

# Identification of an Epithelial-Mesenchymal Transition-Related Long Non-coding RNA Prognostic Signature for Hepatocellular Carcinoma

**Shenglan Huang**

Nanchang University Second Affiliated Hospital

**Dan Li**

Nanchang University Second Affiliated Hospital

**LingLing Zhuang**

Nanchang University Second Affiliated Hospital

**Jian Zhang**

Nanchang University Second Affiliated Hospital

**Jianbing Wu** (✉ [Ndefy93008@ncu.edu.cn](mailto:Ndefy93008@ncu.edu.cn))

Nanchang University Second Affiliated Hospital <https://orcid.org/0000-0001-9851-3135>

---

## Primary research

**Keywords:** hepatocellular carcinoma, epithelial-mesenchymal transition, long non-coding RNA, prognostic signature, nomogram

**Posted Date:** June 24th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-642960/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Introduction:** Hepatocellular carcinoma (HCC) is one of the most common malignant tumors with poor prognosis. Epithelial–mesenchymal transition (EMT) is crucial for cancer progression and associates with a worse prognosis. Thus, we aimed to construct an EMT-related lncRNAs signature for predicting the prognosis of HCC patients.

**Methods:** We built an EMT-related lncRNA risk signature in the training set by using Cox regression and LASSO regression based on TCGA database. Kaplan-Meier survival analysis was conducted to compare the overall survival (OS) in different risk groups. Cox regression was performed to explore whether the signature could be used as an independent factor. A nomogram was built involving the risk score and clinicopathological features. Furthermore, we explored the biological functions and immune states in two groups.

**Results:** 12 EMT-related lncRNAs were obtained for constructing the prognosis model in HCC. The Kaplan-Meier curve analysis revealed that patients in the high-risk group had worse survival than low-risk group. Time-dependent ROC and Cox regression analyses suggested that the signature could predict HCC survival exactly and independently. The prognostic value of the risk model was confirmed in the validation group. The nomogram was built and could accurately predict survival of HCC patients. GSEA results showed that in high-risk group cancer-related pathways were enriched, and exhibited more cell division activity suggested by Gene Ontology (GO) analysis.

**Conclusions:** We established a novel EMT-related prognostic risk signature including 12 lncRNAs and constructed a nomogram to predict the prognosis in HCC patients, which may improve prognostic predictive accuracy for HCC patients and guide the individualized treatment methods for the patients with HCC.

## Introduction

Hepatocellular carcinoma (HCC) is one of the most common aggressive tumors that ranks as the fourth leading cause of cancer-related mortality worldwide [1]. Despite the proven efficacies of several treatment methods, such as surgical resection, liver transplantation, ablation therapy, radiotherapy, targeted therapy, systemic chemotherapy, and immunotherapy [2], the prognosis remains unsatisfactory in HCC patients. According to the World Health Organization estimates, more than 1 million patients will die from HCC by 2030, with the 5-year survival rate being no more than 18% [2, 3]. Such poor survival rates are mainly due to the low rates of early HCC diagnosis, rapid progression of the disease, resistance to the treatment, high heterogeneity, and lack of early prognostic indicators for HCC. Therefore, there is an urgent need to identify highly sensitive and specific prognostic biomarkers for predicting the clinical outcomes for HCC patients.

Epithelial-mesenchymal transition (EMT) is a process that leads to morphological and functional changes, in which the polarized epithelial cells convert from epithelial to mesenchymal-like or

mesenchymal fibroblastoid phenotypes, which is characterized by the dissolution of cell-cell junctions, synthesis of extracellular matrix, and increasing cell motility that consequently result in the cells gaining invasive and metastatic properties [4, 5]. Numerous studies have shown that EMT plays vital roles in the development and progression of diverse types of cancer, including HCC [6, 7]. Transforming growth factor- $\beta$  (TGF- $\beta$ ), one of the most potent inducers of EMT, is able to promote fibrogenesis and carcinogenesis in HCC. Furthermore, it is correlated with the invasion and metastasis of HCC [8]. Epithelial cadherin (E-cadherin), an important marker of EMT, is directly associated with the worse prognosis and short survival rates in HCC patients [9]. In a previous study, it was reported that the downregulation of E-cadherin led to the microvascular invasion and metastasis of cells in 323 HCC patients [10]. Moreover, the EMT-transformed HCC cells stimulated with TGF- $\beta$  were resistant to sorafenib-induced apoptosis [11]. Increased expression of EMT markers may be related to the poor prognosis in HCC patients exhibiting aggressive local recurrence after radiofrequency ablation [5]. Therefore, EMT-related genes might be used as novel and ideal prognostic and therapeutic targets for the treatment of HCC.

Long non-coding RNAs (lncRNAs), which account for 80–90% of all ncRNAs, structurally contain more than 200 nucleotides. Recently, it have attracted wide attention as biomarkers in the early diagnosis and prognosis tumors, thus lncRNA have been regarded as potential therapeutic targets of tumors[12–14]. Interestingly, accumulating evidence suggests that lncRNAs are involved in EMT and metastasis and also regulate these processes in various cancers, including HCC [15]; for example, lncRNAs, such as lncRNA-POIR [16], lncRNA-AB209371 [17], lncRNA colorectal neoplasia differentially expressed (lncRNA CRNDE) [18], lncRNA small nucleolar RNA host gene 3 (lncRNA SNHG3) [19], and lncRNA LOC554202 [20], have been reported to promote the epithelial-mesenchymal transition, accelerate the metastasis of HCC, and suppress the drug sensitivity of cancer cells. Conversely, the lncRNA cancer susceptibility candidate 2 (lncRNA CASC2) and lncRNA miR503HG inhibit EMT and exert anti-metastatic effects in HCC cells [21, 22]. In addition, the EMT-related lncRNA signatures exhibit predictive potency for the survival rates of patients, which have been reported in bladder and kidney cancer (clear cell renal cell carcinoma) [23, 24]. However, the potential values of EMT-related lncRNAs as prognostic indicators and therapeutic targets have not yet been explored in the patients with HCC.

Therefore, in this study, we first built and verified an EMT-related lncRNA prognostic model and nomogram based on the data extracted from The Cancer Genome Atlas (TCGA) and then explored the biological functions and related immune responses of EMT-related lncRNAs in the patients with HCC. A flowchart of this study is shown in Fig. 1.

## Methods And Materials

### Data acquisition and processing

The RNA sequencing data of 424 HCC patients and the corresponding clinical data were obtained from the TCGA data portal based on the fragments per kilobase of transcript per million mapped reads (FPKM)

values (<https://cancergenome.nih.gov/>). The genes with low expression levels with an average value  $< 0.5$  were eliminated and the included data were commented and sorted into 710 lncRNAs and 11282 protein-coding genes according to the ENSEMBL database (<http://asia.ensembl.org/index.html>). After removing the normal samples and less than 30 d of follow-up studies, the remaining data of 343 HCC patients were randomly assigned to the training ( $n = 240$ ) and testing group ( $n = 103$ ). The EMT-related genes were obtained from the Molecular Signatures Database (MSigDB) v.7.4 (hallmark\_epithelial\_mesenchymal\_transition M5930, <http://www.broadinstitute.org/gsea/msigdb/index.jsp>).

### **Screening for the EMT-related lncRNAs in HCC**

A total of 200 EMT-related genes were downloaded from MSigDB v.7.4. Then, 159 EMT-related genes matrix data in HCC was extracted by utilizing the “limma” package of R v.4.0.4 software (<http://www.r-project.org/>). Pearson correlation coefficient analysis was performed to identify the EMT-related lncRNAs according to the criteria:  $|\text{correlation coefficient}| > 0.3$  and  $P < 0.001$ . A total of 449 EMT-related lncRNAs were screened for further verification.

### **Construction and verification of the EMT-related lncRNA prognostic model for HCC**

To extract the prognosis-associated EMT-related lncRNAs, we performed the univariate Cox regression analysis and log-rank tests. We then incorporated the prognostic EMT-related lncRNA candidates into the least absolute shrinkage and selection operator (LASSO) regression and multivariate Cox regression to construct the prognostic risk models in the training set. According to the multivariate Cox regression coefficient and expression value of each lncRNA in the patients with HCC, the risk score of each patient was evaluated using the risk score equation as follows:

$$\text{Risk score} = \text{coef}(\text{lncRNA1}) \times \exp(\text{lncRNA1}) + \text{coef}(\text{lncRNA2}) \times \exp(\text{lncRNA2}) + \dots + \text{coef}(\text{lncRNA}_n) \times \exp(\text{lncRNA}_n).$$

where  $\text{coef}(\text{lncRNA})$  represents the regression coefficient and  $\exp(\text{lncRNA})$  is the expressive value of each EMT-related lncRNA. The patients were divided into high-and low-risk groups based on the median cut-off of risk scores. We then performed the Kaplan-Meier survival analysis to compare the differences in the overall survival (OS) between the two groups in the training and validation cohorts. Meanwhile, the stratification analysis was performed based on the following clinicopathological features: age ( $\leq 65$  and  $> 65$  years), sex (male and female), grade (grade 1–2 and grade 3), and clinical stage (stage I–II and stage III–IV);  $P < 0.05$  was considered to be statistically significant. Time-dependent receiver operating characteristic (time-ROC) curve was sketched to evaluate the predictive accuracy of the prognostic model for 1-year, 3-year, and 5-year survival via “survivalROC” package of R 4.0.4 software (<http://www.r-project.org/>).

Then, univariate and multivariate Cox regression analyses were performed to explore whether the constructed prognostic model could be used as an independent factor to determine the prognosis of HCC patients by integrating the following clinicopathological factors: age, sex, grade, clinical stage;  $P < 0.05$  was considered to be statistically significant. A multi-indicator receiver operating characteristic (multi-

ROC) curve was drawn and the area under the curve (AUC) of the ROC curve was estimated to assess the predictive accuracy of the risk score and other clinicopathological features for the survival of patients.

In addition, the relationship between the prognostic EMT-related lncRNAs and their co-expressed mRNAs was analyzed using a co-expression network. The Cytoscape v.3.8.2 software (<http://www.cytoscape.org/>) and “ggalluvial” R package were used to visualize the co-expression correlation.

### **Construction and validation of the predictive nomogram**

Based on the risk scores and conventional clinical characteristics of the training group, we established an OS-associated nomogram to predict the 1-year and 3-year survival probability of HCC patients by using the “survival” and “rms” R packages. We then performed internal verification to assess the predictive precision of this nomogram. A calibration curve was drawn to appraise the uniformity of the predicted result and the ROC curve was sketched to assess the predictive accuracy of the nomogram model.

### **Enrichment analysis**

The transcriptome data of the entire group from TCGA were selected for gene set enrichment analysis (GSEA) using the GSEA v.4.1.0 software (<http://www.gsea-msigdb.org/gsea/index.jsp>). Based on the prognostic signature, 343 samples were divided into high- and low-risk groups. The `c2.cp.kegg.v7.4.symbols.gmt` downloaded from MSigDB (<http://www.gsea-msigdb.org/gsea/msigdb/collections.jsp>) was used as the reference and a false discovery rate (FDR) of  $q\text{-value} < 0.05$  was used to identify the significantly enriched pathways. Thereafter, the significant differentially expressed genes (DEGs) were extracted from TCGA database depending on the risk of grouping and the statistical significance was defined as  $FDR < 0.05$  and  $|\log_2FC| \geq 1$ . According to the DEGs, Gene Ontology (GO) analysis was used to evaluate the related biological pathways in the two risk groups.

### **Immune cell infiltration and immune-related pathways**

According to the EMT-related lncRNA prognostic model, single-sample gene set enrichment analysis (ssGSEA) was conducted to quantitatively analyze the differences in the infiltrating levels of immune cells as well as the immune-related pathways between the high- and low-risk groups in the HCC patients by using “GSVA” and “GSEABase” R package [25]. The result was visualized with boxplot using the “ggpubr” R package.  $P < 0.05$  was considered statistically significant.

### **Statistical analyses**

Statistical analyses were performed using the R software v.4.0.4 (<https://www.r-project.org/>) and SPSS v.24 (IBM, New York, USA). Chi-square test was used to analyze the baseline characteristics of patients with HCC. LASSO and Cox regression analyses were used to build a prognostic model. First, univariate Cox analysis and log-rank tests were conducted to screen for the prognostic EMT-related lncRNA candidates with  $P < 0.05$ . Then, the LASSO algorithm was used to reduce the data dimensionality. The risk score for the 12 selected lncRNAs was calculated using a multivariate Cox model. Kaplan-Meier analysis was used to evaluate the differences in the survival of the two risk groups. The ROC curve was used to estimate the predictive ability of the risk models. The predictive values of the nomogram for 1-

year and 3-year OS were assessed using calibration and ROC curves. All statistical tests were two-sided and  $P < 0.05$  was considered statistically significant.

## Results

### Identification of EMT-related lncRNAs in HCC

First, the gene expression matrix and clinical information of 374 HCC samples were acquired from TCGA database. After removing the repeated samples, patients with incomplete clinicopathological data and less than 30 d follow-up time, 343 patients with complete follow-up information and 319 patients with complete clinicopathological data were included in this study for subsequent analysis; 343 patients were divided into the training group ( $n = 240$ ), testing group ( $n = 103$ ), and entire group ( $n = 343$ ) at a ratio of 0.7:0.3:1. The clinicopathological features of the HCC patients included in this study are shown in Table 1. Then, 200 EMT-related genes were downloaded from MSigDB and 41 genes were excluded owing to their low expression levels. A total of 159 EMT-related genes were identified in HCC derived from the TCGA database. According to the Pearson correlation coefficient analysis, 449 EMT-related lncRNAs were screened based on the filtering criteria of correlation coefficient  $< 0.3$  and  $P < 0.001$ .

Table 1

Characteristics of hepatocellular carcinoma patients from TCGA dataset.

Characteristics	Training group (n = 224)	Testing group (n = 95)	Entire group (n = 319)	P value
<b>Age(year)</b>				
≤ 60	113	47	160	0.987
> 60	111	48	159	
<b>Gender</b>				
Male	156	63	219	0.842
Female	58	32	100	
<b>Tumor grade</b>				
G1	31	13	44	0.456
G2	117	37	154	
G3	68	41	109	
G4	8	4	12	
<b>Clinical stage</b>				
I	118	42	160	0.66
II	48	28	76	
III	55	25	80	
IV	3	0	3	

### Establishment and validation of the EMT-related lncRNA prognostic model for HCC

Among the 449 EMT-related lncRNAs, 41 lncRNAs were identified as prognosis-associated candidates by univariate Cox regression analysis and log-rank test in the training set with  $P < 0.05$ , which were significantly correlated with the survival of HCC patients (Table 2); the results are presented in **Supplementary Table 1**. Subsequently, LASSO penalty analysis was performed to identify the prognostic lncRNAs (Fig. 2A-B). Twelve lncRNAs were filtered out to construct a prognostic signature with multivariate Cox regression coefficients, including AC103760.1, AC015908.3, LINC02362, LINC02499, F11-AS1, LINC00942, PRRT3-AS1, AC012146.1, AC092171.2, AC099850.3, CASC19, and AL158206.1 (Table 3). Based on the 12 EMT-related lncRNAs and corresponding regression coefficients, the individualized risk score was calculated based on the following formula: Risk score =  $(-0.124 \times \text{AC103760.1 expression}) + (-0.113 \times \text{AC015908.3 expression}) + (-0.027 \times \text{LINC02362 expression}) +$

$(-0.021 \times \text{LINC02499 expression}) + (-0.009 \times \text{F11.AS1 expression}) + (0.02 \times \text{LINC00942 expression}) + (0.03 \times \text{PRRT3.AS1 expression}) + (0.033 \times \text{AC012146.1 expression}) + (0.122 \times \text{AC092171.2 expression}) + (0.142 \times \text{AC099850.3 expression}) + (0.163 \times \text{CASC19 expression}) + (0.313 \times \text{AL158206.1 expression})$ ; the EMT-related lncRNA prognostic signature was built in the training set. Then, we calculated the risk score of HCC patients in the testing and entire groups according to the above-mentioned formula. HCC patients in the three groups were further divided into two groups: low-risk and high-risk groups (median value = 0.8488). In the training set, Kaplan-Meier survival curves for OS indicated that the HCC patients in the high-risk group had shorter survival than those in the low-risk group ( $P < 0.001$ ) as shown in Fig. 2C. Time-ROC curve analysis was performed to evaluate the precision of the prognostic model and the results demonstrated that the risk score model provided a precise predictive effect of 1-year, 3-year, and 5-year OS, the AUC values were 0.828, 0.811, and 0.758, respectively (Fig. 2D). Risk score curves and scatter plots were used to explain the relationship between the risk score and survival status of patients with HCC and the results revealed that the higher the score at risk, the higher the mortality rate observed in the affected patients (Fig. 2E-F). The heatmap also showed that seven lncRNAs (LINC00942, PRRT3-AS1, AC012146.1, AC092171.2, AC099850.3, CASC19, and AL158206.1) were upregulated in the high-risk group, whereas the other five lncRNAs (AC103760.1, AC015908.3, LINC02362, LINC02499, and F11-AS1) were upregulated in the low-risk group (Fig. 2G). In the validation cohort, Kaplan-Meier analysis of the testing and entire groups was the same as that of the training cohort (Fig. 3A and D). The AUC values of the risk score for 1-year, 3-year, and 5-year OS were 0.880, 0.835, and 0.867, respectively, in the testing group and 0.810, 0.770, and 0.735, respectively, in the entire group (Fig. 3B and E). The results of the risk score curves, scatter plots, and heatmap in the validation groups showed similar trends to those observed in the training group (Fig. 3C and F).

Table 2

Univariate Cox regression analysis and log-rank test of EMT-related lncRNA.

<b>LncRNA</b>	<b>K-M</b>	<b>B</b>	<b>SE</b>	<b>HR</b>	<b>HR.95L</b>	<b>HR.95H</b>	<b>P value</b>
AC103760.1	0.003	-0.145	0.068	0.865	0.757	0.988	0.03307
AC015908.3	0.001	-0.393	0.111	0.675	0.543	0.838	0.00038
LINC02362	0.007	-0.078	0.037	0.925	0.861	0.994	0.03252
LINC02499	0.013	-0.064	0.028	0.938	0.888	0.991	0.02339
F11-AS1	0.022	-0.196	0.072	0.822	0.714	0.947	0.0065
LINC00942	0.02	0.039	0.009	1.04	1.022	1.058	1.21E-05
PRRT3-AS1	0.002	0.134	0.032	1.143	1.073	1.217	3.28E-05
AC012146.1	0.025	0.157	0.055	1.17	1.05	1.304	0.0046
AC092171.2	0.004	0.184	0.043	1.202	1.105	1.307	1.71E-05
AC099850.3	0.009	0.188	0.04	1.207	1.117	1.305	1.93E-06
CASC19	0.041	0.217	0.047	1.243	1.133	1.363	3.87E-06
AL158206.1	0.025	0.215	0.108	1.24	1.003	1.533	0.047
Note: HR: hazard ratio; HR.95L: low 95% CI of HR; HR.95H: high 95% CI of HR.							

Table 3

Multivariate Cox regression analysis of EMT-related lncRNA.

lncRNA	Coefficient	HR	HR.95L	HR.95H	pvalue
AC012146.1	0.033	1.034	0.898	1.191	0.645
CASC19	0.163	1.177	1.066	1.301	0.001
AC092171.2	0.122	1.13	1.024	1.246	0.015
PRRT3.AS1	0.03	1.031	0.95	1.118	0.465
AC103760.1	-0.124	0.883	0.771	1.011	0.072
LINC00942	0.02	1.02	0.996	1.045	0.097
LINC02362	-0.028	0.973	0.911	1.039	0.409
AC099850.3	0.142	1.153	1.055	1.26	0.002
F11.AS1	-0.009	0.991	0.885	1.111	0.879
AC015908.3	-0.113	0.893	0.703	1.134	0.354
AL158206.1	0.313	1.368	1.072	1.745	0.012
LINC02499	-0.02	0.979	0.927	1.035	0.457
Note: HR: hazard ratio; HR.95L: low 95% CI of HR; HR.95H: high 95% CI of HR.					

To further explore the prognostic value of the EMT-lncRNA signature in HCC patients, stratification analysis was performed in the entire group based on age ( $\leq 65$  and  $> 65$  years), gender (male and female), grade (grade 1–2 and grade 3–4), and clinical stage (stage I–II and stage III–IV), and the results showed that male patients, patients with grade 1–2, different age, and clinical stage in the high-risk group had worse survival than those in the low-risk group; however, there was no significant difference in female patients and patients with grade 3–4 mainly owing to the small sample, indicating that the prognostic signature was applicable in most different subtypes of HCC (Fig. 4).

Furthermore, univariate and multivariate Cox proportional hazards regression analyses were performed to evaluate whether the risk score model could be independently associated with the OS of HCC. The results showed that the 12 lncRNAs signature was an independent predictor in the training group and two validation groups (Fig. 5A-F). Multi-ROC curve analysis was implemented to identify the accuracy of the risk score model, and the results showed that the risk score model provided a more precise predictive role for survival than other clinical parameters with AUC = 0.825 in the training group, AUC = 0.823 in the testing group, and AUC = 0.794 in the entire group (Fig. 5G-I). The above results demonstrated that the EMT-related lncRNA prognostic model had an accurate predictive ability for prognosis in HCC patients.

We constructed a co-expression network of the 12 EMT-related lncRNA-mRNAs using Cytoscape and Sankey diagrams to visually present the correlation among EMT-related genes, EMT-related lncRNAs, and risk types of lncRNAs (Fig. 6A-B).

### **Construction and evaluation of the predictive nomogram**

We integrated age, gender, grade, and clinical stage with risk scores to establish a nomogram to predict the 1-year and 3-year OS of HCC patients. A comprehensive score was obtained by combining each clinical factor; the higher the total score of patients, the worse the prognosis (Fig. 7A). The predictive consistency and accuracy of the nomogram for 1-year and 3-year OS were assessed using calibration and ROC curves. The calibration curves indicated the prognostic nomogram model nearly in accordance with reality, and ROC curve analysis demonstrated that the nomogram provided an accurate survival prediction with AUC = 0.795 for 1-year OS and 0.773 for 3-years OS as shown in Fig. 7B-E, demonstrating that the nomogram exhibited preferable clinical practicality in predicting the prognosis of HCC.

### **Enrichment analysis**

GSEA analysis was conducted in the entire group to probe the potential biological mechanism of the EMT-related lncRNA prognostic signature. The GSEA results revealed that tumor-related pathways, including bladder cancer, thyroid cancer, pathways in cancer, VEGF signaling pathway, cell cycle, and DNA replication were highly enriched in the high-risk group with FDR (q-value) < 0.05 as shown in Fig. 8A.

To further determine the biofunctions associated with the risk signature, differentially expressed genes were extracted from the TCGA database in the low- and high-risk groups. A total of 969 DEGs (upregulated 832 genes, downregulated 146 genes, fold change > 1,  $P < 0.05$ ) were identified in the high-risk group compared to the low-risk group. We then selected the DEGs for GO enrichment analysis, and the results showed that these DEGs were enriched in terms associated with cell division process (Fig. 8B-C) and suggested that the proliferation ability of HCC cells in the high-risk group was notably more rapid than that in the low-risk group.

### **Immune cell infiltration and immune-related pathways**

Lastly, ssGSEA was conducted to compare the difference in infiltrating levels of immune cells and immune-related pathways between high- and low-risk groups in HCC patients, and we calculated the score of 16 immune cells and 13 immune function-related pathways based on ssGSEA analysis. The results revealed that the infiltration level of immune cells, including aDCs, macrophages, Th2 cells, and Treg cells, were significantly upregulated in the high-risk groups, but natural killer (NK) cells were upregulated in the low-risk group ( $P < 0.05$ ; Fig. 9A). As for the immune-related pathways, antigen-presenting cell (APC) co-stimulation, checkpoint, and major histocompatibility complex-I (MHC-I) were significantly upregulated in the high-risk group; nevertheless, the INF-I and INF-II responses were upregulated in the low-risk group ( $P < 0.05$ ) (Fig. 9B).

## **Discussion**

Hepatocellular carcinoma is one of the most common malignant tumors worldwide, exhibiting high recurrence rates, high mortality rates, and poor prognosis in patients [26]. Thus, identifying potential prognostic biomarkers is critical for predicting the overall survival of patients with HCC and guiding the individual therapeutic strategies for high-risk patients. Previously, most studies emphasized the impact of EMT on the tumor metastatic progression and resistance to therapy, while only a few studies have focused on its prognostic value in cancer. In addition, a wealth of recent studies have shown that lncRNAs could serve as novel and effective biomarkers and therapeutic targets for tumors [27]. Therefore, it will be beneficial to explore the efficacies of the EMT-related lncRNAs as biomarkers for predicting the prognosis in HCC patients.

In this study, we focused on the prognostic value of EMT-related lncRNA signature. Based on public expression data and clinical information from the TCGA dataset, we identified 12 prognostic EMT-related lncRNAs that could efficiently predict the survival of HCC patients. First, the 343 HCC patients with completed follow-up information were randomly divided into training and validation groups. In the training HCC patient cohort, we screened out 12 prognostic EMT-related lncRNAs to construct a prognostic signature for predicting survival using Cox regression and LASSO regression. The risk scores were calculated based on the 12-lncRNA expression profiles and coefficients, and Kaplan–Meier analysis showed that the survival of HCC patients in the high-risk group was worse than that of patients in the low-risk group, and the risk score model could correctly predict the 1-year, 3-years, and 5- years OS with AUC = 0.828, 0.811, and 0.758, respectively. In addition, the prognostic availability of the risk model was validated in the testing and entire groups, and univariate and multivariate Cox regression analysis and ROC curve analysis demonstrated that the EMT-related lncRNA signature provided more sensitivity and specificity, and acted as independent indicators of HCC prognosis compared with other clinical parameters. Next, we established a nomogram based on age, gender, grade, clinical stage, and risk score and verified predictive consistency and accuracy with calibration curve and ROC curve, all of which indicated that the nomogram had a satisfactory uniformity with actual survival and provided better clinical practicality than the traditional tumor grade or clinical stage system.

Among the 12 prognostic EMT-related lncRNAs (AC103760.1, AC015908.3, LINC02362, LINC02499, F11-AS1, LINC00942, PRRT3-AS1, AC012146.1, AC092171.2, AC099850.3, CASC19, AL158206.1) found in HCC, and only LINC02499, F11-AS1, LINC00942, PRRT3-AS1, and CASC19 have been reported in HCC or other cancers. LINC02499 was significantly downregulated in HCC tissues compared to paracancerous tissues, and its low expression was also remarkably correlated with poorer overall survival based on bioinformatic analysis[28]. lncRNA F11-AS1 was under-expressed and attenuated tumor growth and metastasis capacity via the F11-AS1/miR-3146/PTEN axis in HCC[29]. In addition, the low expression of lncRNA F11-AS1 was correlated with poor prognosis in patients with HBV-related HCC via regulation of the miR-211-5p /NR1I3 pathway [30]. A study revealed that LINC00942 acts as an oncogene that promotes cell proliferation and colony formation and inhibits cell apoptosis, subsequently elevating METTL14-mediated m6A methylation levels in breast cancer initiation and progression by LINC00942-METTL14-CXCR4/CYP1B1 signaling axis[31]. The lncRNA CASC19 was widely reported and markedly upregulated in multiple tumor tissues and cells, higher CASC19 expression is associated with tumor

progression and poorer prognoses, including colorectal cancer, advanced gastric cancer, non-small lung cancer cells, and pancreatic cancer [32–35]. Two other studies revealed that a high expression level of CASC19 was positively correlated with radio resistance in cervical cancer and nasopharyngeal carcinoma[36, 37]. However, the other seven prognostic EMT-related lncRNAs (AC103760.1, AC015908.3, LINC02362, AC012146.1, AC092171.2, AC099850.3, and AL158206.1) have been rarely reported, and our research may provide a new perspective for their potential functions. Moreover, most previous studies have emphasized that single lncRNAs function as prognostic indicators and relevant mechanisms; thus far, no research has applied EMT-related lncRNAs to the prognosis of HCC patients. Therefore, the construction of a prognostic EMT-associated lncRNA signature may provide invaluable insights for current prognostic predictions of HCC.

Subsequently, GSEA analysis was performed to investigate the related functional differences and underlying mechanisms of the risk model constructed using the 12 lncRNA signature in HCC. The results demonstrated that various tumor-related, proliferation, and metastasis-related pathways were highly enriched in the high-risk group, including the VEGF signaling pathway, DNA replication, and cell cycle. The VEGF signaling pathway is composed of the VEGF superfamily, VEGF receptors, some signal proteins, and nuclear transcriptional regulatory factors, which can regulate cell biological behavior and participate in tumorigenesis, tumor development, invasion, and metastasis[38]. The expression of VEGF and its receptors, including VEGFR1, VEGFR2, and VEGFR3, is elevated in HCC cell lines and tissues, and increased VEGF expression is associated with high HCC tumor grade, vascular invasion, and portal vein invasion[39]. Hence, VEGF signaling pathway inhibitors have been implicated as a therapeutic strategy for the prevention of HCC, for example, sorafenib for advanced-stage HCC[40]. A study reported that increased expression of p28 in HCC induces EMT by activating the PI3K-AKT-HIF-1 $\alpha$  pathway to promote Twist1, VEGF, and metalloproteinase 2 expression, which consequently leads to tumor angiogenesis, vascular invasion, and metastasis[41]. The VEGF signaling pathway was enriched in the EMT-related lncRNA signature high-risk group, which might be one of the contributors to the poor prognosis. Moreover, according to the GO analysis, DEGs in the high-risk group were focused on cell division activities, which promote the proliferation of cancer cells and accelerate tumor growth, which may be a reason for unfavorable prognosis in the high-risk group.

Given the significant role of the tumor microenvironment in the process of tumorigenesis, involving immune cell infiltration and immune-related pathways, we investigated the roles of immune infiltrating cells and immune function-related pathways of the tumor microenvironment in the prognosis of HCC. ssGSEA analysis revealed that immune cells, including aDCs, macrophages, Th2 cells, and Treg cells, were enriched in high-risk groups, but NK cells were upregulated in the low-risk group. In addition, APC co-stimulation, check-point, and MHC-I were significantly upregulated in the high-risk group; however, the INF-I and INF-II responses were upregulated in the low-risk group. Macrophages play a prominent role in the tumor microenvironment, acting as a cancer promoter in HCC by presenting M1 and M2 polarization, and M2 polarization induces cellular proliferation, angiogenesis, and EMT in HCC [42]. A study found that HIF1 $\alpha$ / IL-1 $\beta$  signaling promotes hepatoma EMT through macrophages in a hypoxic-inflammatory microenvironment [43]. It has been reported that the infiltration of macrophages is proportional to Tregs

and inversely proportional to CD8<sup>+</sup> T cells in HCC, and the absence of NK cells may trend toward M2 macrophages and accelerate liver fibrogenesis[42]. Given the above results, we speculated that the immune response ability in the high-risk group was lower in the low-risk group owing to immunosuppressive cells, such as macrophages and Tregs. Conversely, the anti-tumor immune response pathway was enriched in the low-risk group, which may contribute to a favorable prognosis in this group.

Although we successfully established a prognostic signature model and the nomogram performed well in predicting the survival of HCC patients, our study did have a few limitations. First, the 12 EMT-related lncRNAs prognostic signature was constructed and evaluated using limited data and clinical information from TCGA database and the 12 EMT-related lncRNAs identified in the training set were not completely comprised either in the Gene Expression Omnibus database or in the International Cancer Genome Consortium dataset, so this study was performed without any external validation, which may have limited the practicality and generalizability of our model. In addition, some of the model-building lncRNAs reported in this study have not been previously studied; therefore, further investigations are necessary to elucidate their functions. Finally, it is critical to verify the functional features and molecular mechanisms of the risk signature model by conducting further biological experiments and clinical research.

In conclusion, our study established a novel EMT-related prognostic risk signature including 12 lncRNAs and constructed a nomogram to predict the prognosis in HCC patients, which may aid in improving the prognostic predictive accuracy of HCC and guide the individualized treatment methods for the patients with HCC.

## Abbreviations

APC: antigen-presenting cell;

HCC: hepatocellular carcinoma;

EMT: epithelial-mesenchymal transition;

E-cadherin: epithelial cadherin;

lncRNA: long noncoding RNA;

TCGA: The Cancer Genome Atlas;

FPKM: fragments per kilobase of transcript per million mapped reads;

LASSO: the least absolute shrinkage and selection operator;

HR: hazard ratio;

OS: overall survival;

ROC: receiver operating characteristic;

time-ROC: time-dependent receiver operating characteristic;

multi-ROC: multi-indicator receiver operating characteristic;

AUC: area under the curve;

GSEA: gene set enrichment analysis;

GO: Gene Ontology;

ssGSEA: Single-sample gene set enrichment analysis;

TGF- $\beta$ : Transforming growth factor- $\beta$ ;

MHC-I: major histocompatibility complex-I

## **Declarations**

### **Competing interests**

The authors declare that they have no competing interests.

### **Funding**

This work was supported by a grant from the National Natural Science Foundation of China(Grant No.82060435).

### **Authors' contributions**

Conceptualization, S.-L.H. and J.-B.W.; investigation, S.-L.H.; methodology, L.-L.Z., D.L., and J.Z.; supervision, L.-L.Z. and D.L.; visualization, L.-L.Z. and J.Z.; writing-original draft, S.-L.H.; writing-review and editing, J.-B.W. All authors have read and agreed to the published version of the manuscript.

### **Acknowledgements**

We acknowledge and appreciate our colleagues for their valuable efforts and comments on this paper.

### **Ethical approval statement**

This study was approved by the ethics committee of the Second Affiliated Hospital of Nanchang University.

### **Availability of data and materials**

Processed data generated in this study is contained within the article. Raw data is available from the corresponding author upon request.

## Authors' information

<sup>1</sup> Department of Oncology, The Second Affiliated Hospital of Nanchang University, No. 1, Minde Road, Nanchang 330006, Jiangxi Province, P.R. China.

<sup>2</sup> Department of Gynaecology, The Second Affiliated Hospital of Nanchang University, No. 1, Minde Road, Nanchang 330006, Jiangxi Province, P.R. China.

## References

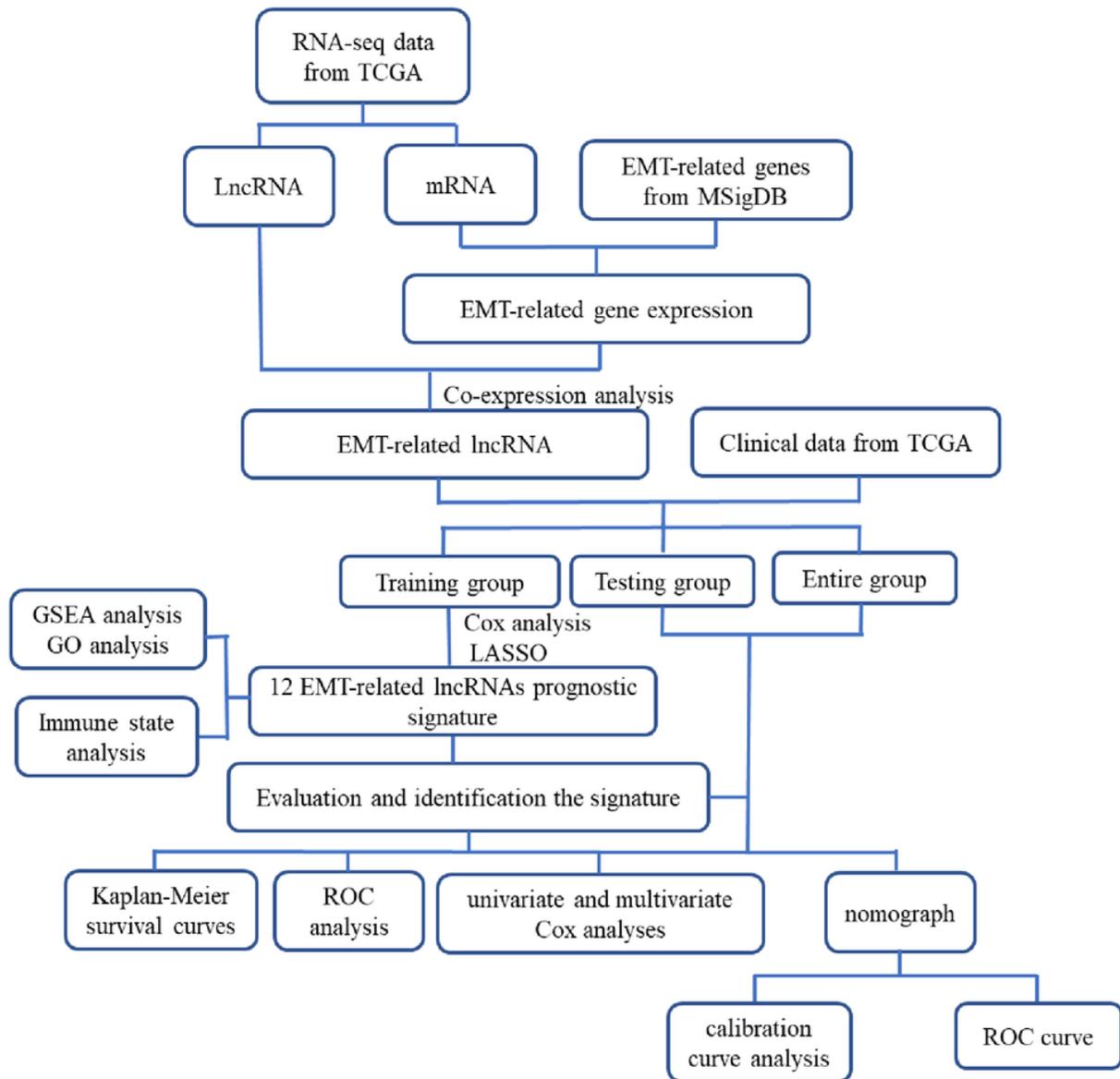
1. Yoon JS, Sinn DH, Lee JH, Kim HY, Lee CH, Kim SW, Lee HY, Nam JY, Chang Y, Lee YB, Cho EJ, Yu SJ, Kim HC, et al. Tumor Marker-Based Definition of the Transarterial Chemoembolization-Refractoriness in Intermediate-Stage Hepatocellular Carcinoma: A Multi-Cohort Study. *Cancers (Basel)*. 2019; 11.
2. Villanueva A, Lango DL. Hepatocellular Carcinoma. *New England Journal of Medicine*. 2019; 380: 1450-62.
3. Jemal A, Ward EM, Johnson CJ, Cronin KA, Ma J, Ryerson B, Mariotto A, Lake AJ, Wilson R, Sherman RL, Anderson RN, Henley SJ, Kohler BA, et al. Annual Report to the Nation on the Status of Cancer, 1975-2014, Featuring Survival. *J Natl Cancer Inst*. 2017; 109.
4. Zeisberg EM, Tarnavski O, Zeisberg M, Dorfman AL, McMullen JR, Gustafsson E, Chandraker A, Yuan X, Pu WT, Roberts AB, Neilson EG, Sayegh MH, Izumo S, et al. Endothelial-to-mesenchymal transition contributes to cardiac fibrosis. *Nat Med*. 2007; 13: 952-61.
5. Iwahashi S, Shimada M, Utsunomiya T, Imura S, Morine Y, Ikemoto T, Takasu C, Saito Y, Yamada S. Epithelial-mesenchymal transition-related genes are linked to aggressive local recurrence of hepatocellular carcinoma after radiofrequency ablation. *Cancer Lett*. 2016; 375: 47-50.
6. Diepenbruck M, Christofori G. Epithelial-mesenchymal transition (EMT) and metastasis: yes, no, maybe? *Curr Opin Cell Biol*. 2016; 43: 7-13.
7. Simeone P, Trerotola M, Franck J, Cardon T, Marchisio M, Fournier I, Salzet M, Maffia M, Vergara D. The multiverse nature of epithelial to mesenchymal transition. *Semin Cancer Biol*. 2019; 58: 1-10.
8. Coulouarn C, Factor VM, Thorgeirsson SS. Transforming growth factor-beta gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. *Hepatology*. 2008; 47: 2059-67.

9. Yamada S, Okumura N, Wei L, Fuchs BC, Fujii T, Sugimoto H, Nomoto S, Takeda S, Tanabe KK, Kodera Y. Epithelial to mesenchymal transition is associated with shorter disease-free survival in hepatocellular carcinoma. *Ann Surg Oncol*. 2014; 21: 3882-90.
10. Zhu K, Dai Z, Pan Q, Wang Z, Yang GH, Yu L, Ding ZB, Shi GM, Ke AW, Yang XR, Tao ZH, Zhao YM, Qin Y, et al. Metadherin promotes hepatocellular carcinoma metastasis through induction of epithelial-mesenchymal transition. *Clin Cancer Res*. 2011; 17: 7294-302.
11. Fernando J, Malfettone A, Cepeda EB, Vilarrasa-Blasi R, Bertran E, Raimondi G, Fabra A, Alvarez-Barrientos A, Fernandez-Salguero P, Fernandez-Rodriguez CM, Giannelli G, Sancho P, Fabregat I. A mesenchymal-like phenotype and expression of CD44 predict lack of apoptotic response to sorafenib in liver tumor cells. *Int J Cancer*. 2015; 136: E161-72.
12. Yuan L, Xu ZY, Ruan SM, Mo S, Qin JJ, Cheng XD. Long non-coding RNAs towards precision medicine in gastric cancer: early diagnosis, treatment, and drug resistance. *Mol Cancer*. 2020; 19: 96.
13. Necula L, Matei L, Dragu D, Neagu AI, Mambet C, Nedeianu S, Bleotu C, Diaconu CC, Chivu-Economescu M. Recent advances in gastric cancer early diagnosis. *World J Gastroenterol*. 2019; 25: 2029-44.
14. Fan Y, Wang YF, Su HF, Fang N, Zou C, Li WF, Fei ZH. Decreased expression of the long noncoding RNA LINC00261 indicate poor prognosis in gastric cancer and suppress gastric cancer metastasis by affecting the epithelial-mesenchymal transition. *J Hematol Oncol*. 2016; 9: 57.
15. Han L, Jia L, Zan Y. Long intergenic noncoding RNA smad7 (Linc-smad7) promotes the epithelial-mesenchymal transition of HCC by targeting the miR-125b/SIRT6 axis. *Cancer Med*. 2020; 9: 9123-37.
16. Chen BW, Zhou Y, Wei T, Wen L, Zhang YB, Shen SC, Zhang J, Ma T, Chen W, Ni L, Wang Y, Bai XL, Liang TB. lncRNA-POIR promotes epithelial-mesenchymal transition and suppresses sorafenib sensitivity simultaneously in hepatocellular carcinoma by sponging miR-182-5p. *J Cell Biochem*. 2021; 122: 130-42.
17. Xiao C, Wan X, Yu H, Chen X, Shan X, Miao Y, Fan R, Cha W. LncRNA-AB209371 promotes the epithelial-mesenchymal transition of hepatocellular carcinoma cells. *Oncol Rep*. 2019; 41: 2957-66.
18. Zhu L, Yang N, Du G, Li C, Liu G, Liu S, Xu Y, Di Y, Pan W, Li X. LncRNA CRNDE promotes the epithelial-mesenchymal transition of hepatocellular carcinoma cells via enhancing the Wnt/ $\beta$ -catenin signaling pathway. *J Cell Biochem*. 2018; 120: 1156-64.
19. Zhang PF, Wang F, Wu J, Wu Y, Huang W, Liu D, Huang XY, Zhang XM, Ke AW. LncRNA SNHG3 induces EMT and sorafenib resistance by modulating the miR-128/CD151 pathway in hepatocellular carcinoma. *J Cell Physiol*. 2019; 234: 2788-94.

20. Yang L, Deng WL, Zhao BG, Xu Y, Wang XW, Fang Y, Xiao HJ. FOXO3-induced lncRNA LOC554202 contributes to hepatocellular carcinoma progression via the miR-485-5p/BSG axis. *Cancer Gene Ther.* 2021.
21. Wang Y, Liu Z, Yao B, Li Q, Wang L, Wang C, Dou C, Xu M, Liu Q, Tu K. Long non-coding RNA CASC2 suppresses epithelial-mesenchymal transition of hepatocellular carcinoma cells through CASC2/miR-367/FBXW7 axis. *Mol Cancer.* 2017; 16: 123.
22. Song S, Qiu X. LncRNA miR503HG inhibits epithelial-mesenchymal transition and angiogenesis in hepatocellular carcinoma by enhancing PDCD4 via regulation of miR-15b. *Dig Liver Dis.* 2021; 53: 107-16.
23. Du Y, Wang B, Jiang X, Cao J, Yu J, Wang Y, Wang X, Liu H. Identification and Validation of a Stromal EMT Related LncRNA Signature as a Potential Marker to Predict Bladder Cancer Outcome. *Front Oncol.* 2021; 11: 620674.
24. Zhong W, Zhang F, Huang C, Lin Y, Huang J. Identification of Epithelial-Mesenchymal Transition-Related lncRNA With Prognosis and Molecular Subtypes in Clear Cell Renal Cell Carcinoma. *Front Oncol.* 2020; 10: 591254.
25. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* 2015; 160: 48-61.
26. Srivastava S, Wong KF, Ong CW, Huak CY, Yeoh KG, Teh M, Luk JM, Salto-Tellez M. A morpho-molecular prognostic model for hepatocellular carcinoma. *Br J Cancer.* 2012; 107: 334-9.
27. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol.* 2021; 22: 96-118.
28. Ma X, Mo M, Tan HJJ, Tan C, Zeng X, Zhang G, Huang D, Liang J, Liu S, Qiu X. LINC02499, a novel liver-specific long non-coding RNA with potential diagnostic and prognostic value, inhibits hepatocellular carcinoma cell proliferation, migration, and invasion. *Hepatol Res.* 2020; 50: 726-40.
29. Du J, Chen M, Liu J, Hu P, Guan H, Jiao X. LncRNA F11-AS1 suppresses liver hepatocellular carcinoma progression by competitively binding with miR-3146 to regulate PTEN expression. *J Cell Biochem.* 2019; 120: 18457-64.
30. Deng Y, Wei Z, Huang M, Xu G, Wei W, Peng B, Nong S, Qin H. Long non-coding RNA F11-AS1 inhibits HBV-related hepatocellular carcinoma progression by regulating NR1I3 via binding to microRNA-211-5p. *J Cell Mol Med.* 2020; 24: 1848-65.
31. Sun T, Wu Z, Wang X, Wang Y, Hu X, Qin W, Lu S, Xu D, Wu Y, Chen Q, Ding X, Guo H, Li Y, et al. LNC942 promoting METTL14-mediated m(6)A methylation in breast cancer cell proliferation and progression. *Oncogene.* 2020; 39: 5358-72.

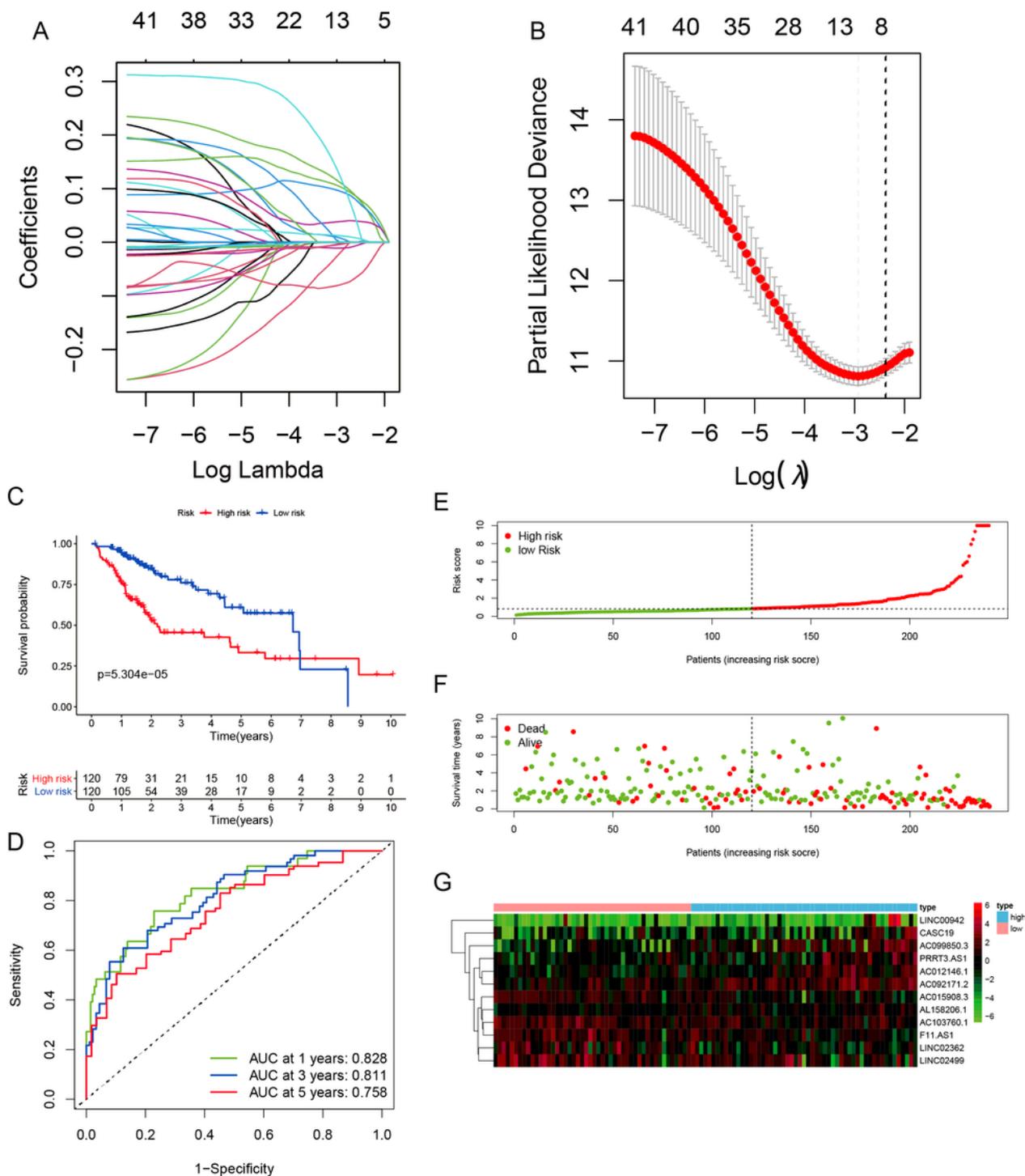
32. Lu T, Wei GH, Wang J, Shen J. LncRNA CASC19 contributed to the progression of pancreatic cancer through modulating miR-148b/E2F7 axis. *Eur Rev Med Pharmacol Sci.* 2020; 24: 10462-71.
33. Wang L, Lin C, Sun N, Wang Q, Ding X, Sun Y. Long non-coding RNA CASC19 facilitates non-small cell lung cancer cell proliferation and metastasis by targeting the miR-301b-3p/LDLR axis. *J Gene Med.* 2020; 22: e3254.
34. Wang WJ, Guo CA, Li R, Xu ZP, Yu JP, Ye Y, Zhao J, Wang J, Wang WA, Zhang A, Li HT, Wang C, Liu HB. Long non-coding RNA CASC19 is associated with the progression and prognosis of advanced gastric cancer. *Aging (Albany NY).* 2019; 11: 5829-47.
35. Wang XD, Lu J, Lin YS, Gao C, Qi F. Functional role of long non-coding RNA CASC19/miR-140-5p/CEMIP axis in colorectal cancer progression in vitro. *World J Gastroenterol.* 2019; 25: 1697-714.
36. Liu H, Zheng W, Chen Q, Zhou Y, Pan Y, Zhang J, Bai Y, Shao C. lncRNA CASC19 Contributes to Radioresistance of Nasopharyngeal Carcinoma by Promoting Autophagy via AMPK-mTOR Pathway. *Int J Mol Sci.* 2021; 22.
37. Liu YJ, Guo RX, Han LP, Gu H, Liu MZ. [Effect of CASC19 on proliferation, apoptosis and radiation sensitivity of cervical cancer cells by regulating miR-449b-5p expression]. *Zhonghua Fu Chan Ke Za Zhi.* 2020; 55: 36-44.
38. Sitohy B, Nagy JA, Dvorak HF. Anti-VEGF/VEGFR therapy for cancer: reassessing the target. *Cancer Res.* 2012; 72: 1909-14.
39. Zhu AX, Duda DG, Sahani DV, Jain RK. HCC and angiogenesis: possible targets and future directions. *Nat Rev Clin Oncol.* 2011; 8: 292-301.
40. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, Lencioni R, Koike K, Zucman-Rossi J, Finn RS. Hepatocellular carcinoma. *Nat Rev Dis Primers.* 2021; 7: 6.
41. Fu J, Chen Y, Cao J, Luo T, Qian YW, Yang W, Ren YB, Su B, Cao GW, Yang Y, Yan YQ, Shen F, Wu MC, et al. p28GANK overexpression accelerates hepatocellular carcinoma invasiveness and metastasis via phosphoinositol 3-kinase/AKT/hypoxia-inducible factor-1 $\alpha$  pathways. *Hepatology.* 2011; 53: 181-92.
42. Li Z, Wu T, Zheng B, Chen L. Individualized precision treatment: Targeting TAM in HCC. *Cancer Lett.* 2019; 458: 86-91.
43. Zhang J, Zhang Q, Lou Y, Fu Q, Chen Q, Wei T, Yang J, Tang J, Wang J, Chen Y, Zhang X, Zhang J, Bai X, et al. Hypoxia-inducible factor-1 $\alpha$ /interleukin-1 $\beta$  signaling enhances hepatoma epithelial-mesenchymal transition through macrophages in a hypoxic-inflammatory microenvironment. *Hepatology.* 2018; 67: 1872-89.

# Figures



**Figure 1**

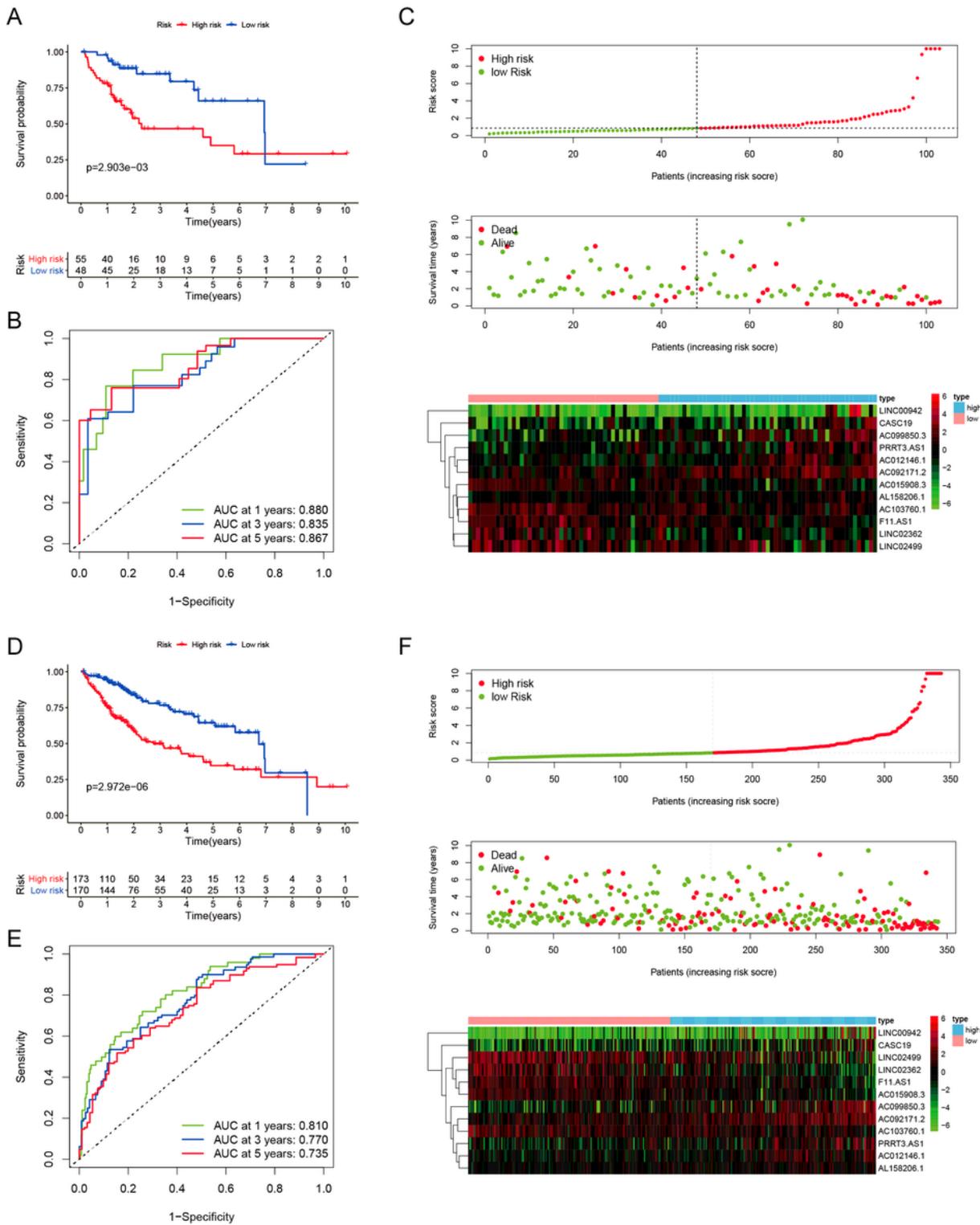
Flowchart of this study. EMT: epithelial-mesenchymal transition; LncRNA: long non-coding RNA; TCGA: The Cancer Genome Atlas; LASSO: the least absolute shrinkage and selection operator; ROC: receiver operating characteristic; GSEA: gene set enrichment analysis; GO: Gene Ontology.



**Figure 2**

Construction of the EMT-related lncRNA signature. (A) LASSO coefficient profiles of the 41 survival-associated EMT-related lncRNAs. (B) Cross-validation for tuning parameter selection in the proportional hazards model. (C) Kaplan-Meier survival analysis of the high-risk and low-risk groups. (D) Time-dependent receiver operating characteristic (time-ROC) curve evaluated the precision of this prognostic model in determining the overall survival (OS) rates of patients in 1-year, 3-years, and 5-years. (E) Risk

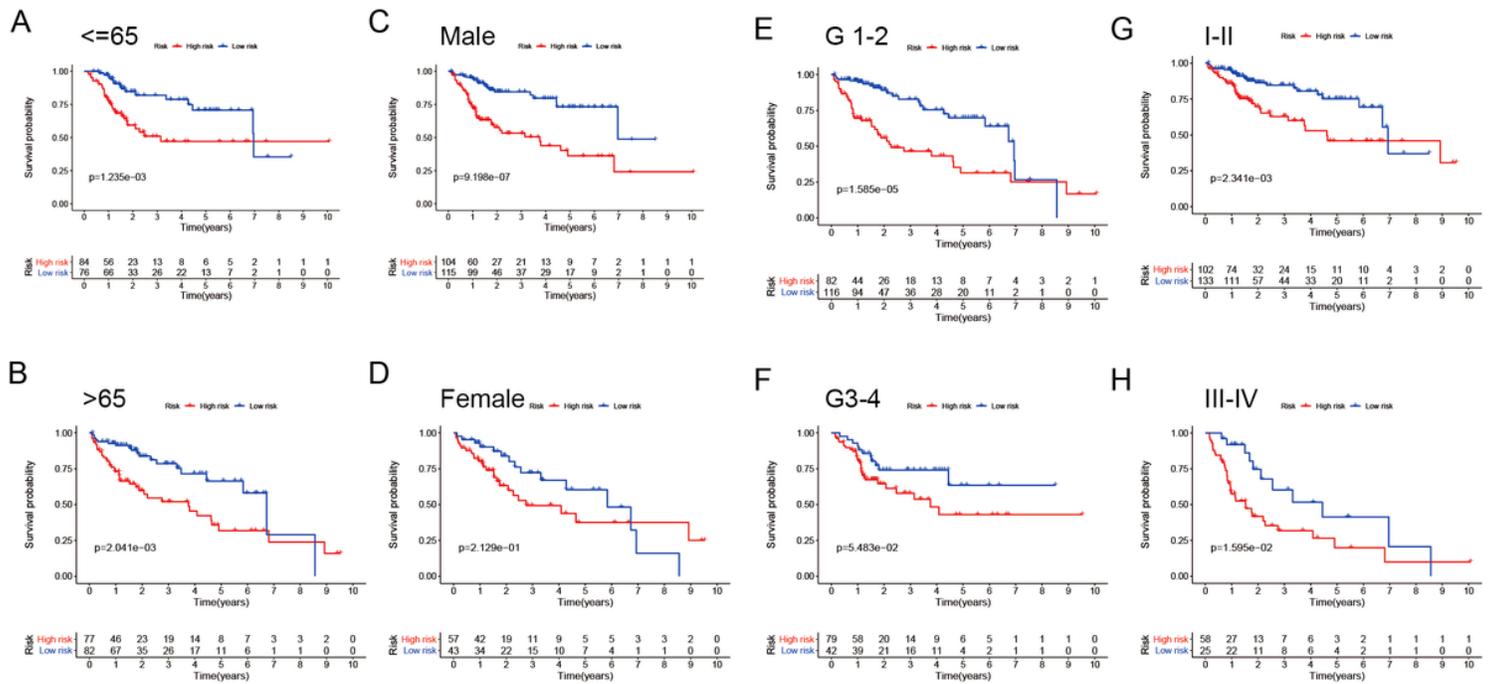
score curves based on the risk score of each patient. (F) Scatter plots based on the survival status of each patient. (G) Heatmap displaying the expression levels of the 12 EMT-related lncRNAs in high-risk and low-risk groups.



**Figure 3**

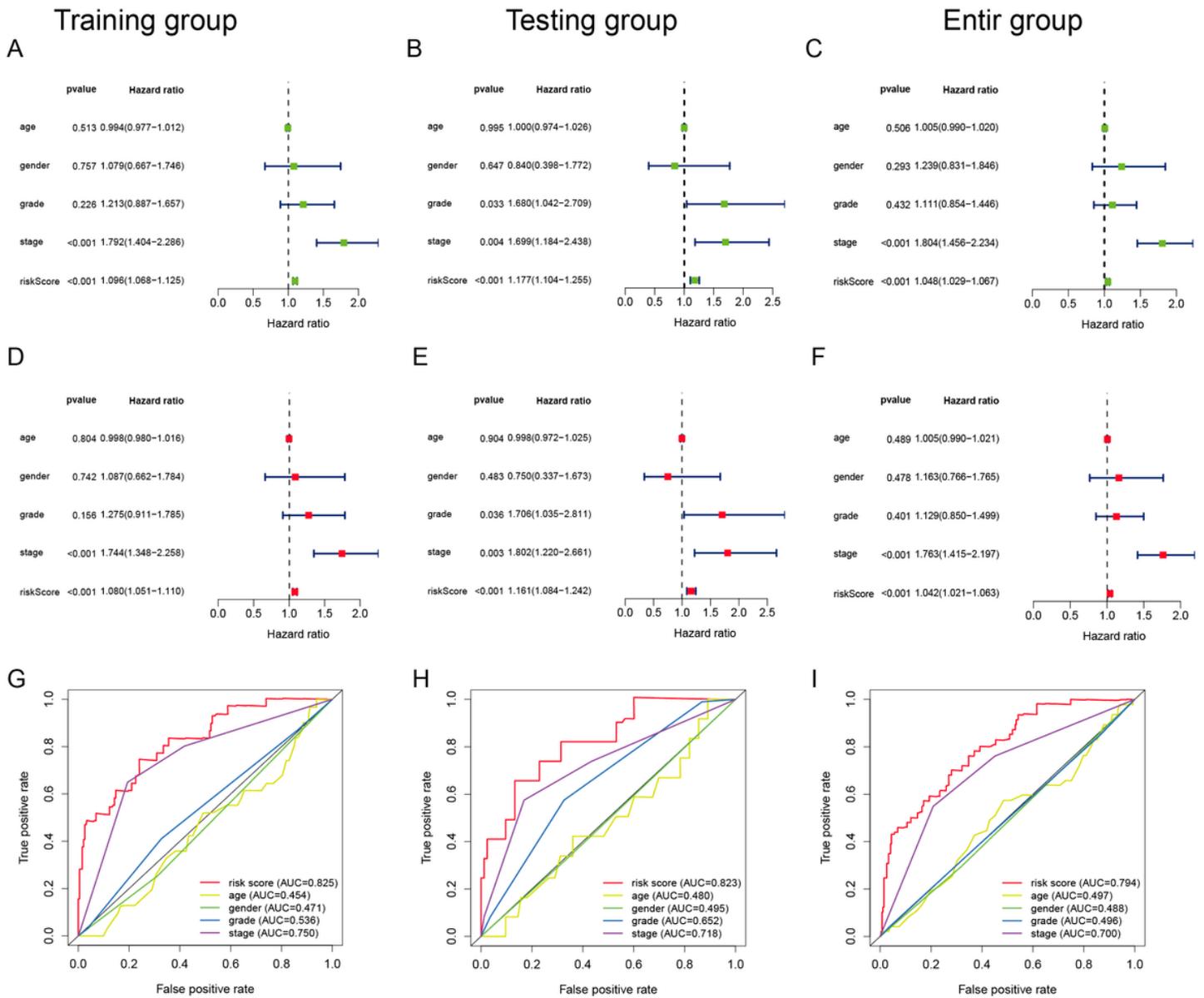
Validation of the EMT-related lncRNA signature in the testing and entire group of patients. (A-C): Kaplan-Meier curve, time-dependent ROC analysis, risk plot, survival status, and heatmap of the mRNA

expression of the lncRNA signature for HCC patients in the testing group. (D-F): Kaplan–Meier curve, time-dependent ROC analysis, risk plot, survival status, and heatmap of the mRNA expression of the lncRNA signature for HCC patients in the entire group.



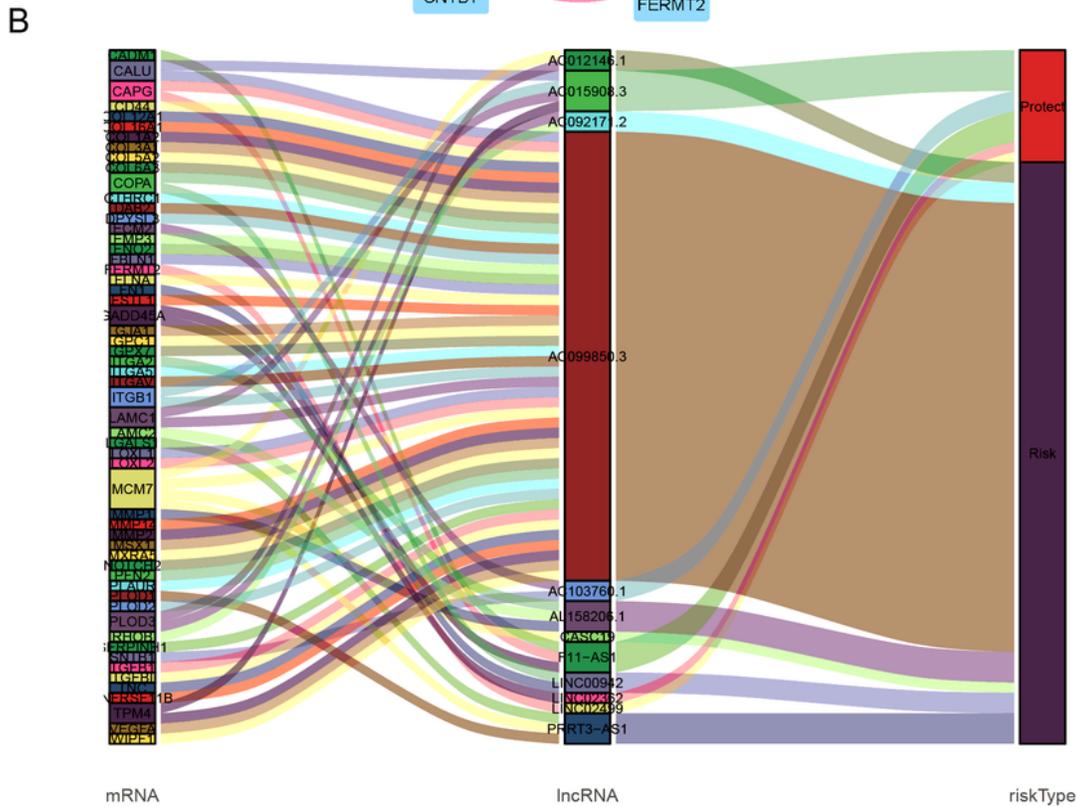
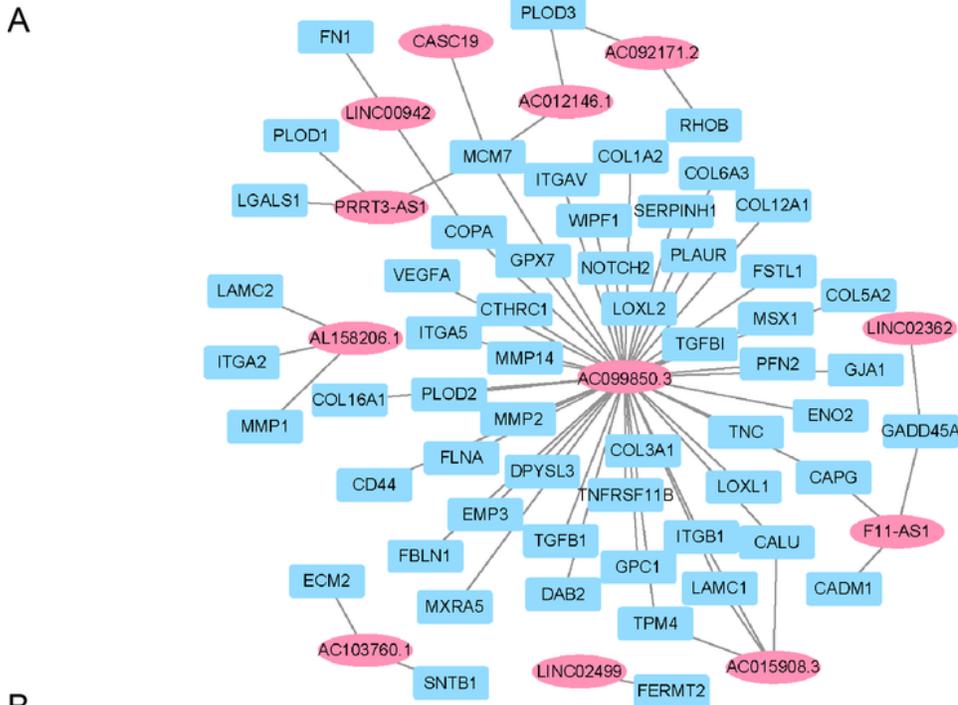
**Figure 4**

Stratification analysis of the correlation between the risk group and survival based on the age and gender of the patients and the grade and clinical stage of HCC. (A) Age  $\leq 60$ ; (B) Age  $> 60$ ; (C) Male; (D) Female; (E) Grade 1–2; (F) Grade 3–4; (G) Clinical stage I–II; (H) Clinical stage III–IV.



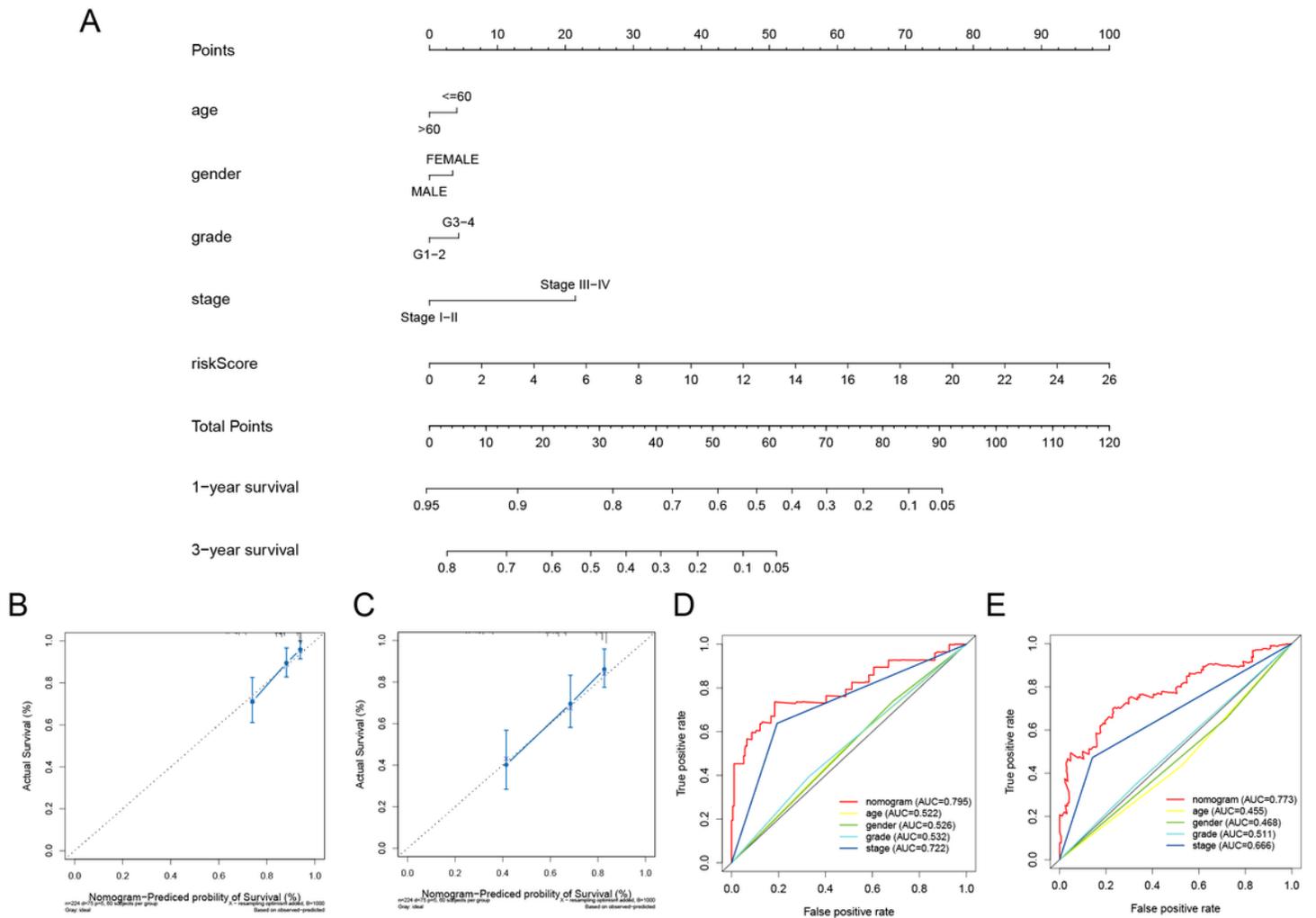
**Figure 5**

Univariate and multivariate Cox regression analyses and multi-ROC curve analysis of the risk score model and clinical factors. (A-C) Univariate Cox regression analysis in the training, testing, and entire groups; (D-F) Multivariate Cox regression analysis in the training, testing, and entire groups; (G-I) Multi-ROC curve analysis in the training, testing, and entire groups. Multi-ROC: Multi-indicator receiver operating characteristic.



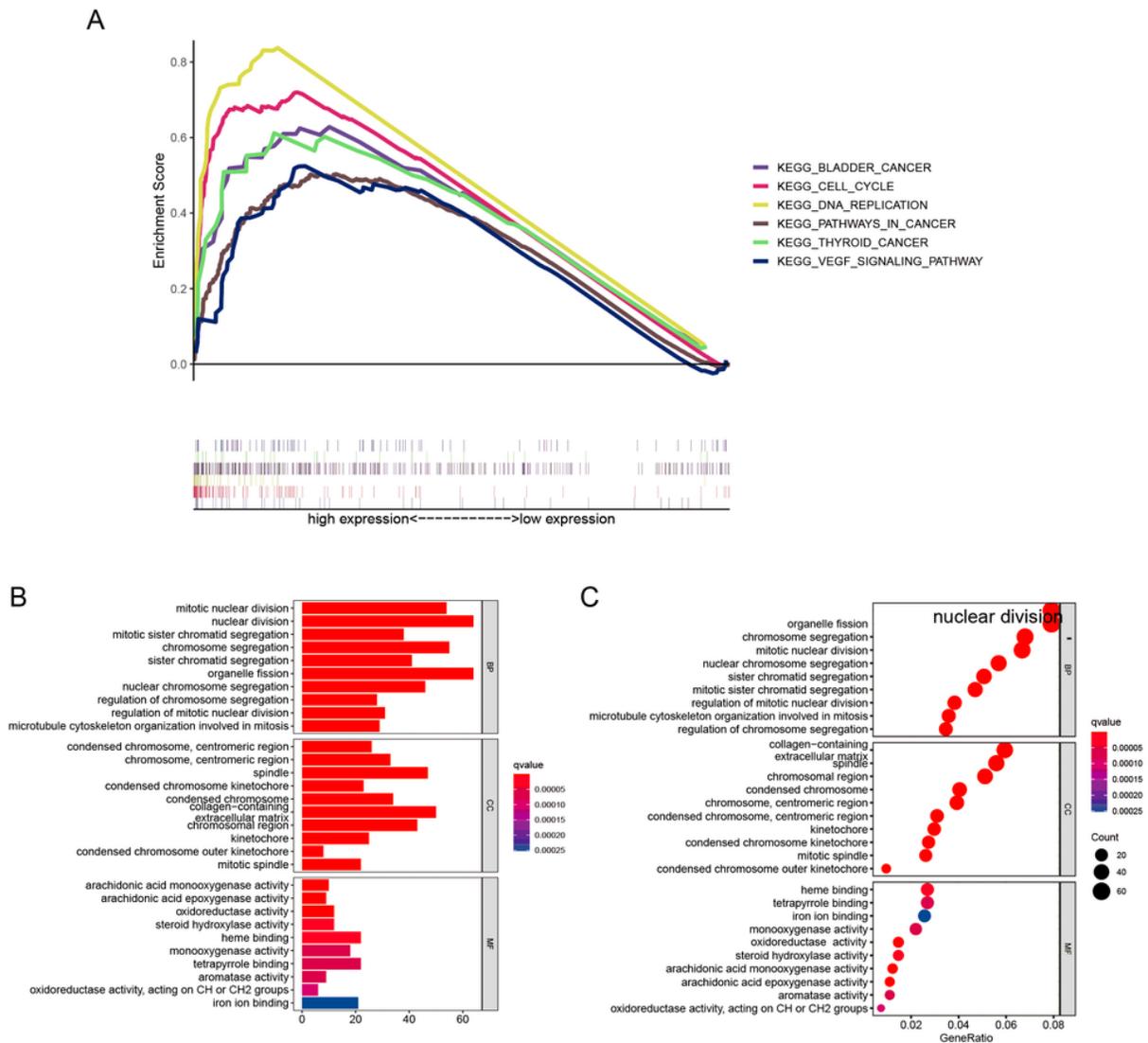
**Figure 6**

Co-expression network of the prognostic EMT-related lncRNAs-mRNAs and visualization by Cytoscape (A) and Sankey diagram (B)



**Figure 7**

Construction of the predictive nomogram for HCC patients from the TCGA database. (A) Nomogram predicting the probability of 1-year and 3-year OS for the prognosis of HCC patients. (B, C) Calibration curve for the predictive accuracy of the nomogram. The Y-axis represents actual survival rate and the X-axis represents nomogram-predicted survival rate. (D, E) ROC curve analysis of the nomogram for 1-year and 3-year OS of HCC patients.



**Figure 8**

Functional analysis between the low- and high-risk patients in the entire group. (A) GSEA based on the risk of grouping in the cohort from TCGA. (B) GO analysis of the differentially expressed genes (DEGs) between high- and low-risk groups.

