Identification of lncRNA prognostic signature and analysis of their functions for HCV-related hepatocellular carcinoma

Kun Wang

Cancer Hospital of China Medical University

Wenxin Li

Cancer Hospital of China Medical University

Yefu Liu

Cancer Hospital of China Medical University

Zhiqiang Hao

Cancer Hospital of China Medical University

Xiangdong Hua (✉ xdh20927@sohu.com)

Cancer Hospital of China Medical University

Zuohong Ma

Cancer Hospital of China Medical University

Zhaoqiang Feng

Cancer Hospital of China Medical University

Research

Keywords: hepatocellular carcinoma; Hepatitis C virus; competing endogenous RNA; co-expression; signature; prognosis

Posted Date: August 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-50007/v1>

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

[Read Full License](#)

Abstract

Background Hepatitis C virus (HCV) infection is a main contribution to the increase in hepatocellular carcinoma (HCC) incidence and patients' death recently, but prognostic biomarkers for HCV-related HCC remain rarely reported. This study was to identify an lncRNA prognostic signature for HCV-HCC patients and explore their underlying function mechanisms.

Methods In total, 102 HCV-HCC samples and 50 normal control samples were obtained from The Cancer Genome Atlas (TCGA) database. Univariate and multivariate Cox regression analysis were conducted to screen an lncRNA signature that could predict overall survival (OS) and then, the risk score was calculated using this signature. The prognostic potential of this risk score was evaluated by drawing Kaplan-Meier, receiver operating characteristic (ROC) curves and performing multivariate Cox regression analyses with clinical variables. Furthermore, a co-expression and competing endogenous RNA (ceRNA) networks were constructed to explore the functional mechanisms of lncRNAs.

Results Multivariate Cox regression showed six lncRNAs (SLC16A1-AS1, ZFPM2-AS1, JARID2-AS1, LINC01426, USP3-AS1 and LYPLAL1-AS1) were significantly associated with OS of HCV-HCC patients. These six lncRNAs were used to establish a risk score model, which displayed a higher prognosis prediction accuracy [area under the ROC curve (AUC) = 0.95 for training set; AUC = 0.885 for testing; AUC = 0.907 for entire set]. Also, this was independent of various clinical variables. The crucial co-expression (LINC01426/SLC16A1-AS1-AURKA/SFN/CCNB1, ZFPM2-AS1/LYPLAL1-AS1/JARID2-AS1-TSSK6) or ceRNA (USP3-AS1-hsa-miR-383-SFN) interaction axes were identified.

Conclusion Our study identified a novel six-lncRNA prognosis signature for HCV-HCC patients and indicated their underlying mechanisms for HCC progression.

Introduction

Hepatocellular carcinoma (HCC) is one of the most frequently observed malignancies worldwide, showing gradually increasing morbidity and mortality due to infection by hepatitis C virus (HCV) [1]. Recently, great progress has been made for treatment of HCC (such as surgical resection, chemoradiotherapy and liver transplantation), which have significantly improved the survival rate of patients [2]. However, due to the high heterogeneity of HCC, clinical outcomes are substantially variable among different patients, even in the same subgroup according to the classification of traditional clinical characteristics (such as stage) [3, 4]. Therefore, there is urgent to screen more effective prognostic biomarkers to early identify high-risk HCV-HCC patients and give them individualized treatment for the improvement of prognosis.

Long non-coding RNAs (lncRNAs) are a batch of newly-discovered noncoding RNAs that are approximately 200 nucleotides in length. Accumulating evidence has revealed lncRNAs play a crucial role in HCV-HCC. For example, Zhang C et al. found lncRNA HEIH expression was increased in serum and exosomes of HCV-related HCC patients compared to patients with chronic hepatitis C [5]. The study of

Zhang Q et al. illustrated lncRNA aHIF and PAR5 were specifically downregulated for HCV-related HCC compared with normal liver specimen [6]. By microarray analysis, Zhang H et al. observed that in comparison with matched non-tumor liver tissues, LINC01419 was significantly overexpressed in HCV-related HCC. Functional predictions indicated that LINC01419 may be a key factor for HCV-related HCC by modulating cell cycle progression [7]. Sur et al. speculated LINC00657 favored uncontrolled cell growth of HCV-HCC cells by forming a complex with miR-373 and then regulating their common target Wee1 [8]. These findings imply lncRNAs may be underlying prognostic biomarkers for HCV-HCC patients.

Although several studies have reported the prognostic value of lncRNAs for HCC patients [9–11], none of them specifically focused on the HCV-HCC. The goal of the present study was, for the first time, to identify a prognostic signature for HCV-HCC patients through using the sequencing and clinical data deposited in The Cancer Genome Atlas (TCGA) database. Moreover, the potential functions of lncRNAs were predicted by constructing a competing endogenous RNA (ceRNA) regulatory network and co-expression network, which are important mechanisms for lncRNAs as described previously [12, 13].

Material And Methods

Data source

Level 3 mRNA- and miRNA-sequencing data from 373 HCC and 50 normal control samples were retrieved from the TCGA database portal (<https://portal.gdc.cancer.gov/>) on April 20, 2019. Only 102 HCC samples were utilized because they met the following inclusion criteria: (1) both of mRNA-and miRNA were sequenced; (2) HCV-positive alone, not consisting of other virus infection; and (3) having survival information. These 102 patients were randomized to a training dataset (n = 51) and a testing dataset (n = 51) groups for the subsequent analysis (Table 1; supplementary file 1).

Table 1
Characteristics of the patients

Clinical characteristics	Training set (N = 51)	Validation set (N = 51)	Entire set (N = 102)
Age (years, mean $\pm$ SD)	58.31 $\pm$ 13.69	58.22 $\pm$ 13.99	58.26 $\pm$ 13.77
Gender (male/female)	39/12	37/14	76/26
Neoplasm histologic grade (G1/G2/G3/-)	10/31/10/0	14/24/12/1	24/55/22/1
Pathologic M (M0/M1/-)	32/0/19	26/0/25	58/0/44
Pathologic N (N0/N1/-)	35/1/15	27/0/24	62/1/39
Pathologic T (T1/T2/T3/T4/-)	16/14/20/0/1	16/16/12/5/2	32/30/32/5/3
Pathologic stage (stage I/II/III/IV/-)	16/13/21/0/1	16/15/17/0/3	32/28/38/0/4
Tumor recurrence (Yes/No/-)	23/21/7	19/28/4	42/49/11
Risk score status (high/low)	25/26	25/26	51/51
Death (dead/alive)	23/28	24/27	47/55
Overall survival days (months, mean $\pm$ SD)	18.89 $\pm$ 17.65	19.36 $\pm$ 18.44	19.13 $\pm$ 17.96
SD, standard deviation; Pathologic TNM, tumor, node, metastasis			

Differentially Expressed Rnas Between Hcv-hcc And Non-cancer Tissues

The lncRNAs and protein-coding mRNAs in RNA-sequencing data were annotated using the HUGO Gene Nomenclature Committee (HGNC; <http://www.genenames.org/>) [14]. The differentially expressed lncRNAs (DELs), genes (DEGs) and miRNAs (DEMs) were screened using the Linear Models for Microarray Data (LIMMA) method (version 3.34.7; <https://bioconductor.org/packages/release/bioc/html/limma.html>) [15] in R package. False discovery rate (FDR) < 0.05 and $|\log_{2}(\text{fold change})| > 1$ were considered as the statistical threshold. Bidirectional hierarchical clustering of DEGs, DELs or DEMs was performed using pheatmap package (version: 1.0.8; <https://cran.r-project.org/web/packages/pheatmap/>) based on centered Pearson's correlation.

Selection Of Survival-related Genes

Univariate Cox regression analysis using the survival package (version 2.41-1; <http://bioconductor.org/packages/survival/>) was performed to identify overall survival (OS)-related

DEGs, DELs and DEMs in the training set. Log-rank p-value < 0.05 was defined as the statistical cut-off point.

Identification Of Prognostic Lncrna Signature

LncRNAs significantly related with OS in univariate Cox regression analysis were then entered into the multivariate analysis to obtain a lncRNA prognostic signature model ($p < 0.05$) which was used to calculate the prognostic risk score according to the following formula:

$$\text{Risk score} = \sum \beta_{\text{lncRNA}} \times \text{Exp}_{\text{lncRNA}}$$

($\text{Exp}_{\text{lncRNA}}$ was the expression level of DELs; β_{lncRNA} was the regression coefficient for the DELs in multivariate Cox hazard model analysis).

Using the median risk score as the threshold, the HCV-HCC patients were divided into low-risk and high-risk groups. Kaplan-Meier survival curves were used to calculate the differences in OS between high- and low-risk groups of training, testing and entire datasets. Receiver operating characteristic (ROC) curve and area under the curve (AUC) were also applied to evaluate the predictive potential of this risk score model in each dataset.

Moreover, univariate and multivariate Cox regression analyses were also carried out to verify whether the prognostic potential of the risk score model was independent of other clinical features (including age, gender, histologic grade, pathologic TNM stage, recurrence and survival) for patients with HCV-HCC.

Construction Of Lncrna Cerna And Co-expression Networks

To explore the molecular mechanisms underlying our prognostic signature genes, lncRNA ceRNA and co-expression networks were constructed.

The DIANA-LncBase (version 2.0; http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=lncbasev2/index-predicted) [16] database was used to predict the interaction relationships between signature DELs and prognostic DEMs. The starBase database (version 2.0; <http://starbase.sysu.edu.cn/>) [17] was used to predict the interactions between prognostic DEMs and prognostic DEGs. Only the interactors with opposite expression direction were left to create the ceRNA network, which was visualized using Cytoscape (version 3.6.1; www.cytoscape.org/) [18].

The correlation between signature DELs and DEGs was assessed by Pearson correlation coefficients which were calculated using `tcor.test` function (<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/cor.test.html>) in R. To prevent the oversize network, only the co-expression pairs with Pearson correlation coefficient (PCC) > 0.6 were selected to create the lncRNA-mRNA co-expression network using Cytoscape (version 3.4; www.cytoscape.org/) [18].

Function Enrichment Analysis

Enrichr online tool (<http://amp.pharm.mssm.edu/Enrichr/>) was used to carry out Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses for DEGs in the lncRNA related co-expression-ceRNA network [19]. FDR < 0.05 was considered statistically significant.

Results

Differential expression analysis

According to the RefSeq ID in the downloaded data, a total of 15,885 mRNAs, 1,241 lncRNAs and 1,046 miRNAs were annotated. Under the threshold of FDR < 0.05 and $|\log_2FC| > 1$, 29 downregulated and 590 upregulated mRNAs; 14 downregulated and 199 upregulated lncRNAs; 54 downregulated and 29 upregulated miRNAs were identified to be differentially expressed in the HCV-HCC tissues compared with those of the normal controls. Volcano plots were used to display the distribution of the DELs-DEGs and DEMs (Fig. 1A). The heat map showed the expression of these DELs, DEGs and DEMs can clearly distinguish HCV-HCC from normal tissues (Fig. 1B).

Identification Of Lncrna Signature With Prognostic Value

Univariate Cox regression analysis identified 214 differentially expressed RNAs that were significantly related with OS in the training set of HCV-HCC patients, including 154 DEGs, 20 DELs and 40 DEMs. These 20 candidate DELs were then entered into multivariate Cox regression, 6 of which (SLC16A1-AS1, ZFPM2-AS1, JARID2-AS1, LINC01426, USP3-AS1 and LYPLAL1-AS1) were shown to be independent predictive biomarkers. Thus, they may be potential signature lncRNAs that can optimally predict the OS of patients with HCV-HCC (Table 2). Subsequently, a risk score system of this 6-lncRNA signature was established for each patient according to the following formula: risk score = $(2.06022) \times \text{Exp}_{\text{SLC16A1-AS1}} + (-4.60301) \times \text{Exp}_{\text{ZFPM2-AS1}} + (-4.24416) \times \text{Exp}_{\text{JARID2-AS1}} + (0.73659) \times \text{Exp}_{\text{LINC01426}} + (1.93958) \times \text{Exp}_{\text{USP3-AS1}} + (-3.73102) \times \text{Exp}_{\text{LYPLAL1-AS1}}$.

Table 2
Prognostic RNAs significantly associated with OS

	ID	logFC	FDR	β	HR	95%CI	P-value
lncRNA	SLC16A1-AS1	1.65	2.23E-10	2.06	7.85	1.65–10.39	4.85E-03
	ZFPM2-AS1	3.74	1.03E-04	-4.60	0.10	0.02–0.62	1.42E-02
	JARID2-AS1	1.09	2.44E-02	-4.24	0.14	0.03–0.65	1.44E-02
	LINC01426	3.07	4.50E-07	0.74	2.09	1.39–4.65	3.55E-02
	USP3-AS1	1.19	1.74E-12	1.94	6.96	1.77–12.88	4.21E-02
	LYPLAL1-AS1	1.56	1.67E-02	-3.73	0.24	0.03–0.90	4.73E-02
mRNA	AURKA	2.01	4.21E-18		1.56		2.70E-02
	CCNB1	2.24	5.11E-15		1.51		7.00E-03
	MSH2	1.25	5.11E-15		1.76		4.80E-02
	SFN	2.14	4.47E-09		1.39		1.05 E-03
	TSSK6	1.66	9.39E-11		0.37		2.50E-02
	TMPRSS9	1.03	1.02E-08		0.53		1.80E-02
miRNA	hsa-miR-383	-2.18	2.48E-12		0.84		4.10E-02
Multivariate analysis results for lncRNAs and univariate analysis results for miRNA and mRNAs. FDR, False discovery rate; HR, hazard ratio; CI, confidence interval.							

Using the median risk score as the cut-off value, the patients were assigned into a high-risk group (n = 26) and a low-risk group (n = 25). The Kaplan-Meier survival curve and log-rank test showed that OS was significantly worse in HCV-HCC patients of the high risk group than that in patients of the low risk group in the training (hazard ratios [HR] = 3.798, 95% confidence intervals [CI] = 1.545–9.332) (Fig. 2A), testing (HR = 2.688, 95% CI = 1.134–6.373) (Fig. 2B) and entire set (HR = 3.206, 95% CI = 1.723–5.966) (Fig. 2C). The AUC of the ROC curve was respectively 0.95, 0.885, and 0.907 for the training, testing and entire set (Fig. 2D). These findings suggested a stronger prediction potential of this 6-lncRNA prognostic model.

Independence Of 6-lncrna Signature For Survival Prediction

Univariate analysis results indicated that neoplasm histologic grade (G1/G2/G3/-), pathologic T (T1/T2/T3/T4/-), pathologic stage (stage I/II/III/IV/-) and risk score status were significant factors associated with prognosis. However, only pathologic stage (stage I/II/III/IV/-) and risk score status were still significant in the multivariate Cox regression analysis (Table 3). These findings further validated this 6-miRNA signature may be an independent prognostic factor. Also, the stratification analysis showed this 6-miRNA signature can further divide the patients in each stage into high and low risk groups which

possessed significantly different OS ratio (Fig. 3), indicating our molecular risk score had higher predictive power for OS than the clinically used stage classification system.

Table 3
Univariate and multivariate analysis of lncRNA signature and survival in different patient sets

Variables	Univariate analysis			Multivariate analysis		
	HR	95%CI	P-value	HR	95%CI	P-value
Training set (N = 51)						
Age (years, mean $\pm$ SD)	0.975	0.947–1.005	1.03E-01	-	-	-
Gender (male/female)	1.086	0.425–2.772	8.64E-01	-	-	-
Neoplasm histologic grade (G1/G2/G3/-)	0.965	0.481–1.940	9.22E-01	-	-	-
Pathologic M (M0/M1/-)	-	-	-	-	-	-
Pathologic N (N0/N1/-)	2.867	0.366–22.43	2.94E-01	-	-	-
Pathologic T (T1/T2/T3/T4/-)	2.232	1.214–4.105	6.76E-03	0.516	0.175–1.523	2.31E-01
Pathologic stage (stage I/II/III/IV/-)	2.542	1.333–4.846	4.61E-03	4.269	1.274–14.308	1.87E-02
Tumor recurrence (Yes/No/-)	0.749	0.289–1.948	5.53E-01	-	-	-
Risk score status (high/low)	3.798	1.545–9.332	1.86E-03	3.583	1.349–9.509	1.04E-02
Testing set (N = 51)						
Age(years, mean $\pm$ SD)	1.005	0.976–1.036	7.34E-01	-	-	-
Gender (male/female)	1.686	0.624–4.557	2.98E-01	-	-	-
Neoplasm histologic grade (G1/G2/G3/-)	1.801	1.014–3.195	4.04E-02	1.21	0.659–2.221	5.38E-01
Pathologic M (M0/M1/-)	-	-	-	-	-	-
Pathologic N (N0/N1/-)	-	-	-	-	-	-
Pathologic T (T1/T2/T3/T4/-)	2.065	1.308–3.258	1.12E-03	2.361	0.509–10.938	2.72E-01

CI, confidence interval; HR, hazard ratio; TNM, tumor, node, metastasis. Bold indicates significance at $p < 0.05$.

Variables	Univariate analysis			Multivariate analysis		
	HR	95%CI	P-value	HR	95%CI	P-value
Pathologic stage (stage I/II/III/IV/-)	2.994	1.511–5.932	5.79E-04	2.579	1.279–5.196	8.05E-03
Tumor recurrence (Yes/No/-)	1.784	0.736–4.324	1.94E-01	-	-	-
Risk score status (high/low)	2.688	1.134–6.373	1.96E-02	2.939	1.119–7.721	2.87E-02
Entire set (N = 102)						
Age (years, mean ± SD)	0.991	0.971–1.012	4.12E-01	-	-	-
Gender (male/female)	1.278	0.715–2.283	4.07E-01	-	-	-
Neoplasm histologic grade G1/G2/G3/-)	1.416	0.915–2.193	1.18E-01	-	-	-
Pathologic M (M0/M1/-)	-	-	-	-	-	-
Pathologic N (N0/N1/-)	2.548	0.342–19.01	3.44E-01	-	-	-
Pathologic T (T1/T2/T3/T4/-)	2.124	1.486–3.035	1.85E-05	1.115	0.319–3.897	8.64E-01
Pathologic stage (stage I/II/III/IV/-)	2.471	1.720–4.365	4.66E-06	2.507	1.600–3.930	6.11E-05
Tumor recurrence (Yes/No/-)	1.175	0.615–2.247	6.25E-01	-	-	-
Risk score status (high/low)	3.206	1.723–5.966	1.06E-04	3.326	1.683–6.574	5.47E-04
CI, confidence interval; HR, hazard ratio; TNM, tumor, node, metastasis. Bold indicates significance at p < 0.05.						

Functional Characteristics Of Prognostic Signature

In total, 10 downregulated prognosis-related DEMs were predicted to interact with 4 upregulated signature DELs (SLC16A1-AS1, USP3-AS1, LINC01426 and ZFPM2-AS1) in the DIANA-LncBase database; while 37 upregulated prognostic DEGs were predicted to interact with 6 downregulated DEMs (hsa-miR-133b, hsa-miR-33b, hsa-miR-383, hsa-miR-424, hsa-miR-153 and hsa-miR-511) by starBase database. These interaction pairs were integrated to construct an lncRNA-related ceRNA network (Fig. 4).

Under the threshold of PCC > 0.6, 5 upregulated signature DELs (SLC16A1-AS1, USP3-AS1, LINC01426, ZFPM2-AS1 and LYPLAL1-AS1) were discovered to be co-expressed with 133 prognostic DEGs, which were used to construct the co-expression network (Fig. 5).

The DEGs in these two networks were uploaded into the Enrichr database to predict their underlying functions. The results showed 18 GO biological process terms and 10 KEGG pathways were enriched (Table 4; Fig. 6). Four (MSH2, AURKA, SFN, CCNB1) of 14 DEGs (MAST2, SLC2A1, UBE2C, MSH2, ROBO1, DBF4, AURKA, EFNA3, CDCA8, SFN, TOP2A, ZWINT, RACGAP1 and CCNB1) that can be commonly regulated by lncRNAs in the co-expression and ceRNA networks were enriched into both GO terms and KEGG pathways, such as mitotic sister chromatid segregation (GO:0000070) (CCNB1), anaphase-promoting complex-dependent catabolic process (GO:0031145) (CCNB1, AURKA), mitotic spindle organization (GO:0007052) (CCNB1, AURKA), regulation of mitotic cell cycle phase transition (GO:1901990) (CCNB1, AURKA), DNA metabolic process (GO:0006259) (MSH2), cellular response to DNA damage stimulus (GO:0006974) (SFN), Cell cycle (CCNB1, SFN), Oocyte meiosis (CCNB1, AURKA), p53 signaling pathway (CCNB1, SFN) and Mismatch repair (MSH2). Accordingly, these four DEGs related ceRNA (USP3-AS1-hsa-miR-383-SFN) and co-expression (LINC01426/ SLC16A1-AS1-AURKA/SFN/CCNB1) relationship pairs may be especially important. In addition, ZFPM2-AS1, LYPLAL1-AS1 and JARID2-AS1 may be involved in HCC by co-expression with TSSK6, although the PCC of ZFPM2-AS1 and LYPLAL1-AS1 was < 0.6 and was not included in the co-expression network. TSSK6 was enriched in phosphorylation pathway. More important, the above crucial relationship pairs were selected because, as expected, the prognosis and expression trend between all these lncRNAs and mRNAs were consistent, but contrast between lncRNAs and miRNAs (Table 2).

Table 4
Function enrichment for genes regulated by lncRNAs

Categories	Term	FDR	Genes
GO BP	sister chromatid segregation (GO:0000819)	1.47E-05	TOP2A; SPAG5; KIFC1; CDCA8; KNSTRN; ZWINT
	mitotic sister chromatid segregation (GO:0000070)	1.47E-05	CCNB1; PSRC1; SPAG5; KIFC1; CDCA8; NCAPD2; KNSTRN; ZWINT
	anaphase-promoting complex-dependent catabolic process (GO:0031145)	1.30E-04	CDC20; CCNB1; PTTG1; UBE2C; UBE2S; CDK1; AURKA
	regulation of ubiquitin protein ligase activity (GO:1904666)	4.94E-04	CDC20; CCNB1; UBE2C; UBE2S; CDK1
	mitotic spindle organization (GO:0007052)	1.02E-03	TPX2; CCNB1; KIFC1; STMN1; BIRC5; AURKA
	DNA replication (GO:0006260)	1.12E-02	RECQL4; DBF4; POLD1; CDK1; MCM6; RHNO1
	positive regulation of ubiquitin protein ligase activity (GO:1904668)	1.12E-02	CDC20; CCNB1; UBE2C; UBE2S; CDK1
	regulation of mitotic cell cycle phase transition (GO:1901990)	1.12E-02	CDC20; TPX2; CCNB1; UBE2C; CEP131; CDK1; AURKA
	DNA metabolic process (GO:0006259)	1.12E-02	TOP2A; RECQL4; DBF4; MSH2; POLD1; CDK1; MCM6; TK1; RHNO1
	cellular response to DNA damage stimulus (GO:0006974)	1.29E-02	TOP2A; RECQL4; MSH2; POLD1; CDK1; SFN; IKBKE; ZBTB40; RHNO1
	G1/S transition of mitotic cell cycle (GO:0000082)	2.35E-02	CDKN2C; DBF4; RRM2; MCM6; CDKN3
	positive regulation of cell cycle process (GO:0090068)	2.35E-02	CCNB1; SPAG5; RACGAP1; AURKA; RHNO1
	regulation of mitotic cell cycle (GO:0007346)	2.38E-02	TPX2; CCNB1; PSRC1; TACC3; CDK1; BIRC5
	peptidyl-lysine modification (GO:0018205)	2.66E-02	TOP2A; CDCA8; BIRC5; P3H4; LOXL2
	phosphorylation (GO:0016310)	2.66E-02	PKN3; MAST2; ADM2; LIMK1; CDK1; TSSK6; BIRC5; IKBKE; AURKA
	G2/M transition of mitotic cell cycle (GO:0000086)	3.42E-02	CCNB2; CCNB1; CEP131; CDK1; AURKA

FDR, false discovery rate; GO, Gene ontology; BP, biological process; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Categories	Term	FDR	Genes
	cell cycle G2/M phase transition (GO:0044839)	3.45E-02	CCNB2;CCNB1;CEP131;CDK1;AURKA
	regulation of G2/M transition of mitotic cell cycle (GO:0010389)	4.95E-02	TPX2; CCNB1; CEP131; CDK1; AURKA
KEGG pathway	Cell cycle	1.18E-07	CDC20; CCNB2; CCNB1; CDKN2C; DBF4; PTTG1; CDK1; SFN; MCM6
	Oocyte meiosis	1.76E-05	CDC20; CCNB2; CCNB1; PTTG1; CDK1; CPEB3; AURKA
	p53 signaling pathway	1.06E-04	CCNB2; CCNB1; RRM2; CDK1; SFN
	Drug metabolism	5.43E-03	RRM2; UGT2B11; NAT2; TK1
	Central carbon metabolism in cancer	8.74E-03	G6PD; SLC2A1; SLC16A3
	Mismatch repair	9.70E-03	MSH2; POLD1
	Human T-cell leukemia virus 1 infection	1.43E-02	CDC20; CCNB2; CDKN2C; PTTG1; SLC2A1
	Pentose and glucuronate interconversions	2.05E-02	AKR1B10; UGT2B11
	DNA replication	2.29E-02	POLD1; MCM6
	Axon guidance	3.06E-02	EFNA3; LIMK1; PLXNA1; ROBO1
FDR, false discovery rate; GO, Gene ontology; BP, biological process; KEGG, Kyoto Encyclopedia of Genes and Genomes.			

Discussion

By using TCGA data and a serial of bioinformatic analysis procedures, a six-lncRNA signature (SLC16A1-AS1, ZFPM2-AS1, JARID2-AS1, LINC01426, USP3-AS1 and LYPLAL1-AS1) was identified as an independent prognosis predictor for HCV-HCC patients in our study. Also, it exhibited better predictive ability, with the AUC of 0.95, 0.885, and 0.907 for the training, testing and entire set, which seemed to be higher than that of other lncRNA signatures identified in overall HCC by other authors, including Liao et al. (13-lncRNA signature, AUC = 0.809, 0.782 and 0.79 for 1-, 3- and 5-year survival, respectively)[20], Sui et al. (4-lncRNA signature, AUC = 0.709)[10], Wu et al. (6-lncRNA signature, AUC = 0.830, 0.739, and 0.852 for the first, third, and fifth year)[21], Wang et al. (4-lncRNA signature, AUC = 0.73)[11], Zhao et al. (5-lncRNA signature, AUC = 0.857)[9] and Yan et al. (7-lncRNA signature, AUC = 0.752)[22]. Furthermore, most of

these lncRNAs were not included in the known prognostic lncRNA signature, except that ZFPM2-AS1 was reported to be associated with OS in patients with HCC [22] and LINC01426 was demonstrated to be an independent predictor for OS in glioma [23] and squamous cell lung carcinoma [24]. These findings suggested our six-lncRNA signature may be a novel, more effective biomarker for predicting the prognosis for HCC patients, especially HCV-HCC.

Wang et al. used the RT-PCR to confirm LINC01426 was expressed at high levels in both glioma tissues and cell lines. Upregulated LINC01426 expression was negatively associated with World Health Organization grade, Karnofsky performance scale scores and OS. *In vitro* mechanistic analysis revealed that silencing of LINC01426 inhibited glioma cells proliferation, invasion and migration, and triggered apoptosis by influencing PI3K/Akt signaling pathway [23]. These findings implied LINC01426 was a proto-oncogene and a high-risk factor for poor prognosis, which was also demonstrated in our study because of logFC = 3.07 and HR = 2.09 (Table 2). However, the mechanisms of LINC01426 remain rarely reported. In this study, we predicted LINC01426 could be involved in the development and progression of HCV-HCC by co-expressing with AURKA (aurora kinase A), SFN (stratifi; 14-3-3σ) or CCNB1 (cyclin B1). Also, as expected, these three genes were also oncogenes in cancer development as described as follows: Chen et al. found AURKA was up-regulated in HCC tissues and correlated with pathological stage and distant metastasis. Overexpression of AURKA promoted cancer metastases by inducing epithelial-mesenchymal transition and cancer stem cell behaviors through PI3K/AKT pathway [25]. AURKA was also reported to mediate the oncogenic activity of c-Myc in HCC [26, 27]. Knock-out of AURKA or use of AURKA inhibitor alisertib suppressed cell proliferation, invasiveness and arrested cell cycle at G₂/M phase *in vitro* and impaired tumor growth *in vivo* [25–28]. Increased AURKA predicted poor prognosis in bladder cancer [29] or non-small cell lung cancer [30]. Using a proteomics approach, Lee et al.[31], Chen et al.[32] and Reis et al. [33] all identified SFN protein was up-regulated in HCC. Boudreau et al. reported SFN mediated cell migration and invasion of breast cancer by directly regulating cytoskeletal dynamics [34]. Suppression of SFN expression by siRNA significantly reduced cell proliferation activity and the S-phase subpopulation of human lung adenocarcinoma cell line A549 and blocked tumor development in mice [35]. Based on analysis of ten ovarian cancer datasets, Hu et al. identified high expressed SFN was associated with significantly worse OS [36]. CCNB1 is a cell cycle gene and thus high levels of CCNB1 were extensively demonstrated to be associated with poor prognosis in HCC patients [37, 38]. Knockdown of CCNB1 by RNA interference significantly inhibited proliferation, migration and invasion in HCC cells [37]. In line with these studies, we also found the expressions of AURKA, SFN and CCNB1 were higher in HCC tissues than those in the normal controls and they were high-risk factors for poor prognosis (Table 2).

There were no studies to explore the roles of SLC16A1-AS1, LYPLAL1-AS1, JARID2-AS1 and USP3-AS1 until now. Our study, for the first time, predicted high expressed SLC16A1-AS1 may be a high-risk factor for poor prognosis in HCV-HCC by co-expressing with AURKA, SFN and CCNB1 which were also upregulated and associated with unfavorable prognosis as above described. High expressed LYPLAL1-AS1 and JARID2-AS1 may be low-risk factors for poor prognosis by co-expressing with TSSK6 which was

also upregulated to result in excellent prognosis. Although rare studies were performed to directly confirm the roles of TSSK6 in cancer, it had been reported that TSSK6 expression was spatiotemporally coincident with γ H2AX formation and loss of TSSK6 blocked phosphorylated histone 2AX (γ -H2AX) formation [39]; γ -H2AX is a DNA damage biomarker. Excessive DNA damage may induce γ -H2AX-positive cells to undergo apoptosis and exert protective roles [40]. Also, Saikosaponin a (SSa) was reported to induce caspase-mediated apoptosis in human colon carcinoma cells by triggering DNA damage, as evidenced by an increase in the levels of γ -H2AX [41]. Thus, TSSK6 may inhibit HCV-HCC progression by upregulating γ -H2AX. High expressed USP3-AS1 was predicted to function as a sponge of hsa-miR-383 to promote the upregulation of SFN. USP3-AS1 and SFN had the similar expressed trend and predicted poor prognosis, while the expression of hsa-miR-383 was downregulated and related with good prognosis. Although no reports directly explored their interactions, the roles of SFN and miR-383 in cancer may indirectly explain our results. The study of Zhu et al. showed the expression of miR-383 was significantly lower in tissues of patients with gastric cancer (especially advanced). Overexpression of miR-383 inhibited the proliferation, induced cell cycle arrest at G1 phase and enhanced the apoptosis of gastric cancer cells by influencing its target gene cyclin E2; while addition of cyclin E2 in miR-383-expressed cells significantly reversed the roles of miR-383 [42]. miR-383 was also proved to inhibit HCC cell proliferation and doxorubicin resistance via targeting TNF superfamily member 13 [43] and eukaryotic translation initiation factor 5A2 [44], respectively.

Kong et al. demonstrated ZFPM2-AS1 expression was higher in gastric tumors compared with normal gastric tissues. Also, increased expression of ZFPM2-AS1 was associated with tumor size, tumor invasion depth, TNM stage and shorter OS in patients with gastric cancer. Functional assays implied ZFPM2-AS1 could promote cell proliferation *in vitro* and tumor growth *in vivo* by controlling macrophage migration inhibitory factor (MIF)/p53 signaling axis [45]. These findings suggested ZFPM2-AS1 may be a high-risk factor for cancer, which was also confirmed in our univariate analysis. However, our multivariate analysis results showed ZFPM2-AS1 was a protective gene for HCV-HCC. Therefore, its roles and mechanisms need further investigation in the future.

However, there were some limitations to this present study. First, only HCV-HCC samples were included from the TCGA data and no other datasets could be obtained, which lead to the fact that sample size was relatively small. Therefore, large HCC cohorts should be collected to confirm the prognosis value of our lncRNA signature. Second, we only preliminarily predicted the functions of lncRNAs and further wet experiments (immunoprecipitation, luciferase assay, overexpression or silencing) should be performed to verify them. Third, due to the limited description about treatment regimens in the TCGA data, we could not further evaluate the difference in OS caused by this heterogeneity.

Conclusion

In present study, we constructed a 6-lncRNA risk score model for survival prediction of HCV-HCC based on RNA-Seq data, with the prediction accuracy more than 85%. The identified co-expression (LINC01426/SLC16A1-AS1-AURKA/SFN/CCNB1, ZFPM2-AS1/LYPLAL1-AS1/JARID2-AS1-TSSK6) or

ceRNA (USP3-AS1-hsa-miR-383-SFN) interaction axes could explain their potential function mechanisms for HCV-HCC.

Declarations

Ethics approval and consent to participate

Not applicable.

Conflict of interest

The authors declare that they have no conflict of interest.

Consent for publication

Not applicable.

Authors' contributions

KW and XDH designed the study, collected the data and performed the statistical analysis. KW drafted the manuscript. WXL, YFL, ZQH, ZHM and ZQF contributed to the interpretation of the results and critically reviewed the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

The mRNA and microRNA sequencing data were obtained from The Cancer Genome Atlas (<https://portal.gdc.cancer.gov/>) using the sample ID in the supplementary file 1.

Acknowledgements

Not applicable.

Funding

None.

References

1. Shiels MS, Engels EA, Yanik EL, McGlynn KA, Pfeiffer RM, O'Brien TR. Incidence of hepatocellular carcinoma among older Americans attributable to hepatitis C and hepatitis B: 2001 through 2013. *Cancer* 2019; 125(15):2621-2630.
2. Guo H, Wu T, Lu Q, Li M, Guo JY, Shen Y, Wu Z, Nan KJ, Lv Y, Zhang XF. Surgical resection improves long-term survival of patients with hepatocellular carcinoma across different Barcelona Clinic Liver Cancer stages. *Cancer Manag Res* 2018; 10:361-9.

3. Kao WY, Su CW, Chau GY, Lui WY, Wu CW, Wu JC. A comparison of prognosis between patients with hepatitis B and C virus-related hepatocellular carcinoma undergoing resection surgery. *World J Surg* 2011; 35(4):858-67.
4. Yoo JJ, Chung GE, Lee JH, Nam JY, Chang Y, Lee JM, Lee DH, Kim HY, Cho EJ, Yu SJ, Kim YJ, Yoon JH. Sub-classification of Advanced-Stage Hepatocellular Carcinoma: A Cohort Study Including 612 Patients Treated with Sorafenib. *Cancer Res Treat* 2018; 50(2):366-73.
5. Zhang C, Yang X, Qi Q, Gao Y, Wei Q, Han S. lncRNA-HEIH in serum and exosomes as a potential biomarker in the HCV-related hepatocellular carcinoma. *Cancer Biomark* 2018; 21(3):651-9.
6. Zhang Q, Matsuura K, Kleiner DE, Zamboni F, Alter HJ, Farci P. Analysis of long noncoding RNA expression in hepatocellular carcinoma of different viral etiology. *J Transl Med* 2016; 14(1):328.
7. Zhang H, Zhu C, Zhao Y, Li M, Wu L, Yang X, Wan X, Wang A, Zhang MQ, Sang X, Zhao H. Long non-coding RNA expression profiles of hepatitis C virus-related dysplasia and hepatocellular carcinoma. *Oncotarget* 2015; 6(41):43770-8.
8. Sur S, Sasaki R, Devhare P, Steele R, Ray R, Ray RB. Association between MicroRNA-373 and Long Noncoding RNA NORAD in Hepatitis C Virus-Infected Hepatocytes Impairs Wee1 Expression for Growth Promotion. *J Virol* 2018; 92(20):pii: e01215-8.
9. Zhao QJ, Zhang J, Xu L, Liu F. Identification of a five-long non-coding RNA signature to improve the prognosis prediction for patients with hepatocellular carcinoma. *World J Gastroenterol* 2018;24(30):3426-39.
10. Sui J MY, Han J, Nan H, Shen B, Zhang X, Zhang Y, Wu Y, Wu W, Liu T, Xu S, Yang S, Yin L, Pu Y, Liang G. Systematic analyses of a novel lncRNA-associated signature as the prognostic biomarker for Hepatocellular Carcinoma. *Cancer Medicine* 2018; 7(7):3240-3256.
11. Wang Z, Miao Y, Feng S, Zhao Y, Tao C. Identification of four prognostic lncRNAs for survival prediction of patients with hepatocellular carcinoma. *PeerJ* 2017; 5:e3575.
12. Huang M, Zhong Z, Lv M, Shu J, Tian Q, Chen J. Comprehensive analysis of differentially expressed profiles of lncRNAs and circRNAs with associated co-expression and ceRNA networks in bladder carcinoma. *Oncotarget* 2016; 7(30):47186-200.
13. Li X, Ao J, Wu J. Systematic identification and comparison of expressed profiles of lncRNAs and circRNAs with associated co-expression and ceRNA networks in mouse germline stem cells. *Oncotarget* 2017; 8(16):26573-90.
14. Povey S, Lovering R, Bruford E, Wright M, Lush M, Wain H. The HUGO Gene Nomenclature Committee (HGNC). *Human Genetics* 2001; 109(6):678-80.
15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43(7):e47.
16. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Reczko M, Maragkakis M, Dalamagas TM, Hatzigeorgiou AG. DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Res* 2013; 41(Database issue):D239.

17. Yang JH, Li JH, Shao P, Zhou H, Chen YQ, Qu LH. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res* 2011; 39(Database issue):D202.
18. Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol* 2011; 696:291-303.
19. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016; 44(Web Server issue):W90-W97.
20. Liao X, Yang C, Huang R, Han C, Yu T, Huang K, Liu X, Yu L, Zhu G, Su H, Wang X, Qin W, Deng J, Zeng X, Ye X, Peng T. Identification of Potential Prognostic Long Non-Coding RNA Biomarkers for Predicting Survival in Patients with Hepatocellular Carcinoma. *Cell Physiol Biochem* 2018;48(5):1854-69.
21. Wu Y, Wang PS, Wang BG, Xu L, Fang WX, Che XF, Qu XJ, Liu YP, Li Z. Genomewide identification of a novel six-LncRNA signature to improve prognosis prediction in resectable hepatocellular carcinoma. *Cancer Med* 2018; 7(12):6219-33.
22. Yan J ZC, Guo K, Li Q, Wang Z. A novel seven-lncRNA signature for prognosis prediction in hepatocellular carcinoma. *J Cell Biochem* 2019; 120(1):213-23.
23. Wang SJ, Wang H, Zhao CD, Li R. Long noncoding RNA LINC01426 promotes glioma progression through PI3K/AKT signaling pathway and serves as a prognostic biomarker. *Eur Rev Med Pharmacol Sci* 2018; 22(19):6358-68.
24. Li J, Wang J, Chen Y, Yang L, Chen S. A prognostic 4-gene expression signature for squamous cell lung carcinoma. *J Cell Physiol* 2017;232(12):3702-13.
25. Chen C, Song G, Xiang J, Zhang H, Zhao S, Zhan Y. AURKA promotes cancer metastasis by regulating epithelial-mesenchymal transition and cancer stem cell properties in hepatocellular carcinoma. *Biochem Biophys Res Commun* 2017;486(2):514-20.
26. Lu L, Han H, Tian Y, Li W, Zhang J, Feng M, Li Y. Aurora kinase A mediates c-Myc's oncogenic effects in hepatocellular carcinoma. *Mol Carcinog* 2015; 54(11):1467-79.
27. Dauch D, Rudalska R, Cossa G, Nault JC, Kang TW, Wuestefeld T, Hohmeyer A, Imbeaud S, Yevsa T, Hoenicke L, Pantisar T, Bozko P, Malek NP, Longerich T, Laufer S, Poso A, Zucman-Rossi J, Eilers M, Zender L. A MYC-aurora kinase A protein complex represents an actionable drug target in p53-altered liver cancer. *Nat Med* 2017; 22(7):744-5.
28. Zhu Q, Yu X, Zhou ZW, Zhou C, Chen XW, Zhou SF. Inhibition of Aurora A Kinase by Alisertib Induces Autophagy and Cell Cycle Arrest and Increases Chemosensitivity in Human Hepatocellular Carcinoma HepG2 Cells. *Curr Cancer Drug Targets* 2017;17(4):386-401.
29. Guo M, Lu S, Huang H, Wang Y, Yang MQ, Yang Y, Fan Z, Jiang B, Deng Y. Increased AURKA promotes cell proliferation and predicts poor prognosis in bladder cancer. *BMC Syst Biol* 2018; 12(Suppl 7):118.

30. Al-Khafaji ASK, Marcus MW, Davies MPA, Risk JM, Shaw RJ, Field JK, Liloglou T. AURKA mRNA expression is an independent predictor of poor prognosis in patients with non-small cell lung cancer. *Oncol Lett* 2017; 13(6):4463-8.
31. Lee IN, Chen CH, Sheu JC, Lee HS, Huang GT, Yu CY, Lu FJ, Chow LP. Identification of human hepatocellular carcinoma-related biomarkers by two-dimensional difference gel electrophoresis and mass spectrometry. *J Proteome Res* 2005; 4(6):2062-9.
32. Chen XL, Zhou L, Yang J, Shen FK, Zhao SP, Wang YL. Hepatocellular carcinoma-associated protein markers investigated by MALDI-TOF MS. *Mol Med Rep* 2010; 3(4):589-96.
33. Reis H, Pütter C, Megger DA, Bracht T, Weber F, Hoffmann AC, Bertram S, Wohlschläger J, Hagemann S, Eisenacher M, Scherag A, Schlaak JF, Canbay A, Meyer HE, Sitek B, Baba HA. A structured proteomic approach identifies 14-3-3Sigma as a novel and reliable protein biomarker in panel based differential diagnostics of liver tumors. *Biochim Biophys Acta* 2015; 1854(6):641-50.
34. Boudreau A, Tanner K, Wang D, Geyer FC, Reis-Filho JS, Bissell MJ. 14-3-3 σ stabilizes a complex of soluble actin and intermediate filament to enable breast tumor invasion. *Proc Natl Acad Sci U S A* 2013; 110(41):E3937-44.
35. Shiba-Ishii A, Kim Y, Shiozawa T, Iyama S, Satomi K, Kano J, Sakashita S, Morishita Y, Noguchi M. Stratifin accelerates progression of lung adenocarcinoma at an early stage. *Mol Cancer* 2015; 14:142.
36. Hu Y, Zeng Q, Li C, Xie Y. Expression profile and prognostic value of SFN in human ovarian cancer. *Biosci Rep* 2019; 39(5):pii: BSR20190100.
37. Gu J, Liu X, Li J, He Y. MicroRNA-144 inhibits cell proliferation, migration and invasion in human hepatocellular carcinoma by targeting CCNB1. *Cancer Cell Int* 2017; 19:15.
38. Zhuang L, Yang Z, Meng Z. Upregulation of BUB1B, CCNB1, CDC7, CDC20, and MCM3 in Tumor Tissues Predicted Worse Overall Survival and Disease-Free Survival in Hepatocellular Carcinoma Patients. *Biomed Res Int* 2018;2018:7897346.
39. Jha KN, Tripurani SK, Johnson GR. TSSK6 is required for γ H2AX formation and the histone-to-protamine transition during spermiogenesis. *J Cell Sci* 2017;130(10):1835-44.
40. Guo Z, Pei S, Si T, Li J, Jiang C, Li S, Zhao J. Expression of the γ -phosphorylated histone H2AX in gastric carcinoma and gastric precancerous lesions. *Oncol Lett* 2015; 9(4):1790-4.
41. Kang SJ, Lee YJ, Kang SG, Cho S Yoon W, Lim JH, Min SH, Lee TH, Kim BM. Caspase-4 is essential for saikosaponin a-induced apoptosis acting upstream of caspase-2 and γ -H2AX in colon cancer cells. *Oncotarget* 2017; 8(59):100433-48.
42. Zhu C, Huang Q, Zhu H. miR-383 Inhibited the Cell Cycle Progression of Gastric Cancer Cells via Targeting Cyclin E2. *DNA Cell Biol* 2019; 38(8):849-856.
43. Chen L, Guan H, Gu C, Cao Y, Shao J, Wang F. miR-383 inhibits hepatocellular carcinoma cell proliferation via targeting APRIL. *Tumour Biol* 2016; 37(2):2497-507.
44. Tu C, Chen W, Wang S, Tan W, Guo J, Shao C, Wang W. MicroRNA-383 inhibits doxorubicin resistance in hepatocellular carcinoma by targeting eukaryotic translation initiation factor 5A2. *J Cell Mol Med*

45. Kong F, Deng X, Kong X, Du Y, Li L, Zhu H, Wang Y, Xie D, Guha S, Li Z, Guan M, Xie K. ZFPM2-AS1, a novel lncRNA, attenuates the p53 pathway and promotes gastric carcinogenesis by stabilizing MIF. *Oncogene* 2018; 37(45):5982-96.

Figures

Figure 1

Differentially expressed RNAs. A, volcano plots were used to display the distribution of the DELs, DEGs and DEMs. Blue dot, downregulated RNAs; pink, upregulated RNAs; Red horizontal dotted line, $FDR < 0.05$; Red vertical dotted line, $|\log_2FC| > 1$. B, hierarchical clustering and heat map analysis of DELs, DEGs and DEMs in HCC and normal tissues. The color to red, high expression; the color to light green, low expression. DEGs, differentially expressed genes; DEMs, differentially expressed miRNAs; DELs, differentially expressed long non-coding RNAs; FDR, false discovery rate; FC, fold change.

Figure 2

Prediction performance evaluation of the risk score model for overall survival. A, Kaplan–Meier curve for patients in the training dataset. B, Kaplan–Meier curve for patients in the validation dataset. C, Kaplan–Meier plots curve for patients in the entire dataset. D, receiver operating characteristic curves constructed based on 6-lncRNA signature. HR, hazard ratio.

Figure 3

Stratification analysis of the survival between high- and low-risk groups according to pathologic stage. HR, hazard ratio.

Figure 4

A competing endogenous RNA network among DELs, DEGs and DEMs. Red, upregulated; green, downregulated. Circular, genes; square, lncRNAs; triangle, miRNAs. DEGs, differentially expressed genes; DEMs, differentially expressed miRNAs; DELs, differentially expressed long non-coding RNAs; FC, fold change.

Figure 6

Function enrichment analysis. A, Gene ontology (GO). FDR, false discovery rate. B, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway.

Supplementary Files

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

- [Supplementaryfile1.xlsx](#)
- [Supplementaryfile1.xlsx](#)
- [Supplementaryfile1.xlsx](#)