Systematic Analysis of the Function and Prognostic Value of RNA-Binding Proteins in Hepatocellular Carcinoma

TONG WU

Capital Medical University Affiliated Beijing Ditan Hospital <https://orcid.org/0000-0001-6760-2418>

Zhiyun Yang

Capital Medical University Affiliated Beijing Ditan Hospital

Yuying Yang

Capital Medical University Affiliated Beijing Ditan Hospital

Yuyong Jiang

Capital Medical University Affiliated Beijing Ditan Hospital

Peipei Meng

Capital Medical University Affiliated Beijing Ditan Hospital

Huimin Liu

Capital Medical University Affiliated Beijing Ditan Hospital

Qiaoli Zhang (✉ zhangqiaoli1009@163.com)

Third Affiliated Hospital, Beijing University of Chinese Medicine <https://orcid.org/0000-0001-5376-2808>

Primary research

Keywords: RNA-binding proteins (RBPs), hepatocellular carcinoma (HCC), prognostic analysis, prediction model

Posted Date: November 25th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-113475/v1>

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

[Read Full License](#)

Abstract

Background: RNA-binding proteins (RBPs) are abnormally expressed in a variety of malignant tumors and are closely related to tumorigenesis, tumor progression, and prognosis. The role of RBPs in hepatocellular carcinoma (HCC) is unclear. Based on the cancer genome atlas (TCGA) database, we conducted a systematic bioinformatics analysis of abnormally expressed RBPs in HCC, with the aim of identifying the prognostic markers and potential therapeutic targets.

Methods: HCC RNA sequencing data downloaded from TCGA database were used to determine the differentially expressed RBPs in liver cancer and normal tissues, followed by performing functional enrichment analysis and visualization of interaction relationships. Univariate and multivariate Cox regression analyses were subsequently used to identify RBPs that were significantly related to the prognosis to construct a prognostic model. The predictive performance of the prognostic model was evaluated by survival analysis and receiver operating characteristic (ROC) curve analysis and verified in the test cohort. Human protein atlas online database was used to verify the expression level of RBPs in the prognostic model.

Results: In total, 82 differentially expressed RBPs were identified, including 55 upregulated and 27 downregulated RBPs. Further functional enrichment and interaction analyses showed that the differentially expressed RBPs were mainly related to regulating of mRNA metabolic process, RNA catabolic, mRNA catabolic process, and macromolecule methylation. Five RBP genes, *LIN28B*, *SMG5*, *PPARGC1A*, *LARP1B*, and *ANG* were identified as prognostic-related genes and used to construct the prognostic model. The predictive ability of the prognostic model was verified in the test cohort. ROC curve analysis showed that the prognostic model had good sensitivity and specificity. Independent prognostic analysis showed that the risk score may be an independent prognostic factor for HCC.

Conclusion: This study constructed a reliable prognostic prediction model by analyzing the differentially expressed RBPs of HCC, facilitating the identification of HCC prognostic biomarkers and therapeutic targets.

Introduction

Hepatocellular carcinoma (HCC) is one of the most common malignant tumors worldwide. A global malignant tumor report in 2018 showed that there were approximately 841,000 new cases of HCC, ranking 6th among new malignant tumors and involved in approximately 782,000 deaths. HCC is the 4th most common cause of cancer-related death, accounting for 46.6% new cases worldwide and 47.1% of global deaths ^[1]. The American Cancer Society estimates that there will be 42,810 new cases and 30,160 deaths of HCC in 2020 ^[2]; therefore, active prevention and treatment are of great significance. Due to the insidious and rapid progress of HCC, only 30% of patients benefit from surgical resection and liver transplantation. Most HCC patients lose the opportunity to have surgery when they are diagnosed and are mainly treated by trans-arterial chemoembolization (TACE), radiofrequency ablation, cryoablation,

molecular targeted therapies, and other adjuvant treatments. However, these therapies have not significantly improved the 5-year survival rate of patients with HCCs [3]. Hence, actively searching for molecular markers and therapeutic targets for predicting the prognosis of patients with primary HCC is important for improving prognosis.

RNA-binding proteins (RBPs) play important roles in post-transcriptional modification. In the human genome, more than 1,500 RBP genes have been identified through whole-genome sequencing. RBPs are a class of proteins interacting with a variety of RNAs such as, ribosomal RNAs, non-coding RNAs (ncRNAs), small nuclear RNAs (snRNAs), microRNAs (miRNAs), messenger RNAs (mRNAs), and transfer RNAs (tRNAs) [4]. In recent years, related studies have shown that abnormal functions of RBPs in tumor tissues due to genome changes, regulation of transcription and post-transcriptional modification, and post-transcriptional modification [5–7], affecting the conversion of mRNA to protein and involved in tumorigenesis and tumor progression [8–9]. For example, RBPs affect the activation of tumor growth-related signaling pathways and the expression of growth-related target genes, leading to tumor proliferation [10–11]. In addition, RBPs promote tumor metastasis by regulating the epithelial–mesenchymal transition and ncRNA [12–13]. Many studies have shown that the expression of RBP in cancer tissues is significantly different from that in tumor-adjacent normal tissues and is closely related to the prognosis of cancer patients [14–16]. Thus, further in-depth studies of RBPs will not only help to reveal the pathogenesis of tumors but will also have significance for identifying tumor treatment targets and prognosis. Survival cancer models based on the expression of RBPs have been established to assess prognosis and identify treatment targets [17–18]. However, no relevant research on HCC is available.

The goal of this study was to identify independent prognostic markers to better guide the clinical treatment of HCC. To this end, we downloaded RNA sequencing and clinical data related to HCC from the cancer genome atlas (TCGA) database, identified RBPs differentially expressed between tumor and normal tissues through bioinformatics analysis, and systematically explored their potential functions and molecular mechanism to construct a prognostic prediction model in order to evaluate the prognosis of HCC patients.

Materials And Methods

Data collection and analysis of differential expression

RNA-seq FPKM data and clinical data of 374 cases of HCC and 50 cases of non-tumor samples were downloaded from the TCGA database (<https://portal.gdc.cancer.gov>; as of June 7, 2020) for analysis. The Limma package of R software was used to perform differential expression analysis. The Wilcoxon signed-rank test was used to identify RBPs that were differentially expressed in tumor tissues and normal tissues [4], with the cut-off values of false discovery rate (FDR) < 0.05 and $|\log_2 \text{FC}| > 1$. The pheatmap software package was used to prepare heat maps in this study. Ethical approval was not required in this study because of our strictly adherence to the publishing guidelines provided by the TCGA database.

Gene ontology and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analyses

To fully analyze the biological functions of these differentially expressed RBPs, the ClusterProfiler package in the R software was used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. GO has three categories: biological process (BP), cellular component (CC), and molecular function (MF). $P < 0.05$ and $FDR < 0.05$ were used as statistically significant thresholds in this study.

Protein–protein interaction network construction and module selecting

To analyze the correlation between differentially expressed RBPs, the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (<https://string-db.org/>) database was used to establish the interaction between differential proteins, The Cytoscape (version 3.6.1) was used for network visualization. The key modules of PPI network were identified by the Molecular Complex Detection (MCODE) plug-in with both MCODE score and node count number > 4 .

Prognostic model construction and verification

Univariate Cox regression analysis was performed on RBPs in the protein–protein interaction (PPI) to identify RBPs related to patient survival. Furthermore, samples with complete clinical information were included as the entire cohort and divided into training and test cohorts. Multivariate Cox regression was performed on survival-related RBPs in the training cohort to construct a prognostic model and to calculate the risk score, thereby evaluating the prognosis of the patients. The formula for calculating the risk score of each sample is as follows:

$$\text{Risk score} = \sum_{i=1}^n (\text{Exp}_i * \beta_i),$$

where β represents the regression coefficient, and Exp represents the gene expression value.

To evaluate and verify the predictive power of the prognostic model, we divided the patients in the training and test cohorts into low-risk and high-risk groups, according to the median risk of the training cohort, and performed Kaplan–Meier survival analysis to compare the overall survival rates of the two groups. Receiver operating characteristic (ROC) curves were used to evaluate the reliabilities of the prognostic model. The ROC curve with the area under the curve (AUC) > 0.6 was considered an acceptable model. Univariate and multivariate Cox regression analyses with risk score and other clinical variable (e.g., age, gender, pathological grade in HCC, tumor stage) were used to determine whether the risk score could be used as an independent prognostic factor. In addition, a nomogram based on the RBPs in the prognostic model were prepared.

Genetic alteration analysis and verification of expression levels

The cBioPortal (<https://www.cbioportal.org/>) was used to analyze the genetic alternation of RBPs in the risk model. The human protein atlas (HPA) online database (<http://www.proteinatlas.org/>) was used to detect the expression level of RBPs in the prognostic model.

Results

Identification of differently expressed RBPs in HCC patients

A flowchart of our study design is shown in Fig. 1. Clinical information and gene expression data of 374 HCC samples and 50 non-tumor tissue samples were obtained from the TCGA database. A total of 1,343 RBPs were collected in this study. The Lamma package of R software was used after the differential expression analysis to identify 82 differentially expressed RBPs ($FDR < 0.05$ and $|\log_2 FC| > 1$), including 55 upregulated RBPs and 27 downregulated RBPs. The expression distribution of these differently expressed RBPs is shown in Fig. 2.

Functional enrichment analysis of the differentially expressed RBPs

To analyze the biological functions and related signaling pathways of the differentially expressed RBPs, GO and KEGG pathway enrichment analyses were performed. The results of GO enrichment analysis showed that the upregulated differentially expressed RBPs in BP were significantly enriched in regulating the mRNA metabolic process, RNA and mRNA catabolic processes, and macromolecule methylation. The downregulated differentially expressed RBPs were significantly enriched in the RNA catabolic process, regulation of translation, nucleic acid phosphodiester bond hydrolysis, and regulation of the cellular amide metabolic process. In MF, the upregulated differentially expressed RBPs were mainly and significantly enriched in catalytic activity, acting on RNA, mRNA 3'-UTR binding, and RNA helicase activity; and the downregulated differentially expressed RBPs were mainly and significantly enriched in ribonuclease activity, nuclease activity, and catalytic activity, acting on RNA. In the results of CC analysis, the upregulated differentially expressed RBPs were mainly enriched in cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, and cytoplasmic stress granule, and the downregulated differentially expressed RBPs were mainly enriched in mRNA cap binding complex, RNA cap binding complex, CCR4-NOT complex, and apical dendrite (Fig. 3A, B).

In addition, the results of KEGG pathway enrichment analysis showed that the upregulated differentially expressed RBPs were significantly enriched in the mRNA surveillance pathway, miRNAs in cancer, RNA transport, and other signaling pathways. The downregulated differentially expressed RBPs were significantly enriched in signal pathways, such as hepatitis C and influenza A (Table 1).

Table 1
KEGG pathway enrichment analysis of aberrantly expressed RBPs.

	Term	Pvalue
Upregulated RBPs	mRNA surveillance pathway	2.26E-06
	MicroRNAs in cancer	0.000796199
	RNA transport	0.000891865
	RNA degradation	0.000925152
	DNA replication	0.003526771
	Cysteine and methionine metabolism	0.006457051
Downregulated RBPs	Hepatitis C	0.001748179
	Influenza A	0.002278063

PPI network construction and module selection

To further study the interaction between these differentially expressed RBPs, the STRING database was used to construct a PPI network, which included 64 nodes and 126 edges, followed by use of the Cytoscape for visualization (Fig. 4A). The screening was performed using the median of two topological features, degree and betweenness to obtain six pivotal genes: *GSPT2*, *DDX39A*, *ELAVL2*, *IGF2BP1*, *BOP1*, and *OASL*. The PPI network was processed through the MCODE plug-in, and two key modules were found (Fig. 4B, C, D). Module 1 contained 11 nodes and 27 edges, and module 2 contained 5 nodes and 9 edges. The results of functional enrichment analysis showed that module 1 was mainly enriched in the defense response to virus, nuclear-transcribed mRNA catabolic process, and RNA catabolic process (Table 2); module 2 was mainly enriched in DNA alkylation, DNA methylation or demethylation and DNA modification (Table 3).

Table 2
Functional enrichment analysis of critical module1 associated with HCC

Category	ID	Term	<i>P</i> value
BP	GO:0051607	defense response to virus	1.40E-07
BP	GO:0060337	type I interferon signaling pathway	2.02E-07
BP	GO:0071357	cellular response to type I interferon	2.02E-07
BP	GO:0034340	response to type I interferon	2.39E-07
BP	GO:0009615	response to virus	6.67E-07
CC	GO:0022625	cytosolic large ribosomal subunit	5.04E-06
CC	GO:0022626	cytosolic ribosome	2.85E-05
CC	GO:0015934	large ribosomal subunit	3.59E-05
CC	GO:0044391	ribosomal subunit	0.000137288
CC	GO:0044445	cytosolic part	0.000297502
MF	GO:0003725	double-stranded RNA binding	1.18E-05
MF	GO:0003735	structural constituent of ribosome	0.000225975
If there were more than five terms in this category, the first five terms selected on the <i>P</i> value.			

Table 3
Functional enrichment analysis of critical module2 associated with HCC

Category	ID	Term	Pvalue
BP	GO:0006305	DNA alkylation	5.47E-07
BP	GO:0006306	DNA methylation	5.47E-07
BP	GO:0044728	DNA methylation or demethylation	1.23E-06
BP	GO:0034968	histone lysine methylation	2.20E-06
BP	GO:0006304	DNA modification	2.32E-06
CC	GO:0098687	chromosomal region	5.36E-05
CC	GO:0000781	chromosome, telomeric region	6.52E-04
CC	GO:0000782	telomere cap complex	3.29E-03
CC	GO:0000783	nuclear telomere cap complex	0.003292637
CC	GO:0045120	pronucleus	0.003798426
MF	GO:0140097	catalytic activity, acting on DNA	1.01E-07
MF	GO:0008757	S-adenosylmethionine-dependent methyltransferase activity	6.51E-06
MF	GO:0008168	methyltransferase activity	1.74E-05
MF	GO:0016741	transferase activity, transferring one-carbon groups	2.02E-05
MF	GO:1990841	promoter-specific chromatin binding	1.66E + 08
If there were more than five terms in this category, the first five terms selected on the <i>P</i> value.			

Prognosis-related RBP selection

This study identified 64 differentially expressed RBPs in the PPI. To study the prognostic significance of these RBPs, univariate Cox regression was performed to obtain 22 prognostic related RBPs: LIN28B, NR0B1, IGF2BP3, CELF3, EXO1, EZH2, RNASEH2A, DNMT3B, BOP1, SMG5, EIF5A2, PRIM1, DNMT1, DHX34, BARD1, SNRPB, PPARGC1A, LARP1B, NSUN6, ANG, RCL1, and AZGP1 (Fig. 5).

Construction and verification of the prognostic model

Multivariate Cox regression analysis on the prognosis-related RBPs in the training cohort was performed to obtain five RBPs for the construction of the prognostic model: LIN28B, SMG5, PPARGC1A, LARP1B, and ANG (Fig. 6).

According to the median risk score, 186 patients in the training cohort were divided into two groups for survival analysis: low-risk and high-risk groups. The results showed that compared with patients in the low-risk group, the survival rate of the patients in the high-risk group was poorer (Fig. 7A). ROC analysis showed that the model could be used to well predict the prognosis of HCC patients. The AUC values of the ROC curves of 1/3/5-year survival rates were 0.718, 0.731, and 0.657, respectively (Fig. 7B). The risk score curve, survival status distribution, and heat map of RBP gene expression of the patients in the high- and low-risk groups are shown in Fig. 7C. To further verify the accuracy of the prognostic model, survival and ROC analyses in the test cohort consisting of 184 patients were performed. The results showed a significant difference between the survival rates of the high- and low-risk groups. The AUC values of the ROC curves of the 1-, 3-, and 5-year survival rates were 0.732, 0.642, and 0.627, respectively, suggesting that the prognostic model had good sensitivity and specificity (Fig. 8).

In addition, univariate and multivariate Cox regression analyses were used to evaluate the prognostic values of the risk score and other clinical features of the prognostic model in the training cohort and test cohort. Univariate Cox regression analysis suggested that the tumor stage and risk score were related to the prognosis of HCC patients. Multivariate Cox regression analysis suggested that the risk score could be used as an independent risk factor (Fig. 9).

Construction of a nomogram based on RBPs in the model

To better predict the survival of HCC patients quantitatively, the gene expression of RBPs in the prognostic model was used to construct a nomogram (Fig. 10). The expression of each RBP in the prognostic model corresponded to 1 point. The total points of the patients were calculated by adding all points and were subsequently compared with each prognostic axis to obtain the survival rates of the patient at 1, 2, and 3 years.

Genetic alteration analysis and verification of expression levels

The cBioPortal was used to analyze genetic alternation of genes significantly related to HCC prognosis including *LIN28B*, *SMG5*, *PPARGC1A*, *LARP1B*, and *ANG*. The results showed that alterations occurred in five genes in 72 (20%) of the 366 HCC samples. The alternations were mainly mRNA upregulation (Fig. 11A, B). To verify the differences in the expression of these five genes, immunohistochemical results of the five corresponding genes in HCC were obtained through the HPA database. The results indicated that the expression of *LARP1B* was positive in HCC tissues and negative in normal tissues. *LIN28B* expression was negative in both HCC and normal tissues (Fig. 12).

Discussion

RBPs interact with other proteins or RNA to form ribonucleoprotein complexes, regulate RNA processing, translation, export, and localization, thereby maintaining the stability of the intracellular environment. Abnormal expression levels and changes in activity can lead to various diseases, including tumors^[19–21].

Studies have shown that the post-transcription of RBPs are involved in the tumorigenesis and tumor development. However, their roles in tumors have not been completely revealed [22]. Thus, further research on the differential expression and interaction of RBPs in various tumors may reveal new mechanism for tumor progression and new targets for anti-tumor therapy. HCC is one of the most common malignant tumors of the digestive tract and has a poor prognosis. Hence, the development of effective and early screening and prognostic markers has positive significance for early treatment, intervention, and prognostic judgement.

This study collected the gene expression and clinical information data of 374 HCC samples and 50 non-tumor tissue samples from the TCGA database and identified 82 differentially expressed RBPs, including 55 upregulated and 27 downregulated RBPs, followed by performing systematic analysis of the related biological pathways to construct PPI network of some RBPs. In addition, univariate and multivariate Cox regression analyses of the RBPs were performed to construct a risk model for predicting the prognosis of HCC based on five *RBPs* genes, followed by performing cohort verification based on a test cohort.

The results of enrichment analyses showed that in biological functions, various differentially expressed RBPs were greatly enriched in BP, CC, and MF terms. These differentially expressed RBPs were significantly enriched in mRNA surveillance pathway, miRNAs in cancer, RNA transport, and hepatitis C- and influenza A-related signaling pathways. RBPs interact with miRNAs, mRNAs, long ncRNAs, and cRNAs to form ribonucleoprotein complexes, thereby improving the stability of target genes, promoting gene expression, and playing a key role in the tumorigenesis and progression of many tumors [23–27]. For example, the abnormal expression of RBP-eIF3c promotes the proliferation of HCC, which is positively correlated with KRAS, vascular endothelial growth factor, and Hedgehog signaling pathways [28]. A previous study showed that the effects of RBPs in HCC on biological functions are mainly focused on RNA splicing, translation, transcription termination, RNA localization and transport, RNA surveillance and degradation, RNA modification, ribosome, tRNA, among others [29]. These findings are consistent with the results of this study.

In addition, this study constructed a PPI network of differentially expressed RBPs, including 64 key RBPs to also collect 6 hub proteins including GSPT2, DDX39A, ELAVL2, IGF2BP1, BOP1, and OASL. GSPT2 is highly expressed in HCC, promoting the progression of HCC by affecting cell cycles [30]. DDX39A is upregulated in HCC tissue and cells. High DDX39A expression is positively correlated with advanced clinical stage and DDX39A activates the Wnt/ β -catenin signaling pathway through β -catenin to promote HCC growth, invasion, and metastasis [31]. ELAVL2 activates endogenous proto-oncogenes, causing the progression of a variety of tumors [32] and is involved in tumor resistance to chemotherapy. High ELAVL2 expression may be an independent risk factor for poor chemotherapy response in patients with esophageal squamous cell carcinoma [33]. IGF2BP1 regulates the expression of some important mRNA targets required for tumor cell proliferation, growth, invasion, and chemotherapy resistance and is related to the overall survival rate and metastatic rate of various human cancers [34]. BOP1 regulates the epithelial-mesenchymal transition, leading to the invasion and migration of HCC cells [35] and playing an

important role in the metastasis of colorectal cancer, with mechanism being related to the regulation of Wnt/ β -catenin signaling pathway^[36]. In addition to the core RBPs, a variety of other RBPs play an important role in tumorigenesis and tumor progression. For example, TERT is an important catalytic subunits of telomerase activation. It upregulates the transcription/activity in 80–90% of malignant tumors and is closely related to cell proliferation, tumor invasion, and transformation^[37]. *TRIM71* promotes the proliferation of non-small cell lung cancer (NSCLC) cells by inhibiting the kappa B/nuclear factor kappa B pathway. Upregulated expression of *TRIM71* is related to the tumor size, lymph node metastasis, tumor-node-metastasis staining, and poor prognosis of NSCLC^[38]. *NR0B1* protein is detected in more than 50% of the human lung adenocarcinoma tissues and is highly expressed in the poorly differentiated cancer tissues in male^[39]. It promotes HCC proliferation, invasion, and metastasis by regulating the Wnt/ β -catenin signaling pathway^[40]. By analyzing the key modules in the PPI network, this study showed that these modules were mainly related to defense response to virus, nuclear-transcribed mRNA catabolic process, and DNA modification.

Subsequently, univariate Cox regression analysis was used to obtain 22 prognosis-related RBPs in this study. The multivariate Cox regression analysis of the TCGA training cohort was performed to obtain a construction of 5-RBP (i.e., LIN28B, SMG5, PPARGC1A, LARP1B, and ANG) prognostic risk model. LIN28B is highly expressed in liver cells and is related to the level of alpha-fetoprotein, which promotes the tumorigenesis and progression of HCC. The overall survival rate of HCC patients with high LIN28B expression is significantly shortened, which is related to the multidrug resistance of HCC^[41]. SMG5 is highly expressed in malignant tumors, especially prostate cancer and it is a potential molecular marker for the early diagnosis of cancers^[42]. A survey of HCC in the Han population in eastern China has shown that PPARGC1A is associated with the risk of HCC^[43]. LARP1B belongs to one of the La ribonucleoprotein 1, translational Regulator (*LARP1*) genes. A study showed that high LARP1 protein levels in HCC tissue increased the risk of patient's death by approximately 35% (compared to the low protein levels), and are related to tumor size, survival time, and the Child-Pugh score^[44]. However, the existing research mostly focuses on LARP1A, and only few studies in LARP1B are available. One report has showed that ANG-2 is highly expressed in HCC tissues. Related survival analysis has also shown that ANG-2 is a useful tumor marker for differential diagnosis of patients with HCC from those with chronic liver diseases or healthy subjects and prognosis prediction of HCC^[45].

Further verification of the reliability and stability of the model showed that our model accurately predicted the prognosis of HCC patients. ROC curve analysis also indicated that the prognostic model had good diagnostic capabilities to identify the HCC patients with poor prognosis. Subsequently, survival analysis and ROC curve analysis in the TCGA test cohort showed that the results supported the above conclusion, and our model independently predict the prognosis of the patients with HCC. A nomogram based on this model was constructed to help us more intuitively predict the 1-, 3-, and 5-year prognosis of the HCC patients. The HPA database was used to verify the expression of 5 RBPs in immunohistochemistry. The results showed that the expression of LARP1B was positive in tumor tissues and negative in normal tissues. Consistent with the results of this study, LIN28B was negative expressed in the normal and tumor

tissues. No relevant data on SMG5, PPARGC1A, and ANG were available, which may be related to the limitations of the data included in the database. In brief, our prognostic prediction model was relatively reliable and could be used to identify HCC patients with poor prognosis, which is conducive to the early intervention and treatment of the patients.

This study systematically explored the expression and prognostic value of differentially expressed RBPs in HCC through a series of bioinformatics analyses. This study also constructed a prognostic model based on 5-RBP encoding genes that better predicted the prognosis of the patients with HCC and helped physicians to make clinical decision. However, this study still has some limitations. First, the prognostic model constructed in this study was only based on the TCGA database. It needs to be further verified by a large cohort of the patients. Second, *in vitro* and *in vivo* experiments are needed to further reveal the mechanism of the selected RBPs' action in HCC.

Conclusion

This study conducted a systematic analysis of the key role and prognostic value of RBPs in HCC. Through univariate and multivariate Cox regression analyses of the differentially expressed RBPs in tumor tissues and normal tissues, a prognostic model was constructed by identifying five RBPs that were significantly related to the prognosis. Survival analysis and independent prognostic analyses showed that the prognostic model of this study was an independent predictor of the prognosis of HCC. This study helped further reveal the pathogenesis of HCC and had certain guiding significance for the prognostic molecular of HCC and the development of new therapeutic targets for HCC in clinical practice.

Abbreviations

RBPs: RNA-binding proteins; HCC: Hepatocellular carcinoma; TCGA: The Cancer Genome Atlas; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene ontology; ROC: Receiver operating characteristic; AUC: Area under the curve; BP: Biological process; MF: Molecular function; CC: Cellular component.

Declarations

Acknowledgements

We thank LetPub(www.ltpub.com) for its linguistic assistance during the preparation of this manuscript.

Authors'contributions

TW and QZ designed the study;ZY,YY,YJ,PM and HL downloaded and organized the clinical, RNA-binding proteins data;TW performed data analysis and wrote the paper;QZ critically revised the article for essential intellectual content and administrative support.All authors read and approved the final manuscript.

Funding

This work was supported by Beijing Municipal Natural Science Foundation (No.7202122).

Availability of data and materials

The authors declare that the data supporting the findings of this study are available in the TCGA database.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 2018;68(6):394-424.
2. Rebecca L, Siegel MPH, Kimberly D, et al. Cancer statistics, 2020. *CA Cancer J Clin*, 2020;70(1):7-30.
3. Xiao-Tian Zhou, Jia Ding, Hui-Yan, et al. Hedgehog signalling mediates drug resistance through targeting TAP1 in hepatocellular carcinoma. *J Cell Mol Med*, 2020;24(7):4298-4311.
4. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet*. 2014;15:829–45.
5. Sbestyen E, B Singh, B Minana, et al. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res*, 2016;26(6): 732-44.
6. Jiang Y, M Zhang, Y Qian, et al. Rbm24, an RNA-binding protein and a target of p53, regulates p21 expression via mRNA stability. *J Biol Chem*, 2014;289(6):3164-75.
7. Nakka K, N Chaudhary, S Joshi, et al. Nuclear matrix-associated protein SMAR1 regulates alternative splicing via HDAC6-mediated deacetylation of Sam68. *Proc Natl Acad Sci USA*, 2015;112(26):E3374-83.
8. Legrand N, Dixon DA, Sobolewski C. AU-rich element-binding proteins in colorectal cancer. *World J Gastrointest Oncol*. 2019;11:71–90.
9. Chatterji P, Rustgi AK. RNA Binding Proteins in Intestinal Epithelial Biology and Colorectal Cancer. *Trends Mol Med*. 2018; 24:490–506.

10. Huang X,H Zhang,X Guo,et al.Insulin-like growth factor 2 mRNA-binding protein J(IGF2BP1)in cancer.J Hematol Oncol,2018,11(1):88.
11. Palanichamy J K,T M Tran,J M Howard,et al.RNA-binding protein IGF2BP3 targeting of oncogenic transcripts promotes hematopoietic progenitor proliferation.J ClinInvest,2016,126(4): 1495-511.
12. Zhao Q,O L Caballero,I D Davis,et al.Tumor-specific isoform switch of the fibroblast growth factor receptor 2 underlies the mesenchymal and malignant phenotypes of clear cell renal cell carcinomas.Clin Cancer Res,2013,19(9):2460-72.
13. Zheng L,Z Zhang,S Zhang,et al. RNA Binding Protein RNPCI Inhibits Breast Cancer Cell Metastasis via Activating STARD13-Related ceRNA Network.Mol Pharm,2018,15(6): 2123-2132.
14. Hopkins T G,M Mura,H A A1-Ashtal,et al.The RNA-binding protein LARP1 is a post-transcriptional regulator of survival and tumorigenesis in ovarian cancer.Nucleic Acids Res,2016.44(3):1227-46.
15. Wurth L,P Papasaikas,D Olmeda,et al.UNR/CSDEI Drives a Post-transcriptional Program to Promote Melanoma Invasion and Metastasis.Cancer Cell,2016,30(5):694–707.
16. Sanchcz-Jimenez C,M D Ludena,J M Izquierdo.T-cell intracellular antigens function as tumor suppressor genes.Cell Death Dis,2015,6:e1669.
17. Wei Li,Li-Na Gao,Pei-Pei Song,et al.Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma.Aging(Albany NY),2020,12(4):3558-3573.
18. Ke Wang,Ling Li,Liang Fu et al.Integrated Bioinformatics Analysis the Function of RNA Binding Proteins(RBPs) and Their Prognostic Value in Breast Cancer.Front Pharmacol,2019,10:140. doi: 10.3389/fphar.2019.00140.
19. Stefanie Gerstberger,Markus Hafner, Manuel Ascano,et al. Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. Adv Exp Med Biol,2014,825:1-55.
20. Stefanie Gerstberger, Markus Hafner, and Thomas Tuschl. A Census of Human RNA-Binding Proteins. Nature Reviews Genetics,2014,15(12):829-45.
21. Masuda, K., and Kuwano, Y. Diverse roles of RNA–binding proteins in cancer traits and their implications in gastrointestinal cancers.Wiley Interdiscip Rev RNA,2019,10(3):e1520.
22. Bruno Pereira,Marc Billaud,and Raquel Almeida.RNA-Binding Proteins in Cancer: Old Players and New Actors. Trends Cancer,2017,3(7):506-528.
23. Siang DT, Lim YC, Kyaw AM,et al. The RNA-binding protein HuR is a negative regulator in adipogenesis. Nat Commun. 2020,11:213.
24. Jain A, Brown SZ, Thomsett HL,et al. Evaluation of Post-transcriptional Gene Regulation in Pancreatic Cancer Cells: Studying RNA Binding Proteins and Their mRNA Targets. Methods Mol Biol.2019,1882:239-52.
25. Luo M, Li Z, Wang W, Zeng Y, Liu Z, Qiu J. Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression. Cancer Lett 2013,333: 213-21.

26. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. *Cancer Cell*,2016,29: 452-63.
27. Conn SJ, Pillman KA, Toubia J et al. The RNA binding protein quaking regulates formation of circRNAs. *Cell*,2015,160: 1125-34.
28. Tangjian Li,Shengli Li,Di Chen,et al. Transcriptomic analyses of RNA-binding proteins reveal eIF3c promotes cell proliferation in hepatocellular carcinoma.*Cancer Sci*,2017,108(5):877-885.
29. Bin Zhang,Kamesh R Babu,Chun You Lim,et al. A Comprehensive Expression Landscape of RNA-Binding Proteins (RBPs) across 16 Human Cancer Types.*RNA Biol*,2020,17(2):211-226.
30. Man Li,Jian Wang,Lei Yang,et al.eRF3b, a biomarker for hepatocellular carcinoma, influences cell cycle and phosphorylation status of 4E-BP1.*PLoS One*,2014,9(1):e86371.
31. Tong Zhang,Zhenjiang Ma,Lijuan Liu,et al.DDX39 promotes hepatocellular carcinoma growth and metastasis through activating Wnt/ β -catenin pathway.*Cell Death Dis*,2018,9(6):675.
32. Tomoyuki Hatanaka,Fumihiko Higashino,Kanchu Tei,et al.The neural ELAVL protein HuB enhances endogenous proto-oncogene activation.*Biochem Biophys Res Commun*,2019,517(2):330-337.
33. Wen-Si Zhao,Wan-Pu Yan,Dong-Bo Chen,et al.Genome-scale CRISPR activation screening identifies a role of ELAVL2-CDKN1A axis in paclitaxel resistance in esophageal squamous cell carcinoma.*Am J Cancer Res*,2019,9(6):1183-1200.
34. Xinwei Huang,Hong Zhang,Xiaoran Guo,et al.Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1) in cancer.2018,11(1):88.
35. Kit-Ying Chung,Ibis K-C Cheng,Arthur K-K Ching,et al.Block of proliferation 1 (BOP1) plays an oncogenic role in hepatocellular carcinoma by promoting epithelial-to-mesenchymal transition.*Hepatology*,2011,54(1):307-18.
36. Qi J, Yu Y, Akilli Öztürk Ö,et al. New Wnt/ β -catenin target genes promote experimental metastasis and migration of colorectal cancer cells through different signals. *Gut*.2016; 65:1690-701.
37. Qi J, Yu Y, Akilli Öztürk Ö,et al. New Wnt/ β -catenin target genes promote experimental metastasis and migration of colorectal cancer cells through different signals. *Gut*.2016; 65:1690-701.
38. Hongjiu Ren,Yitong Xu,Qiongzi Wang,et al.E3 ubiquitin ligase tripartite motif-containing 71 promotes the proliferation of non-small cell lung cancer through the inhibitor of kappaB- α /nuclear factor kappaB pathway.*Oncotarget*,2017,9(13):10880-10890.
39. Yongjie Lu,Yunqiang Liu,Shunyao Liao,et al.Epigenetic modifications promote the expression of the orphan nuclear receptor NR0B1 in human lung adenocarcinoma cells.*Oncotarget*,2016,7(28):43162-43176.
40. Jiang HL, Xu D, Yu H, et al. DAX-1 inhibits hepatocellular carcinoma proliferation by inhibiting β -catenin transcriptional activity. *Cell Physiol Biochem*. 2014,34:734-742.
41. Nan Tian,Wenbing Shangguan,Zuolin Zhou,et al.Lin28b is involved in curcumin-reversed paclitaxel chemoresistance and associated with poor prognosis in hepatocellular carcinoma.*J Cancer*,2019,10(24):6074-6078.

42. Kalpana Kannan,Liguo Wang,Jianghua Wang,et al.Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing.Proc Natl Acad Sci U S A,2011,108(22):9172-9177.
43. Sheng Zhang, Jiakai Jiang,Zhan Chen,et al.Relationship of PPARG, PPARGC1A, and PPARGC1Bpolymorphisms with susceptibility to hepatocellular carcinoma in an eastern Chinese Han population.Onco Targets Ther,2018,11:4651-4660.
44. Chara Stavraka,Sarah,Blagden.The La-Related Proteins, a Family with Connections to Cancer.Biomolecules,2015,5(4):2701-22.
45. Yuxin Chen,Yanping Wu,Xiao Zhang,et al.Angiopoietin-2 (Ang-2) is a useful serum tumor marker for liver cancer in the Chinese population.Clin Chim Acta,2018,478:28-27.

Figures

Figure 1

Whole procedures for analyzing RBPs in hepatocellular carcinoma.

Figure 1

Whole procedures for analyzing RBPs in hepatocellular carcinoma.

Figure 2

Differentially expressed RBPs in hepatocellular carcinoma. (A) Heat map; (B) Volcano plot.

Figure 2

Differentially expressed RBPs in hepatocellular carcinoma. (A) Heat map; (B) Volcano plot.

Figure 3

GO enrichment analysis of aberrantly expressed RBPs. (A) GO enrichment analysis of upregulated expressed RBPs; (B) GO enrichment analysis of downregulated expressed RBPs.

Figure 3

GO enrichment analysis of aberrantly expressed RBPs. (A) GO enrichment analysis of upregulated expressed RBPs; (B) GO enrichment analysis of downregulated expressed RBPs.

Figure 4

PPI network and modules analysis. (A) PPI network of differentially expressed RBPs; (B) critical module from PPI network; (C) critical module 1 from the PPI network; (D) critical module 2 from PPI network. Green circles: downregulation with a fold change of more than 2, red circles: up-regulation with fold change of more than 2.

Figure 4

PPI network and modules analysis. (A) PPI network of differentially expressed RBPs; (B) critical module from PPI network; (C) critical module 1 from the PPI network; (D) critical module 2 from PPI network. Green circles: downregulation with a fold change of more than 2, red circles: up-regulation with fold change of more than 2.

Figure 5

Univariate Cox regression analysis for identification of RBPs in the training dataset.

Figure 5

Univariate Cox regression analysis for identification of RBPs in the training dataset.

Figure 6

Multivariate Cox regression analysis to identify prognosis-related hub RBPs.

Figure 6

Multivariate Cox regression analysis to identify prognosis-related hub RBPs.

Figure 7

Risk score analysis of prognostic model in the TCGA training cohort. (A) Survival curve for low- and high-risk subgroups; (B) ROC curves for forecasting OS based on risk score; (C) Expression heat map, survival status, and risk score distribution.

Figure 7

Risk score analysis of prognostic model in the TCGA training cohort. (A) Survival curve for low- and high-risk subgroups; (B) ROC curves for forecasting OS based on risk score; (C) Expression heat map, survival status, and risk score distribution.

Figure 8

Risk score analysis of prognostic model in the TCGA test cohort. (A) Survival curve for low- and high-risk subgroups; (B) ROC curves for forecasting OS based on risk score; (C) Expression heat map, survival status, and risk score distribution.

Figure 8

Risk score analysis of prognostic model in the TCGA test cohort. (A) Survival curve for low- and high-risk subgroups; (B) ROC curves for forecasting OS based on risk score; (C) Expression heat map, survival status, and risk score distribution.

Figure 9

(A) Univariate Cox regression analysis for prognosis model in the training cohort; (B) Multivariate Cox regression analysis for prognosis model in the training cohort; (C) Univariate Cox regression analysis for prognosis model in the test cohort; (D) Multivariate Cox regression analysis for prognosis model in the test cohort.

Figure 9

(A) Univariate Cox regression analysis for prognosis model in the training cohort; (B) Multivariate Cox regression analysis for prognosis model in the training cohort; (C) Univariate Cox regression analysis for prognosis model in the test cohort; (D) Multivariate Cox regression analysis for prognosis model in the test cohort.

Figure 10

Nomogram for predicting 1-, 3-, and 5-year OS of HCC patients.

Figure 10

Nomogram for predicting 1-, 3-, and 5-year OS of HCC patients.

Figure 11

Five prognosis-related RBPs expression and alteration analysis in HCC. (A) Total alteration frequency; (B) genetic alteration of each gene.

Figure 11

Five prognosis-related RBPs expression and alteration analysis in HCC. (A) Total alteration frequency; (B) genetic alteration of each gene.

Normal

Tumor

(A) LIN28B

(B) LARP1B

Figure 12

Verification of RBPs expression in HCC and normal liver tissue using the HPA database. (A) LIN28B, (B) LARP1B.

Figure 12

Verification of RBPs expression in HCC and normal liver tissue using the HPA database. (A) LIN28B, (B) LARP1B.

Supplementary Files

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

- [clinical.xls](#)
- [clinical.xls](#)