RNA binding protein-based risk score model for prognosis prediction of patients with hepatocellular carcinoma

Shengjun Xu[1], Ning Ding[2], Sheng Zheng[2], Jun Chen[1], Yafeng Wan[1], Sunbin Ling[1], Xiao Xu[3]

[1]Key Laboratory of Integrated Oncology and Intelligent Medicine of Zhejiang Province, Department of Hepatobiliary and Pancreatic Surgery, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310006, China;
[2]Department of Gastroenterology, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310003, China;
[3]Institute of Organ Transplantation, Zhejiang University, Hangzhou, Zhejiang 310003, China.

To the Editor: Liver cancer is estimated to be the sixth most lethal tumor, representing approximately 5.0% of all cancer-related deaths.[1] Mainly due to the high rate of infection with hepatitis B virus, liver cancer remains a serious health concern in China, severely limiting the extension of human life expectancy. In China, the number of new cases and deaths from liver cancer is projected to exceed 400,000 in 2022.[2] Hepatocellular carcinoma (HCC) is the major histological type of liver cancer and is a complex and highly heterogeneous disease. Due to the lack of early symptoms, patients with HCC are usually diagnosed at an advanced stage, thus losing the opportunity for surgical resection or liver transplantation. Tumor recurrence and metastasis are major bafflements that limit the efficiency of surgical therapeutics. It is urgent to explore the molecular mechanism of HCC to develop effective methods for early diagnosis and identify patients at high risk of tumor recurrence and metastasis.

RNA binding proteins (RBPs) have been reported to participate in the tumorigenesis and progression of various cancers and have gained increasing attention. A total of 1542 RBPs have been experimentally validated in humans, representing 7.5% of the protein-coding genes.[3] RBPs can dynamically interact with proteins and coding and non-coding RNAs to form ribonucleoprotein complexes, which have crucial functions in post-transcriptional progress and influence the fate of the RNAs.[4] However, there is a lack of systematic studies analyzing the role of RBPs in HCC. Hence, a systematic analysis was executed to explore the role of RBPs in HCC, with the schema illustrated in Supplementary Figure 1, http://links.lww.com/CM9/B124.

First, RNA sequencing and clinical information from 374 HCC samples and 50 normal liver tissues were downloaded from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/). Eighty-two differentially expressed RBPs between normal and HCC samples were identified using the limma package in R software, with a false discovery rate <0.05 and $|\log_2 FC| > 1.0$ [Figure 1A and 1B; Supplementary Tables 1 and 2, http://links.lww.com/CM9/B124]. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyzes were performed to identify possible signal transduction pathways involved in the differentially expressed RBPs [Supplementary Table 3, http://links.lww.com/CM9/B124]. GO enrichment terms are shown in Supplementary Figure 2A, http://links.lww.com/CM9/B124, which shows the enrichment in processes associated with the RNA catabolic process and regulation of translation. KEGG pathway analysis showed that upregulated RBPs were mainly enriched for the mRNA surveillance pathway, microRNAs in cancer and RNA transport [Supplementary Figure 2B, http://links.lww.com/CM9/B124 and Supplementary Table 3, http://links.lww.com/CM9/B124]. Downregulated RBPs were associated with hepatitis C and influenza A virus infections [Supplementary Table 3, http://links.lww.com/CM9/B124].

Univariate Cox regression analysis of key RBPs in the training dataset was performed using the R survival package, and the data were visualized using forest plots [Supplementary Figure 3A, http://links.lww.com/CM9/B124]. The least absolute shrinkage and selection operator (LASSO) regression test was then executed to obtain prognosis-related hub RBPs and their coefficients [Supplementary Figure 3B and 3C, http://links.lww.com/CM9/B124]. Subsequently, eight prognosis-related RBPs were identified, including enhancer of zeste 2 polycomb repressive complex 2 subunit (EZH2), SMG5 nonsense mediated mRNA decay factor (SMG5), RNA terminal phosphate cyclase like 1 (RCL1), breast cancer type 1 susceptibility protein (BRCA1) associated RING domain 1 (BARD1), eukaryotic translation initiation factor 5A2

Access this article online

Quick Response Code:

Website: www.cmj.org

DOI: 10.1097/CM9.0000000000002232

Shengjun Xu and Ning Ding contributed equally to this work.

Correspondence to: Dr. Xiao Xu, Institute of Organ Transplantation, Zhejiang University, Hangzhou, Zhejiang 310003, China
E-Mail: zjxu@zju.edu.cn

Chinese Medical Journal 2022;135(23)

Received: 22-01-2022; **Online:** 01-08-2022 **Edited by:** Yuanyuan Ji

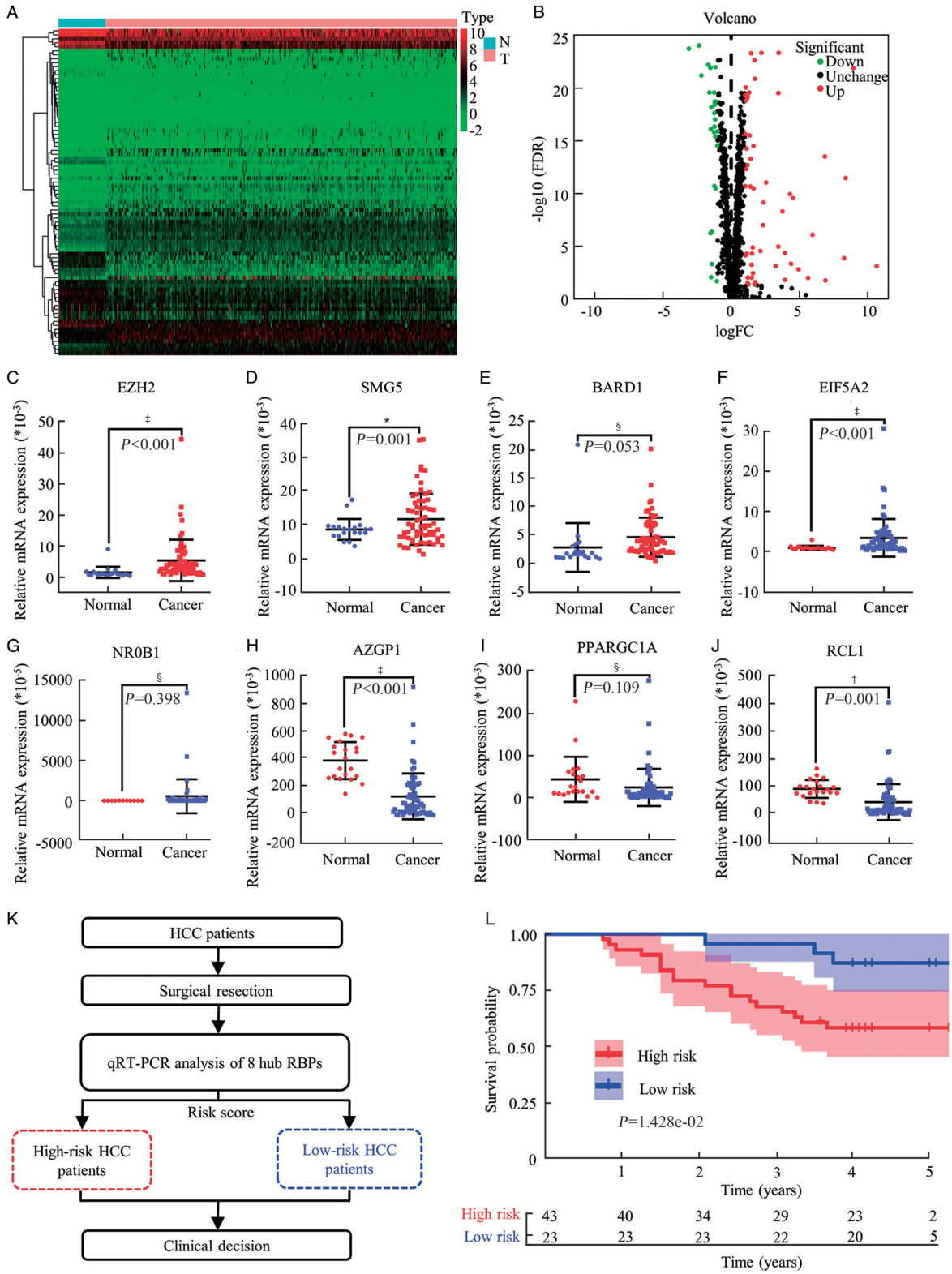

Figure 1: Screen and validation of hub RBPs in HCC. (A, B) Heat map (A) and volcano plot (B) of differentially expressed RBPs between HCC and normal liver tissues. Validation of eight hub RBPs with prognostic significance for patients with HCC. (C–J) Expression of hub RBPs in surgically resected specimens of patients with HCC. Due to the extremely low expression, the mRNA expression of some samples was undetected. For EIF5A2 expression (C), there were two normal liver samples and two cancer samples undetected. For NR0B1 expression (G), there were ten normal liver samples and 19 cancer samples undetected. For PPARGC1A expression (I), there were four cancer samples undetected. (K) Illustration of the validation of eight hub RBPs in the clinical cohort. (L) Survival analysis according to the grouping of risk scores in the clinical cohort. The unpaired two-tailed Student *t* test was applied to assess the statistical significance between the two groups. Statistical significance was defined as $^{*}P < 0.05$, $^{\dagger}P < 0.01$, $^{\ddagger}P < 0.001$, and $^{\S}P > 0.05$. EIF5A2: Eukaryotic translation initiation factor 5A2; HCC: Hepatocellular carcinoma; NR0B1: Nuclear receptor subfamily 0 group B member 1; PPARGC1A: PPARG coactivator 1 alpha; qRT-PCR: Quantitative real-time polymerase chain reaction; RBPs: RNA binding proteins.

(EIF5A2), alpha-2-glycoprotein 1 (AZGP1), PPARG coactivator 1 alpha (PPARGC1A), and nuclear receptor subfamily 0 group B member 1 (NR0B1). Based on these hub RBPs and their corresponding coefficients, a risk score formula was constructed, as shown in Eq. (1):

$$\begin{aligned}\text{Risk score} = {} & [0.221 \times \text{Exp}(EZH2)] + [0.190 \times Exp(SMG5)] \\ & + [-0.0170 \times Exp(RCL1)] + [0.154 \times Exp(BARD1)] \\ & + [0.130 \times Exp(EIF5A2)] + [-0.0148 \times Exp(AZGP1)] \\ & + [-0.151 \times Exp(PPARGC1A)] + [0.222 \times Exp(NR0B1)] \end{aligned} \quad (1)$$

where Exp represents the expression level of the RBPs. The risk score for each patient with HCC was calculated using Eq. (1). Based on the median score of 370 HCC patients in the training cohort, patients were divided into high-risk and low-risk subgroups. Kaplan–Meier analysis revealed that patients in the high-risk subgroup had a significantly lower overall survival (OS) rate than those in the low-risk subgroup [Supplementary Figure 4A and 4C, http://links.lww.com/CM9/B124]. In the time-dependent receiver operating characteristic (ROC) analysis, the area under the curve (AUC) of the ROC curve for OS in the training cohort was 0.730 [Supplementary Figure 4E, http://links.lww.com/CM9/B124]. The results of the validation assay using the GSE14520 dataset from the Gene Expression Omnibus database were consistent with those of the training cohort [Supplementary Figure 4B, 4D, and 4F, http://links.lww.com/CM9/B124]. Survival analysis also showed that the expression of the eight hub RBPs was associated with OS in patients with HCC [Supplementary Figure 5, http://links.lww.com/CM9/B124]. Patients with high expression of EZH2, SMG5, BARD1, EIF5A2, and NR0B1 had a relatively low OS rate, while AZGP1, PPARGC1A, and RCL1 worked as HCC-suppressing RBPs.

Furthermore, eight-gene risk scores, tumor-node-metastasis (TNM) classification, and age of patients with HCC were integrated to establish a nomogram [Supplementary Figure 6A, http://links.lww.com/CM9/B124]. The survival rate of each patient was predicted based on the total points by summing the points for all variables (risk score, tumor stage, and age). Additionally, a time-dependent ROC curve to predict 1-, 3-, and 5-year OS rates is shown in Supplementary Figure 6B, http://links.lww.com/CM9/B124. The AUC values for 1-, 3-, and 5-year OS rates were 0.762, 0.761, and 0.752, respectively, suggesting moderate sensitivity and specificity of the nomogram. Calibration plots were constructed to evaluate the predictive accuracy of the nomogram and it showed high consistency between the predicted and observed outcomes [Supplementary Figure 6C, http://links.lww.com/CM9/B124].

Surgically resected specimens from HCC patients, containing 66 HCC samples and 21 normal liver tissue samples, were collected and quantitative real-time polymerase chain reaction analysis was performed to analyze the mRNA expression of RBPs, with the primers presented in Supplementary Table 4, http://links.lww.com/CM9/B124. As shown in Figure 1C–J, there was significantly upregulated mRNA expression of EZH2, SMG5, and EIF5A2 and downregulated expression of AZGP1 and RCL1, which was consistent with our findings from the TCGA database. There were no statistically significant differences between HCC samples and paracancerous samples for BARD1, NR0B1, and PPARGC1A, probably due to the extremely low expression in both cancer and paracancerous tissues and the limited sample size. According to the mRNA expression derived from the HCC samples and Eq. (1), patients with HCC undergoing surgical resection were divided into high-risk and low-risk subgroups [Figure 1K]. The OS of HCC patients in the low-risk group were significantly better than that of the high-risk group (mean OS, 37.3 ± 10.8 months *vs.* 24.8 ± 11.0 months) [Figure 1L]. There were no significant differences in the baseline characteristics of patients with HCC between the high-risk and low-risk subgroups [Supplementary Table 5, http://links.lww.com/CM9/B124]. For HCC patients with high-risk scores, postoperative treatment should be considered. However, limited by the extremely low expression of some RBPs (EIF5A2, NR0B1, and PPARGC1A) in both cancer and paracancerous tissue and the sample size, the reliability and accuracy of the prognostic nomogram have not been further verified.

In conclusion, we systemically investigated the role of RBPs in HCC and identified eight differently expressed RBPs to construct a risk score model and prognostic nomogram for HCC patients. Our work not only helped clinicians identify HCC patients with poor prognosis after HCC resection, but also discovered some RBPs (SMG5, PPARGC1A, and RCL1), whose biological function in HCC is unclear.

Acknowledgements

The authors express their deepest gratitude to Dr. Kunkai Su for his advice in the data curation.

Funding

This work was supported in part by the grants from the National Science and Technology Major Project of China (No. 2017ZX10203205), the Major Research Plan of the National Natural Science Foundation of China (No. 92159202), Key Program of National Natural Science Foundation of China (No. 81930016), Key Research and Development Plan of Zhejiang Province (No. 2019C03050), and Hangzhou Municipal Social Development Independent Declaration Project (No. 20180533B34).

Conflicts of interest

None.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin 2022;72:7–33. doi: 10.3322/caac.21708.
2. Xia C, Dong X, Li H, Cao M, Sun D, He S, *et al*. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. Chin Med J 2022;135:584–590. doi: 10.1097/cm9.0000000000002108.
3. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. Nat Rev Genet 2014;15:829–845. doi: 10.1038/nrg3813.
4. Pereira B, Billaud M, Almeida R. RNA-binding proteins in cancer: old players and new actors. Trends Cancer 2017;3:506–528. doi: 10.1016/j.trecan.2017.05.003.

How to cite this article: Xu S, Ding N, Zheng S, Chen J, Wan Y, Ling S, Xu X. RNA binding protein-based risk score model for prognosis prediction of patients with hepatocellular carcinoma. Chin Med J 2022;135:2890–2892. doi: 10.1097/CM9.0000000000002232