Original Paper

Genome-Wide Analysis of Prognostic lncRNAs, miRNAs, and mRNAs Forming a Competing Endogenous RNA Network in Hepatocellular Carcinoma

Peng Lin^a Dong-yue Wen^a Qing Li^a Yun He^a Hong Yang^a
Gang Chen^b

^aDepartment of Medical Ultrasonics, First Affiliated Hospital of Guangxi Medical University, Nanning,

^bDepartment of Pathology, First Affiliated Hospital of Guangxi Medical University, Nanning, China

Key Words

Hepatocellular carcinoma • lncRNA • miRNA • mRNA • TCGA • Competing endogenous RNA

Abstract

Background/Aims: Hepatocellular carcinoma (HCC) is the most prevalent subtype of primary liver tumor worldwide. Growing evidence has led to a consensus that long non-coding RNAs (lncRNAs) have considerable influence on tumorigenesis and tumor progression of HCC via the mechanism of competing endogenous RNAs (ceRNAs). **Methods:** Here, we systematically investigated the expression landscape and clinical prognostic value of lncRNAs, miRNAs, and mRNAs from The Cancer Genome Atlas. Differentially expressed RNAs were submitted to Cox regression analysis and the construction of prognostic indexes. A lncRNA-miRNA-mRNA regulatory network was then constructed based on interaction information derived from miRcode, TargetScan, miTarBase, and miRDB. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses were performed to reveal and determine the functional roles of the ceRNA network in the prognosis of HCC. **Results:** We detected 77 differentially expressed lncRNAs, 29 differentially expressed miRNAs, and 1014 differentially expressed mRNAs in HCC, which were significantly associated with the overall survival of patients with HCC. We developed three prognostic prediction models that showed moderate predicting prognosis performance and were highly correlated with tumor burden, histological grade and pathological stage. Additionally, 10 survival-related lncRNAs, 6 survival-related miRNAs, and 31 survival-related mRNAs were included to develop a ceRNA network. Further functional enrichment analysis suggested that the ceRNA network was associated with a dismal prognosis for patients with HCC by disturbing the homeostasis of the cell cycle. **Conclusion:** Together, our study highlights the significant roles of lncRNAs in the development

Hong Yang

Department of Medical Ultrasonics, First Affiliated Hospital of Guangxi Medical University
Nanning, Guangxi, Zhuang Autonomous Region 530021 (China)
Tel. 0771-5356777, Fax 0771-5356777, E-Mail yanghong@gxmu.edu.cn

and implementation of monitoring surveillance and prognosis of HCC and provides a deeper understanding of the lncRNA-related ceRNA regulatory mechanism in the pathogenesis of HCC.

© 2018 The Author(s)
Published by S. Karger AG, Basel

Introduction

Hepatocellular carcinoma (HCC), the most prevalent subtype of primary liver tumor, is a fearsome disease, and its mortality rate, which has been increasing in the United States, has risen by approximately 3% per year in the past few years [1]. Worldwide, HCC risk factors mainly include hepatitis B or C virus infection, aflatoxin B1, alcohol, cirrhosis, and metabolic disorders [2-4]. The process underlying the onset and progression of HCC is complex and involves many factors, but the precise molecular events that underlie the poor prognosis of HCC remain obscure. To date, the continuous advancement in medical technology, including liver transplantation, surgical resection, and radiofrequency ablation, is still unable to elevate the dismal 5-year survival rate of patients with HCC, which is lower than 20% [5, 6]. The undesirable prognosis of HCC is largely due to limited knowledge of the molecular factors involved in HCC, which leads to the lack of effective interventions [7, 8]. Thus, further characterization of the genetic mechanism and identification of reliable biomarkers to predict HCC prognosis are important steps toward developing more rational, specific, and effective treatments.

Long non-coding RNAs (lncRNAs) are a class of noncoding functional RNAs that have recently attracted increasing research attention [9]. Initially, lncRNAs were regarded as “dark matter” in the genome. The advancement of high-throughput RNA-sequencing (RNA-seq) led to the discovery of thousands of lncRNAs. Indeed, lncRNAs exert significant regulatory roles in diverse cellular processes, especially in various types of tumor [10-12]. Recently, the mechanism of competing endogenous RNAs (ceRNAs) was proposed as a specific regulatory pathway of lncRNAs to explain how they exert their influence on protein levels [13, 14]. The central concept is that ceRNAs control the large-scale transcriptome by interacting with microRNA (miRNA) response elements (MREs). In recent years, several studies reported that perturbations of lncRNAs could have consequences for the biological process of HCC through the mechanism of ceRNAs, thus affecting the prognosis of HCC [15, 16]. Well-described examples of functionally important lncRNAs, such as the highly up-regulated *NEAT1*, *HULC*, all could function by various ceRNA mechanisms to exert indispensable roles in the initiation and progression of HCC [17, 18]. However, most of these studies focused only on the mechanism of a certain lncRNA-miRNA-mRNA axis. Although Zhang et al [19] constructed a ceRNA network in HCC, they did not proceed from the perspective of prognosis. Therefore, it is important to study comprehensively the effect of a ceRNA network in the context of the poor prognosis of HCC. Through a better understanding of how lncRNAs exert their functions in hepatocarcinogenesis, we might find solutions for the most urgent challenges faced in the management of liver cancer.

To address these challenges, in this study, we comprehensively analyzed and provided an overview of the prognostic roles of different types of RNA (lncRNAs, miRNAs and mRNAs) in HCC. By integrating lncRNAs, miRNAs and mRNAs, we developed a multi-RNA-based individualized prognostic index (PI) to stratify HCC patients effectively into two groups with a significant difference in overall survival (OS). Furthermore, we constructed a prognostic-related ceRNA network in HCC and concluded with functional enrichment analyses to reveal the underlying mechanism. Via these efforts, we aimed to lay a foundation for the regulatory mechanism of lncRNAs in the prognosis of HCC to identify novel biomarkers that can be used to predict clinical outcome effectively.

Materials and Methods

Data resources and pretreatment

RNA-seq data and clinical information of HCC patients were downloaded from The Cancer Genome Atlas (TCGA) dataset (<https://cancergenome.nih.gov/>). On the basis of annotation information provided by the Ensembl database (<http://asia.ensembl.org/index.html>), we identified 14,367 lncRNAs and 19,667 protein coding genes. To analyze mature miRNAs rather than pre-miRNAs, we downloaded miRNA expression profiles from UCSC Xena (<http://xena.ucsc.edu/>). Through mapping miRNA sequencing, we extracted 372 mature miRNA profiles in the form of $\log_2(\text{reads per million [RPM]} + 1)$. lncRNAs and mRNAs with an expression raw count value >10 and miRNAs with $\log_2(\text{RPM} + 1) > 1$ in more than 75% of samples were retained to remove low-abundance genes.

Differential analysis

Differentially expressed lncRNAs and mRNAs between HCC and adjacent normal liver tissues were screened by using the edgeR package of R software. $|\text{Fold change}| > 2$ and $P\text{-value} < 0.05$ were set as the thresholds. After normalization by edgeR, the expression profiles of lncRNAs and mRNAs were further converted to $\log_2$ (normalized value + 1) transformation to be used for the next operation. Similarly, differentially expressed miRNAs were selected by the limma package of R software according to the following criteria: $|\text{Fold change}| > 2$ and $P\text{-value} < 0.05$.

Survival analysis and gene network construction

The clinical parameters of the liver hepatocellular carcinoma cohort were downloaded and extracted from TCGA. Only patients with at least 90 days of OS were included in the survival study. The associations of lncRNAs, miRNAs, and mRNAs with OS were assessed by univariate Cox regression. Only those genes with a $P\text{-value} < 0.001$ were identified as candidate prognosis related biomarkers. Then, those candidate genes were applied by multivariate COX regression to remove any genes that might not be independent factors in prognosis prediction. Finally, lncRNA, miRNA and mRNA signature scores were constructed, respectively, by multiplying the expression levels of independent biomarkers ($P\text{-value} < 0.01$) by the coefficient of multivariate Cox regression. The "SurvivalROC" package in R, which allows for time-dependent receiver-operator characteristic (ROC) curve estimation, was performed to obtain the area under the curve (AUC) of the ROC curve for each model.

CeRNA network construction

The identified prognosis-related lncRNAs, miRNAs and mRNAs were selected to develop the ceRNA network. The miRcode (<http://www.mircode.org/>) database was used to predict lncRNA-miRNA interactions. miRNA-mRNA interactions were predicted using TargetScan (<http://www.targetscan.org/>), miRTarBase (<http://mirtarbase.mbc.nctu.edu.tw/php/index.php>), and miRDB (<http://www.mirdb.org/>). MiRNA-mRNA interactions that were predicted by the three databases simultaneously were used to construct the ceRNA network. Combining lncRNA-miRNA interactions with miRNA-target gene interactions, a ceRNA regulatory network based on lncRNA-miRNA-mRNA axes was established. Cytoscape software [20] was utilized to draw this network.

Functional analysis

We also performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses to reveal and determine the functional roles of the ceRNA network in the prognosis of HCC. These gene functional enrichment analyses were conducted by using the clusterProfiler package of R software. GO terms or KEGG pathways were identified as significantly enriched by these genes when $P\text{-value} < 0.05$. The GOplot package of R software was utilized to display the results of the GO and KEGG analyses.

Statistical processing

All statistical analyses were conducted using SPSS 24.0 (SPSS, Inc., Chicago, IL, USA). R software (3.3.3 version) and GraphPad Prism 5 were used to plot the figures. The relationships between the gene expression profiles and clinical or demographic characteristics of patients with HCC were calculated using independent sample t-tests. Survival curves were plotted by Kaplan-Meier (K-M) analysis, and differences in survival rates were assessed using a log-rank test.

Results

Differentially expressed lncRNAs, miRNAs, and mRNAs in HCC

We detected differentially expressed genes in 374 HCC tissues and 50 adjacent normal liver tissues. In total, 244 lncRNAs (Fig. 1A), 106 miRNAs (Fig. 1B) and 2389 mRNAs (Fig. 1C) showed different expression profiles between HCC and non-tumor tissues. Among them, 178 up-regulated lncRNAs and 66 down-regulated lncRNAs were obtained; 1598 mRNAs were up-regulated, and 791 were down-regulated. For the miRNA microarray assay, 106 miRNAs met the selection criteria between the HCC and non-tumor tissues, including 25 up-regulated and 81 down-regulated miRNAs (Fig. 1C). For the heatmap shown in Fig. 2, the numerical data represent the expression profiles of differentially expressed lncRNAs (Fig. 2A), miRNAs (Fig. 2B), and mRNAs (Fig. 2C), respectively.

Fig. 1. Volcano plot of differentially expressed lncRNAs (A), miRNAs (B) and mRNAs (C). The red point in the plot represents up-regulated RNAs and blue point represents down-regulated RNAs with statistical significance.

Fig. 2. Heatmap of differentially expressed RNAs between HCC and non-tumor tissues. (A) lncRNAs; (B) miRNAs; (C) mRNAs.

Fig. 3. Forest plots of hazard ratios (HR) of most significant survival associated RNAs in HCC. Red represents the risk factors (HR>1) and blue represents protective factors (HR<1). (A) HR of top 20 survival associated lncRNAs. (B) HR of top 20 survival associated miRNAs. (C) HR of top 20 survival associated mRNAs.

Survival associated lncRNAs, miRNAs, and mRNAs in HCC

The associations between the differentially expressed lncRNAs, miRNAs, and mRNAs with OS were studied in detail in 329 patients with HCC with at least 90 days of OS. We conducted univariate survival analysis to select OS-related RNAs. As a result, we detected a total of 77 survival associated lncRNAs, 29 survival associated miRNAs, and 1014 survival associated mRNAs in HCC (P-value < 0.05). The top 20 most significant survival associated genes are presented in Fig. 3. The most significant survival associated mRNAs in HCC (P-value < 0.001) were piped to STRING to generate gene interaction networks, and Cytoscape analysis of the gene networks revealed the important cancer pathways, including hub genes, such as *CDK1*, *CCNB1*, and *PLK1* (Fig. 4).

Fig. 4. Protein-protein interaction (PPI) network of most survival associated mRNAs in HCC generated by Cytoscape. PPI data obtained from the online database STRING (<http://string-db.org/>). The size and brightness of the circle represents the degree of connection. The larger and brighter circles are hub genes in the network. The thickness of the lines represents the combined score.

Derivation of prognostic signatures for HCC patients

These most significant survival associated lncRNAs, miRNAs and mRNAs were further submitted to multivariate COX regression analysis to identify the independent prognostic predictors for HCC. Next, the PIs for predicting survival were constructed based on 3 lncRNAs, 3 miRNAs, and 3 mRNAs. The formulae for the three prognostic signatures were as follows: lncRNA-based PI = [expression value of TMCC1-AS1 × 0.402] + [expression value of LINC01138 × 0.278] + [expression value of AC009005.1 × 0.196]; miRNA-based PI = [expression value of miR-9-5p × 0.113] + [expression value of miR-326 × 0.359] + [expression value of miR-139-5p × -0.287]; mRNA-based PI signature scores = [expression value of SOCS2 × -0.285] + [expression value of SLC16A11 × -0.193] + [expression value of KPNA2 × 0.451]. By using the three signatures, we calculated a risk score for each patient individually

and ranked them according to increased risk score. HCC patients were separated into low- or high-risk groups according to the median risk score. K-M analysis showed a significant difference in OS between the two groups with all three models (Fig. 5). ROC curves were also applied to compare the efficiency of these prognostic models. The three proposed PIs all showed moderate prognostic assessment ability with time-dependent AUC values of 0.728, 0.719, and 0.724, respectively (Fig. 6). As shown in Fig. 7, the scores assigned to each patient provide a good assessment of prognosis. The expression pattern of these RNAs in the high- and low-risk groups are also displayed (Fig. 7G-I).

Building a prognostic predictor that includes different types of RNA could reveal the potential molecular mechanism of HCC and increase prediction accuracy. Finally, the multiple types of RNA-based PI for HCC OS were selected, including the above nine RNAs. To determine the value of the PI in predicting HCC prognosis, we performed K-M analysis, which revealed that patients in the high-risk group had a shorter OS (hazard ratio [HR] = 4.030, 95% confidence interval [CI]: 2.753-5.898, P-value < 0.001; Fig. 8A). Furthermore, the AUC of the

Fig. 5. Kaplan-Meier (K-M) curves for prognostic index (PI) in TCGA HCC cohort. (A) K-M survival curves showing overall survival (OS) outcomes according to relative high-risk and low-risk patients who were grouped by lncRNA-based PI. (B) K-M survival curves showing OS outcomes according to relative high-risk and low-risk patients who were grouped by miRNA-based PI. (C) K-M survival curves showing OS outcomes according to relative high-risk and low-risk patients who were grouped by mRNA-based PI.

Fig. 6. Receiver operating characteristic (ROC) curves for prognostic index (PI) in TCGA HCC cohort. (A) Time-dependent ROC curves analysis for survival prediction by lncRNA-based PI. (B) Time-dependent ROC curves analysis for survival prediction by miRNA-based PI. (C) Time-dependent ROC curves analysis for survival prediction by mRNA-based PI.

Fig. 7. Performance of prognostic signatures in distinguishing patients into high- and low-risk groups. (A) The distribution of lncRNA-based risk score. (B) The distribution of miRNA-based risk score. (C) The distribution of mRNA-based risk score. The Y axis of panel A-C represent the risk score of each patient. The X axis of panel A-C represent the rank of each patient based on risk score. (D) lncRNA-based risk score for the distribution of patients' survival status. (E) miRNA-based risk score for the distribution of patients' survival status. (F) mRNA-based risk score for the distribution of patients' survival status. The Y axis of panel D-F represent the survival time of each patients. The X axis of panel D-F represent the rank of each patient based on risk score. (G) TMCC1-AS1, LINC01138 and AC009005.1 are significantly up-regulated in the high-risk group according to lncRNA-based risk score. (H) Mir-9-5P and mir-326 are significantly up-regulated in the high-risk group, while mir-139-5P is down-regulated in the high-risk group according to miRNA-based risk score. (I) KPNA2 is significantly up-regulated in the high-risk group, while SOCS2 and SLC16A11 is down-regulated in the high-risk group according to mRNA-based risk score.

Fig. 8. Time-dependent receiver operating characteristic (ROC) curves and Kaplan-Meier (K-M) survival based on the integrated classifier in the TCGA datasets. (A) K-M plots revealed patients in high-risk group that were assigned by prognostic index with a short overall survival (OS). (B) Time-dependent ROC curve with AUC of prognostic index built by different types of RNAs. (C) K-M plots revealed patients in advanced TNM stage with a short OS. (D) Time-dependent ROC curve with AUC of TNM stage.

Table 1. Univariate and multivariate Cox regression analysis of overall survival in HCC. Age, pathologic stage, tumor stage, histologic grade and prognostic index were coded as continuous variable. Specifically, pathologic stage was coded as I=1, II=2, III/IIIA=3, IIIB=3.5, IIIC=4, IV=5. Tumor stage was coded as T1=1, T2=2, T3=3, T4=4. Histologic grade was coded as G1=1, G2=2, G3=3, G4=4

Variables	Univariate analysis		Multivariate analysis	
	Hazard ratio (95%CI)	P-value	Hazard ratio (95%CI)	P-value
Age	1.006 (0.992-1.021)	0.390	1.013 (0.989-1.038)	0.282
Gender (male/female)	0.863 (0.588-1.266)	0.452	1.199 (0.643-2.235)	0.568
Pathologic stage	1.710 (1.419-2.060)	<0.001	0.322 (0.080-1.289)	0.109
Tumor stage	1.803 (1.484-2.190)	<0.001	4.563 (1.180-17.652)	0.028
Node stage (Whether lymphatic metastasis had occurred or not)	2.478 (0.604-10.167)	0.208	34.200 (1.358-861.388)	0.032
Metastasis stage (Whether distant metastasis had occurred or not)	4.914 (1.536-15.719)	0.007	2.961 (0.351-25.004)	0.319
Histologic grade	1.151 (0.893-1.483)	0.278	0.932 (0.624-1.392)	0.730
Tumor burden (With tumor/Tumor free)	3.372 (2.194-5.183)	<0.001	2.428 (1.227-4.805)	0.011
Prognostic Index	2.521 (2.066-3.075)	<0.001	2.358 (1.776-3.131)	<0.001

time-dependent ROC curve was 0.776 (Fig. 8B). To display the prognostic efficiency of PI more intuitively, we also generated a K-M plot of tumor stage and the corresponding ROC curve. Tumor, node, metastasis (TNM) stage also divided the HCC patients into two groups with a significant difference in prognosis (HR=3.897, 95%CI: 2.399-6.271, P-value < 0.001; Fig. 8C). The AUC of the ROC curve based on TNM stage was 0.668 (Fig. 8D). Although TNM stage is commonly used for prognosis prediction in the clinical setting, its prognostic AUC value is limited. Multivariate regression analysis showed that the PI was still significantly associated with survival when adjusted for age, gender, stage, grade, tumor stage, node stage, metastasis stage, histologic grade and tumor burden in TCGA (HR = 2.358, 95%CI: 1.776-3.131, P-value < 0.001; Table 1).

Clinical significance of prognostic predictors

As prognostic genes involved in the onset and prognosis in HCC, it was most likely that these RNAs could accelerate the development of HCC and be of great significance. Therefore, we analyzed the expression of these genes under various clinical parameters and investigated their associations with clinical progress. We specifically analyzed several major clinical parameters, including age (over or under 60 years), gender (male or female), tumor status (with tumor or tumor free), pathologic stage (Stage III, IV or Stage I, II), T stage (Stage III, IV or Stage I, II), N stage (whether lymphatic metastasis had occurred or not), M stage (whether distant metastasis had occurred or not), and histological grade (III, IV or I, II). The nine genes included in our risk score showed significant correlation with HCC tumor burden, histological grade, and stage (Fig. 9, Tables 2-4). These results revealed that these genes can be used for effective risk stratification in HCC.

Fig. 9. Clinical significance of prognostic RNAs in HCC. Correlations between clinical parameters and prognostic RNAs. The numbers in each block indicate the P-value. The deeper red the block, the smaller the P-value.

Table 2. Relationships between three lncRNAs expression and clinical parameters in TCGA. M±SD: Mean ± Std. Deviation

Parameters		N	AC009005.1 expression value			TMCC1-AS1 expression value			LINC01138 expression value		
			M±SD	t value	P-value	M±SD	t value	P-value	M±SD	t value	P-value
Age (n=370)	≥60	201	5.001±1.503	-0.804	0.422	5.305±0.9737	-1.904	0.058	7.141±0.990	-0.717	0.474
	<60	169	5.126±1.452			5.510±1.097			7.221±1.158		
Gender (n=371)	Male	250	4.927±1.482	-2.598	0.010	5.422±1.061	0.676	0.499	7.199±1.013	0.447	0.655
	Female	121	5.351±1.451			5.344±0.981			7.146±1.184		
Tumor status (n=352)	With tumor	151	5.278±1.414	2.753	0.006	5.522±1.072	2.076	0.039	7.362±1.010	2.562	0.011
	Tumor free	201	4.842±1.511			5.290±1.005			7.077±1.051		
Histological grade (n=366)	G3-G4	134	5.413±1.490	3.397	0.001	5.725±1.121	4.471	<0.001	7.470±0.932	3.971	<0.001
	G1-G2	232	4.871±1.459			5.212±0.936			7.016±1.117		
Pathologic stage (n=347)	III-IV	90	5.321±1.676	1.918	0.056	5.604±1.1404	1.986	0.048	7.254±1.206	0.703	0.483
	I-II	257	4.973±1.406			5.353±0.995			7.162±1.007		
T stage (n=368)	T3-T4	93	5.334±1.677	1.976	0.049	5.574±1.092	1.866	0.063	7.222±1.224	0.347	0.782
	T1-T2	275	4.983±1.409			5.343±1.013			7.177±1.018		
N stage (n=256)	N1-3	4	5.257±1.397	0.241	0.810	5.684±1.681	0.485	0.628	7.415±0.660	0.439	0.661
	N0	252	5.078±1.471			5.420±1.074			7.178±1.080		
M stage (n=270)	M1	4	6.683±0.976	2.084	0.038	5.325±0.978	-0.165	0.869	8.025±0.961	1.562	0.119
	M0	266	5.078±1.535			5.415±1.093			7.153±1.110		

Table 3. Relationships between three miRNAs expression and clinical parameters in TCGA. M±SD: Mean ± Std. Deviation

Parameters		N	miR-326 expression value			miR-9-5p expression value			miR-139-5p expression value		
			M±SD	t value	P-value	M±SD	t value	P-value	M±SD	t value	P-value
Age (n=367)	≥60	197	2.076±1.021	0.228	0.820	8.189±2.221	1.001	0.318	6.366±1.301	-1.003	0.31
	<60	170	2.052±0.965			7.957±2.211			6.496±1.167		
Gender (n=368)	Male	250	2.011±0.958	-1.552	0.121	8.051±2.245	-0.465	0.642	6.508±1.191	1.951	0.052
	Female	118	2.183±1.059			8.166±2.160			6.238±1.329		
Tumor status (n=349)	With tumor	150	2.162±1.014	2.025	0.044	8.229±2.333	1.714	0.088	6.258±1.227	-2.484	0.013
	Tumor free	199	1.949±0.935			7.830±2.011			6.592±1.256		
Histological grade (n=364)	G3-G4	136	2.037±0.950	-0.384	0.701	8.366±2.338	1.898	0.058	6.142±1.176	-3.372	0.001
	G1-G2	228	2.078±1.027			7.912±2.131			6.587±1.241		
Pathologic stage (n=345)	III-IV	89	2.274±1.195	2.596	0.010	8.219±2.384	0.631	0.529	6.072±1.372	-2.899	0.004
	I-II	256	1.959±0.904			8.046±2.171			6.544±1.175		
T stage (n=365)	T3-T4	93	2.349±1.189	2.868	0.005	8.328±2.451	1.228	0.220	6.071±1.352	-3.007	0.002
	T1-T2	272	1.962±0.902			8.001±2.128			6.523±1.176		
N stage (n=256)	N1-3	3	1.820±0.569	-0.469	0.639	7.547±0.928	-0.347	0.729	6.555±1.837	0.204	0.839
	N0	253	2.092±1.003			8.003±2.269			6.413±1.196		
M stage (n=270)	M1	4	1.566±1.875	-0.938	0.349	10.174±3.612	1.840	0.067	5.416±0.848	-1.548	0.123
	M0	266	2.049±1.009			8.053±2.269			6.386±1.248		

Table 4. Relationships between three mRNAs expression and clinical parameters in TCGA. M±SD: Mean ± Std. Deviation

Parameters	N	SLC16A11 expression value			SOCS2 expression value			KPNA2 expression value			
		M±SD	t value	P-value	M±SD	t value	P-value	M±SD	t value	P-value	
Age (n=370)	≥60	201	8.631±2.295	-	0.764	8.817±1.511	0.105	0.916	11.093±0.844	-	0.114
	<60	169	8.701±2.143	0.301		8.800±1.564			11.235±0.873	1.584	
Gender (n=371)	Male	250	8.839±2.143	2.191	0.029	8.833±1.562	0.442	0.659	11.216±0.877	1.784	0.075
	Female	121	8.301±2.343			8.758±1.470			11.047±0.814		
Tumor status (n=352)	With tumor	151	8.430±2.036	-	0.101	8.592±1.581	-	0.020	11.306±0.830	3.058	0.002
	Tumor free	201	8.824±2.357	1.643		8.976±1.479	2.342		11.028±0.856		
Histological grade (n=366)	G3-G4	134	8.313±2.1872	-	0.015	8.404±1.568	-	<0.001	11.429±0.865	4.540	<0.001
	G1-G2	232	8.892±2.191	2.437		9.033±1.477	3.838		11.016±0.820		
Pathologic stage (n=347)	III-IV	90	8.371±2.509	-	0.124	8.425±1.459	-	0.007	11.384±0.893	2.808	0.005
	I-II	257	8.827±2.072	1.548		8.936±1.568	2.707		11.091±0.837		
T stage (n=368)	T3-T4	93	8.340±2.494	-	0.122	8.410±1.371	-	0.004	11.424±0.878	3.345	0.001
	T1-T2	275	8.752±2.113	1.551		8.934±1.567	2.877		11.087±0.829		
N stage (n=256)	N1-3	4	7.397±3.778	-	0.276	8.964±2.571	0.262	0.794	10.727±1.020	-	0.258
	N0	252	8.640±2.236	1.092		8.762±1.516			11.216±0.855	1.132	
M stage (n=270)	M1	4	7.581±2.041	-	0.367	8.507±0.807	-	0.745	10.957±0.442	-	0.526
	M0	266	8.616±2.276	0.904		8.768±1.596	0.326		11.235±0.874	0.635	

Fig. 10. Prognostic competing endogenous RNA (ceRNA) network in HCC. The red indicates the risky RNAs, and the blue indicates the protective RNAs in HCC. The diamond represents lncRNAs, triangle represents miRNAs, and circle represents mRNAs.

Survival-related ceRNA network in HCC

Combining the lncRNA-miRNA interactions with the miRNA-mRNA interactions, an integrated lncRNA-miRNA-mRNA network was established, consisting of 47 molecules and 51 interactions (Fig. 10). This ceRNA network included some familiar star molecules, such as SNHG3, mir-101-3p and FOXO1. Of course, there are some molecules that are not well understood in HCC. Our ceRNA network provide some novel insights into the diagnosis, surveillance, and prognosis of HCC. Considering the regulatory role of lncRNAs, K-M plots generated by Cutoff Finder [21] were used to show their value in predicting prognosis (Fig. 11). According to the ceRNA paradigm, lncRNAs compete with protein-coding genes for the pool of shared miRNAs. Thus, we performed a comprehensive integrating analysis to assess whether the mRNAs included in the ceRNA network correlated with the prognosis of HCC patients as a whole. It showed that the overexpression of these mRNAs was significantly associated with poor OS (pooled HR=1.18, 95% CI: 1.08–1.29, P-value < 0.001, random effect; Fig. 12).

On the basis of enrichment analysis of the network, we found that the genes included in the ceRNA network were highly associated with five KEGG pathways, such as “cell cycle,” “P53 signaling pathway,”

Fig. 11. Kaplan-Meier survival curves for ten lncRNAs associated with overall survival. (Horizontal axis: overall survival time: days, Vertical axis: survival proportion). (A) AL51845.1; (B) FAM99A; (C) LINC00261; (D) MYLK-AS1; (E) SNHG1; (F) SNHG3; (G) SNHG6; (H) SNHG7; (I) SNHG12; (J) TSPER-AS1.

and “microRNAs in cancer” (Fig. 13). For the GO functional categories, these genes were mainly involved in several biological processes (BP) that are closely related to the tumor progression. In the cellular component (CC) process, the target genes were significantly clustered into items, including PCG protein complex, chromatin and heterochromatin. Chromatin, transcription factor binding, beta-catenin binding, and other vital molecular functions (MF) were significantly related to these genes (P-value < 0.05). The GO enrichment networks of BP, CC and MF for these genes are shown in Fig. 14.

Discussion

It is becoming increasingly clear that many complex diseases, particularly cancers, can rarely be ascribed to one or a few genomic variations alone [22]. Recently, the ceRNA hypothesis has been considered a novel measure of gene regulation at the post-transcriptional level through miRNA competition in HCC [16, 23, 24]. In the present study, to gain insights into the molecular events relevant to HCC prognosis, we took advantage of the molecular resolution provided by TCGA and identified distinct lncRNAs, miRNAs, and mRNAs that together specify biologically relevant HCC prognosis. Particularly, we developed a prognostic ceRNA network according to specific criteria outlined by a large sample of HCC tissues from TCGA, which has provided a deeper understanding of HCC at the integration level, as well as new dimensions to predict the prognosis of individual patients with HCC.

HCC is an overwhelming experience for patients due to its inferior prognosis. Hence, identifying effective prognostic biomarkers and exploring the underlying regulatory network are indispensable steps toward the development of effective therapies and tailored treatment plans. Furthermore, prognostic signatures based on multiple genes have contributed to more accurate predictions of survival compared to single gene analysis. Previous studies mainly focused on one type of RNA. Various prognostic risk models based on lncRNAs [25], miRNAs [26, 27] and mRNAs [28] have been proposed by different scholars. However, the prognostic

Fig. 12. Forest plot of hazard ratios of mRNAs included in the ceRNA network.

Fig. 13. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of genes included in the ceRNA network. (A) Hierarchical clustering of the gene expression profiles in each KEGG pathways; (B) Chord plot displays of the relationship between genes and KEGG pathways.

value of a PI including different kinds of RNA is still not clear. In the present study, we proposed 3 distinct prognostic signatures based on lncRNAs, miRNAs and mRNAs, respectively, which could successfully predict the survival of patients with HCC. It is worth noting that the genes that were submitted to survival analysis are differentially expressed in HCC; our findings indicated that these genes are expressed during the onset and progression of HCC. Finally, a novel PI that included multiple RNAs was proposed, which was validated as an independent predictor for the OS of patients with HCC. In particular, to obtain more reliable and accurate results, we removed patients with less than 90 days of OS. Additionally, low-abundance genes were also removed when analyzing the relationships between survival and the gene expression profiles. These additional measures provided more accurate results. Furthermore, genes included in our model were closely related with tumor burden, histological grade, and stage. Importantly, these findings provided several clues that the PI may be particularly useful for the assessment of treatment response and recurrence surveillance.

Through analysis of survival, we identified several biomarkers with significant clinical significance, some of which have not been reported previously. Three lncRNAs (TMCC1-AS1, LINC01138 and AC009005.1) are rarely reported in HCC. Cui et al [29]. reported that SOCS2 inhibited HCC cell metastasis *in vitro* and *in vivo*. Furthermore, it has been reported that HCC patients with low SOCS2 expression have significantly inferior OS [30]. Up-regulated KPNA2 expression has also been identified as a biomarker of poor prognosis in HCC [31, 32]. These reports facilitated the identification of additional biomarkers and improved the accuracy of prognosis prediction.

Several studies have revealed that some dysregulated lncRNAs may act as potential onco- or tumor-suppressor RNAs in numerous cancer types, including gastric cancer [33], pancreatic cancer [34] and HCC [35]. The ceRNA hypothesis provides a new framework to understand complicated RNA regulatory relationships. Cross-talk is thought to occur between lncRNAs, miRNAs, and mRNAs during all aspects of the HCC process, including metastasis [36], proliferation and invasion [37, 38]. We proposed a prognostic-related ceRNA network in HCC, aimed to explore the underlying mechanism of the disturbed ceRNA axis, and identified effective prognostic indicators from a comprehensive perspective. In this novel ceRNA network, several oncogenes and tumor suppressors were identified that participate in the development of HCC. The identified lncRNAs showed moderate prognostic value; among them were the lncRNAs SNHG1 [39, 40], SNHG3 [41], SNHG6 [42], SNHG12

Fig. 14. Gene Ontology (GO) terms of genes included in the ceRNA network. (A) Biological process; (B) Cellular component; (C) Molecular function.

[43], which have been validated previously as oncogenes and to promote the inferior survival of patients with HCC. Host genes may affect small nucleolar RNAs (snoRNAs) expression through co-transcriptional regulation [44]; snoRNAs are non-coding RNAs of 60-300 nucleotides in length [45]. Up-regulated LINC00261 has also been reported to suppress the proliferation and invasion of HCC cells [46]. These basic findings may, in part, hint that snoRNA host genes may have a profound impact on the prognosis of HCC, emphasizing that the prognostic value of lncRNAs warrants future research.

Considering that mRNAs are the implementers of molecular function in the ceRNA network, GO and KEGG analyses were performed to discover pathways that were disturbed by the molecular cluster. Remarkably, bioinformatics analysis hinted that destruction of the homeostasis of the cell cycle may be the predominant approach influence of the ceRNA network on the prognosis of HCC. The precise factors involved in the proliferative signals of HCC are understood poorly [47, 48]. Our results demonstrated that the cell cycle may be the most affected pathway, which can guide individualized therapy choices.

However, several limitations should be discussed. First, the HCC patient information provided by TCGA should be confirmed by other experimental methods. Second, several novel lncRNAs, with significant clinical significance in HCC need to be explored further to determine the underlying molecular mechanism. Finally, despite the limited power of detecting individual events, the PI that we proposed has promising implications in clinical practice. In summary, our study delineates a ceRNA network that may improve the poor prognosis of HCC and identified functional implications for regulating the cell cycle. These findings have provided several new potential prognostic markers, and identified novel therapeutic targets for HCC.

Acknowledgements

This study was supported by the Guangxi Science and Technology Program (AB17195020), the Fund of National Natural Science Foundation of China (NSFC81060202, NSFC81260222), the Guangxi Medical University Training Program for Distinguished Young Scholars (2017) and Innovation Project of Guangxi Graduate Education (YCSW2018104).

Disclosure Statement

The authors declare that there are no conflicts of interest.

References

- 1 Siegel RL, Miller KD, Jemal A: Cancer Statistics, 2017. *CA Cancer J Clin* 2017;67:7-30.
- 2 Singal AG, El-Serag HB: Hepatocellular Carcinoma From Epidemiology to Prevention: Translating Knowledge into Practice. *Clin Gastroenterol Hepatol* 2015;13:2140-2151.
- 3 McGlynn KA, Petrick JL, London WT: Global epidemiology of hepatocellular carcinoma: an emphasis on demographic and regional variability. *Clin Liver Dis* 2015;19:223-238.
- 4 Ledda C, Loreto C, Zammit C, Marconi A, Fago L, Matera S, Costanzo V, Fuccio Sanza G, Palmucci S, Ferrante M, Costa C, Fenga C, Biondi A, Pomara C, Rapisarda V: Noninfective occupational risk factors for hepatocellular carcinoma: A review (Review). *Mol Med Rep* 2017;15:511-533.
- 5 Bosetti C, Turati F, La Vecchia C: Hepatocellular carcinoma epidemiology. *Best Pract Res Clin Gastroenterol* 2014;28:753-770.
- 6 Kudo M: Systemic Therapy for Hepatocellular Carcinoma: 2017 Update. *Oncology* 2017;93 Suppl 1:135-146.

- 7 Li D, Liu X, Zhou J, Hu J, Zhang D, Liu J, Qiao Y, Zhan Q: Long noncoding RNA HULC modulates the phosphorylation of YB-1 through serving as a scaffold of extracellular signal-regulated kinase and YB-1 to enhance hepatocarcinogenesis. *Hepatology* 2017;65:1612-1627.
- 8 Ghanem I, Riveiro ME, Paradis V, Faivre S, de Parga PM, Raymond E: Insights on the CXCL12-CXCR4 axis in hepatocellular carcinoma carcinogenesis. *Am J Transl Res* 2014;6:340-352.
- 9 Quinn JJ, Chang HY: Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* 2016;17:47-62.
- 10 Schmitt AM, Chang HY: Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* 2016;29:452-463.
- 11 Slaby O, Laga R, Sedlacek O: Therapeutic targeting of non-coding RNAs in cancer. *Biochem J* 2017;474:4219-4251.
- 12 He T, Zhang L, Kong Y, Huang Y, Zhang Y, Zhou D, Zhou X, Yan Y, Zhang L, Lu S, Zhou J, Wang W: Long non-coding RNA CASC15 is upregulated in hepatocellular carcinoma and facilitates hepatocarcinogenesis. *Int J Oncol* 2017;51:1722-1730.
- 13 Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP: A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011;146:353-358.
- 14 Thomson DW, Dinger ME: Endogenous microRNA sponges: evidence and controversy. *Nat Rev Genet* 2016;17:272-283.
- 15 Li C, Chen J, Zhang K, Feng B, Wang R, Chen L: Progress and Prospects of Long Noncoding RNAs (lncRNAs) in Hepatocellular Carcinoma. *Cell Physiol Biochem* 2015;36:423-434.
- 16 Shi X, Sun M, Liu H, Yao Y, Song Y: Long non-coding RNAs: a new frontier in the study of human diseases. *Cancer Lett* 2013;339:159-166.
- 17 Wang Y, Chen F, Zhao M, Yang Z, Li J, Zhang S, Zhang W, Ye L, Zhang X: The long noncoding RNA HULC promotes liver cancer by increasing the expression of the HMGA2 oncogene via sequestration of the microRNA-186. *J Biol Chem* 2017;292:15395-15407.
- 18 Fang L, Sun J, Pan Z, Song Y, Zhong L, Zhang Y, Liu Y, Zheng X, Huang P: Long non-coding RNA NEAT1 promotes hepatocellular carcinoma cell proliferation through the regulation of miR-129-5p-VCP-IkappaB. *Am J Physiol Gastrointest Liver Physiol* 2017;313:G150-G156.
- 19 Zhang J, Fan D, Jian Z, Chen GG, Lai PB: Cancer Specific Long Noncoding RNAs Show Differential Expression Patterns and Competing Endogenous RNA Potential in Hepatocellular Carcinoma. *PLoS One* 2015;10:e0141042.
- 20 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498-2504.
- 21 Budczies J, Klauschen F, Sinn BV, Györfy B, Schmitt WD, Darb-Esfahani S, Denkert C: Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. *PLoS One* 2012;7:e51862.
- 22 Nibbe RK, Chowdhury SA, Koyuturk M, Ewing R, Chance MR: Protein-protein interaction networks and subnetworks in the biology of disease. *Wiley Interdiscip Rev Syst Biol Med* 2011;3:357-367.
- 23 Xie CR, Wang F, Zhang S, Wang FQ, Zheng S, Li Z, Lv J, Qi HQ, Fang QL, Wang XM, Yin ZY: Long Noncoding RNA HCAL Facilitates the Growth and Metastasis of Hepatocellular Carcinoma by Acting as a ceRNA of LPTM4B. *Mol Ther Nucleic Acids* 2017;9:440-451.
- 24 Sui J, Yang X, Qi W, Guo K, Gao Z, Wang L, Sun D: Long Non-Coding RNA Linc-USP16 Functions As a Tumour Suppressor in Hepatocellular Carcinoma by Regulating PTEN Expression. *Cell Physiol Biochem* 2017;44:1188-1198.
- 25 Wang Z, Wu Q, Feng S, Zhao Y, Tao C: Identification of four prognostic lncRNAs for survival prediction of patients with hepatocellular carcinoma. *PeerJ* 2017;5:e3575.
- 26 Zhen Y, Xinghui Z, Chao W, Yi Z, Jinwen C, Ruifang G, Chao Z, Min Z, Chunlei G, Yan F, Lingfang D, Long S, Wenzhi S, Xiaohu L, Rong X: Several microRNAs could predict survival in patients with hepatitis B-related liver cancer. *Sci Rep* 2017;7:45195.
- 27 Shi KQ, Lin Z, Chen XJ, Song M, Wang YQ, Cai YJ, Yang NB, Zheng MH, Dong JZ, Zhang L, Chen YP: Hepatocellular carcinoma associated microRNA expression signature: integrated bioinformatics analysis, experimental validation and clinical significance. *Oncotarget* 2015;6:25093-25108.

- 28 Li B, Feng W, Luo O, Xu T, Cao Y, Wu H, Yu D, Ding Y: Development and Validation of a Three-gene Prognostic Signature for Patients with Hepatocellular Carcinoma. *Sci Rep* 2017;7:5517.
- 29 Cui M, Sun J, Hou J, Fang T, Wang X, Ge C, Zhao F, Chen T, Xie H, Cui Y, Yao M, Li J, Li H: The suppressor of cytokine signaling 2 (SOCS2) inhibits tumor metastasis in hepatocellular carcinoma. *Tumour Biol* 2016;37:13521-13531.
- 30 Qiu X, Zheng J, Guo X, Gao X, Liu H, Tu Y, Zhang Y: Reduced expression of SOCS2 and SOCS6 in hepatocellular carcinoma correlates with aggressive tumor progression and poor prognosis. *Mol Cell Biochem* 2013;378:99-106.
- 31 Yin L, Chang C, Xu C: G2/M checkpoint plays a vital role at the early stage of HCC by analysis of key pathways and genes. *Oncotarget* 2017;8:76305-76317.
- 32 Jiang P, Tang Y, He L, Tang H, Liang M, Mai C, Hu L, Hong J: Aberrant expression of nuclear KPNA2 is correlated with early recurrence and poor prognosis in patients with small hepatocellular carcinoma after hepatectomy. *Med Oncol* 2014;31:131.
- 33 Hao NB, He YF, Li XQ, Wang K, Wang RL: The role of miRNA and lncRNA in gastric cancer. *Oncotarget* 2017;8:81572-81582.
- 34 Huang X, Zhi X, Gao Y, Ta N, Jiang H, Zheng J: LncRNAs in pancreatic cancer. *Oncotarget* 2016;7:57379-57390.
- 35 Klingenberg M, Matsuda A, Diederichs S, Patel T: Non-coding RNA in hepatocellular carcinoma: Mechanisms, biomarkers and therapeutic targets. *J Hepatol* 2017;67:603-618.
- 36 Wang H, Huo X, Yang XR, He J, Cheng L, Wang N, Deng X, Jin H, Wang N, Wang C, Zhao F, Fang J, Yao M, Fan J, Qin W: STAT3-mediated upregulation of lncRNA HOXD-AS1 as a ceRNA facilitates liver cancer metastasis by regulating SOX4. *Mol Cancer* 2017;16:136.
- 37 Wang Y, Hu Y, Wu G, Yang Y, Tang Y, Zhang W, Wang K, Liu Y, Wang X, Li T: Long noncoding RNA PCAT-14 induces proliferation and invasion by hepatocellular carcinoma cells by inducing methylation of miR-372. *Oncotarget* 2017;8:34429-34441.
- 38 Liu D, Zhu Y, Pang J, Weng X, Feng X, Guo Y: Knockdown of long non-coding RNA MALAT1 inhibits growth and motility of human hepatoma cells via modulation of miR-195. *J Cell Biochem* 2018;119:1368-1380.
- 39 Jin YX, Zheng Z, Yu XF, Zhang JB, Namgoong S, Cui XS, Hyun SH, Kim NH: Autophagy and ubiquitin-mediated proteolysis may not be involved in the degradation of spermatozoon mitochondria in mouse and porcine early embryos. *Zygote* 2016;24:31-41.
- 40 Zhang M, Wang W, Li T, Yu X, Zhu Y, Ding F, Li D, Yang T: Long noncoding RNA SNHG1 predicts a poor prognosis and promotes hepatocellular carcinoma tumorigenesis. *Biomed Pharmacother* 2016;80:73-79.
- 41 Zhang T, Cao C, Wu D, Liu L: SNHG3 correlates with malignant status and poor prognosis in hepatocellular carcinoma. *Tumour Biol* 2016;37:2379-2385.
- 42 Chang L, Yuan Y, Li C, Guo T, Qi H, Xiao Y, Dong X, Liu Z, Liu Q: Upregulation of SNHG6 regulates ZEB1 expression by competitively binding miR-101-3p and interacting with UPF1 in hepatocellular carcinoma. *Cancer Lett* 2016;383:183-194.
- 43 Lan T, Ma W, Hong Z, Wu L, Chen X, Yuan Y: Long non-coding RNA small nucleolar RNA host gene 12 (SNHG12) promotes tumorigenesis and metastasis by targeting miR-199a/b-5p in hepatocellular carcinoma. *J Exp Clin Cancer Res* 2017;36:11.
- 44 Filipowicz W, Pogacic V: Biogenesis of small nucleolar ribonucleoproteins. *Curr Opin Cell Biol* 2002;14:319-327.
- 45 Jorjani H, Kehr S, Jedlinski DJ, Gumienny R, Hertel J, Stadler PF, Zavolan M, Gruber AR: An updated human snoRNAome. *Nucleic Acids Res* 2016;44:5068-5082.
- 46 Zhang HF, Li W, Han YD: LINC00261 suppresses cell proliferation, invasion and Notch signaling pathway in hepatocellular carcinoma. *Cancer Biomark* 2018;21:575-582.
- 47 Qin LX, Tang ZY: The prognostic molecular markers in hepatocellular carcinoma. *World J Gastroenterol* 2002;8:385-392.
- 48 Nahon P, Zucman-Rossi J: Single nucleotide polymorphisms and risk of hepatocellular carcinoma in cirrhosis. *J Hepatol* 2012;57:663-674.