



# Featured lncRNA-based signature for discriminating prognosis and progression of hepatocellular carcinoma

Huiyuan Qiu<sup>1,2</sup> · Bo Jiang<sup>3</sup> · Yinqi Chen<sup>1</sup> · Zhaoyi Lin<sup>1</sup> · Wenjie Zheng<sup>1,2</sup> · Xiaolei Cao<sup>1</sup>

Received: 1 December 2023 / Revised: 16 January 2024 / Accepted: 20 January 2024 / Published online: 12 February 2024  
© The Author(s), under exclusive licence to Institute of Plant Genetics Polish Academy of Sciences 2024, corrected publication 2024

## Abstract

Long non-coding RNAs (lncRNAs) have been implicated in carcinogenesis and progression of hepatocellular carcinoma (HCC). This study aimed to identify a robust lncRNA signature for predicting the survival of HCC patients. We performed an integrated analysis of the lncRNA expression profiling in The Cancer Genome Atlas (TCGA)—liver hepatocellular carcinoma database to identify the prognosis-related lncRNA for the HCC. The HCC cohort was randomly divided into a training set ( $n = 250$ ) and a testing set ( $n = 113$ ). Following a two-step screening, we identified an 18-lncRNA signature risk score. The high-risk subgroups had significantly shorter survival time than the low-risk group in both the training set ( $P < 0.0001$ ) and the testing set ( $P = 0.005$ ). Stratification analysis revealed that the prognostic value of the lncRNA-based signature was independent of the tumor stage and pathologic stage. The area under the receiver operating characteristic curve (AUROC) of the 18-lncRNA signature risk score was 0.826 (95%CI, 0.764–0.888), 0.817 (95%CI, 0.759–0.876), and 0.799 (95%CI, 0.731–0.867) for 1-year, 3-year, and 5-year follow-up, respectively. Bioinformatics analyses indicated that the 18 lncRNA might mediate cell cycle, DNA replication processes, and canonical cancer-related pathways, in which MCM3AP-AS1 was a potential target for HCC. In conclusion, the 18-lncRNA signature was a robust predictive biomarker for the prognosis and progression of HCC.

**Keywords** lncRNA · Hepatocellular carcinoma · Prognosis · Biomarker · MCM3AP-AS1

## Introduction

According to the statistics, about 906,000 cases were newly diagnosed with liver cancer with 830,000 deaths annually worldwide (Sung et al. 2021). Hepatocellular carcinoma (HCC), accounting for approximately 75% of all liver cancer

cases, could be caused by multiple factors, including hepatitis B virus (HBV) or hepatitis C virus (HCV) infection, non-alcoholic steatohepatitis (NASH), autoimmune diseases (Wallace et al. 2015; Goh et al. 2015). For HCC patients at early stages, surgical resection and liver transplantation are conventional treatments. For decades, HCC patients have benefited from novel approaches, including radiofrequency ablation (RFA), transcatheter arterial chemoembolization (TACE), targeted therapy, T cell-based therapy, and gene therapy (Hsu et al. 2012). However, the overall survival of HCC patients, especially for cases diagnosed at advanced stages, remains dismal (Llovet et al. 2016). Although various kinds of molecules have been evaluated as potential targets or markers, they lack widely acknowledged biomarkers to efficiently predict the overall survival of HCC patients. As is well known, intricate genetic and epigenetic transformations contribute to the genesis and development of HCC. Hence, it is of great importance to explore novel genes as robust biomarkers and potential targets for HCC.

According to the transcriptome sequencing, approximately 70% of the human genome has been identified as

---

Communicated by: Ewa Ziętkiewicz

---

Huiyuan Qiu and Bo Jiang contributed equally to this work.

---

✉ Wenjie Zheng  
wenjiezhang@ntu.edu.cn

✉ Xianglei Cao  
caoxiaolei@126.com

<sup>1</sup> Medical School of Nantong University, Nantong 226001, China

<sup>2</sup> Research Center of Clinical Medicine, Affiliated Hospital of Nantong University, Nantong, Jiangsu, China

<sup>3</sup> Department of Gastroenterology, Suqian First People's Hospital, Suqian, Jiangsu, China

non-coding RNAs (ncRNAs), which could be further stratified into long ncRNAs (lncRNAs) and small ncRNAs based on their transcript size (Yan et al. 2015). lncRNAs, a form of ncRNAs longer than 200 nucleotides (nt) in length, exert regulatory effects on gene transcription by epigenetic activities with refrained protein coding activities (Iyer et al. 2015). Accumulating evidence has suggested the critical roles of lncRNAs in multiple physiological and pathological processes (Fatica and Bozzoni 2014; Ulitsky and Bartel 2013; Zhao et al. 2020). Current studies demonstrate that dysregulated expression of lncRNAs is widely involved in the initiation and progression of various cancer types and is recognized as a hallmark feature (Schmitt and Chang 2016; Grixti and Ayers 2020). For HCC, lncRNAs could regulate the expression of key oncogenes and subsequently facilitate malignant behaviors, including proliferation, chemoresistance, and invasive capacity. Therefore, the lncRNAs may serve as potential anti-tumor targets (Davalos and Esteller 2019). In addition, based on the expression features of serum or tissues in specific status, lncRNAs have been evaluated as biomarkers of diagnosis and prognosis for HCC. Interestingly, lncRNAs enriched in serum exosomes derived from HCC patients have also been considered novel prognostic markers and therapeutic targets for HCC (Lee et al. 2019; Huang et al. 2020). Although single detection of the lncRNA has shown certain significance, a risk model established by multiple lncRNAs may have better performance in predicting the overall survival of HCC patients (Pea et al. 2020). In the current study, we conducted a genome-wide analysis of lncRNA expression profile integrating clinical information of 363 HCC patients from The Cancer Genome Atlas (TCGA) datasets. Based on the screening and validation, we identified a robust 18-lncRNA signature to predict the survival of HCC patients.

## Materials and methods

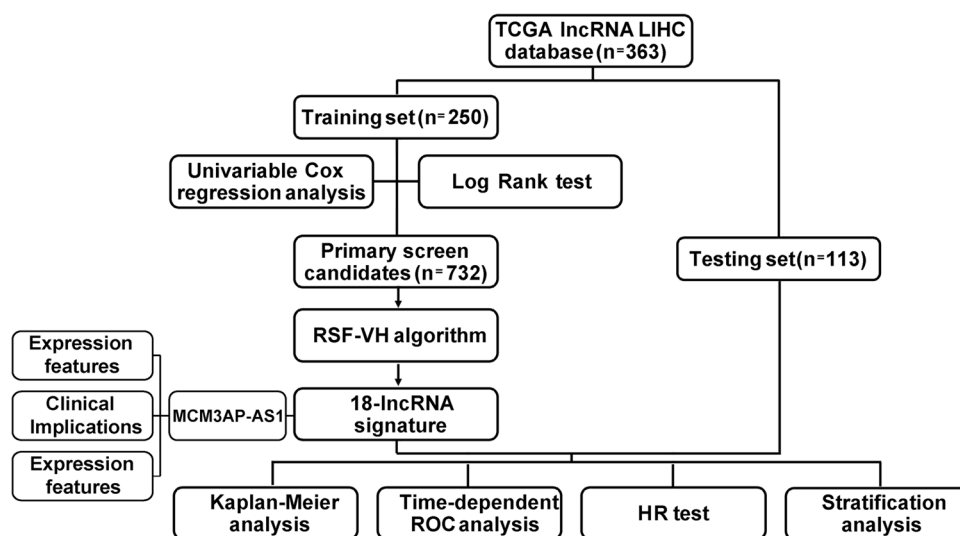
### Acquisition of lncRNA expression and clinical information

The RNA transcriptional data as well as survival and clinical information were downloaded from hepatocellular carcinoma (LIHC) datasets of TCGA. After matching the data sets with corresponding prognostic information and excluding the normal samples, a total of 363 samples were included in this study with 60,483 transcriptional RNAs and 15,121 lncRNAs-related targets. The total set was randomly divided into a training cohort (250 samples) and a testing cohort (113 samples) for further analysis. The clinical characteristics of HCC patients in TCGA cohorts are presented in Table S1. Furthermore, we extracted expression data of lncRNAs from three GEO datasets (GSE55092, GSE76247, and GSE36411). The expression of the lncRNAs in blood samples-derived exosomes was obtained from exoRBase 2.0 (<http://www.exorbase.org/exoRBaseV2/toIndex>) (Lai et al. 2022).

### Identification of the lncRNA-based signature

The identification of the lncRNA signature was based on the two-step procedure (Fig. 1). For the initial step, univariate Cox regression and log-rank test were conducted to screen potential genes associated with HCC prognosis. The lncRNAs with  $P$  values  $< 0.05$  from the analyses aforementioned were intersected to obtain the primary prognosis-related lncRNAs. Subsequently, the primary lncRNAs were further calculated by random survival forest-variable hunting (RSF-VH) algorithm (R package Random Forest SRC) in the variable selection function of BRB-Array Tools. Based

**Fig. 1** The flowchart of the present study. TCGA, the Cancer Genome Atlas; ROC, receiver operating characteristics; HR, hazard ratio



on 100 random iterations, lncRNAs with  $P$  values  $< 0.05$  were considered significantly associated with the prognostic classification and enrolled into the risk signature. Then, the formula of the lncRNA risk score was established based on the expression of each lncRNA weighted by regression coefficients in the univariate Cox regression analysis. The risk score of each patient was calculated for further analysis.

### Predictive accuracy evaluation of the risk model

Based on the median value of the lncRNA-based risk score, the training or testing cohort was divided into two groups, namely the high-risk group and the low-risk group, respectively. Prognostic differences between high- and low-risk groups were analyzed by Kaplan–Meier curves with log-rank tests. Stratified analysis was performed to evaluate the lncRNA signature as an independent predictor for survival of HCC patients. Besides, the predictive performance of the lncRNA signature was evaluated by receiver operating characteristic (ROC) curves by using R Package “risksetROC.”

### Functional enrichment

Based on the high-risk and low-risk groups, we performed gene set enrichment analysis (GSEA, 3.0 software, MA, USA; <http://www.broad.mit.edu/gsea>) regarding the Molecular Signatures Database (<http://software.broadinstitute.org/gsea/msigdb>). The possible pathways and functions enriched in each group were sorted according to  $P$  value and normalized enrichment score (NES). Functional enrichment in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) was assessed with the hypergeometric test by using the R package “clusterProfiler.”

Co-expression analysis was conducted using expression profiles of lncRNAs, miRNAs, and mRNAs in the entire TCGA-LIHC dataset. lncRNA-related mRNAs and miRNAs were identified by calculating the Pearson correlation coefficients. Genes significantly correlated with the signature lncRNAs were included in the analysis. The potential interactions among lncRNA signature, mRNA, and miRNA were visualized by using Cytoscape software.

### Statistical analysis

Statistical analyses were performed by using IBM SPSS software (version 22.0), R software (version 3.5.1), and Bioconductor (version 3.8). Univariate Cox regression analysis was conducted with clinical variables and lncRNA-signature risk score to calculate hazard ratios (HRs) with a 95% confidence interval (CI). Kaplan–Meier analysis was performed to compare the overall survival between high- or low-risk groups with a log-rank test. ROC analysis was performed to evaluate the prognostic prediction of lncRNA signature

and other clinicopathological variables by assessing the area under the curves (AUC).  $P$  value  $< 0.05$  was considered statistically significant.

## Results

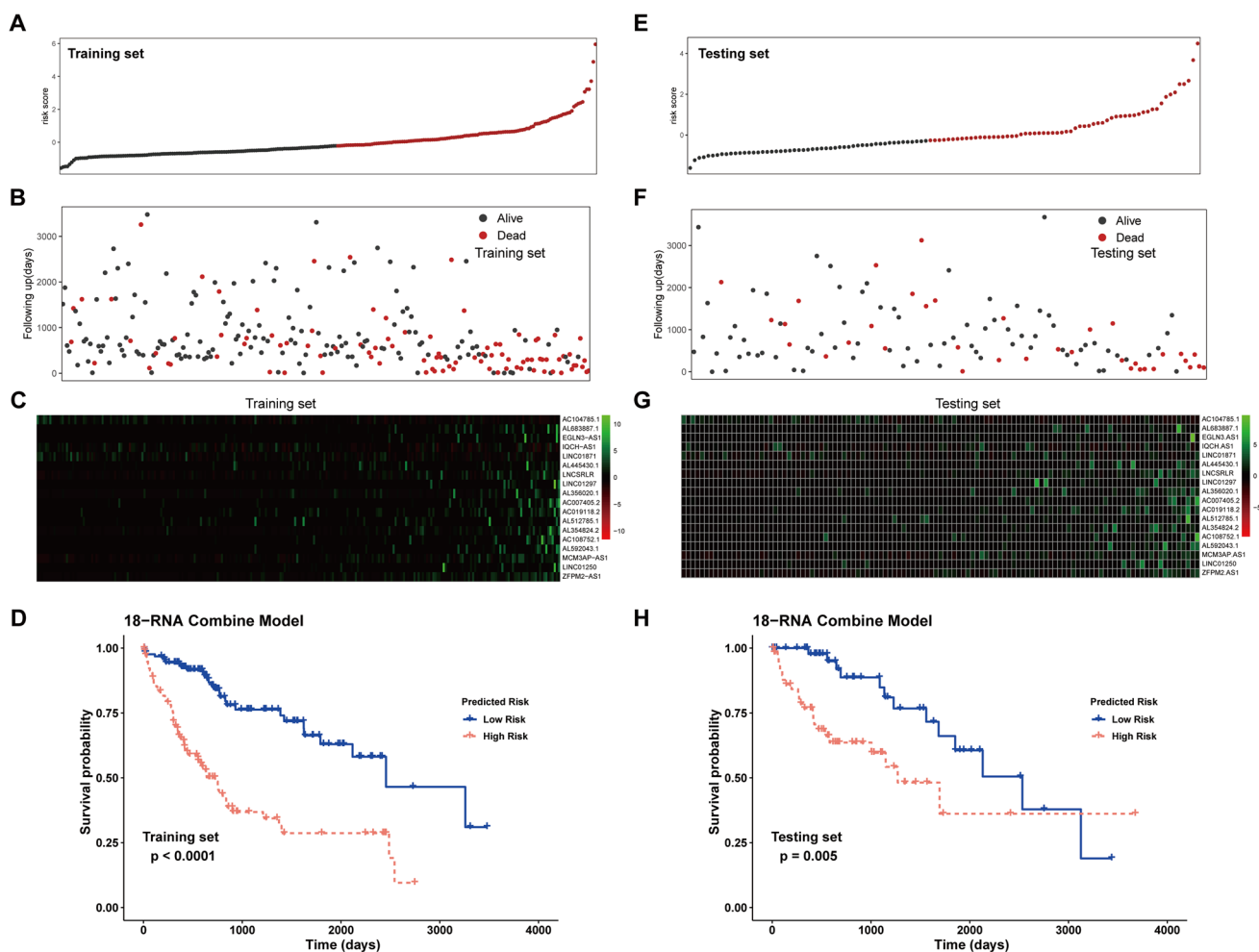
### Identification of prognostic lncRNAs in training set

Two procedures were performed to initially screen candidate lncRNAs with prognostic value in the training set ( $n = 250$ ). First, we identified 2363 prognosis-related lncRNAs as potential candidates by using single-variable Cox regression ( $P < 0.05$ ). Subsequently, based on the median expression value of each lncRNA, the training set was divided into a high-expression group and a low-expression group. Then, 1306 lncRNAs were correlated with the survival of HCC patients by log Rank test ( $P < 0.05$ ). By intersecting the two analyses above, a total of 732 lncRNAs were enrolled into the RSF-VH algorithm with 100 iterations. Following that, 18 lncRNAs were identified as the prognostic signatures (Table S2).

### Establishment and validation of the formula for risk score

To comprehensively investigate the correlation of these 18 identified lncRNAs with the prognosis of HCC, we established a formula for risk score to predict the overall survival of HCC patients based on the expression of these lncRNAs and Cox coefficients. The formula was presented as follows: Risk score =  $(1.551245352 \times AL354824.2) + (1.193980091 \times LNCSRLR) + (18.69307557 \times AC019118.2) + (0.373287815 \times ZFPM2-AS1) + (13.99500262 \times AL356020.1) + (1.869701407 \times AL683887.1) + (5.343933289 \times LINC01250) + (-0.261739147 \times LINC01871) + (0.704823146 \times EGLN3-AS1) + (3.049698013 \times AL445430.1) + (1.665645142 \times LINC01297) + (1.92682538 \times AL512785.1) + (-0.739161379 \times IQCH-AS1) + (1.933011211 \times MCM3AP-AS1) + (1.847129616 \times AL592043.1) + (6.017576414 \times AC007405.2) + (0.555545078 \times AC108752.1) + (-0.414728141 \times AC104785.1)$ . Herein, positive coefficients suggested poorer survival of HCC patients, while negative coefficients indicated longer survival.

Then, the risk-score formula was applied to calculate for each patient in the training set. The distribution of risk scores, survival status, and expression features of the 18-lncRNA signature risk score in the training set were shown in Fig. 2A–C, respectively. The patients were divided into high-risk group and low-risk group based on



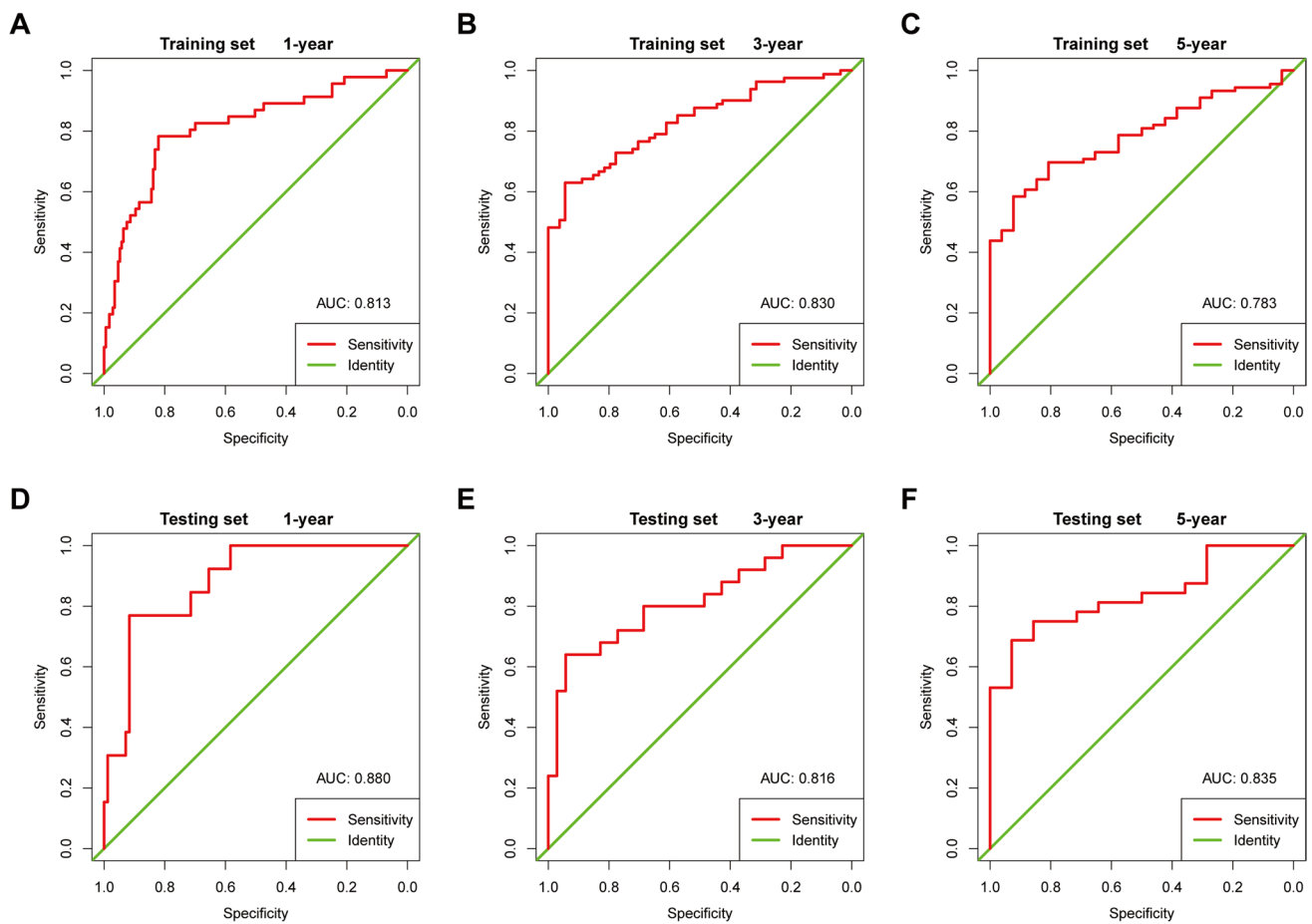
**Fig. 2** Construction and validation of the 18-lncRNA signature risk score. **A** and **E** lncRNA signature risk score distribution in training set and testing set. **B** and **F** survival status of patients at each group in training set and testing set. **C** and **G** Heat map of the lncRNA expres-

sion profiles in training set and testing set. **D** and **H** The Kaplan-Meier curves of cases in high-risk group and low-risk group in training set and testing set

the median risk score as the cut-off value. For the training set, the Kaplan-Meier analysis indicated that patients in the high-risk subgroup had significantly shorter overall survival compared with those cases in the low-risk group ( $P < 0.0001$ , Fig. 2D). Subsequently, the testing cohort was also divided into high- and low-risk groups based on the risk core (Fig. 2E). Survival status and expression features of the lncRNA signatures in the testing set were shown in Fig. 2F and G, respectively. Consistent with the observation in the training set, the high-risk score was significantly correlated with poorer prognosis than the low-risk score in the testing cohort ( $P = 0.005$ , Fig. 2H).

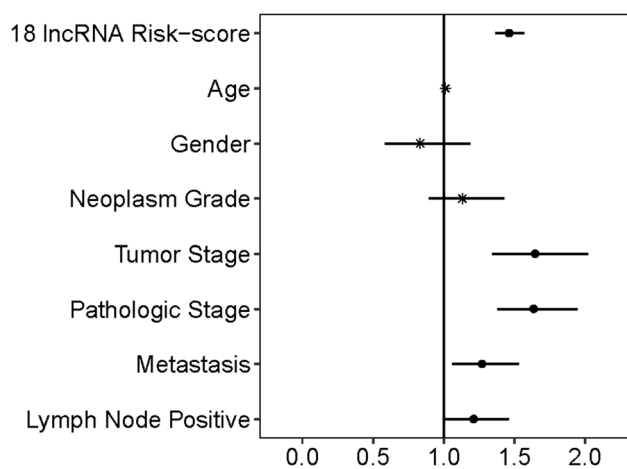
### Validation of the 18-lncRNA prognostic signature for survival prediction

Given the prognostic values of the 18-lncRNA signature for HCC, then we further evaluated its predictive performance for OS in the training set and testing set by AUC. As shown in Fig. 3A–C, the AUC for the 18-lncRNA signature in predicting 1-year, 3-year, and 5-year OS for HCC patients in the training set were 0.813, 0.830, and 0.783, respectively. Consistently, the 18-lncRNA signature also showed predictive performance in the testing set, for which the AUC of 1-year, 3-year, and



**Fig. 3** ROC analysis of the 18-lncRNA signature risk score. **A** and **D** ROC analysis was conducted to evaluate the survival prediction of the 18-lncRNA signature risk score within 1 year for HCC patients

in training set and testing set. **B** and **E** 3-year ROC analysis for HCC patients in training set and testing set; **C** and **F** 5-year ROC analysis for HCC patients in training set and testing set



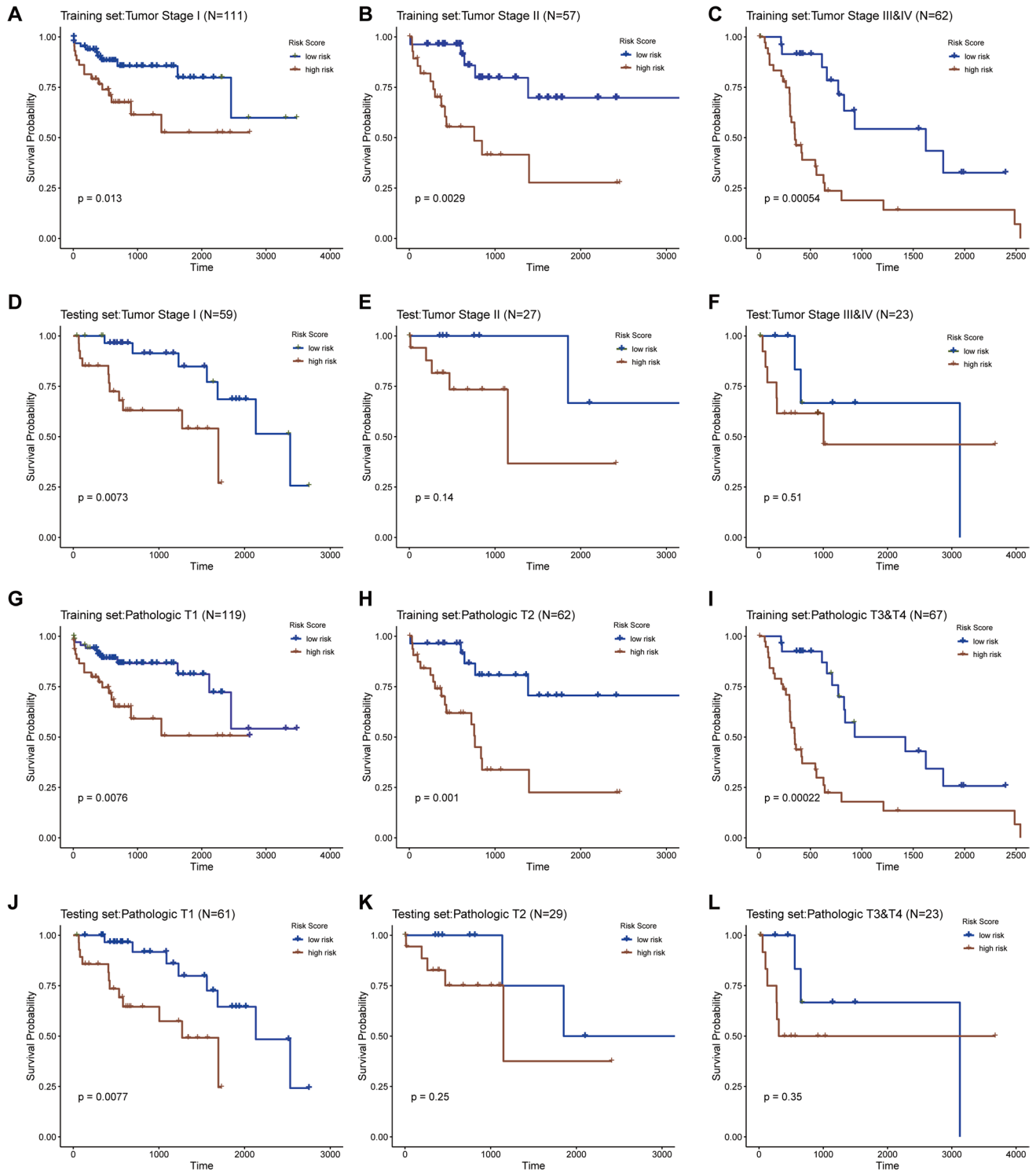
**Fig. 4** HR test of the 18-lncRNA signature risk score and clinical variables. Univariate Cox regression analysis was conducted to evaluate the prognostic value of 18-lncRNA signature risk score and clinical variables

5-year OS was 0.880, 0.816, and 0.835, respectively (Fig. 3D–F). In addition, 18-lncRNA prognostic signature was also beneficial for the prediction of disease-free survival (DFS, Figure S1) and progression-free survival (PFS, Figure S2) of HCC patients. These results indicated that the 18-lncRNA-based signature had excellent

**Table 1** HR test of the 18-lncRNA risk score and clinical variables

Variables	HR	95% CI	P value
18 lncRNA risk score	1.466	1.369–1.569	4.90E-28
Age	1.013	0.999–1.027	0.066
Gender	0.834	0.585–1.190	0.318
Neoplasm grade	1.132	0.896–1.430	0.299
Tumor stage	1.650	1.346–2.023	1.48E-06
Pathologic stage	1.639	1.380–1.947	1.82E-08
Metastasis	1.274	1.059–1.532	0.010
Lymph node status	1.214	1.010–1.461	0.0392

HR, hazard ratio; CI, confidential interval



**Fig. 5** Stratification analyses based on tumor stages and pathologic stages. **A–C** Kaplan–Meier curves of high-risk group and low-risk group at different tumor stages in training set. **D–E** Kaplan–Meier curves of high-risk group and low-risk group at different tumor stages

in testing set. **F–H** Kaplan–Meier curves of high-risk group and low-risk group at different pathologic stages in training set. **I–K** Kaplan–Meier curves of high-risk group and low-risk group at different pathologic stages in testing set

**Table 2** ROC analysis of the 18-lncRNA risk score and clinical variables

Factors	AUROC	95%CI	
		Low	High
1-year prediction			
18-lncRNA risk score	0.826	0.764	0.888
Tumor stage	0.650	0.572	0.728
Metastasis	0.510	0.448	0.572
Lymph node status	0.544	0.484	0.604
Pathologic stage	0.664	0.587	0.740
3-year prediction			
18-lncRNA risk score	0.817	0.759	0.876
Tumor stage	0.664	0.592	0.738
Metastasis	0.559	0.496	0.622
Lymph node status	0.549	0.486	0.612
Pathologic stage	0.669	0.599	0.739
5-year prediction			
18-lncRNA risk score	0.799	0.731	0.867
Tumor stage	0.639	0.549	0.730
Metastasis	0.618	0.550	0.686
Lymph node status	0.603	0.532	0.674
Pathologic stage	0.657	0.571	0.742

AUROC, area under receiver operating characteristic curve; CI, confidential interval

sensitivity and specificity in predicting the survival of HCC patients.

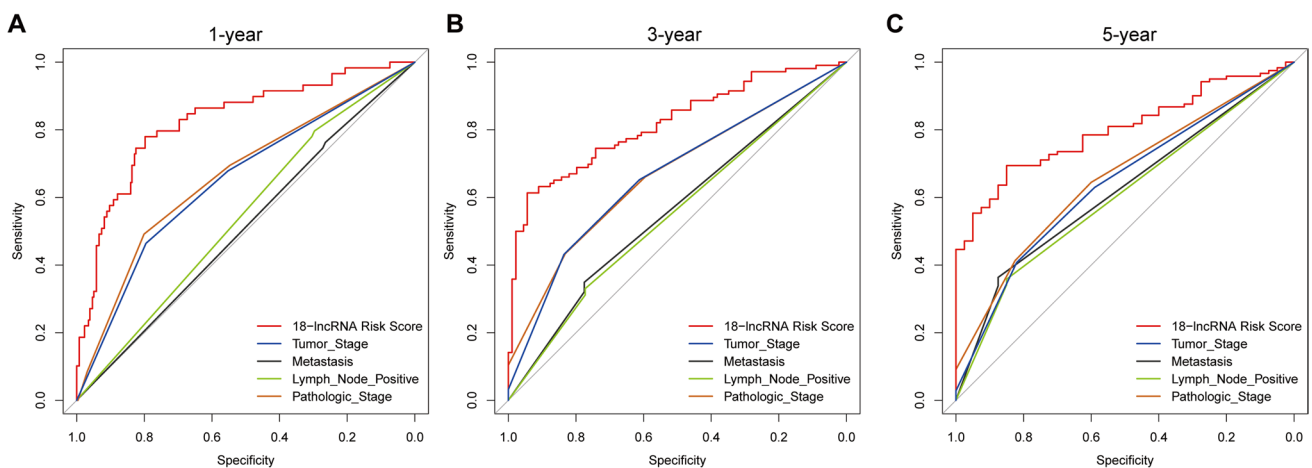
### Validation of the 18-lncRNA risk score by Cox regression analysis

Next, we investigated the prognostic value of the 18-lncRNA risk score as well as clinical variables by Cox

regression analysis in the entire set. As is shown in Fig. 4, the 18-lncRNA signature risk score was a prognostic candidate for HCC patients (HR = 1.466, 95%CI, 1.368–1.569,  $P < 0.001$ ). In addition, tumor stage (HR = 1.650, 95%CI, 1.346–2.023,  $P < 0.001$ ), pathological stage (HR = 1.639, 95%CI, 1.380–1.947,  $P < 0.001$ ), metastasis (HR = 1.274, 95%CI, 1.059–1.532,  $P = 0.01$ ), and lymph node status (HR = 1.214, 95%CI, 1.010–1.461,  $P = 0.039$ ) were also potential OS-related factors for HCC. However, other variables like age, gender, and neoplasm grade were not statistically significant (Table 1). To further test the efficacy of the risk signature for the prognosis of HCC patients, we randomly performed a log-rank test of the TCGA LIHC cohort (one-third of entire samples/ per time). Consistent with the results above, the risk signature presented excellent performance in all the 10-time random tests (Table S3).

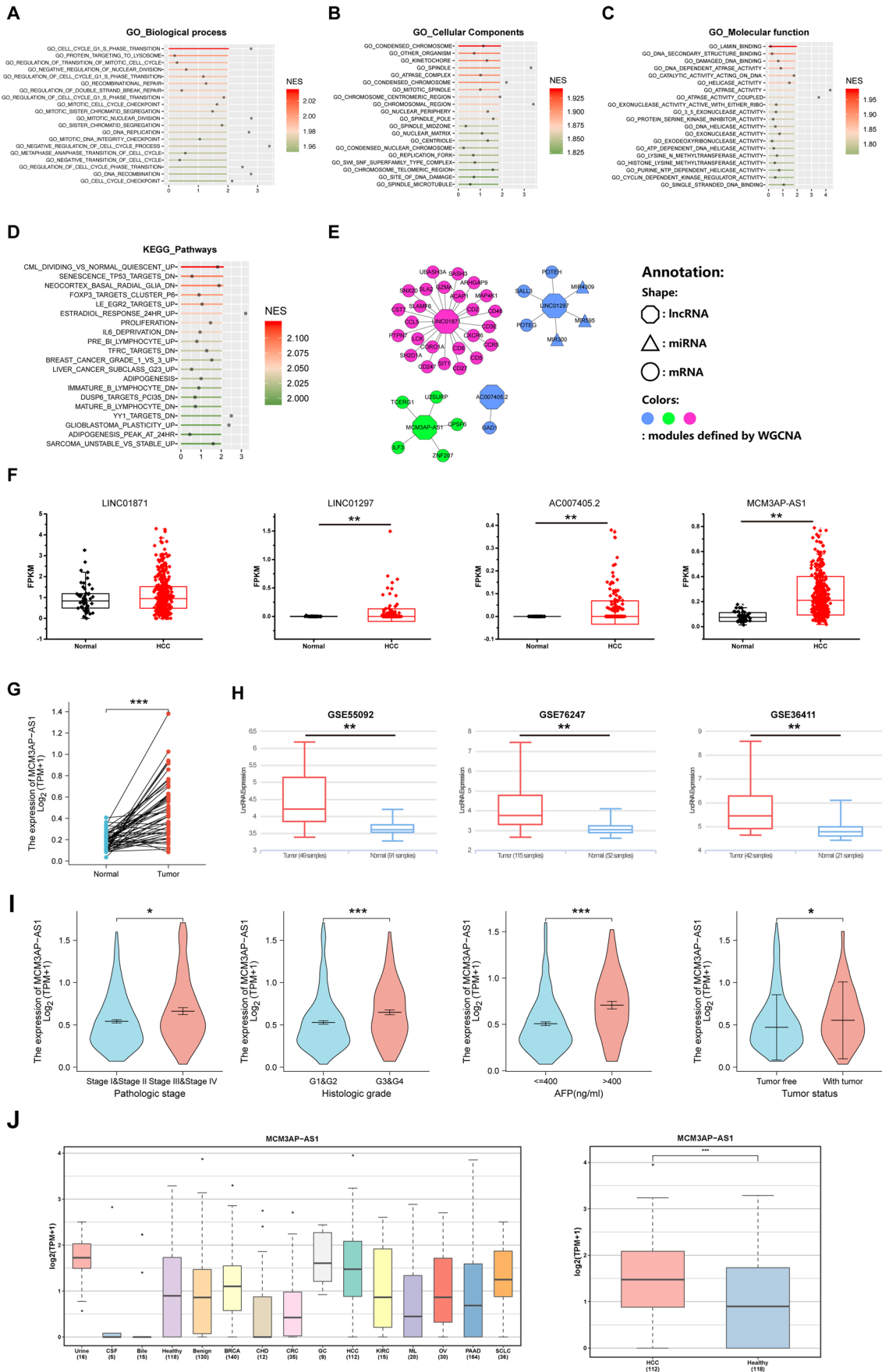
### Evaluate 18-lncRNA signature by stratification analyses

Then, we performed stratified analysis based on the clinical variables aforementioned with high HR. According to the Cox regression test above, the 18-lncRNA risk score was further evaluated in sub-groups with different tumor stages. Thus, the training set was stratified into three subgroups, including stage I ( $n = 111$ ), stage II ( $n = 57$ ), and stage III&IV ( $n = 62$ ). As shown in Fig. 5A–C, Kaplan–Meier analyses demonstrated that the high-risk score was correlated with poor survival in patients at stage I ( $P = 0.013$ ), stage II ( $P = 0.029$ ), and stage III&IV ( $P = 0.00054$ ). In addition, we got consistent results in the testing set (Fig. 5D–F) for HCC patients at stage I ( $n = 59$ ,  $P = 0.0073$ ). However, it



**Fig. 6** ROC analyses of the 18-lncRNA signature risk score and clinical variables in entire set. **A** The ROC analysis was conducted to evaluate the accuracy of the survival prediction within 1 year for

18-lncRNA signature risk score and clinical variables. **B** The 3-year ROC curves. **C** The 5-year ROC curves





**Fig. 7** Function prediction and expression features of the signature lncRNAs. **A–D** The 18-lncRNAs-mediated functions and pathways were predicted by GSEA, including biological process, cellular components, molecular function, and KEGG pathways. **E** The potential interactions of the 18 lncRNAs with miRNAs and mRNAs were visualized by Cytoscape. **F** The expression of the 4 lncRNAs in TCGA dataset. **G** The expression of MCM3AP-AS1 in 50 pairs of HCC tissues and normal tissues in TCGA dataset. **H** The expression of MCM3AP-AS1 in HCC tissues and normal tissues from GSE55092, GSE76247, and GSE36411. **I** The expression levels in subgroups of HCC cases in TCGA. **J** The expression of MCM3AP-AS1 in healthy control and tumor patients-derived exosomes was analyzed in exoR-Base 2.0 database. GO, Gene Ontology; NES, normalized enrichment score. \*\* $P < 0.01$

was not statistically significant for cases at stage II ( $n = 27$ ,  $P = 0.14$ ) or stage III&IV ( $n = 27$ ,  $P = 0.51$ ), which might be attributed to the small size in the two subgroups. Generally, these results indicated that the 18-lncRNA signature could robustly predict the OS of HCC patients at different tumor stages. Subsequently, the 18-lncRNA risk score was also evaluated in sub-groups with different pathologic stages. As presented in Fig. 5G–I, high-risk patients in the training set showed shorter survival time in the T I subgroup ( $n = 119$ ,  $P = 0.0076$ ), T II subgroup ( $n = 62$ ,  $P = 0.001$ ), and T III&IV subgroup ( $n = 67$ ,  $P = 0.0022$ ). For the testing set (Fig. 5J–L), the high-risk score was associated with poor survival in T I ( $n = 59$ ,  $P = 0.0073$ ). Though the high-risk score was associated with shorter survival time, the difference was not statistically significant due to the small size in the T II subgroup ( $n = 29$ ,  $P = 0.25$ ) and T III&IV subgroup ( $n = 23$ ,  $P = 0.35$ ).

### Evaluating survival prediction of 18-lncRNA signature and other factors

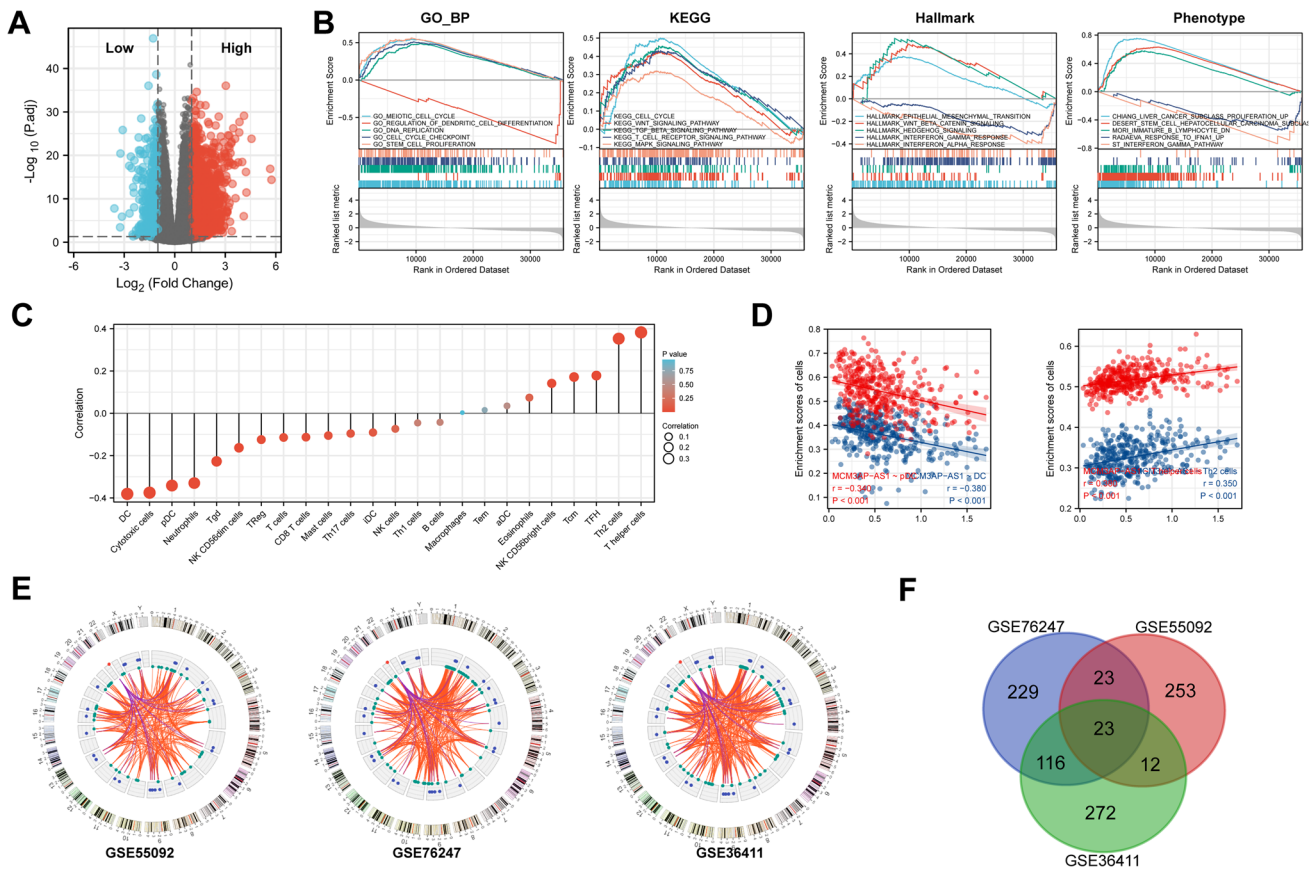
To evaluate the independence of lncRNA signature in survival prediction, time-dependent ROC analysis was performed to assess the sensitivity and specificity in classifying death and survival on the follow-up of 1-, 3-, and 5-year in the entire datasets. Multiple risk clinicopathologic characteristics (statistically significant in the univariable analysis above) were enrolled in this test, including the 18-lncRNA signature risk score, tumor stage, metastasis, lymph node status, and pathologic stage (Table 2). As shown in Fig. 6A–C, the area under the receiver operating characteristic curve (AUROC) of 18-lncRNA signature risk score was 0.826 (95%CI, 0.764–0.888), 0.817 (95%CI, 0.759–0.876), and 0.799 (95%CI, 0.731–0.867) for 1-year, 3-year, and 5-year follow-up, respectively. The results showed that the 18-lncRNA signature exhibited a more efficient predictive performance than any other risk factor.

### The potential functions mediated by the 18 lncRNAs

To explore the molecular functions related to the 18 lncRNAs, we identified the differentially expressed genes between high- and low-risk groups stratified by the median value of the risk score. Then, we performed the GSEA analysis based on the genes enriched in the high-risk group. GO enrichment consisted of three sections, namely biological process (BP), cellular component (CC), and molecular functions (MF). The top-involved BP included regulation of cell cycle, lysosome targets, recombinational repair, and DNA replication (Fig. 7A); the top-involved CC included kinetochore, spindle, and ATPase complex (Fig. 7B); The MF included Lamin binding, DNA secondary structure binding, and damaged DNA binding (Fig. 7C). In addition, pathway analyses indicated that the 18 lncRNA might be correlated with senescence-TP53 pathway, FOXP3 signaling, proliferation-related pathway, and EGR2 signaling, which were reported to facilitate malignant phenotypes of HCC (Fig. 7D). Moreover, we conducted the co-expression analysis based on the transcriptional information of the 18 lncRNA, related miRNA, and mRNA. As shown in Fig. 7E, according to the sub-network visualized by Cytoscape, 4 lncRNAs were co-expressed with featured miRNA or mRNA. Of them, obvious upregulation of MCM3AP-AS1 was found in HCC tissues in entire TCGA LIHC cases and paired HCC cases (Fig. 7F and G). Then, we validated the expression of the expression of MCM3AP-AS1, AC007405.2, and LINC01297 in three lncRNA-related GEO datasets (GSE55092, GSE76247, and GSE36411). In all three GEO datasets, MCM3AP-AS1 showed significantly higher expression in HCC tissues compared with normal liver tissues (Fig. 7H). Additionally, MCM3AP-AS1 was elevated in subgroups of HCC patients with advanced stages/grades, high levels of AFP, and existing tumor status (Fig. 7I). Interestingly, HCC patients showed relatively higher exosomal MCM3AP-AS1 levels in contrast to other cancer types. In addition, compared with healthy control, MCM3AP-AS1 was dramatically elevated in HCC blood-derived exosomes (Fig. 7J).

### Expression features and function of MCM3AP-AS1

MCM3AP-AS1 expression was significantly correlated with pathologic stage, adjacent inflammation, and AFP levels (Table S4). Additionally, high expression of MCM3AP-AS1 was correlated with poor survival of HCC patients (Figure S3). Subsequently, we further investigated the potential biological processes and mechanisms modulated by MCM3AP-AS1. 1320 DEGs were detected



**Fig. 8** Potential functions and CeRNA networks of MCM3AP-AS1. **A** The heatmap elucidating the differential expressional genes between high and low expression of MCM3AP-AS1 in TCGA LIHC dataset. **B** GSEA analysis of the functions and mechanisms mediated by MCM3AP-AS1. **C** ssGSEA demonstrated the correlation of

MCM3AP-AS1 with intratumoral immune cells. **D** Pearson correlation analysis of the MCM3AP-AS1 with DCs, Th2, and T helper cells. **E** KEGG and CeRNA prediction of MCM3AP-AS1 in three datasets above. **F** 23 co-existed genes were presented in the CeRNA networks of MCM3AP-AS1 in the GEO datasets. \*\* $P < 0.01$

between cases with high- and low- MCM3AP-AS1 expression (Fig. 8A). Further GSEA analysis demonstrated that MCM3AP-AS1 might be involved in tumor-related phenotypes and pathways, including cell cycle, proliferation, cancer stem cell, EMT, TGF- $\beta$  pathway, MAPK signaling, and Wnt signaling. It might be also implicated in immune-related processes, including IFN $\alpha/\gamma$  response, DC cell differentiation, and T cell receptor pathway (Fig. 8B). Thus, we further investigated the correlation of MCM3AP-AS1 with 24 types of immune cells in TCGA dataset by performing ssGSEA (Fig. 8C). MCM3AP-AS1 was significantly correlated with Th2 cells and T helper cells, while it was negatively correlated with DC and pDC (Fig. 8D). As is known, lncRNAs always function as a component in competing endogenous RNA (ceRNA) network. Then, we predicted the and pathways and potential ceRNA network regarding MCM3AP-AS1 in the three datasets (Fig. 8E). After the intersection of results, a total of 23 candidates were predicted as MCM3AP-AS1-related ceRNA networks (Fig. 8F). Key genes like TOP2A, TTK, CENPF, IPO9,

FANCD2, and related miRNAs were observed in these CeRNAs (Table S5)

## Discussion

In addition to the vital roles in multiple biological and pathological processes, lncRNAs have shown its clinical significance as robust biomarkers for prognosis or diagnosis (Nemeth et al. 2023; Ahmad et al. 2023). In contrast to single gene, gene-based signatures have gained significant significance in oncology as prognostic markers due to enhanced specificity and accuracy (Pinto-Marques et al. 2022). In the present study, we identified 18 robust lncRNAs with prognostic value for HCC patients, namely AL354824.2, LNCsRLR, AC019118.2, ZFPM2-AS1, AL356020.1, AL683887.1, LINC01250, LINC01871, EGLN3-AS1, AL445430.1, LINC01297, AL512785.1, IQCH-AS1, MCM3AP-AS1, AL592043.1, AC007405.2, AC108752.1, and AC104785.1. Then, the

18-lncRNA-based risk score was established based on the coefficients calculated by the univariate Cox regression. Kaplan–Meier analysis showed that HCC patients with high-risk score had significantly shorter survival than those with low-risk score. The correlation of high-risk score with poor OS of HCC patients was also confirmed by univariate Cox regression with other clinical parameters. Additionally, we randomly performed a log-rank test in one-third of the entire LIHC cohort 10 times, which also exhibited excellent performance in all tests. To evaluate the independence of the 18-lncRNA signature model, we further conducted the stratification analyses in the training and testing set. As expected, high-risk patients tended to acquire shorter survival at different tumor stages or pathological stages in sub-groups of the training set. Though the differences were not statistically significant in some sub-groups in testing set due to small sample size, a shorter survival time was observed in high-risk patients. The evidence demonstrates that the 18-lncRNA signature risk score is a robust predictor for the overall survival of HCC patients.

Previous studies have constructed several lncRNA-based signatures derived from HCC datasets. For example, a recent seven-lncRNA signature model showed prognostic potential for HCC patients with AUC up to 0.754 (Yan et al. 2019). Interestingly, a previous study constructed a 5-lncRNA signature with 0.769 in 5-year AUC for HCC patients, exhibiting better predictive capacity after comparing with two existing lncRNA-models with 0.701 and 0.721 (Zhao et al. 2018). For the current model, the 18-lncRNA signature had more excellent AUC (almost higher than 0.8) in predicting 1-year, 3-year, and 5-year OS in both the training set and testing set. Remarkably, the AUC of the 18-lncRNA signature for 1, 3, and 5 years were 0.826, 0.817, and 0.799 in the entire set, which was significantly higher than any of other clinical factors. Thus, the risk signature identified in this study may have superior performance than previous models. The improvements might be attributed to the screening approaches. The previous models were mainly established on the pre-screening of the differentially expressed lncRNAs between HCC tissues and normal tissues. As we know, it might lose some key prognostic factors with fewer fold changes or no significant *P* values. Therefore, we used a combination of log-rank test and univariable Cox regression test to obtain the initial candidates for further RSV algorithm, eventually got the 18-lncRNA signature risk model.

Given the prognostic performance, we further explored the function and pathways mediated by the 18 lncRNAs. GO analyses found that the lncRNAs were involved in cell cycle, DNA replication, and DNA recombination, which might be correlated with the proliferation of HCC cells. In addition, these lncRNA-mediated genes were

enriched in canonical cancer-related pathways, including P53 signaling, FOXP3, and IL-6 signaling. Indeed, most of lncRNAs aforementioned have not been investigated to date. Furthermore, we predicted the interaction of the 18 lncRNAs with miRNAs and mRNAs. 4 lncRNA were co-expressed with mRNA and miRNA in TCGA datasets. LINC01871 could interact with chemokine family members (CCL5, CXCR6, and CCR5) and CD family members (CD2, CD5, CD48, and CD247), which play crucial roles in tumor microenvironment of HCC. For MCM3AP-AS1, it exerted inhibitory effects on proliferation in cervical squamous cell carcinoma, suggesting its distinct roles in different cancer types (Lan et al. 2020). In contrast, it was recently implicated in provoking tumor growth of HCC by binding to miR-194-5p as ceRNA (Song et al. 2019). Consistent with this, our current study found that the over-expression of MCM3AP-AS1 was observed in the TCGA datasets and three GEO datasets. Interestingly, MCM3AP-AS1 was also enriched in HCC blood-derived exosomes, suggesting the potential detection and targeted significance. In addition, MCM3AP-AS1 might contribute to the initiation and progression of HCC by regulating various tumor-related pathways, including MAPK signaling, Wnt pathway, and TGF- $\beta$  signaling. Moreover, MCM3AP-AS1 was also implicated in tumor microenvironment by regulating immune-related pathways and altering the activities of the immune cells. According to previous studies, lncRNA may localize to nuclear to couple transcription, splicing, and maturation of key immune genes, thereby regulating activation and differentiation of immune cells. The prognostic performance of the lncRNA-based risk signature needs validation in external datasets and large HCC cohorts. In addition, in light of the potential targeted value, more experimental assays should be conducted to investigate the roles and mechanisms of these lncRNAs in HCC progression.

## Conclusions

In conclusion, this study identified an 18-lncRNA signature, which is significantly associated with the prognosis of HCC patients. Compared with other clinical risk factors, the risk signature presented a more efficient performance in predicting the survival time of HCC. In addition, the function analyses suggested that MCM3AP-AS1 might be implicated in cancer-related processes and pathways. Future studies should validate the lncRNA-based signature in more HCC cohorts and further investigate the molecular mechanisms.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s13353-024-00836-5>.

**Author contribution** W.Z. and X.C. designed and supervised this study. H.Q., Z.L., and Y.C. conducted the analysis. H.Q. drafted the manuscript. B.J. performed the revision. All authors approved the final version of the manuscript.

**Funding** This study was supported by grants from National Natural Science Foundation (81702419).

**Data availability** The datasets presented in this study can be found in online repositories.

## Declarations

**Conflict of interest** The authors declare no competing interests.

## References

- Ahmad M, Weiswald LB, Poulain L, Denoyelle C, Meryet-Figuere M (2023) Involvement of lncRNAs in cancer cells migration, invasion and metastasis: cytoskeleton and ECM crosstalk. *J Exp Clin Cancer Res* 42(1):173
- Davalos V, Esteller M (2019) Disruption of long noncoding RNAs targets cancer hallmark pathways in lung tumorigenesis. *Cancer Res* 79(12):3028–3030
- Fatica A, Bozzoni I (2014) Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* 15(1):7–21
- Goh GB, Chang PE, Tan CK (2015) Changing epidemiology of hepatocellular carcinoma in Asia. *Best Pract Res Clin Gastroenterol* 29(6):919–928
- Grixti JM, Ayers D (2020) Long noncoding RNAs and their link to cancer. *Noncoding RNA Res* 5(2):77–82
- Hsu CY, Hsia CY, Huang YH, Su CW, Lin HC, Pai JT et al (2012) Comparison of surgical resection and transarterial chemoembolization for hepatocellular carcinoma beyond the Milan criteria: a propensity score analysis. *Ann Surg Oncol* 19(3):842–849
- Huang Z, Zhou JK, Peng Y, He W, Huang C (2020) The role of long noncoding RNAs in hepatocellular carcinoma. *Mol Cancer* 19(1):77
- Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y et al (2015) The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 47(3):199–208
- Lai H, Li Y, Zhang H, Hu J, Liao J, Su Y et al (2022) exoRBase 2.0: an atlas of mRNA, lncRNA and circRNA in extracellular vesicles from human biofluids. *Nucleic Acids Res* 50(D1):D118–D128
- Lan L, Liang Z, Zhao Y, Mo Y (2020) LncRNA MCM3AP-AS1 inhibits cell proliferation in cervical squamous cell carcinoma by down-regulating miRNA-93. *Biosci Rep* 40(2):BSR20193794
- Lee YR, Kim G, Tak WY, Jang SY, Kweon YO, Park JG et al (2019) Circulating exosomal noncoding RNAs as prognostic biomarkers in human hepatocellular carcinoma. *Int J Cancer* 144(6):1444–1452
- Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M et al (2016) Hepatocellular Carcinoma *Nat Rev Dis Primers* 2:16018
- Nemeth K, Bayraktar R, Ferracin M, Calin GA (2023) Non-coding RNAs in disease: from mechanisms to therapeutics. *Nat Rev Genet*. <https://doi.org/10.1038/s41576-023-00662-1>
- Pea A, Jamieson NB, Braconi C (2021) Biology and clinical application of regulatory RNAs in hepatocellular carcinoma. *Hepatology* 73(Suppl 1):38–48
- Pinto-Marques H, Cardoso J, Silva S, Neto JL, Goncalves-Reis M, Proenca D et al (2022) A gene expression signature to select hepatocellular carcinoma patients for liver transplantation. *Ann Surg* 276(5):868–874
- Schmitt AM, Chang HY (2016) Long noncoding RNAs in cancer pathways. *Cancer Cell* 29(4):452–463
- Song M, Zhong A, Yang J, He J, Cheng S, Zeng J et al (2019) Large-scale analyses identify a cluster of novel long non-coding RNAs as potential competitive endogenous RNAs in progression of hepatocellular carcinoma. *Aging (albany NY)* 11(22):10422–10453
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 71(3):209–249
- Ulitsky I, Bartel DP (2013) lincRNAs: genomics, evolution, and mechanisms. *Cell* 154(1):26–46
- Wallace MC, Preen D, Jeffrey GP, Adams LA (2015) The evolving epidemiology of hepatocellular carcinoma: a global perspective. *Expert Rev Gastroenterol Hepatol* 9(6):765–779
- Yan X, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD et al (2015) Comprehensive genomic characterization of long non-coding RNAs across human cancers. *Cancer Cell* 28(4):529–540
- Yan J, Zhou C, Guo K, Li Q, Wang Z (2019) A novel seven-lncRNA signature for prognosis prediction in hepatocellular carcinoma. *J Cell Biochem* 120(1):213–223
- Zhao QJ, Zhang J, Xu L, Liu FF (2018) Identification of a five-long non-coding RNA signature to improve the prognosis prediction for patients with hepatocellular carcinoma. *World J Gastroenterol* 24(30):3426–3439
- Zhao Y, Teng H, Yao F, Yap S, Sun Y, Ma L (2020) Challenges and strategies in ascribing functions to long noncoding RNAs. *Cancers (Basel)* 12(6):1458

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.