

Development and validation of a 4-lncRNA combined prediction model for patients with hepatocellular carcinoma

Cui Xu^{B,D}, Xiangxiu Qi^{A,F}

Department of General Surgery, Shengjing Hospital of China Medical University, Shenyang, China

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation;

D – writing the article; E – critical revision of the article; F – final approval of the article

Advances in Clinical and Experimental Medicine, ISSN 1899–5276 (print), ISSN 2451–2680 (online)

Adv Clin Exp Med. 2022;31(10):1087–1097

Address for correspondence

Xiangxiu Qi

E-mail: lnszlyysurgeon@sohu.com

Funding sources

None declared

Conflict of interest

None declared

Received on October 24, 2021

Reviewed on December 12, 2021

Accepted on May 19, 2022

Published online on August 1, 2022

Abstract

Background. Hepatocellular carcinoma (HCC) is one of the most common and lethal cancers worldwide. Therefore, it is necessary to develop and validate a novel prognostic model for HCC patients.

Objectives. To establish an innovative and valuable prediction model of long non-coding RNAs (lncRNAs) for HCC.

Materials and methods. Transcriptome and clinical data from The Cancer Genome Atlas (TCGA) were analyzed globally using bioinformatic approaches. We used Cox and least absolute shrinkage and selection operator (LASSO) regression analyses to screen for prognostic lncRNAs, while receiver operating characteristic (ROC) and Kaplan–Meier curve analyses were used to evaluate the effectiveness of the models. Clinical data from our center were used as a validation set.

Results. In the training set, a prediction model was established based on the expression of AP000844.2, LINC00942, SRGAP3-AS2, and AC010280.2. Hepatocellular carcinoma patients were divided into 2 groups (high-risk group and low-risk group) according to their risk score, and differences in survival were compared between the groups. The clinical data from our center served as a validation set to re-evaluate the effectiveness of the predictive model. The model had an excellent performance. The area under the curve (AUC) of 3-year survival was 0.771, while for 5-year survival it was 0.741, and the concordance index (C-index) was 0.756 (standard error (SE) = 0.023, 95% confidence interval (95% CI) = 0.620–0.891).

Conclusions. The 4-lncRNA combination model is critically important in evaluating the prognosis of HCC. It is an effective independent prognostic factor, although prospective, multi-center studies are needed to validate our findings.

Key words: prognosis, survival analysis, hepatocellular carcinoma, least absolute shrinkage and selection operator

Cite as

Xu C, Qi X. Development and validation of a 4-lncRNA combined prediction model for patients with hepatocellular carcinoma. *Adv Clin Exp Med.* 2022;31(10):1087–1097. doi:10.17219/acem/150256

DOI

10.17219/acem/150256

Copyright

Copyright by Author(s)

This is an article distributed under the terms of the Creative Commons Attribution 3.0 Unported (CC BY 3.0) (<https://creativecommons.org/licenses/by/3.0/>)

Background

Hepatocellular carcinoma (HCC) is one of the most frequent malignancies globally and ranks 3rd for morbidity and mortality rates, respectively.¹ Although advances have been made in multidisciplinary treatment and targeted medicine, the 5-year survival rate of HCC patients remains relatively low due to the poor performance of diagnostic techniques and lack of comprehensive treatment.² Hepatocellular carcinoma diagnosis and treatment still pose a considerable challenge. Patients undergoing surgical treatment miss the opportunity for intervention due to the atypical recurrence of symptoms if there is no active postoperative follow-up.^{3,4} Therefore, a reasonable estimation of the patient's survival time and paying attention to the time window for possible recurrence are vital to improve the prognosis. Thus, it is critical to deepen our understanding of the pathogenesis of HCC development and to seek novel biological markers or therapeutic targets.⁵

Recently, it was detected that the abnormal expression of long non-coding RNAs (lncRNAs) correlate with the occurrence and development of tumors and other diseases.^{6,7} They are valuable in prognostic evaluation, early diagnosis and clinical treatment of many malignant tumors.⁸ More notably, lncRNAs have a critical impact on abnormal cellular regulation and tumorigenesis. At present, in the post-genomic era, lncRNAs are also a vital research hotspot. They regulate genes and play critical biological functions in cell development and multiple levels, such as transcription, post-transcription, translation, and post-translational modification.⁹ Therefore, it is crucial to characterize the structure and interaction of lncRNA for understanding its cellular mechanism of action.¹⁰ Some studies have found that lncRNAs are closely related to tumors, and their expression and regulation are found to be abnormal in breast cancer, colon cancer and stomach cancer.¹¹ Studies on the functions of lncRNAs have shown that the abnormal expression of lncRNAs may be used for the prognosis and treatment of HCC.¹² The lncRNA UPK1A-AS1 could work as a scaffold to reinforce the binding of EZH2 and SUZ12 in order to induce the chemoresistance of HCC cells.¹³ Furthermore, lncRNA LL22NC03-N14H11.1 promotes mitochondrial fission and induces the epithelial–mesenchymal transition of HCC through the MAPK pathway.¹⁴ These data indicate that the aberrant expression of lncRNAs may directly or indirectly adjust the malignant phenotype of HCC and affect its prognosis. In addition, as it is commonly known, genes are highly interactive. Therefore, it is necessary to select appropriate lncRNAs and establish multiple lncRNA prediction models. Such actions could play an essential role in assisting the evaluation of HCC prognosis.

Transcriptome data from the Cancer Genome Atlas (TCGA) and clinical data from Shengjing Hospital of China Medical University were analyzed. We have developed and

validated an effective multi-lncRNA combined survival prediction model using Cox regression and least absolute shrinkage and selection operator (LASSO) regression¹⁵; this model might be helpful in the prognosis of HCC patients.

Objectives

This study aims to establish a prediction model for HCC based on differentially expressed lncRNAs (DELs) and evaluate its effectiveness based on clinical data.

Materials and methods

Data acquisition

The raw RNA sequencing data were downloaded from TCGA (<https://portal.gdc.cancer.gov/>) through the RTCGA Toolbox package (TCGA-Liver Hepatocellular Carcinoma (LIHC)) and matched survival data were obtained the same way. The LIHC data were used as the training set to build a predictive model, as described below. The training set consisted of 424 samples, including 374 cancer tissue samples and 50 normal tissue samples. The transcriptome data were captured using the Illumina HiSeq RNA-Seq platform (<https://portal.gdc.cancer.gov/projects/TCGA-LIHC>).

Bioinformatics analysis

We used the edgeR package (<https://bioconductor.org/packages/release/bioc/html/edgeR.html>) to obtain DELs based on the RNA expression profile data. First, we homogenized the expression of each lncRNA in each sample. Then, we compared the expression value of each lncRNA between the cancer tissue group and the normal tissue group, and the multiple of difference was expressed as fold change (FC) value, with $p < 0.05$ considered statistically different. The value of $p < 0.05$ and $|FC| \geq 4$ ($|\log_2 FC| \geq 2$) were determined as the cut-off values. Any lncRNA that met the above 2 conditions was classified as DEL. Then, the “survival” package was used to perform a univariate proportional hazards model (Cox) regression analysis. The meaningful DELs in univariate Cox regression analysis were enrolled to construct the LASSO (“glmnet” package) regression, and the “survminer” package was adopted for visualization. The effectiveness of lncRNA combined prediction model was evaluated in terms of receiver operating characteristic (ROC) and concordance index (C-index). Subjects were divided into the high-expression group and the low-expression group based on the prediction model, and the Kaplan–Meier survival curve described the clinical prognostic significance of the model. Subsequently, clinical data from our center were used as a validation set to evaluate the effectiveness of the model (the workflow process is described in Fig. 1A).

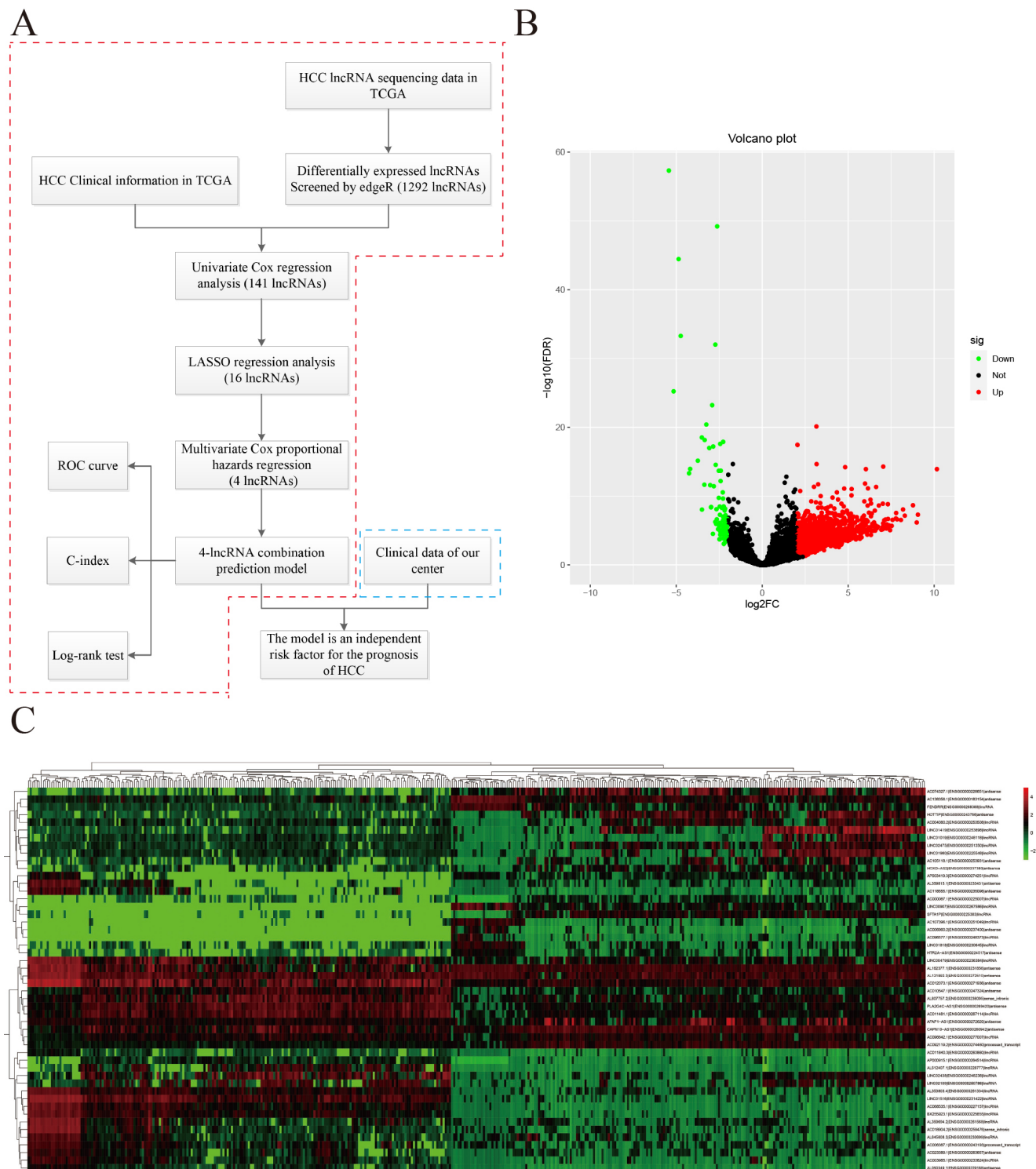


Fig. 1. Differentially expressed long non-coding RNAs (lncRNAs) (DELs) in The Cancer Genome Atlas-Liver Hepatocellular Carcinoma (TCGA-LIHC).

A. The workflow of the study. We obtained the hepatocellular carcinoma (HCC) transcriptome and survival data in TCGA as the training set, and from this we established the DELs prediction model. The clinical data from our center were used as a validation set to evaluate the effectiveness of the model (red box: training set; blue box: validation set); B. Volcano plots showing the expression of DELs screened using edgeR; C. Heatmap showing the expression of the top 50 DELs

LASSO – least absolute shrinkage and selection operator; ROC – receiver operating characteristic; C-index – concordance index.

Ethical statement and tissue samples

We obtained 100 tumor samples and matched non-tumor tissue samples from HCC patients undergoing surgical

resection at the ShengJing Hospital of China Medical University in 2010–2015. All specimens were pathologically confirmed as HCC. The Ethics Committee of ShengJing Hospital of China Medical University approved this study

(approval No. 20191215) and all patients signed informed consent prior to surgery. The follow-up deadline was January 31, 2020. These 100 patients were used as a validation set to evaluate the effectiveness of the model.

Cell culture

The human HCC cell lines (Huh7 and Hep3B) and a hepatocellular cell line (THLE-3) were acquired from China Medical University (Shenyang, China). Cell culture was performed using RPMI-1640 medium (Gibco, Carlsbad, USA) containing 10% fetal bovine serum (FBS) at 37°C with 5% CO₂.

Reverse transcription polymerase chain reaction (RT-PCR)

The TRIzol extraction kits (Invitrogen, Waltham, USA) were adopted for total RNA extraction, and optical density (OD) values were obtained at 260–280 nm using an ultraviolet spectrophotometer. The RNA was used for subsequent quantitative polymerase chain reaction (qPCR) quantification if its OD260/OD280 ratio was >1.8. Next, the reverse transcription of RNA into cDNA was performed using PrimeScript RT kits, with a system of 10 µL, according to the manufacturer's instruction. The reaction conditions

were 25°C for 30 min, 45°C for 30 min and 85°C for 5 min. Using cDNA as a template, quantitative fluorescence PCR was carried out with 2×TaqMan™ Universal PCR Master Mix (Thermo Fisher, Waltham, USA) under the reaction conditions of 95°C for 3 min, cycling 5 times (94°C for 20 s, 63°C for 30 s, 72°C for 30 s) and cycling 40 times (95°C for 15 s, 60°C for 30 s), with U6 being an internal reference. Three wells and negative controls without a template were set up for all reactions. The quantitative analysis was carried out using the 2^{-ΔΔCt} method. All primers were purchased from Sangon Biotech (Shanghai, China).

Statistical analyses

The IBM SPSS v. 21.0 statistical software (IBM Corp., Armonk, USA) and GraphPad Prism v. 8.0 (GraphPad, San Diego, USA) were used for data processing. Each experiment was repeated 3 times. Measurement data were expressed as mean ± standard deviation (SD). The Shapiro–Wilk test was used to check data normality. The expression of lncRNAs in cells conformed to a normal distribution, so an independent t-test was used to compare the expression of lncRNAs between cells. The expression of lncRNAs in tissues did not conform to a normal distribution, so the Mann–Whitney U test was performed to compare the expression of lncRNAs between tumor

Table 1. Results of Shapiro–Wilk test

Item	Statistic	dif	sig
lncRNAs expression in cells			
AP000844.2Huh7	0.996	3	0.878
AP000844.2Hep3B	0.990	3	0.806
AP000844.2THLE3	1.000	3	0.973
SRGAP3AS2Huh7	1.000	3	1.000
SRGAP3AS2Hep3B	0.997	3	0.902
SRGAP3AS2THLE3	1.000	3	0.972
LINC00942Huh7	1.000	3	0.960
LINC00942Hep3B	0.998	3	0.925
LINC00942THLE3	0.942	3	0.537
AC010280.2Huh7	0.997	3	0.900
AC010280.2Hep3B	1.000	3	1.000
AC010280.2THLE3	1.000	3	1.000
lncRNAs expression in tissues			
AP000844.2Cancer	0.971	100	0.025
AP000844.2Adj	0.916	100	0.000
SRGAP3AS2Cancer	0.968	100	0.016
SRGAP3AS2Adj	0.962	100	0.005
LINC00942Cancer	0.927	100	0.000
LINC00942Adj	0.961	100	0.005
AC010280.2Cancer	0.920	100	0.000
AC010280.2Adj	0.952	100	0.001

lncRNAs – long non-coding RNAs.

tissues and non-tumor tissues. Table 1 presents the results of the Shapiro–Wilk test. The Kaplan–Meier method and a log-rank test were used to measure overall survival (OS). Multivariate models of prognostic factors were carried out using Cox regression. The LASSO regression analysis was performed to reduce overfitting caused by univariate Cox regression. Getting the corresponding number of variables by the minimum lambda value of $p < 0.05$ was considered a statistically significant difference.

Results

lncRNAs have differential expression in HCC

The RNA transcriptome data were obtained from TCGA, including 374 HCC tissues and 50 normal tissues (TCGA-LIHC) as the training set. Then, treating the $p < 0.05$ and $|\log_2 FC| \geq 2$ as the cut-off values, the edgeR package was performed to distinguish the DELs. A volcano plot of the distribution of DELs was drawn with the log values of FC and false discovery rate (FDR) as the horizontal and vertical axes, respectively. A total of 1212 up-regulated (in red) and 80 downregulated (in green) DELs were recognized (Fig. 1B). The DELs are listed in Supplementary Table 1 (available at: <https://doi.org/10.5281/zenodo.6794063>). Moreover, the top 50 DELs are displayed in a heatmap (Fig. 1C).

A prediction model based on the co-expression of 4-lncRNAs

It was necessary to evaluate the clinical significance of DELs in HCC. First, univariate Cox regression showed that a total of 141 DELs contributed to the survival of HCC (Supplementary Table 2 (available at: <https://doi.org/10.5281/zenodo.6794063>), $p < 0.05$). Subsequently, we extracted the expression data of DELs in HCC patients ($n = 141$) and obtained the corresponding clinical data. The LASSO regression analysis was performed to further evaluate these data; this analysis could reduce overfitting caused by univariate Cox regression. With a continuous lambda increase, the absolute value of the regression coefficient was correspondingly compressed, and some relatively unimportant variables were compressed to 0. This allowed an expression curve between regression coefficients and lambda values to be obtained (Fig. 2A). When the lambda value reached a specific size, increasing the number of model-independent variables and reducing the lambda value could not significantly improve the model performance. Therefore, we obtained the smallest lambda value using LASSO regression and got the corresponding number of variables. The analysis showed that 16 out of the 141 DELs may be associated with HCC prognosis (Fig. 2B). Furthermore, a multivariate regression analysis

was performed on these 16 DELs, and we discovered that only 4 DELs might be independent risk factors for HCC prognosis (Fig. 2C and Fig. 2D). We adopted these 4 DELs for modeling and assigned scores according to their respective weights in the multi-Cox analysis.

Then, each HCC sample got a risk score based on the expression level of the 4-DEL combination model. A cumulative distribution function (CDF) map was built based on the risk score value of each sample. With the risk score value = 1 as the cut-off value (\log_2 risk score = 0), we divided the samples into the high-risk (red) group and the low-risk (green) group (Fig. 3A). To assess the relationship between the risk score and patient survival, a scatter plot was drawn, with survival time measured in years (Fig. 3B; red: deceased, green: alive). With the increased risk score, the number of surviving patients decreased gradually and the number of deceased patients increased. We used a heatmap to demonstrate the score of 4 DELs in each sample (Fig. 3C).

Then, we evaluated the effectiveness of the model in predicting HCC prognosis. Traditionally, both ROC and C-index have been important indices for a prediction model. We applied these 2 indicators to evaluate the predictive ability of the 4-DEL prediction model. The area under the curve (AUC) of 3-year and 5-year survival were 0.771 and 0.741, respectively (Fig. 3D). In addition, the C-index was 0.756 (standard error (SE) = 0.023, 95% confidence interval (95% CI) = 0.620–0.891). As both indices were greater than 0.7, it suggested that this model was predictive for the prognosis of patients with HCC. The patients were divided into the high-risk group and the low-risk group, according to the risk score. The 5-year survival rate was significantly reduced in the high-risk group (Fig. 3E, $p = 0.0001$). Based on the above analysis, we obtained the 4-DEL combined prediction model for HCC: “Risk score = $1.14 \times \text{AP000844.2} + 1.12 \times \text{LINC00942} + 1.20 \times \text{SRGAP3-AS2} - 0.84 \times \text{AC010280.2}$ ”, and prepared for further model validation (cut-off value = 0.945).

The expression of AP000844.2, LINC00942, SRGAP3-AS2, and AC010280.2 in HCC

We obtained a predictive model based on the molecular expression of 4 lncRNAs, namely AP000844.2, LINC00942, SRGAP3-AS2, and AC010280.2 using a bioinformatics analysis with a public database as the training set. To further demonstrate the authenticity and validity of the analysis, we confirmed the expression of these 4 molecules in HCC. The human HCC cell lines (Huh7 and Hep3B) and the hepatocellular cell line (THLE-3) were used to check the expression at the cell level using RT-PCR. We found that AP000844.2, LINC00942 and SRGAP3-AS2 were overexpressed in HCC cells, as compared with hepatocellular cells (Fig. 4A–C, $p < 0.05$). On the contrary, AC010280.2 was weakly expressed in HCC cells but it was upregulated in THLE-3

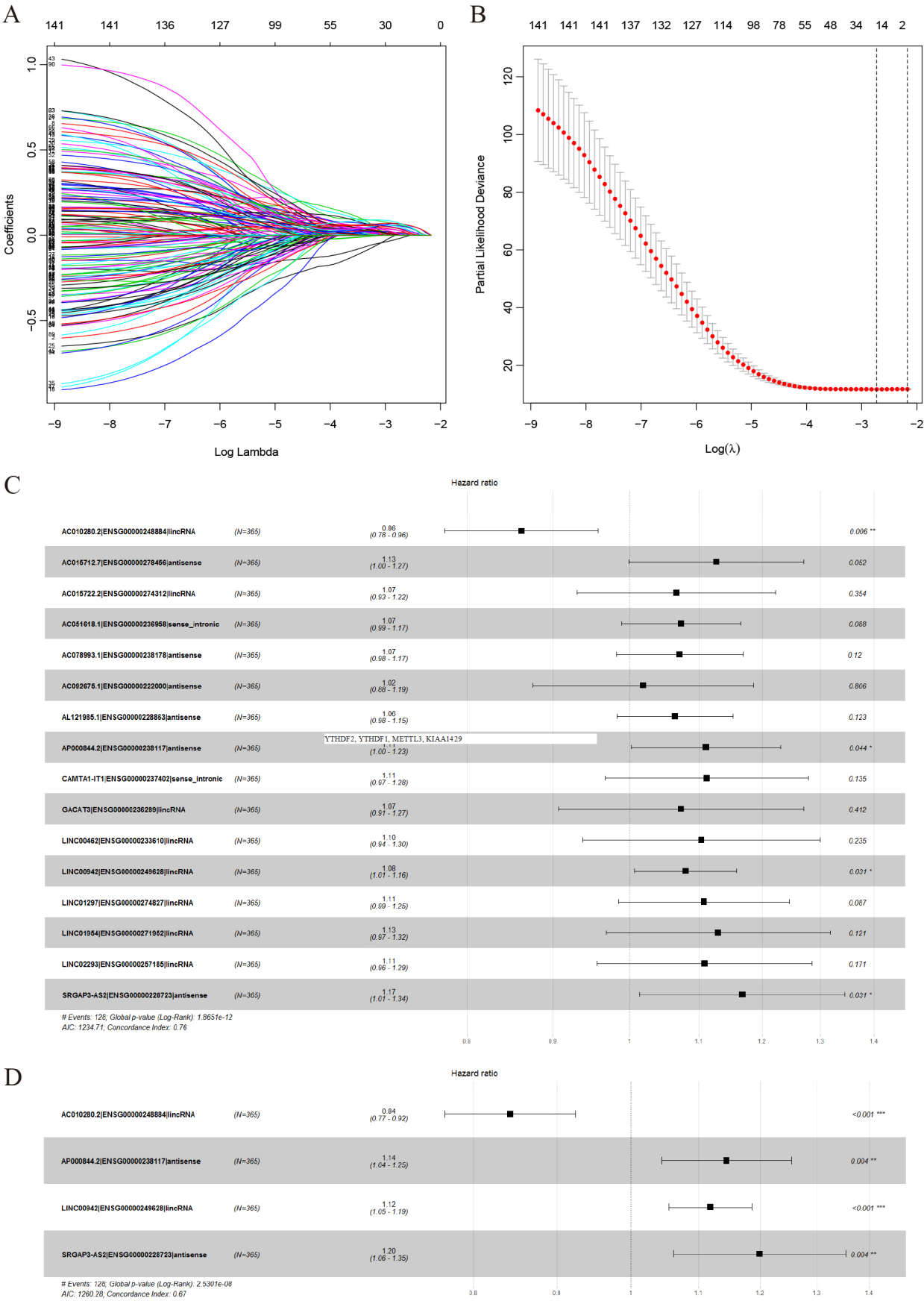


Fig. 2. Long non-coding RNAs (lncRNAs) screened by least absolute shrinkage and selection operator (LASSO) and Cox regression analyses. A. LASSO coefficient values of the 4 prognosis-related lncRNAs in hepatocellular carcinoma (HCC) cohort; B. L1-penalty of LASSO-Cox regression. The hatched vertical lines are at optimal log (lambda) value; C. Forest plot demonstrating the correlations between the 16 lncRNAs and survival; D. Forest plot showing the correlations between the 4 lncRNAs and survival

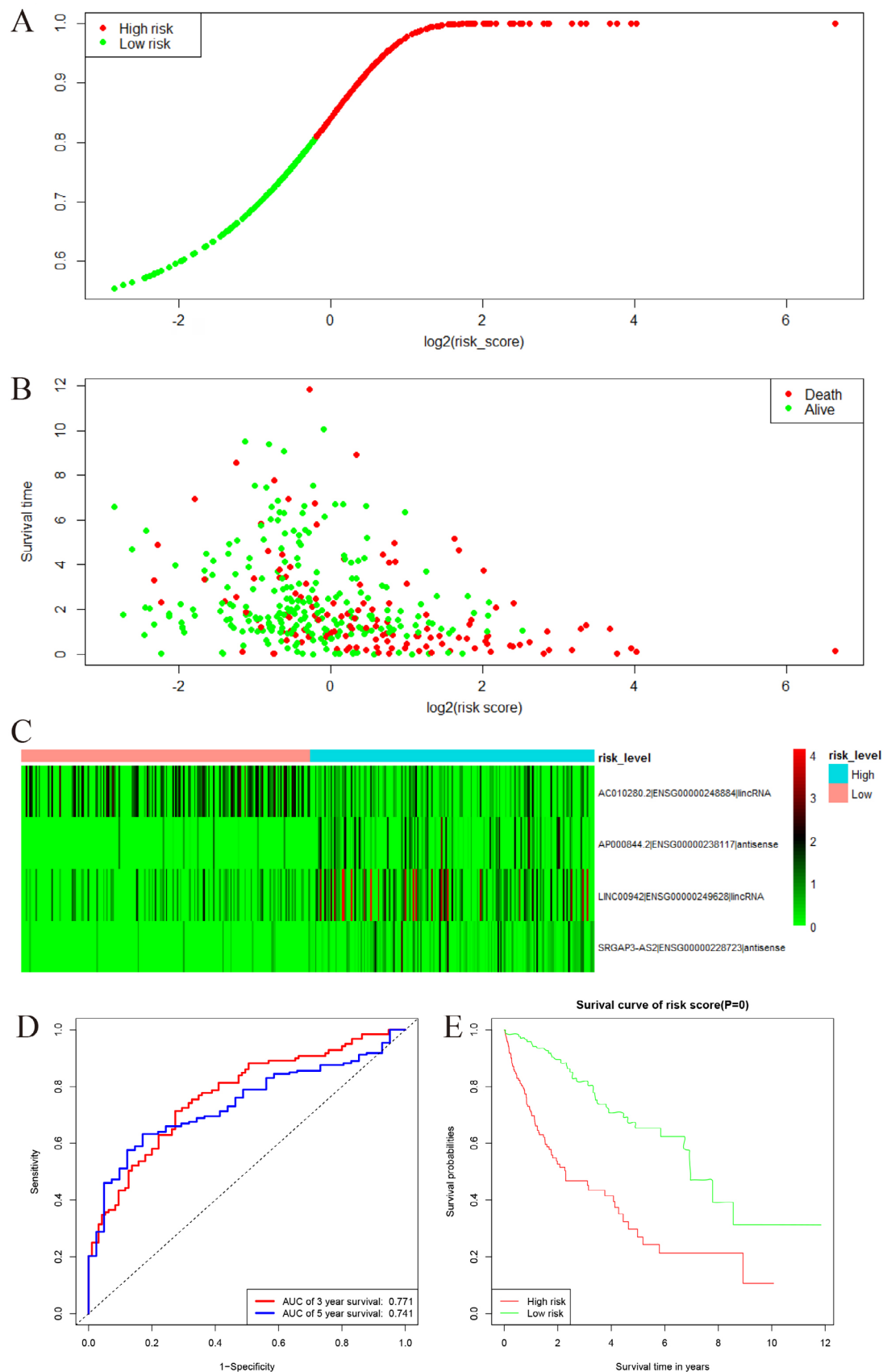


Fig. 3. Characteristics of the combination of 4 long non-coding RNAs (lncRNAs) in The Cancer Genome Atlas (TCGA) cohort. A. A cumulative distribution function (CDF) map built based on the risk score of each sample (the low-risk group: green; the high-risk group: red); B. Survival time in years of each sample (red: deceased; green: alive); C. The heatmap of the 4 lncRNAs expression (blue: high-risk group; pink: low-risk group); D. Receiver operating characteristic (ROC) curve evaluated the predictive effectiveness of the model; E. The high-risk group in this model had lower overall survival (OS)

AUC – area under the curve.

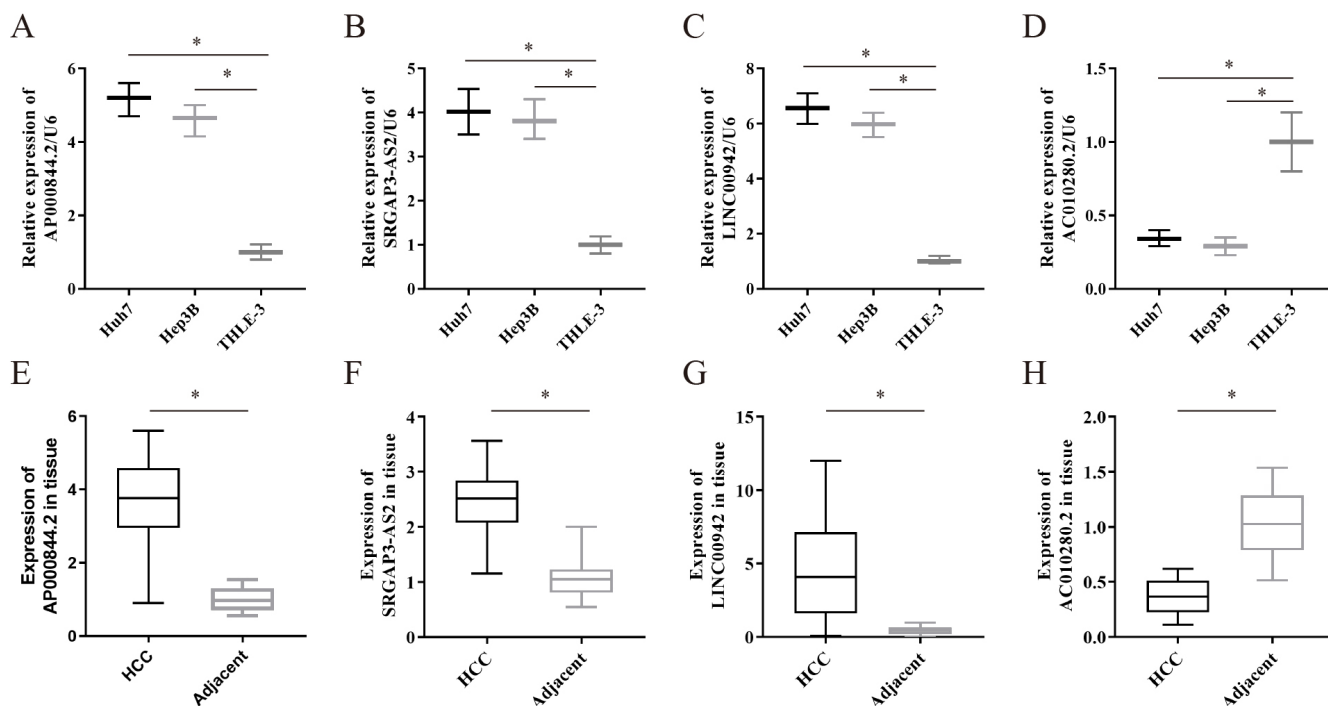


Fig. 4. The expression of AP000844.2, LINC00942, SRGAP3-AS2, and AC010280.2 in hepatocellular carcinoma (HCC) tissues and cells. All 4 long non-coding RNAs (lncRNAs) were overexpressed in both cells and tissues (AP000844.2: A and E; LINC00942: B and F; SRGAP3-AS2: C and G; AC010280.2: D and H)

(Fig. 4D, $p < 0.05$). The RT-PCR results were consistent with the data obtained in the previous bioinformatics analysis. Subsequently, we re-evaluated the expression of the 4 lncRNAs in HCC tissues and matched normal controls in our center ($n = 100$). As expected, AP000844.2, LINC00942 and SRGAP3-AS2 were all overexpressed in HCC tissues when compared to controls. The AC010280.2 was weakly expressed in HCC tissues but overexpressed in normal hepatic tissues (Fig. 4E–H, $p < 0.05$). Consistent results were also obtained for clinical samples and cell samples, which further corroborated the results of our bioinformatics analysis. It is worth emphasizing that LINC00942 showed the most significant difference at both cell and tissue levels.

Verification of the prognostic effect of the 4-lncRNA combination model

The disease-free survival (DFS) ranged from 5 to 90 months, and the OS ranged from 8 to 90 months. We found that 81 out of 100 patients died before the end of the follow-up in the validation set. The expressions of AP000844.2, LINC00942, SRGAP3-AS2, and AC010280.2 were tested in 100 tissue samples (Fig. 4E–H). Then, the correlation between the risk score and survival was assessed. The tumor number, tumor stage, vascular invasion, capsule, distant metastasis, and prediction model all contributed to the poor DFS and OS (Table 2, $p < 0.05$). The high-risk group in the 4-lncRNA combined prediction model demonstrated an unequivocally poor prognosis as evidenced by DFS (25.18 compared to 52.85, $p < 0.01$, Fig. 5A) and OS (31.42 compared to 57.31, $p < 0.01$, Fig. 5B) in the validation

set. The tumor stage and prediction model were independent prognostic risk factors for HCC (Table 3, $p < 0.05$).

Discussion

Hepatocellular carcinoma has high morbidity and mortality rates.¹⁶ Individual treatment and precision medicine are important ways of improving the prognosis. Through the use of genomics, proteomics and transcriptomics, and further development of related technologies, molecular stratification theory has become a powerful tool for the in-depth understanding of tumors. It brings oncology from a discipline that simply describes macro-information, such as size and quantity, to a more in-depth molecular analysis. Furthermore, the discovery of molecular therapies and prognostic biomarkers could bring hope to HCC treatment. Hence, it is necessary to explore novel therapeutic strategies and corresponding molecular targets. The lncRNAs, which have a limited protein-coding ability, play critical roles in cancer progression and metastasis.¹⁷ Clinical prediction tools based on lncRNAs have been rapidly developing, including diagnosis, prognostic biomarkers and potential therapeutic targets.⁸ The lncRNA UPK1A-AS1 promoted HCC development and indicated poor prognosis.¹³ Furthermore, lncRNA CASC9 is a potential diagnostic and prognostic biomarker for HCC.¹⁸ However, these studies only assessed the prognostic value of a single biomarker. Since genes are interactive, a single biomarker may not be enough to accurately predict the prognosis of HCC.

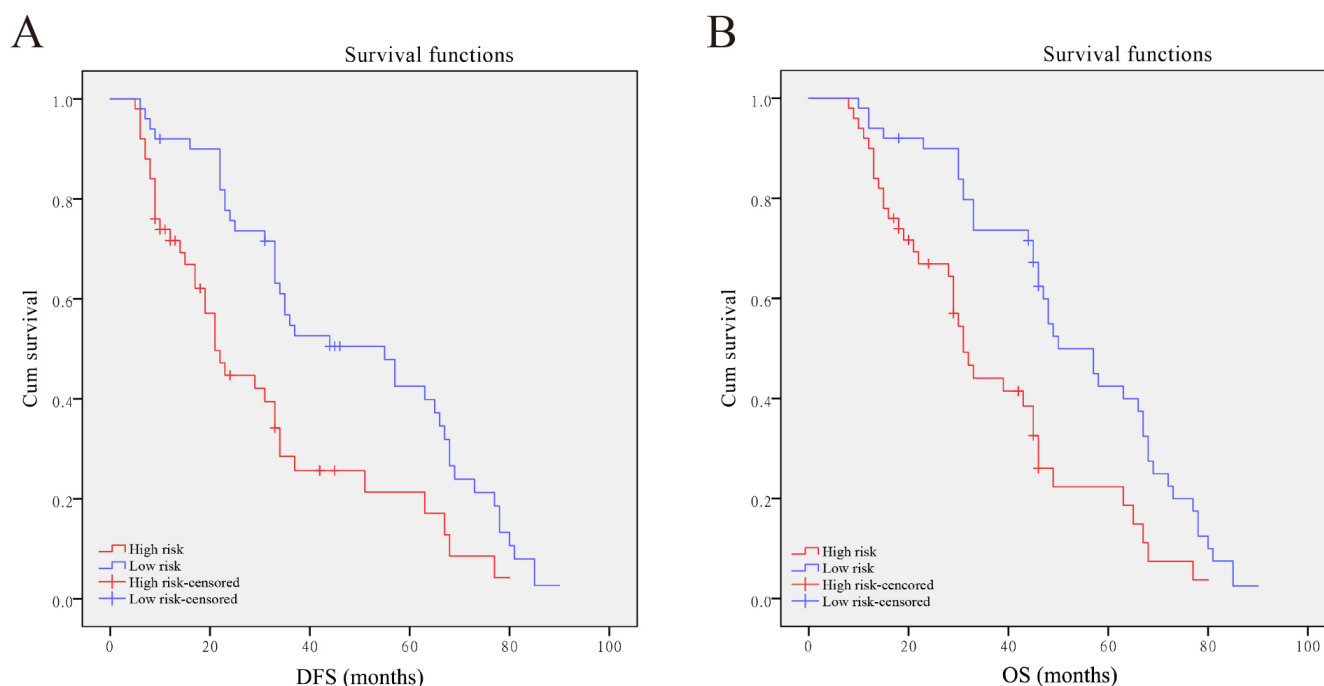


Fig. 5. Kaplan–Meier curves for disease-free survival (DFS) and overall survival (OS). A. DFS curves of 120 hepatocellular carcinoma (HCC) patients stratified by 4-long non-coding RNAs (lncRNAs) expression model ($p = 0.01$); B. OS curves of 120 HCC patients stratified by 4-lncRNA expression model ($p = 0.01$)

In this study, transcriptome and survival data of patients with HCC were obtained from TCGA and used as a training set. In order to improve the predictive accuracy of the regression model, Cox and LASSO regressions were used to evaluate the correlation between the expression of lncRNAs and survival of HCC patients. Finally, a 4-lncRNA combined prediction model was obtained, using lncRNAs AP000844.2, LINC00942, SRGAP3-AS2, and AC010280.2. Clinical data from patients in our hospital were used as a validation set. The AC010280.2 and LINC00942 are long intergenic non-coding RNAs. It has been previously reported that LINC00942 could act as an oncogene that promoted METTL14-mediated m⁶A methylation in breast cancer.¹⁹ The *LINC00942* gene had been recorded to be missing in an autism spectrum disorder patient.²⁰ It had been found that AC010280.2 could participate in establishing the HCC prognosis model.²¹ The AP000844.2 and SRGAP3-AS2 are antisense lncRNAs, and AP000844.2 might be the component of prostate cancer²² and hepatitis virus-positive HCC²³ prognostic models. The SRGAP3-AS2 is also expressed in lung adenocarcinoma²⁴ and could serve as a potential predictive biomarker for that disease.

Both the AUC and C-index demonstrated that our model has good predictive efficacy. The validation set also confirmed the prognostic validity of the model. Although some studies have used LASSO regression to analyze the prognosis of HCC data and obtained a prediction model, the predictive efficacy of this model has not been verified by a validation set.²¹ For the first time, we established a prediction model of HCC-related lncRNAs based on the LASSO analysis and verified its clinical efficacy. The LASSO analysis is a classic regression analysis method

related to statistics and machine learning^{25,26} aimed at improving the prediction accuracy and interpretability through variable selection and regularization compared with other regression methods. The evaluation process of LASSO regression includes the relation with ridge regression, the selection of optimal subset, and the relation between LASSO coefficient estimation and soft threshold.^{27,28} The validation set confirmed that the 4-lncRNA combined prediction model was an independent risk factor for HCC prognosis.

Limitations

This study has 3 limitations. First, it is limited to a single center, and it is still necessary to expand the sample size in order to verify our results in multiple centers. Second, our follow-up research goal was to integrate our model with clinical data to form a comprehensive quantitative index. Finally, the molecular mechanism of each lncRNA needs to be further explored in subsequent experiments.

Conclusions

Through the analysis of public databases and verification of clinical data from our center, we have obtained a 4-lncRNA combined prediction model. The model could effectively evaluate the prognosis of HCC patients. Currently, the TNM staging system is still the most important indicator for evaluating tumor malignancy and prognosis. However, the role of the molecular signature of tumors

Table 2. Patient characteristics and log-rank (Mantel–Cox) analysis

Characteristics	N	DFS			OS		
		month	p-value	F	month	p-value	F
Age							
≥55 <55	57 43	38.36 43.61	0.436	0.606	44.64 47.75	0.417	0.59
Gender							
Male Female	57 43	38.89 42.74	0.654	0.201	44.46 47.70	0.534	0.386
AFP [μg/L]							
<20 ≥20	52 48	46.29 35.36	0.144	2.130	50.15 42.47	0.200	1.644
HbsAg							
Positive Negative	64 36	38.49 44.72	0.316	1.005	44.56 48.68	0.308	1.039
Cirrhosis							
Present Absent	56 44	38.73 42.84	0.645	0.213	43.94 48.25	0.485	0.487
Tumor size							
≥5 cm <5 cm	53 47	34.93 46.12	0.015	5.939	40.42 51.32	0.013	6.123
Tumor number							
Multiple Solitary	43 57	32.07 44.99	0.010	6.718	40.44 48.90	0.010	6.683
Tumor stage							
III–IV I–II	62 38	28.23 56.60	<0.001	26.480	34.49 60.83	<0.001	30.534
Vascular invasion							
Yes No	41 59	32.14 46.55	0.019	5.511	37.71 51.56	0.010	6.556
Capsule							
Absence Presence	36 64	47.73 35.11	0.023	5.199	52.91 40.35	0.008	6.946
Distant metastasis							
Absence Presence	50 50	30.21 49.10	0.001	11.609	36.09 54.07	<0.001	14.364
Model							
High risk Low risk	49 51	25.18 52.85	<0.001	30.760	31.42 57.31	<0.001	33.837

DFS – disease-free survival; OS – overall survival; AFP – alpha-fetoprotein; HbsAg – hepatitis B surface antigen.

Table 3. Multivariate Cox regression analysis of significant prognostic factor for survival in HCC patients


Variables	DFS			OS		
	p-value	HR	95% CI	p-value	HR	95% CI
Tumor number	0.939	0.979	0.572–1.77	0.547	1.189	0.678–2.085
Tumor stage	<0.001	0.330	0.179–0.607	<0.001	0.276	0.145–0.526
Vascular invasion	0.123	0.686	0.424–1.108	0.051	0.619	0.383–1.002
Capsule	0.571	1.169	0.681–2.009	0.385	1.274	0.737–2.202
Distant metastasis	0.085	0.615	0.354–1.069	0.029	0.540	0.311–0.940
Risk score	<0.001	0.359	0.209–0.616	<0.001	0.301	0.172–0.527

DFS – disease-free survival; OS – overall survival; HR – hazard ratio; 95% CI – 95% confidence interval; HCC – hepatocellular carcinoma.

cannot be ignored. For example, Ki-67 indicates proliferation and Her-2 indicates the degree of malignancy. Microsatellite instability (MSI) and tumor mutational burden (TMB) are also beacons for treatment options. We believe that a further improvement of molecular stratification and the application of prognostic markers can provide valuable information for tumor treatment. In addition, even though the expression of molecules still needs to be obtained from tissue samples, if, with the advancement of liquid biopsy technology, we can detect these molecules in the body fluid, their expression changes may indicate tumor recurrence and metastasis.

ORCID iDs

Xiangxiu Qi  <https://orcid.org/0000-0001-7569-165X>

Cui Xu  <https://orcid.org/0000-0002-2251-6150>

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7–30. doi:10.3322/caac.21590
2. Tsuchiya N. Biomarkers for the early diagnosis of hepatocellular carcinoma. *World J Gastroenterol*. 2015;21(37):10573–10583. doi:10.3748/wjg.v21.i37.10573
3. Pillai A, Ahn J, Kulik L. Integrating genomics into clinical practice in hepatocellular carcinoma: The challenges ahead. *Am J Gastroenterol*. 2020;115(12):1960–1969. doi:10.14309/ajg.0000000000000843
4. Trevisan França de Lima L, Broszczak D, Zhang X, Bridle K, Crawford D, Punyadeera C. The use of minimally invasive biomarkers for the diagnosis and prognosis of hepatocellular carcinoma. *Biochim Biophys Acta Rev Cancer*. 2020;1874(2):188451. doi:10.1016/j.bbcan.2020.188451
5. Singal AG, Lok AS, Feng Z, Kanwal F, Parikh ND. Conceptual model for the hepatocellular carcinoma screening continuum: Current status and research agenda. *Clin Gastroenterol Hepatol*. 2022;20(1):9–18. doi:10.1016/j.cgh.2020.09.036
6. Xiao Y, Hu J, Yin W. Systematic identification of non-coding RNAs. In: Li X, Xu J, Xiao Y, Ning S, Zhang Y, eds. *Non-Coding RNAs in Complex Diseases*. Vol 1094. Advances in Experimental Medicine and Biology. Singapore, Singapore: Springer Singapore; 2018:9–18. doi:10.1007/978-981-13-0719-5_2
7. Delaunay S, Frye M. RNA modifications regulating cell fate in cancer. *Nat Cell Biol*. 2019;21(5):552–559. doi:10.1038/s41556-019-0319-0
8. Evans JR, Feng FY, Chinnaiyan AM. The bright side of dark matter: lncRNAs in cancer. *J Clin Invest*. 2016;126(8):2775–2782. doi:10.1172/JCI84421
9. Chu C, Lei X, Li Y, et al. High expression of miR-222-3p in children with *Mycoplasma pneumoniae* pneumonia. *Ital J Pediatr*. 2019;45(1):163. doi:10.1186/s13052-019-0750-7
10. Qian X, Zhao J, Yeung PY, Zhang QC, Kwok CK. Revealing lncRNA structures and interactions by sequencing-based approaches. *Trends Biochem Sci*. 2019;44(1):33–52. doi:10.1016/j.tibs.2018.09.012
11. Peng WX, Koirala P, Mo YY. lncRNA-mediated regulation of cell signaling in cancer. *Oncogene*. 2017;36(41):5661–5667. doi:10.1038/nc.2017.184
12. Xie C, Li SY, Fang JH, Zhu Y, Yang JE. Functional long non-coding RNAs in hepatocellular carcinoma. *Cancer Lett*. 2021;500:281–291. doi:10.1016/j.canlet.2020.10.042
13. Zhang DY, Sun QC, Zou XJ, et al. Long noncoding RNA UPK1A-AS1 indicates poor prognosis of hepatocellular carcinoma and promotes cell proliferation through interaction with EZH2. *J Exp Clin Cancer Res*. 2020;39(1):229. doi:10.1186/s13046-020-01748-y
14. Yi T, Luo H, Qin F, et al. lncRNA LL22NC03-N14H11.1 promoted hepatocellular carcinoma progression through activating MAPK pathway to induce mitochondrial fission. *Cell Death Dis*. 2020;11(10):832. doi:10.1038/s41419-020-2584-z
15. Gao H, Li L, Xiao M, et al. Elevated DKK1 expression is an independent unfavorable prognostic indicator of survival in head and neck squamous cell carcinoma. *Cancer Manag Res*. 2018;10:5083–5089. doi:10.2147/CMAR.S177043
16. Wu Q, Pi L, Le Trinh T, et al. A novel vaccine targeting glypican-3 as a treatment for hepatocellular carcinoma. *Mol Ther*. 2017;25(10):2299–2308. doi:10.1016/j.jymthe.2017.08.005
17. Sahu A, Singhal U, Chinnaiyan AM. Long noncoding RNAs in cancer: From function to translation. *Trends Cancer*. 2015;1(2):93–109. doi:10.1016/j.trecan.2015.08.010
18. Zeng YL, Guo ZY, Su HZ, Zhong FD, Jiang KQ, Yuan GD. Diagnostic and prognostic value of lncRNA cancer susceptibility candidate 9 in hepatocellular carcinoma. *World J Gastroenterol*. 2019;25(48):6902–6915. doi:10.3748/wjg.v25.i48.6902
19. Sun T, Wu Z, Wang X, et al. LNC942 promoting METTL14-mediated m6A methylation in breast cancer cell proliferation and progression. *Oncogene*. 2020;39(31):5358–5372. doi:10.1038/s41388-020-1338-9
20. Silva IMW, Rosenfeld J, Antoniuk SA, Raskin S, Sotomaior VS. A 1.5Mb terminal deletion of 12p associated with autism spectrum disorder. *Gene*. 2014;542(1):83–86. doi:10.1016/j.gene.2014.02.058
21. Li W, Chen QF, Huang T, Wu P, Shen L, Huang ZL. Identification and validation of a prognostic lncRNA signature for hepatocellular carcinoma. *Front Oncol*. 2020;10:780. doi:10.3389/fonc.2020.00780
22. Liu S, Wang W, Zhao Y, Liang K, Huang Y. Identification of potential key genes for pathogenesis and prognosis in prostate cancer by integrated analysis of gene expression profiles and the cancer genome atlas. *Front Oncol*. 2020;10:809. doi:10.3389/fonc.2020.00809
23. Huang ZL, Li W, Chen QF, Wu PH, Shen LJ. Eight key long non-coding RNAs predict hepatitis virus positive hepatocellular carcinoma as prognostic targets. *World J Gastrointest Oncol*. 2019;11(11):983–997. doi:10.4251/wjgo.v11.i11.983
24. Yang Z, Li H, Wang Z, et al. Microarray expression profile of long non-coding RNAs in human lung adenocarcinoma: lncRNA expression in LAD. *Thorac Cancer*. 2018;9(10):1312–1322. doi:10.1111/1759-7714.12845
25. Santosa F, Symes WW. Linear inversion of band-limited reflection seismograms. *SIAM J Sci Stat Comput*. 1986;7(4):1307–1330. doi:10.1137/0907087
26. Tibshirani R. Regression shrinkage and selection via the lasso: A retrospective. *J Royal Statist Soc B Stat Methodol*. 2011;73(3):273–282. doi:10.1111/j.1467-9868.2011.00771.x
27. Tibshirani R. The Lasso method for variable selection in the Cox model. *Statist Med*. 1997;16(4):385–395. doi:10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3
28. Liu Y, Wu W, Hong S, et al. Lasso proteins: Modular design, cellular synthesis, and topological transformation. *Angew Chem Int Ed*. 2020;59(43):19153–19161. doi:10.1002/anie.202006727