Identification of an Eight-LncRNA Signature as the Prognostic LncRNA Markers in Hepatocellular Carcinoma Patients

(Pengenalpastian Tandatangan Lapan-LncRNA sebagai Penanda LncRNAPrognostik dalam Pesakit Karsinoma Hepatosel)

DONGFANG WANG, TIANCHI LIU, WEIPING ZENG, SIXU LI, KEAI SINN TAN,
WEN TAN & LI GU*

ABSTRACT

Long non-coding RNA (lncRNA) signature has been reputable for the predetermination of cancer prognosis. In the present study, we constructed a lncRNA model to predict the survival outcomes for hepatocellular carcinoma (HCC) patients. Using the transcriptome data from TCGA HCC samples, we identified that NEAT1 and MALAT1 were highly expressed in HCC and other tumor subtypes compared to the adjacent normal tissues. Based on the LASSO Cox regression model, we identified an eight-lncRNA signature that significantly correlated with the overall survival and disease-free survival of the HCC training group. The prognostic value of this signature was validated using the test group. Further analysis suggested that this signature was associated with the clinicopathological parameters such as vascular tumor invasion, pathological stage, and tumor grade. Integrated functional analysis showed that these eight-lncRNAs were involved in the cell cycle, metabolic process, and immune response. In conclusion, we constructed an applicable eight-lncRNA signature that is robust and reliable for the prognosis of HCC. This signature may provide an efficient clinical prediction for HCC patients, and further study is required to uncover the function of the identified lncRNAs better.

Keywords: Cancer genome atlas; gene set enrichment analysis; hepatocellular carcinoma; long non-coding RNA; prognosis biomarkers

ABSTRAK

Tandatangan RNA bukan pengekodan panjang (lncRNA) telah terkenal untuk penentuan awal prognosis kanser. Dalam kajian ini, kami membina model lncRNA untuk meramalkan hasil kelangsungan hidup untuk pesakit karsinoma hepatosel (HCC). Menggunakan data transkrip daripada sampel TCGA HCC, kami mengenal pasti bahawa NEAT1 dan MALAT1 sangat dinyatakan dalam HCC dan sub-jenis tumor lain berbanding dengan tisu normal bersebelahan. Berdasarkan model regresi LASSO Cox, kami mengenal pasti tandatangan lapan-lncRNA yang berkorelasi dengan ketara kemandirian keseluruhan dan kelangsungan hidup bebas penyakit kumpulan latihan HCC. Nilai prognostik tandatangan ini telah disahkan menggunakan kumpulan ujian. Analisis lanjut mencadangkan bahawa tandatangan ini dikaitkan dengan parameter klinikopatologi seperti pencerobohan tumor vaskular, peringkat patologi dan gred tumor. Analisis fungsi bersepadu mendedahkan bahawa lapan-lncRNA ini terlibat dalam kitaran sel, proses metabolik dan tindak balas imun. Kesimpulannya, kami membina tandatangan lapan-lncRNA yang terpakai yang teguh dan boleh dipercayai untuk prognosis HCC. Tandatangan ini mungkin memberikan ramalan klinikal yang cekap untuk pesakit HCC dan kajian lanjut diperlukan untuk mendedahkan fungsi lncRNA yang dikenal pasti dengan lebih baik.

Kata kunci: Analisis pengayaan set gen; atlas genom kanser; biopenanda prognosis; karsinoma hepatosel; RNA bukan pengekodan panjang

INTRODUCTION

Liver cancer leads to a predictable 782,500 new cases and 745,500 deaths in 2012 (Chen et al. 2015). Among those, China alone accounting for ~ 50% of the total number of cases and deaths (Chen et al. 2015; Torre et al. 2012). It is the top 5 cancers and the second deadliest malignancy worldwide. Nearly 70-90% of primary liver cancers are diagnosed as hepatocellular carcinoma (HCC) (Chen

et al. 2015). Despite the improvement in diagnostic technology and therapeutic methods, the mortality of HCC is high, and the 5-year survival rate of this cancer in the USA is dismal at 15%, due to its heterogeneity, metastasis, and reoccurrence after surgical resection (Forner et al. 2012; Howlader et al. 2014; Singh et al 2014). Although some critical etiological factors of HCC are known, such as hepatitis B virus (HBV), hepatitis C virus (HCV), alcohol, aflatoxin-intake, obesity, and metabolic disorder (Dutta & Mahato 2017), a lack of predictive biomarkers and therapeutic targets is crucial to the poor outcome. Therefore, identifying novel reliable indicators is essential for the patients' prognostic prediction, which can improve the quality of life and survival for HCC patients.

lncRNA has been identified with diverse functions, including transcriptional regulation of protein-coding genes, organization of chromatins and regulation of proteins or RNA molecules, and the development of transcriptome sequencing technologies (Kopp & Mendell 2018). Studies have also validated the dysregulation of lncRNA in various diseases, including cancer (Xue et al. 2017). Increasing evidence showed that lncRNA has an essential contribution in cancer tumorigenesis, development, and metastasis with multiple mechanisms, indicating their promising candidates as innovative biomarkers for cancer diagnosis, prognosis, and therapy (Chiu et al. 2018).

lncRNAs are potential prognostic candidates for HCC. Therefore, an accurate lncRNA signature might have significant prognostic value for the management of HCC. Several studies have explored the expression profile of lncRNA and their complex competing endogenous RNA (ceRNA) network in HCC, providing a treasured source for further documentation of novel biomarkers and therapeutic targets among HCC patients (Cui et al. 2017; Espoti et al. 2016; Yang et al. 2017). Zhao et al. (2018) found a five-lncRNA signature would advance the prognosis prediction for HCC, and Gu et al. (2019) developed a six-lncRNA signature related to HCC recurrence. Referring to the HCC research publications, Sui et al. (2018) identified lncRNAs, including the LINC00261, TRELM3P, GBP1P1, and CDKN2B-AS1, which may be beneficial biomarkers of the prognosis for HCC patients. Although several prognostic biomarkers for HCC have been identified, many more potential effective lncRNA signatures are imperative to be exposed to advance the clinical outcome of HCC patients.

Here, to search for a more effective lncRNA signature for HCC prognosis, we executed a comprehensive lncRNAexpression profile analysis via the data from The Cancer Genome Atlas (TCGA) database. We found that both NEAT1 and MALAT1 could serve as a diagnostic marker for many cancer subtypes. We identified an eight-lncRNA prognostic signature using the LASSO Cox regression model and verified that this signature was associated with the clinicopathological parameters, suggesting its usefulness. We also showed the integrated regulatory network for the eight lncRNAs. Our results suggest the potential roles of this lncRNA signature in HCC progression as well as prognosis.

MATERIALS AND METHODS

RNA SEQUENCING DATA OF HCC

Data of RNA sequencing and corresponding clinical data of HCC, containing 374 tumor tissues and 50 adjacent normal tissues of the liver, were downloaded from the TCGA database. Illumina HiSeq 2000 platform was used to generate the data, and subsequently, the data were annotated to the hg19 genome (Cancer Genome Atlas Research Network et al. 2017). A total of 365 samples with survival information was included and further randomly assigned to a training group (n = 180) and a testing group (n = 185). The original counts for more than 60,000 transcripts were available. The value of FPKM (Fragment Per Kilobase of Exon per Million fragments mapped) of each sample was obtained for further analyses using the formula: FPKM= (raw read counts)/ (total mapped reads (Millions)*length of lncRNA/gene).

IDENTIFICATION OF DYSREGULATED LNCRNAS IN HCC

Differentially expressed transcripts between HCC and normal tissues were screened using the Cyber-T model (T-testing) (Luo et al. 2017). |logFC|>1, p < 0.05 and FDR < 0.05 were used as the threshold, and lncRNAs were identified based on the GENCODE. The interquartile range (IQR) model was used to identify each lncRNA's various levels, and we selected top 500 differently expressed lncRNAs for advanced analysis. Hierarchical cluster analysis and heatmap were conducted on these lncRNAs to differentiate them better.

CONSTITUTION OF LASSO COX REGRESSION MODEL FOR SURVIVAL ANALYSIS

Based on TCGA HCC patients' survival information, univariate Cox regression analysis was carried out to identify prognostic markers using the differentially expressed lncRNAs, and lncRNAs were considered significant with a cutoff of $P < 0.001$. Then, Glmnet (Friedman et al. 2010) package in R was employed to operate the LASSO Cox regression model (Tibshirani

1997) to select prognostic markers for HCC patients. Cox proportional hazard regression model was used as the algorithm method, and Leave-One-Out Cross-Validation (LOO-CV) was carried out to enhance the solidity and accurateness of our result. The potential prognostic markers were then selected, and the regression coefficients of all markers were derived from this model.

TESTING OF SURVIVAL ANALYSIS

The risk score of each sample was calculated out based on a linear combination of the regression coefficient derived from the LASSO Cox regression model coefficients multiplied with its lncRNAs expression level: Risk score = $\Sigma_t\,\omega_i\chi_i$, while ω_i represents the coefficient of every marker and χ_i represents the expression value of each marker, according to the constructed model. The cutoff value was calculated using X-tile plots (Camp et al. 2004) based on the patients' status, survival months, and the risk scores of patients, in which it was used to classify samples from high-risk to low-risk groups. If one sample's risk score is larger than the cutoff value, it was regarded as a high-risk group; otherwise, it was in the low-risk group. Survival curves were conducted using the Kaplan-Meier method, and survival difference was evaluated by the log-rank test based on the downloaded survival information. $P < 0.05$ was considered statistically significant.

GENE SET ENRICHMENT ANALYSIS (GSEA)

To govern if the members of a given gene set were associated with candidate lncRNA, GSEA was accomplished using GSEA software V2.2 (Subramanian et al. 2005). Gene set used for the enrichment analysis was C5: BP: GO biological process, containing 4436 gene sets. After carrying 1000 random sample permutations, gene set with an FDR less than 0.01 were termed 'enriched'. The integrative analysis was done using the top20 positive/negative-enriched GO terms of all the lncRNAs, and the pathways that arose for more than two candidate lncRNAs were charted using the pheatmap package in R based on its normalized enrichment score (NES).

STATISTICAL ANALYSIS

Unless otherwise noted, all values were presented as median with interquartile range, and the Mann-Whitney test was employed to assess differential expression of lncRNA among two groups. Fisher Exact test was carried out to analyze the relationship between the eight-lncRNA signature-based risk score and the clinicopathological factors. $P < 0.05$ was considered statistically significant, and the statistical analysis was carried out using GraphPad Prism 5.

RESULTS

IDENTIFICATION OF DYSREGULATED LNCRNAS IN HCC

LncRNA played significant roles in cancer development and potentially function in the diagnosis, therapy and prognosis of cancer (Xue et al. 2017). We hypothesis that the lncRNA could serve as the prognostic marker for HCC. There are 424 HCC samples of the RNA-seq data, containing 374 tumor and 50 normal tissues of liver, in which was downloaded from TCGA database. A total of 19,560 transcripts were screened out using Cyber-T model (Luo et al. 2017), and 3231 of them were identified as lncRNAs. Based on the interquartile range (IQR) statistic, the top500 lncRNA was selected for further analysis. The hierarchical cluster analysis of these differently expressed lncRNAs is showed in Figure 1(A). A total of 479 lncRNAs were up-regulated while 21 lncRNAs were down-regulated.

We further constructed a M-versus-A (MA) plot (log2 fold change (M) versus mean expressing (A) showing differentially expressed lncRNAs comparing HCC to adjacent normal tissue samples. Among them, we found that NEAT1 and MALAT1 (Figure 1(B)), two highly expressed human lncRNAs with adjacent chromosomal location in 11q13.1 (Hutchinson et al. 2007), are overexpressed in HCC (Figure 1(C)). Following that, we further investigated the expression of NEAT1 and MALAT1 in other tumor subtypes from TANRIC database (Li et al. 2015) and found that both NEAT1 and MALAT1 exhibited a higher expression in tumor subtypes such as the BRCA, HNSC, KICH, KIRC, KIRP, PRAD and STAD, but not LUAD or THCA (Figure 1(D)). Our findings suggested that NEAT1 and MALAT1 may serve as common diagnostic markers for most of the cancers. We further validated the expression of the other highly expressed lncRNAs (Figure 1(E)). The AC079466.1 and BAIAP2-DT were overexpressed in HCC, while the expression of FAM99A was down-regulated in HCC (Figure 1(E)). However, their functions in the progression of HCC need to be further investigated. In agreement with the other groups, we also found that MEG3 (He et al. 2017), MIR4435-2HG (Kong et al. 2019) and LINC01093 (He et al. 2019) were exceedingly found in HCC (Figure S1), suggesting the reliability of our analysis.
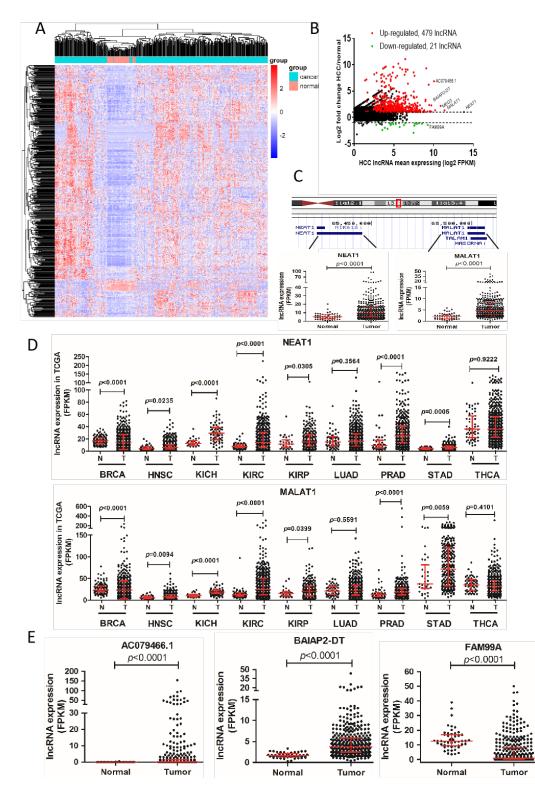
FIGURE 1. Differentially expressed lncRNAs in TCGA HCC samples. (A) Heatmap of 500 different expressed lncRNAs between HCC samples and normal liver tissues. Column labels represent the type of samples: Blue: cancer; Red: normal. (B) A MA plot (log2 fold change (M) versus mean expressing (A)) showing differentially expressed lncRNAs comparing HCC to adjacent normal ones. Among them, significantly up-regulated lncRNAs are represented by red dots, while significantly down-regulated lncRNAs are represented by green dots. (C) The expression of NEAT1 and MALAT1 in HCC. Top: Schematic representations the chromosomal location of NEAT1 and MALAT1 in 11q13.1. Bottom: Expression of NEAT1 and MALAT1 between normal liver tissues and HCC samples. (D) Expression of NEAT1 and MALAT1 in other subtypes of tumor from TANRIC database. N represents normal and T represents tumor. Tumor subtypes with less than 20 adjacent normal tissue samples have been excluded. BRCA: Breast Invasive Carcinoma, HNSC: Head and Neck Squamous Cell Carcinoma, KICH: kidney chromophobe, KIRC: Kidney renal clear cell carcinoma, KIRP: kidney renal papillary cell carcinoma, LUAD: lung adenocarcinoma, PRAD: prostate adenocarcinoma, STAD: stomach adenocarcinoma, THCA: thyroid carcinoma. (E) 3 different expressed lncRNAs (AC079466.1, BAIAP2-DT and FAM99A) between HCC and normal tissues

IDENTIFICATION AND VERIFICATION OF AN EIGHT-LNCRNA SIGNATURE AS THE PROGNOSTIC MARKER FOR HCC PATIENTS

To evaluate the prognostic value of lncRNAs for HCC, a total of 365 samples with survival information was included and further randomly assigned to a training group (n = 180) and a testing group (n = 185). The differentially expressed lncRNAs (Figure 1(A)) were subjected to the LASSO Cox regression model (Shahraki et al. 2015) to identify those lncRNAs that were associated with overall survival (OS) (Figure S2). Eight lncRNAs (including AC145207.5, AC131009.1, SNHG4, CTB-175P5.4, CTD-2012J19.3, RP11-114M1.1, FIRRE, and RP11-359G22.2) were screened out, and the risk score of each sample was calculated based on Risk score=$\Sigma\iota$ $\omega i \chi i$, while ωi represents the coefficient of each marker (Table 1) and χi represents the expression value of each lncRNA.

TABLE 1. The information and coefficient of 8 selected prognostic lncRNA markers for HCC

Gene ID	Gene symbol	Coefficient
ENSG00000263731.1	AC145207.5	0.020863124
ENSG00000255992.1	AC131009.1	0.106454878
ENSG00000281398.1	SNHG4	0.119527412
ENSG00000269416.4	CTB-175P5.4	0.038946385
ENSG00000271119.1	CTD-2012J19.3	0.174812816
ENSG00000228561.2	RP11-114M1.1	0.019652506
ENSG00000213468.3	FIRRE	0.037341095
ENSG00000230400.1	RP11-359G22.2	0.002542046

Based on all the 180 samples' risk scores, we used X-tile plots to calculate the cutoff value and then divided samples from high-risk to low-risk groups. The scores of 62 samples were higher than the cutoff value, and they were regarded as the high-risk group, while the other 118 samples were in the low-risk group (< cutoff value) (Figure 2(A)). Kaplan-Meier survival analysis showed that high-risk patients exhibited a remarkably shorter OS and disease-free survival (DFS) than the low-risk group (log-rank $p < 0.0001$ and $p < 0.0001$; Figure 2(B)). The median survival times of patients with low- and high-risk were 84.4 and 23.1 months, respectively, and the cumulative survival rates for 5-year among the low- and high-risk groups were 57.48 and 18.3%, respectively (Figure 2(B)). The clinical outcome of disease-free survival was improved in the low-risk subgroup compared to that in the high-risk subgroup.

To verify our lncRNA signature model's reliability, we used the same model in the training group to calculate the risk score of samples for the test group, which contains the other 185 samples. Seventy were grouped as high risk, while 115 were grouped as low risk (Figure 2(C)). Similar results were also observed in high-risk group patients, in which they exhibited a worse outcome in both OS and DFS compared to that in the low-risk group (log-rank $p < 0.0001$ and $p = 0.0021$, respectively; Figure 2(D)). Besides, patients with tumors exhibiting high risk had a shorter median overall survival month (33.5 versus 82.87) and median recurrence-free survival months (14.13 versus 35.58) than the patients in low risk, and the 5-year survival rate was much lower in the high-risk group (Figure 2(D)). Additionally, ROC analysis was carried out to compare our lncRNA model's sensitivity and specificity with the TNM stage in predicting the overall survival of HCC patients. Our lncRNA model showed a more robust power than the TNM stage in both the training and test groups and had a better predictive ability to foretell the survival state of HCC patients (Figure 2(E)). Thus, these results indicate that our lncRNA model showed a remarkable performance to envisage HCC patients' outcomes and might be a reliable model for HCC patients' prognosis.

CORRELATION AMONG THE EIGHT-LNCRNA SIGNATURE AND CLINICOPATHOLOGICAL FACTORS

The association of eight-lncRNA signature with clinicopathological factors in the TCGA HCC patients was carried out using the Fisher Exact test. The clinical characteristics include age, gender, hepatic inflammation, Child-Pugh grading, vascular tumor cell

type, pathological stage, TNM stage, fibrosis Ishak score, and radiation status. The Fisher Exact test showed that our eight-lncRNA signature-based risk score is associated with vascular tumor cell type, pathological phase, and tumor stage ($p < 0.05$, Table 2). Patients without vascular tumor invasion seem to have a higher probability of being divided into the low-risk group than patients with vascular tumor invasion. In contrast, patients in the advanced pathological and tumor stages are more concentrated in the high-risk group when compared to those patients in an early stage.
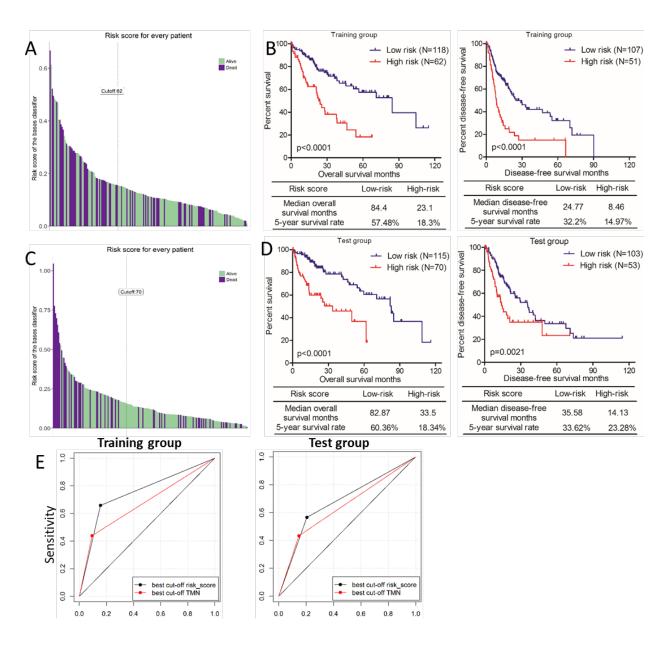

FIGURE 2. The eight-lncRNA expression signature predicted the overall survival (OS) and disease-free survival (DFS) of HCC patients. (A) The distribution of the eight-lncRNA signature-based risk score for HCC training patients. Color labels represent the patients' survival status: Light green: alive; Purple: dead. (B) Kaplan-Meier survival analysis of the OS (left panel) and DFS (right panel) of HCC training group patients. The patients were stratified by low risk versus high risk based on the eight-lncRNA signature. (C) The distribution of the risk score for HCC test patients. Color lables represent the patients' survival status: Light green: alive; Purple: dead. (D) Kaplan-Meier survival analysis of the OS (left panel) and DFS (right panel) of HCC test group patients. (E) Receiver operating characteristic (ROC) analysis of the sensitivity and specificity of the overall survival prediction by the eight-lncRNA risk score and TNM stage in TCGA HCC training group (left panel) and test group (right panel)

TABLE 2. Fisher Exact test between the 8-lncRNA signature-based risk score and clinical characteristics

Characteristics	Variables	high	low	p value
age	<=64	78	138	1
	>64	54	95	
gender	female	48	71	0.295617979
	male	84	162	
hepatic inflammation	NO	34	83	0.886054825
	YES	35	80	
Child-Pugh	A	65	151	1
	B,C	6	16	
vascular tumor cell type	Macro, Micro	44	62	**0.010221211**
	none	55	150	
pathologic stage	I,II	77	177	**0.001087872**
	III,IV	44	43	
tumor stage	T1,T2	86	185	**0.001616755**
	T3,T4	46	45	
node stage	N0	89	159	0.139503672
	N1	3	1	
metastasis stage	M0	97	166	0.556026
	M1	0	3	
fibrosis ishak score	0	16	58	0.25392791
	1,2,3,4,5,6	40	95	
radiation	NO	75	163	1
	YES	1	3	

All the 365 patients were stratified by vascular tumor invasion, pathological stage, and tumor stage and further categorized into high- and low-risk groups based on the eight-lncRNA signature model. While the Kaplan-Meier survival analysis demonstrated that patients without vascular tumor invasion, as well as those in the high-risk group, exhibited a significantly poorer survival when compared to the low-risk group, and a similar result was observed in these patients with vascular tumor invasion (log-rank $p = 0.0017$, and $p = 0.0002$; Figure 3(A)). Also, our lncRNA model is available in the subgroup of early-stage and advanced stage (Figure 3(B)). Regardless of the patients' tumor stage, the variance between high- and low-risk is noticeable (Figure 3(C)). These results showed that the present eight-lncRNA signature-based risk score could servers as an independent prognostic indicator for HCC patients.

THE EXPRESSION OF THE EIGHT-LNCRNAS IN TCGA HCC PATIENTS

To further explore the expression of all the eight-lncRNAs in the TCGA HCC samples, all these lncRNAs are overexpressed in the HCC sample when compared to the adjacent normal tissues (Figure 4(A)). We also assessed their expression among the two risk groups and found that these lncRNAs in the high-risk group are remarkably higher than those in the low-risk group (Figure 4(B)). These data indicate that the patient with a higher expression of the eight-lncRNA had a higher risk, coinciding with our lncRNA model.

INTEGRATIVE ANALYSIS OF THE REGULATORY NETWORK FOR THE EIGHT-LNCRNAS

LncRNAs regulate mRNA expression via translational inhibition, mRNA degradation, and chromatin modifiers' recruitment (Akhade et al. 2017). Considering these mechanisms, we assessed the relationship between our eight lncRNA and mRNA abundance and further investigated the involved potential signal pathways using GSEA. We identified putative regulatory pathways of all the eight-lncRNAs according to the lncRNA and mRNA profiles in the TCGA HCC dataset, and integrative analysis was done using the top20 positive/negative-regulated pathways of all the lncRNAs (Figure 5). As expected, these lncRNAs give the impression to interrelate positively with genes
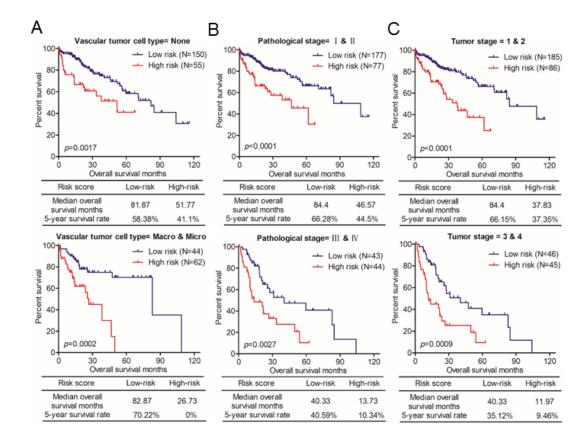
FIGURE 3. Kaplan-Meier survival analysis of the overall survival for the stratified TCGA HCC patients using the eight-lncRNA signature-based risk score. Patients were stratified by vascular tumor cell type, pathological stage and tumor stage. (A) Kaplan-Meier curves for patients without vascular tumor invasion (top panel) and with vascular tumor invasion (bottom panel). (B) Kaplan-Meier curves for patients in early stage (top panel) and advanced stage (bottom panel). (C) Kaplan-Meier curves for patients in T1/2 (top panel) and T3/4 (bottom panel)

FIGURE 4. The expression of the eight-lncRNAs in TCGA HCC patients. (A) LncRNAs expression between TCGA HCC tumor samples and adjacent normal tissues. N represents normal tissues and T represents tumor tissues. (B) LncRNAs expression between the low-risk group samples and high-risk group samples

enriched in the Gene Ontology (GO) biological processes such as cell cycle, nuclear division, organelle fission, chromosome organization, and spindle assembly (Figure 5), all of which would promote the carcinogenesis of the liver. Conversely, they interact negatively with genes associated with the metabolic/biosynthetic process, drug catabolic process, protein activation cascade, complement activation, and humoral immune response (Figure 5), which would destruct the normal function of the liver, inhibit the antitumor effect of patients and stimulate the progress of the tumor. This integrated map shows that these eight lncRNAs have essential liver cancer functions and maybe potential new therapeutic targets against liver cancer.
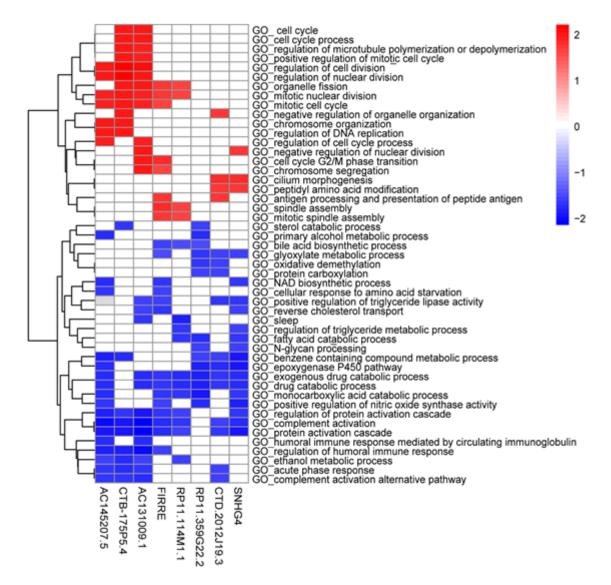

FIGURE 5. Integrative analysis of the regulatory network for the eight-lncRNAs in TCGA HCC samples. Chart shows Gene Ontology (GO) biological process that were enriched for genes involved in Gene Set Enrichment Analysis (GSEA). Each column indicates one lncRNAs of our signature, and each row represents a GO term. A red bar indicates that the GO term is significantly positively enriched in the candidate regulated gene of one or more lncRNAs, while a blue bar indicates that the GO term is negatively enriched. The top20 positive/negative-regulated GO terms were integrated and only pathways that have been enriched by at least two lncRNAs are shown

DISCUSSION

HCC is one of the top malignancies and represents the second leading cause of cancer-caused death globally. In China, HCC is the fourth most commonly diagnosed cancer (466,100) and the third leading cause of cancer death (422,100) (Chen et al. 2016). Though a significantly declining occurrence and mortality trend was observed for HCC in China, the population growth and aging result in a large and growing number of new cases (Chen et al. 2016). The OS rate for 5-year HCC patients remains disgruntled, ranging from 28% for the localized disease to 3% for metastatic patients (Singh et al. 2014). Although most of the risk factors of this cancer are known, and various strategies have been implemented to improve the therapeutic outcome of HCC patients, the specific biomarkers for its prognostic evaluation are crucial.

Currently, an increasing number of comprehensive investigations are related to the human genome. While only 2% of genome sequences encode proteins, most of the transcripts comprise non-coding RNAs (ncRNAs) that lack protein-coding abilities. Depending on the transcript size, there are long non-coding RNAs (lncRNAs, > 200 nt) and small ncRNAs (≤ 200 nt, including miRNA and siRNA) (Ponting et al. 2009). Abnormal ncRNA expression has been demonstrated in many cancers, which often participate in tumorigenesis and tumor progression with various mechanisms. For example, lncRNA TUG1 can accelerate cell metastasis and epithelial-mesenchymal transition by promoting the expression of KIAA1199 via miR-600 in colorectal cancer (Sun et al. 2018), miR-708-5p obstructs lung cancer stem cell-like phenotypes through downregulation of DNMT3a (Liu et al. 2018), and overexpression of lncRNA AGAP2-AS1 causes trastuzumab resistance by activating the NF-κB signaling pathway and epigenetic regulation of MyD88 in breast cancer (Dong et al. 2018). LncRNAs also have been implicated in the diagnosis and prognosis of cancer, like TSLNC8 in HCC (Zhang et al. 2018) and four lncRNAs in acute myeloid leukemia (Beck et al. 2018). More of their underlying mechanisms and potential applications in cancer are still waiting to be explored.

LncRNA signatures have also been investigated in HCC. Cui et al. (2017) showed an all-inclusive genome-wide analysis of lncRNA expression in HCC, several of which accompany HCC patients' OS. Esposti et al. (2016) found both known oncogenic lncRNAs and novel lncRNAs associated with cell cycle and liver metabolism in HCC, using a data of RNA-seq from 23 liver tissues. Yang et al. (2017) identified the recurrently deregulated lncRNAs of HCC samples and summarized their potential function in HCC tumorigenesis and metastasis through 60 clinical samples' RNA-seq data. Several biomarkers for survival prediction have also been identified in HCC, such as the six-gene metastasis signature (Yuan et al. 2017), the five-miRNA expression signature (Liu et al. 2017), the five-lncRNA signature (Zhao et al. 2018), and lncRNACACNA1G-AS1 (Yang et al. 2019). All these studies provided a valuable resource of lncRNA function and biomarkers for HCC development, and more prognostic biomarkers are imperative to be exposed to enhance the accuracy for the prediction of the outcome in HCC patients.

To identify potential prognostic lncRNA markers for HCC, we analyzed the transcriptomic data with clinicopathological information of TCGA HCC samples. In some early studies, the number of TCGA HCC samples was 363, as published previously (Cancer Genome Atlas Research Network 2017). However, the number of samples used in this study was 374 as in 2018. In this study, several highly expressed lncRNAs are dysregulated in HCC, including NEAT1 and MALAT1. These two lncRNAs, both located in the chromosome 11q13.1, have similar transcriptional mechanisms and exhibit colocalization to many human genomic loci, suggesting complementary binding and function (Jen et al. 2017; West et al. 2014). We further investigated their expression in all the TCGA cancer types, and our results showed that NEAT1 and MALAT1 could be the diagnostic biomarkers for most cancers.

Then, the differentially expressed lncRNAs were employed to construct the prediction model using the LASSO Cox regression model. A prognostic formula for HCC patients' outcome was calculated according to the expression of eight-lncRNAs, including AC145207.5, AC131009.1, SNHG4, CTB-175P5.4, CTD-2012J19.3, RP11-114M1.1, FIRRE and RP11-359G22.2. In both the training and test groups, patients in a high-risk group based on the eight-lncRNA signatures exhibited a remarkably shorter OS rate and DFS rate when compared to the low-risk group. Patients' median survival times and the cumulative 5-year survival rates were both worse in the high-risk group. And the effect of this lncRNA model is better than the TNM grade. These results indicate our prognostic model based on the eight-lncRNA signature is reliable for predicting HCC patients.

The relationship of our lncRNA signature with clinicopathological factors was assessed in the HCC patients, and our finding demonstrated that the lncRNA signature is accompanying vascular tumor cell type,

pathological stage, and tumor grade. We further performed stratification analyses and demonstrated that our model of eight-lncRNA signature persisted significant and reliable in HCC subgroups even stratified by vascular tumor invasion, pathological stage, and tumor grade. These data suggested that our model based on the eight-lncRNA signature possibly become an independent prognostic indicator for HCC outcome.

The expression and function of these eight-lncRNAs were further investigated, and the lncRNAs were overexpressed in the tumor samples compared to the adjacent normal tissues. Their expression in the high-risk group was also higher than the low-risk group. Gene Ontology (GO) analysis showed that these lncRNAs positively correlated with genes enriched in the cell cycle, organelle fission, and chromosome organization. In contrast, they are negatively correlated with genes related to metabolic process, drug catabolic process, and immune response, showing that these eight lncRNAs have an essential function for the initiation and progression of HCC through these pathways. Further research is warranted to uncover the functions of these eight-lncRNAs.

In summary, we identify an eight-lncRNA signature to envisage the outcome of HCC patients. The prognostic value of this signature is robust, reliable, and associated with their clinical characteristics. This innovative signature may serve as potential new therapeutic targets against HCC. Further studies will be carried to validate our findings and the functional effects of these identified lncRNAs.

REFERENCES

Akhade, V.S., Pal, D. & Kanduri, C. 2017. Long noncoding RNA: Genome organization and mechanism of action. *Advances in Experimental Medicine and Biology* 1008: 47-74.

Beck, D., Thoms, J.A.I., Palu, C., Herold, T., Shah, A., Olivier, J., Boelen, L., Huang, Y., Chacon, D., Brown, A., Babic, M., Hahn, C., Perugini, M., Zhao, X., Huntly, B.J., Schwarzer, A., Klusmann, J.H., Berdel, W.E., Wormann, B., Buchner, T., Hiddemann, W., Bohlander, S.K., To, L.B., Scott, H.S., Lewis, I.D., D'Andrea, R.J., Wong, J.W.H. & Pimanda, J.E. 2018. A four-gene LincRNA expression signature predicts risk in multiple cohorts of acute myeloid leukemia patients. *Leukemia* 32(2): 263-272.

Camp, R.L., Dolled-Filhart, M. & Rimm, D.L. 2004. X-tile: A new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clinical Cancer Research* 10(21): 7252-7259.

Cancer Genome Atlas Research Network. 2017. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* 169(7): 1327-1341.e23.

Chen, W., Zheng, R., Baade, P.D., Zhang, S., Zeng, H., Bray, F., Jemal, A., Yu, X.Q. & He, J. 2016. Cancer statistics in China, 2015. *A Cancer Journal for Clinicians* 66(2): 115-132.

Chiu, H.S., Somvanshi, S., Patel, E., Chen, T.W., Singh, V.P., Zorman, B., Patil, S.L., Pan, Y., Chatterjee, S.S. Cancer Genome Atlas Research Network, Sood, A.K., Gunaratne, P.H. & Sumazin, P. 2018. Pan-cancer analysis of lncRNA regulation supports their targeting of cancer genes in each tumor context. *Cell Reports* 23(1): 297-312.e212.

Cui, H., Zhang, Y., Zhang, Q., Chen, W., Zhao, H. & Liang, J. 2017. A comprehensive genome-wide analysis of long noncoding RNA expression profile in hepatocellular carcinoma. *Cancer Medicine* 6(12): 2932-2941.

Dong, H., Wang, W., Mo, S., Chen, R., Zou, K., Han, J., Zhang, F. & Hu, J. 2018. SP1-induced lncRNA AGAP2-AS1 expression promotes chemoresistance of breast cancer by epigenetic regulation of MyD88. *Journal of Experimental & Clinical Cancer Research* 37(1): 202.

Dutta, R. & Mahato, R.I. 2017. Recent advances in hepatocellular carcinoma therapy. *Pharmacology & Therapeutics* 173: 106-117.

Esposti, D.D., Hernandez-Vargas, H., Voegele, C., Fernandez-Jimenez, N., Forey, N., Bancel, B., Le Calvez-Kelm, F., McKay, J., Merle, P. & Herceg, Z. 2016. Identification of novel long non-coding RNAs deregulated in hepatocellular carcinoma using RNA-sequencing. *Oncotarget* 7(22): 31862-31877.

Forner, A., Llovet, J.M. & Bruix, J. 2012. Hepatocellular carcinoma. *Lancet* 379(9822): 1245-1255.

Friedman, J., Hastie, T. & Tibshirani, R. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33(1): 1-22.

Gu, J.X., Zhang, X., Miao, R.C., Xiang, X.H., Fu, Y.N., Zhang, J.Y., Liu, C. & Qu, K. 2019. Six-long non-coding RNA signature predicts recurrence-free survival in hepatocellular carcinoma. *World Journal of Gastroenterology* 25(2): 220-232.

He, J., Zuo, Q., Hu, B., Jin, H., Wang, C., Cheng, Z., Deng, X., Yang, C., Ruan, H., Yu, C., Zhao, F., Yao, M., Fang, J., Gu, J., Zhou, J., Fan, J., Qin, W., Yang, X. & Wang, H. 2019. A novel, liver-specific long noncoding RNA LINC01093 suppresses HCC progression by interaction with IGF2BP1 to facilitate decay of GLI1 mRNA. *Cancer Letter* 450: 98-109.

118

He, Y., Luo, Y., Liang, B., Ye, L., Lu, G. & He, W. 2017. Potential applications of MEG3 in cancer diagnosis and prognosis. *Oncotarget* 8(42): 73282-73295.

Howlader, N.N.A., Krapcho, M., Garshell, J., Neyman, N., Altekruse, S.F., Kosary, C.L., Yu, M., Ruhl, J., Tatalovich, Z., Cho, H., Mariotto, A., Lewis, D.R., Chen, H.S., Feuer, E.J. & Cronin, K.A. 2014. *SEER Cancer Statistics Review, 1975-2010*. Bethesda, Maryland: National Cancer Institute.

Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B. & Chess, A. 2007. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* 8: 39.

Jen, J., Tang, Y.A., Lu, Y.H., Lin, C.C., Lai, W.W. & Wang, Y.C. 2017. Oct4 transcriptionally regulates the expression of long non-coding RNAs NEAT1 and MALAT1 to promote lung cancer progression. *Molecular Cancer* 16(1): 104.

Kong, Q., Liang, C., Jin, Y., Pan, Y., Tong, D., Kong, Q. & Zhou, J. 2019. The lncRNA MIR4435-2HG is upregulated in hepatocellular carcinoma and promotes cancer cell proliferation by upregulating miRNA-487a. *Cellular & Molecular Biology Letters* 24: 26.

Kopp, F. & Mendell, J.T. 2018. Functional classification and experimental dissection of long noncoding RNAs. *Cell* 172(3): 393-407.

Li, J., Han, L., Roebuck, P., Diao, L., Liu, L., Yuan, Y., Weinstein, J.N. & Liang, H. 2015. TANRIC: An Interactive open platform to explore the function of lncRNAs in cancer. *Cancer Research* 75(18): 3728-3737.

Liu, G., Wang, H., Fu, J.D., Liu, J.Y., Yan, A.G. & Guan, Y.Y. 2017. A five-miRNA expression signature predicts survival in hepatocellular carcinoma. *Apmis* 125(7): 614-622.

Liu, T., Wu, X., Chen, T., Luo, Z. & Hu, X. 2018. Downregulation of DNMT3A by miR-708-5p inhibits lung cancer stem cell-like phenotypes through repressing Wnt/beta-catenin signaling. *Clinical Cancer Research* 24(7): 1748-1760.

Luo, W., Wang, L., Luo, M.H., Huang, Y.Z., Yang, H., Zhou, Y., Jia, H.T. & Wang, X.X. 2017. hsa-mir-3199-2 and hsa-mir-1293 as novel prognostic biomarkers of papillary renal cell carcinoma by COX ratio risk regression model screening. *Journal of Cellular Biochemistry* 118(10): 3488-3494.

Ponting, C.P., Oliver, P.L. & Reik, W. 2009. Evolution and functions of long noncoding RNAs. *Cell* 136(4): 629-641.

Shahraki, H.R., Salehi, A. & Zare, N. Survival Prognostic factors of male breast cancer in southern Iran: A LASSO-Cox regression approach. *Asian Pacific Journal of Cancer Prevention* 16(15): 6773-6777.

Singh, S., Singh, P.P., Roberts, L.R. & Sanchez, W. 2014. Chemopreventive strategies in hepatocellular carcinoma. *Nature Reviews Gastroenterology & Hepatology* 11(1): 45-54.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. & Mesirov, J.P. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* 102(43): 15545-15550.

Sui, J., Miao, Y., 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. 2018. Systematic analyses of a novel lncRNA-associated signature as the prognostic biomarker for hepatocellular carcinoma. *Cancer Medicine* 7(7): 3240-3256.

Sun, J., Hu, J., Wang, G., Yang, Z., Zhao, C., Zhang, X. & Wang, J. 2018. LncRNA TUG1 promoted KIAA1199 expression via miR-600 to accelerate cell metastasis and epithelial-mesenchymal transition in colorectal cancer. *Journal of experimental & Clinical Cancer Research* 37(1): 106.

The Cancer Genome Atlas Research Network, Wheeler, D.A. & Roberts, L.R. 2017. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* 169(7): 1327-1341.e1323.

Tibshirani, R. 1997. The lasso method for variable selection in the Cox model. *Statistics in Medicine* 16(4): 385-395.

Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J. & Jemal, A. 2015. Global cancer statistics, 2012. *A Cancer Journal for Clinicians* 65(2): 87-108.

West, J.A., Davis, C.P., Sunwoo, H., Simon, M.D., Sadreyev, R.I., Wang, P.I., Tolstorukov, M.Y. & Kingston, R.E. 2014. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Molecular Cancer* 55(5): 791-802.

Xue, M., Zhuo, Y. & Shan, B. 2017. MicroRNAs, long noncoding RNAs, and their functions in human disease. *Methods in Molecular Biology* 1617: 1-25.

Yang, J., Li, C. & Li, H.E.C. 2019. LncRNA CACNA1G-AS1 facilitates hepatocellular carcinoma progression through the miR-2392/C1orf61 pathway. *Journal of Cellular Physiology* 234(10): 18415-18422.

Yang, Y., Chen, L., Gu, J., Zhang, H., Yuan, J., Lian, Q., Lv, G., Wang, S., Wu, Y., Yang, Y.T., Wang, D., Liu, Y., Tang, J., Luo, G., Li, Y., Hu, L., Sun, X., Wang, D., Guo, M., Zi, Q., Xi, J., Wang, H., Zhang, M.Q. & Lu, Z.J. 2017. Recurrently deregulated lncRNAs in hepatocellular carcinoma. *Nature Communications* 8: 14421.

Yuan, S., Wang, J., Yang, Y., Zhang, J., Liu, H., Xiao, J., Xu, Q., Huang, X., Xiang, B., Zhu, S., Li, L., Liu, J., Liu, L. & Zhou, W. 2017. The prediction of clinical outcome in hepatocellular carcinoma based on a six-gene metastasis signature. *Clinical Cancer Research* 23(1): 289-297.

Zhang, J., Li, Z., Liu, L., Wang, Q., Li, S., Chen, D., Hu, Z., Yu, T., Ding, J., Li, J., Yao, M., Huang, S., Zhao, Y. & He, X. 2018. Long noncoding RNA TSLNC8 is a tumor suppressor that inactivates the interleukin-6/STAT3 signaling pathway. *Hepatology* 67(1): 171-187.

Zhao, Q.J., Zhang, J., Xu, L. & Liu, F.F. 2018. Identification of a five-long non-coding RNA signature to improve the prognosis prediction for patients with hepatocellular carcinoma. *World Journal of Gastroenterology* 24(30): 3426-3439.

Dongfang Wang, Tianchi Liu, Weiping Zeng & Keai Sinn Tan
School of Pharmacy
Jinan University
Guangzhou 510632, Guangdong
China

Dongfang Wang, Tianchi Liu, Keai Sinn Tan & Wen Tan
Zhuhai Yuanzhi Health Technology Co. Ltd
Hengqin New Area
Zhuhai 519000, Guangdong
China

Sixu Li
Institute of Biomedical and Pharmaceutical Sciences
Guangdong University of Technology
Guangzhou 510006
China

Li Gu*
Food and Health Engineering Research Center of State Education Ministry
Sun Yat-Sen University
Guangzhou 510275
China

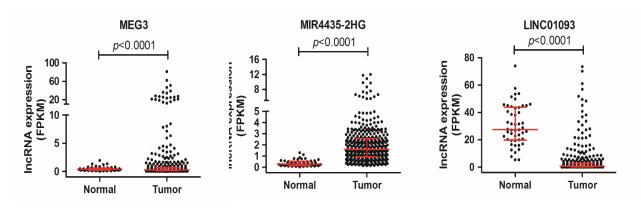*Corresponding author; email: guli9@mail.sysu.edu.cn

FIGURE S1. Differentially expressed lncRNAs between HCC and normal tissues.
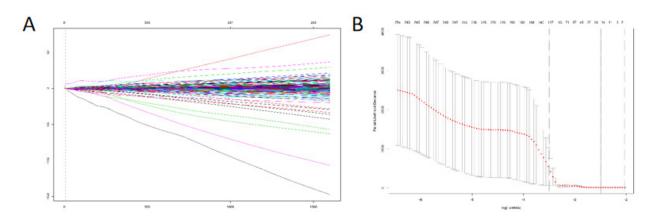(A) MEG3 (B) MIR4435-2HG (C) LINC01093.

FIGURE S2. LASSO profiles of the 500-candidate prognostic lncRNAs in HCC. (A) LASSO
coefficient profiles of the 500-prognostic lncRNAs in HCC. (B) Lasso deviance profiles of
the 500-prognostic lncRNAs in HCC