcode database⁶⁵ to predict the interaction pairs between lncRNAs and miRNAs. mRNAs targeted by the miRNAs were retrieved from the miRDB, miRTarBase, and TargetScan databases.^{66–68} Candidate mRNAs were those identified by all three databases. Based on miRNA-lncRNA and miRNA-mRNA interaction pairs, the ceRNAs network was constructed and visualized using Cytoscape software (version 3.7.3; <https://www.cytoscape.org/>).

Expression level of seven ceRNA network genes in cell lines as detected by a qRT-PCR assay

Total RNA was extracted from the SNU-739 and SNU-368 HCC cell lines as well as the normal human hepatocellular cell line (MIHA cells). Total RNA was prepared using Total RNA Kit I (Omega, R6834-01), and cDNA was synthesized using a reverse transcription kit (TaKaRa Biotechnology, Shiga, Japan) according to the manufacturer’s protocol. Quantitative real-time PCR (qRT-PCR) was performed using SYBR Green Mix (TaKaRa RR820A, Japan) and a C1000 system (Bio-Red, Hercules, CA). Target gene expression values were normalized to human GAPDH, and target miRNA expression values were normalized to human U6.

Statistical analysis

Statistical analyses were performed using SPSS software version 26.0 (Armonk, NY) and R software version 4.0.2 (Auckland City, New Zealand). The χ^2 test or Fisher’s exact test was used to evaluate the qualitative variables as appropriate. All comparisons between the high-risk and low-risk groups were analyzed using Wilcoxon tests, as variables were either not normally distributed or variances were not equally distributed between groups. For all tests applied, differences were considered significant when $p < 0.05$.

DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available within the article.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.omtn.2021.11.006>.

ACKNOWLEDGMENTS

This study was supported partly by grants from the Natural Science Basic Research Program of Shaanxi (Program No. 2021JZ-35) and National Natural Science Foundation of China (81572699).

AUTHOR CONTRIBUTIONS

H.-M.Z. and X.-X.W. designed the study and contributed to study materials and consumables. H.-M.Z., X.-X.W., and L.A. conducted the study. X.-X.W., L.-H.W., L.A., W.P., J.-Y.R., and Q. Z. collected data. H.-M.Z., X.-X.W., and L.-H.W. performed the statistical analyses and interpreted the data. X.-X.W. and L.-H.W. wrote the manuscript. All authors approved the final manuscript.

DECLARATION OF INTERESTS

All authors declared no competing interests and approved the final manuscript.

REFERENCES

- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424.
- de Martel, C., Georges, D., Bray, F., Ferlay, J., and Clifford, G.M. (2020). Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob. Health* 8, e180–e190.
- Roayaie, S., Obeidat, K., Sposito, C., Mariani, L., Bhoori, S., Pellegrinelli, A., Labow, D., Llovet, J.M., Schwartz, M., and Mazzaferro, V. (2013). Resection of hepatocellular cancer ≤ 2 cm: results from two Western centers. *Hepatology* 57, 1426–1435.
- Diaz-Gonzalez, A., Reig, M., and Bruix, J. (2016). Treatment of hepatocellular carcinoma. *Dig. Dis.* 34, 597–602.
- Chan, A., Zhong, J., Berhane, S., Toyoda, H., Cucchetti, A., Shi, K., Tada, T., Chong, C., Xiang, B.D., Li, L.Q., et al. (2018). Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection. *J. Hepatol.* 69, 1284–1293.
- Kim, J., Piao, H.L., Kim, B.J., Yao, F., Han, Z., Wang, Y., Xiao, Z., Siverly, A.N., Lawhon, S.E., Ton, B.N., et al. (2018). Long noncoding RNA MALAT1 suppresses breast cancer metastasis. *Nat. Genet.* 50, 1705–1715.
- Li, C.H., and Chen, Y. (2016). Insight into the role of long noncoding RNA in cancer development and progression. *Int. Rev. Cell Mol. Biol.* 326, 33–65.

8. Goodall, G.J., and Wickramasinghe, V.O. (2021). RNA in cancer. *Nat. Rev. Cancer* 21, 22–36.
9. Momen-Heravi, F., and Bala, S. (2018). Emerging role of non-coding RNA in oral cancer. *Cell Signal.* 42, 134–143.
10. Zhang, Y., Li, Z., Chen, M., Chen, H., Zhong, Q., Liang, L., and Li, B. (2019). Identification of a new eight-long noncoding RNA molecular signature for breast cancer survival prediction. *DNA Cell Biol.* 38, 1529–1539.
11. Majidinia, M., and Yousefi, B. (2016). Long non-coding RNAs in cancer drug resistance development. *DNA Repair (Amst)* 45, 25–33.
12. Hu, X., Sood, A.K., Dang, C.V., and Zhang, L. (2018). The role of long noncoding RNAs in cancer: the dark matter matters. *Curr. Opin. Genet. Dev.* 48, 8–15.
13. Carpenter, S., and Fitzgerald, K.A. (2018). Cytokines and long noncoding RNAs. *Cold Spring Harb. Perspect. Biol.* 10: a028589.
14. Zhang, L., Xu, X., and Su, X. (2020). Noncoding RNAs in cancer immunity: functions, regulatory mechanisms, and clinical application. *Mol. Cancer* 19, 48.
15. Wei, L., Wang, X., Lv, L., Liu, J., Xing, H., Song, Y., Xie, M., Lei, T., Zhang, N., and Yang, M. (2019). The emerging role of microRNAs and long noncoding RNAs in drug resistance of hepatocellular carcinoma. *Mol. Cancer* 18, 147.
16. Sun, Y., Zhang, F., Wang, L., Song, X., Jing, J., Zhang, F., Yu, S., and Liu, H. (2019). A five lncRNA signature for prognosis prediction in hepatocellular carcinoma. *Mol. Med. Rep.* 19, 5237–5250.
17. Yan, J., Zhou, C., Guo, K., Li, Q., and Wang, Z. (2019). A novel seven-lncRNA signature for prognosis prediction in hepatocellular carcinoma. *J. Cell Biochem.* 120, 213–223.
18. Yang, Z., Yang, Y., Zhou, G., Luo, Y., Yang, W., Zhou, Y., and Yang, J. (2020). The prediction of survival in hepatocellular carcinoma based on A four long non-coding RNAs expression signature. *J. Cancer* 11, 4132–4144.
19. Gu, J., Zhang, X., Miao, R., Ma, X., Xiang, X., Fu, Y., Liu, C., Niu, W., and Qu, K. (2018). A three-long non-coding RNA-expression-based risk score system can better predict both overall and recurrence-free survival in patients with small hepatocellular carcinoma. *Aging (Albany NY)* 10, 1627–1639.
20. Gu, J.X., Zhang, X., Miao, R.C., Xiang, X.H., Fu, Y.N., Zhang, J.Y., et al. (2019). Six-long non-coding RNA signature predicts recurrence-free survival in hepatocellular carcinoma. *World J. Gastroenterol.* 25, 220–232.
21. Zhang, Q., Ning, G., Jiang, H., Huang, Y., Piao, J., Chen, Z., et al. (2020). 15-lncRNA-Based classifier-clinicopathologic nomogram improves the prediction of recurrence in patients with hepatocellular carcinoma. *Dis. Markers* 2020, 9180732.
22. Zhang, Z., Weng, W., Huang, W., Wu, B., Zhou, Y., Zhang, J., et al. (2020). A novel molecular-clinicopathologic nomogram to improve prognosis prediction of hepatocellular carcinoma. *Aging (Albany NY)* 12, 12896–12920.
23. Crump, L.S., Wyatt, G.L., Rutherford, T.R., Richer, J.K., Porter, W.W., and Lyons, T.R. (2021). Hormonal regulation of semaphorin 7a in ER(+) breast cancer drives therapeutic resistance. *Cancer Res.* 81, 187–198.
24. Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D.R., Albright, A., Cheng, J.D., Kang, S.P., Shankaran, V., et al. (2017). IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* 127, 2930–2940.
25. Forner, A., Reig, M., and Bruix, J. (2018). Hepatocellular carcinoma. *Lancet* 391, 1301–1314.
26. Pinyol, R., Montal, R., Bassaganyas, L., Sia, D., Takayama, T., Chau, G.Y., Mazzaferro, V., Roayaie, S., Lee, H.C., Kokudo, N., et al. (2019). Molecular predictors of prevention of recurrence in HCC with sorafenib as adjuvant treatment and prognostic factors in the phase 3 STORM trial. *Gut* 68, 1065–1075.
27. Fu, Q., Yang, F., Xiang, T., Huai, G., Yang, X., Wei, L., Yang, H., and Deng, S. (2018). A novel microRNA signature predicts survival in liver hepatocellular carcinoma after hepatectomy. *Sci. Rep.* 8, 7933.
28. Klingenberg, M., Matsuda, A., Diederichs, S., and Patel, T. (2017). Non-coding RNA in hepatocellular carcinoma: mechanisms, biomarkers and therapeutic targets. *J. Hepatol.* 67, 603–618.
29. Zhao, Q.J., Zhang, J., Xu, L., and Liu, F.F. (2018). Identification of a five-long non-coding RNA signature to improve the prognosis prediction for patients with hepatocellular carcinoma. *World J. Gastroenterol.* 24, 3426–3439.
30. Wang, Z., Wu, Q., Feng, S., Zhao, Y., and Tao, C. (2017). Identification of four prognostic lncRNAs for survival prediction of patients with hepatocellular carcinoma. *PeerJ* 5, e3575.
31. Liao, X., Yang, C., Huang, R., Han, C., Yu, T., Huang, K., Liu, X., Yu, L., Zhu, G., Su, H., et al. (2018). Identification of potential prognostic long non-coding RNA biomarkers for predicting survival in patients with hepatocellular carcinoma. *Cell Physiol. Biochem.* 48, 1854–1869.
32. Gao, J., Kwan, P.W., and Shi, D. (2010). Sparse kernel learning with LASSO and Bayesian inference algorithm. *Neural Netw.* 23, 257–264.
33. McNeish, D.M. (2015). Using lasso for predictor selection and to assuage overfitting: a method long overlooked in behavioral sciences. *Multivariate Behav. Res.* 50, 471–484.
34. Ye, J., Zhang, J., Lv, Y., Wei, J., Shen, X., Huang, J., Wu, S., and Luo, X. (2019). Integrated analysis of a competing endogenous RNA network reveals key long non-coding RNAs as potential prognostic biomarkers for hepatocellular carcinoma. *J. Cell Biochem.* 120, 13810–13825.
35. Yue, C., Ren, Y., Ge, H., Liang, C., Xu, Y., Li, G., and Wu, J. (2019). Comprehensive analysis of potential prognostic genes for the construction of a competing endogenous RNA regulatory network in hepatocellular carcinoma. *Oncotargets Ther.* 12, 561–576.
36. Li, Y.Y., Zhang, S., Wang, H., Zhang, S.X., Xu, T., Chen, S.W., Zhang, Y., and Chen, Y. (2021). Identification of crucial genes and pathways associated with atherosclerotic plaque in diabetic patients. *Pharmgenomics Pers. Med.* 14, 211–220.
37. Wei, W., Peng, J., and Shen, T. (2019). Rosuvastatin alleviates ischemia/reperfusion injury in cardiomyocytes by downregulating hsa-miR-24-3p to target upregulated uncoupling protein 2. *Cell Reprogram.* 21, 99–107.
38. Demirsoy, I.H., Ertural, D.Y., Balci, S., Cinkir, U., Sezer, K., Tamer, L., and Aras, N. (2018). Profiles of circulating miRNAs following metformin treatment in patients with type 2 diabetes. *J. Med. Biochem.* 37, 499–506.
39. Uwatoko, H., Hama, Y., Iwata, I.T., Shirai, S., Matsushima, M., Yabe, I., Utsumi, J., and Sasaki, H. (2019). Identification of plasma microRNA expression changes in multiple system atrophy and Parkinson's disease. *Mol. Brain* 12, 49.
40. Camps, C., Saini, H.K., Mole, D.R., Choudhry, H., Reczko, M., Guerra-Assuncao, J.A., Tian, Y.M., Buffa, F.M., Harris, A.L., Hatzigeorgiou, A.G., et al. (2014). Integrated analysis of microRNA and mRNA expression and association with HIF binding reveals the complexity of microRNA expression regulation under hypoxia. *Mol. Cancer* 13, 28.
41. Wang, S., Pan, Y., Zhang, R., Xu, T., Wu, W., Zhang, R., Wang, C., Huang, H., Calin, C.A., Yang, H., et al. (2016). Hsa-miR-24-3p increases nasopharyngeal carcinoma radiosensitivity by targeting both the 3'UTR and 5'UTR of Jab1/CSN5. *Oncogene* 35, 6096–6108.
42. Chen, Q.L., Xie, C.F., Feng, K.L., Cui, D.Y., Sun, S.L., Zhang, J.C., Xiong, C.M., Huang, J.H., and Chong, Z. (2020). microRNAs carried by exosomes promote epithelial-mesenchymal transition and metastasis of liver cancer cells. *Am. J. Transl. Res.* 12, 6811–6826.
43. Mei, J., Liu, G., Wang, W., Xiao, P., Yang, D., Bai, H., and Li, R. (2020). OIP5-AS1 modulates epigenetic regulator HDAC7 to enhance non-small cell lung cancer metastasis via miR-140-5p. *Oncol. Lett.* 20, 7.
44. Bo, L.J., Wei, B., Li, Z.H., Wang, Z.F., Gao, Z., and Miao, Z. (2015). Bioinformatics analysis of miRNA expression profile between primary and recurrent glioblastoma. *Eur. Rev. Med. Pharmacol. Sci.* 19, 3579–3586.
45. Zhai, H., Fesler, A., Ba, Y., Wu, S., and Ju, J. (2015). Inhibition of colorectal cancer stem cell survival and invasive potential by hsa-miR-140-5p mediated suppression of Smad2 and autophagy. *Oncotarget* 6, 19735–19746.
46. Li, Q.G., Xiao, T., Zhu, W., Yu, Z.Z., Huang, X.P., Yi, H., Lu, S.S., Tang, Y.Y., Huang, W., and Xiao, Z.Q. (2020). HDAC7 promotes the oncogenicity of nasopharyngeal carcinoma cells by miR-4465-EphA2 signaling axis. *Cell Death Dis.* 11, 322.
47. Freese, K., Seitz, T., Dietrich, P., Lee, S., Thasler, W.E., Bosserhoff, A., and Hellerbrand, C. (2019). Histone deacetylase expressions in hepatocellular carcinoma

- and functional effects of histone deacetylase inhibitors on liver cancer cells in vitro. *Cancers (Basel)* *11*: 1587.
48. Banerjee, S., Kalyani, Y.S., and Karunakaran, D. (2020). Identification of mRNA and non-coding RNA hubs using network analysis in organ tropism regulated triple negative breast cancer metastasis. *Comput. Biol. Med.* *127*, 104076.
 49. Salmani, T., Ghaderian, S., Hajiesmaeili, M., Rezaeimirghaed, O., Hoseini, M.S., Rakhshan, A., Nasiri, M.J., Ghaedi, H., and Akbarzadeh, R. (2021). Hsa-miR-27a-3p and epidermal growth factor receptor expression analysis in glioblastoma FPPE samples. *Asia Pac. J. Clin. Oncol.* *17*: e185-e190.
 50. Luo, T., Yi, X., and Si, W. (2017). Identification of miRNA and genes involving in osteosarcoma by comprehensive analysis of microRNA and copy number variation data. *Oncol. Lett.* *14*, 5427–5433.
 51. Xu, W., Yu, S., Xiong, J., Long, J., Zheng, Y., and Sang, X. (2020). CeRNA regulatory network-based analysis to study the roles of noncoding RNAs in the pathogenesis of intrahepatic cholangiocellular carcinoma. *Aging (Albany NY)* *12*, 1047–1086.
 52. Ghofrani, J., Lucar, O., Dugan, H., Reeves, R.K., and Jost, S. (2019). Semaphorin 7A modulates cytokine-induced memory-like responses by human natural killer cells. *Eur. J. Immunol.* *49*, 1153–1166.
 53. Kinehara, Y., Nagatomo, I., Koyama, S., Ito, D., Nojima, S., Kurebayashi, R., Nakanishi, Y., Suga, Y., Nishijima-Futami, Y., Osa, A., et al. (2018). Semaphorin 7A promotes EGFR-TKI resistance in EGFR mutant lung adenocarcinoma cells. *JCI Insight* *3*:e123093.
 54. Elder, A.M., Tamburini, B., Crump, L.S., Black, S.A., Wessells, V.M., Schedin, P.J., Borges, V.F., and Lyons, T.R. (2018). Semaphorin 7A promotes macrophage-mediated lymphatic remodeling during postpartum mammary gland involution and in breast cancer. *Cancer Res.* *78*, 6473–6485.
 55. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* *462*, 108–112.
 56. Lin, Y., Pan, X., Chen, Z., Lin, S., and Chen, S. (2020). Identification of an immune-related nine-lncRNA signature predictive of overall survival in colon cancer. *Front. Genet.* *11*, 318.
 57. Iasonos, A., Schrag, D., Raj, G.V., and Panageas, K.S. (2008). How to build and interpret a nomogram for cancer prognosis. *J. Clin. Oncol.* *26*, 1364–1370.
 58. Damian, D., and Gorfine, M. (2004). Statistical concerns about the GSEA procedure. *Nat. Genet.* *36*, 663.
 59. Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Trevino, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* *4*, 2612.
 60. Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* *14*, 7.
 61. Ott, P.A., Bang, Y.J., Piha-Paul, S.A., Razak, A., Bennouna, J., Soria, J.C., Rugo, H.S., Cohen, R.B., O'Neil, B.H., Mehnert, J.M., et al. (2019). T-Cell-Inflamed gene-expression profile, programmed death ligand 1 expression, and tumor mutational burden predict efficacy in patients treated with pembrolizumab across 20 cancers: KEYNOTE-028. *J. Clin. Oncol.* *37*, 318–327.
 62. Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kaminska, B., Huelsken, J., Omberg, L., Gevaert, O., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* *173*, 338–354.
 63. Lian, H., Han, Y.P., Zhang, Y.C., Zhao, Y., Yan, S., Li, Q.F., Wang, B.C., Wang, J.J., Meng, W., Yang, J., et al. (2019). Integrative analysis of gene expression and DNA methylation through one-class logistic regression machine learning identifies stemness features in medulloblastoma. *Mol. Oncol.* *13*, 2227–2245.
 64. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* *12*, 453–457.
 65. Jeggari, A., Marks, D.S., and Larsson, E. (2012). miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics* *28*, 2062–2063.
 66. Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife* *4*:e05005.
 67. Chou, C.H., Chang, N.W., Shrestha, S., Hsu, S.D., Lin, Y.L., Lee, W.H., Yang, C.D., Hong, H.C., Wei, T.Y., Tu, S.J., et al. (2016). miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.* *44*, D239–D247.
 68. Wong, N., and Wang, X. (2015). miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* *43*, D146–D152.