Observational Study

Medicine®

OPEN

A novel RNA binding protein-associated prognostic model to predict overall survival in hepatocellular carcinoma patients

Ye Liu, MD[a], Xiaohong Liu, MD[b], Yang Gu, MD[b], Haofeng Lu, MD[a,*]

Abstract

Hepatocellular carcinoma (HCC) is 1 of the deadliest malignancies worldwide. Despite significant advances in diagnosis and treatment, the mortality rate from HCC persists at a substantial level. Construction of a prognostic model that can reliably predict HCC patients' overall survival is urgently needed.

Two RNA-seq dataset (the Cancer Genome Atlas and International Cancer Genome Consortium) and 1 microarray dataset (GSE14520) were included in our study. RNA-binding proteins (RBPs) in HCC patients was examined by differentially expressed genes analysis, functional enrichment analysis and protein-protein interaction network analysis. Subsequently, the Cancer Genome Atlas dataset was randomly divided into training and testing cohort with a prognostic model developed in the training cohort. In order to evaluate the prognostic value of the model, a comprehensive survival assessment was conducted.

Five RBPs (ribosomal protein L10-like, enhancer of zeste homolog 2 (EZH2), peroxisome proliferator-activated receptor gamma coactivator 1 alpha (PPARGC1A), zinc finger protein 239, interferon-induced protein with tetratricopeptide repeats 1) were used to construct the model. The model accurately predicted the prognosis of liver cancer patients in both the training cohort and validation cohort. HCC patients could be assigned into a high-risk group and a low-risk group by this model, and the overall survival of these 2 groups was significantly different ($P < .05$). Furthermore, the risk scores obtained by this model were highly correlated with immune cell infiltration.

The prognostic model helps to identify HCC patients at high risk of mortality, which optimizes decision-making for individualized treatment.

Abbreviations: AUC = area under the curve, BP = biological process, EZH2 = enhancer of zeste homolog 2, GEO = gene expression omnibus, HCC = hepatocellular carcinoma, IFIT1 = interferon-induced protein with tetratricopeptide repeats 1, KEGG = Kyodo Encyclopedia of Genes and Genomes, OS = overall survival, PPARGC1A = peroxisome proliferator-activated receptor gamma coactivator 1 alpha, PPI = protein-protein interaction, RPL10L = ribosomal protein L10-like, ROC = receiver operating characteristic, TCGA = the Cancer Genome Atlas, ZNF239 = zinc finger protein 239.

Keywords: hepatocellular carcinoma, overall survival, prognostic model, RNA binding proteins

1. Introduction

Liver cancer is the fourth most prevalent cause of cancer-related mortality, which has the sixth highest incidence in the world.[1] Approximately 90% of liver cancers originated from hepatocellular carcinoma (HCC).[2,3] WHO listed in their scientific report that a wide range of risk factors for tumor development in HCC, such as nonalcoholic fatty liver, hepatitis B and C virus, alcohol consumption and diabetes mellitus.[4] The 5-year survival rate of HCC is 18%, which is the next most deadly malignancy following the pancreatic carcinoma. Moreover, the 5-year survival rate for patients from Asian countries, such as China, has been reported to be as low as about 12%.[5] HCC is prone to recur after hepatectomy, with a 5-year recurrence rate of up to 70%.[6] Despite great efforts to improve the early detection rate and develop new treatment strategies, the prognosis of patients with HCC is still poor, especially in patients with advanced metastatic HCC. At present, the accuracy of histopathological diagnosis in clinical prognosis prediction is insufficient, which limits the treatment of HCC. Therefore, it is critical to develop a high-precision molecular prediction model in the future clinical practice.

It is generally acknowledged and accepted that RNA-binding proteins (RBPs) act to bind RNA via 1 or more spherical RNA-binding domains and modify the destiny or function of the bound RNAs.[7,8] RBPs can act on diverse kinds of RNA types, such as mRNAs, tRNAs, snoRNAs, snRNAs and ncRNAs. It is assumed that roughly 50% of RBPs contribute directly or indirectly to the post-transcriptional modulation in gene expression.[9] RBPs can form different ribonucleoprotein complexes to regulate RNA splicing, localization, stability, translation, polyadenylation, and degradation.[10] Given the critical role of RBPs in post-transcrip-

Editor: Christine Pocha.

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

[a] Department of Hepatobiliary and Pancreatic, The First Affiliated Hospital of Yangtze University, Jingzhou, Hubei, China, [b] Department of General Surgery, Wuhan University Zhongnan Hospital, Wuhan, China.

[*] Correspondence: Haofeng Lu, Department of Hepatobiliary and Pancreatic, The First Affiliated Hospital of Yangtze University, Jingzhou, Hubei, 434000, China (e-mail: 40633000@qq.com).

How to cite this article: Liu Y, Liu X, Gu Y, Lu H. A novel RNA binding protein-associated prognostic model to predict overall survival in hepatocellular carcinoma patients. Medicine 2021;100:29(e26491).

Received: 14 December 2020 / Received in final form: 16 April 2021 / Accepted: 8 June 2021

http://dx.doi.org/10.1097/MD.0000000000026491

tional modulation, it is unsurprising that RBPs are highly associated with a variety of biological functions and diseases.[11] By modulating the mRNAs of many oncogenes, growth factors, and cell cycle modulators, RBPs affect the expression patterns of cancer-related genes, as is well established.[12] Liang et al. reported that overexpression of IFITM3 was associated with poor outcome in HCC cases, and inhibition of IFITM3 could restrain HCC cell proliferation, migrations, invasion and apoptosis.[13] Dong et al. found that RBM3 overexpression indicated that HCC patients had short relapse free survival and poor overall survival (OS).[14]

While several studies have investigated the relationship between RBPs and outcomes in patients with HCC, most have focused on the impact of a single gene on prognosis. In order to explore the important role of RBPs in HCC in a more comprehensive and in-depth way, we developed a reliable prognostic model of prognostic risk to determine the outcome of patients with HCC. Moreover, we examined the association between risk factors (risk scores and risk genes) derived from the model and clinical characteristics.

2. Materials and methods

2.1. Data source

The expression profile data of RNA-binding proteins in HCC patients and the respective clinical data had been obtained from the Cancer Genome Atlas (TCGA) databank (https://portal.gdc.cancer.gov/repository). A RNA-seq dataset (ICGC) and a microarray dataset (GSE14520) used for validation were derived from the International Cancer Genome Consortium (https://dcc.icgc.org/) and the Gene Expression Omnibus (GEO) databank (https://www.ncbi.nlm.nih.gov/geo/), respectively. Patients with a total survival time of less than 1 month and incomplete clinical information were excluded. Immune infiltration data of B cells, CD4+ T cells, CD8+ T cells, neutrophils, dendritic cells and macrophages was derived from the Cistrome project (http://www.cistrome.org/).[15,16] 1542 RBPs summarized by predecessors were included in our study.[17] Our research is based on 3 public databases: TCGA, GEO, and ICGC. Patients included in the database have obtained ethical approval. Users can download the data for free to do research and publish relevant articles. So there are no ethical issues or other conflicts of interest.

2.2. Identification of differentially expressed RBPs

The RNA-seq data was annotated by human gene annotation files (GRCh38.99), which had been obtained from the ENSEMBL (http://asia.ensembl.org/index.html). The expression of mRNAs were analyzed and normalized with the "edge" R package.[18] The differentially expressed mRNAs with $|\log 2$ fold change > 1 and FDR < 0.05 were considered significant. The heat map and the volcano map were drawn with pheatmap package and ggplot2 package.

2.3. Functional enrichment analysis and Protein-protein interaction (PPI) network construction of differentially expressed RBPs

With the Kyodo Encyclopedia of Genes and Genomes (KEGG) pathway assays and GO enrichment analysis, we investigated the possible biological functions of these differentially expressed RBPs. Functional enrichment analysis results were obtained by clusterProfiler package and org.Hs.eg.db package. The enrichment results must meet the requirement that both *P* value was less than .05 to be considered statistically significant. The differently expressed RBPs were uploaded to the String database (https://string-db.org/) to explore their interactions, and the PPI network was subsequently visualized with Cytoscape software.[19,20]

2.4. Establishment of the prognostic risk model

The whole TCGA cohort was randomly divided into a training and a testing cohort. Initially, univariate Cox regression model was performed to select the prognosis-associated RBPs, with *P* value $< .05$. Then, Lasso regression was employed to further screen out the prognostic-associated RBPs and delete prognostic-associated RBPs that correlated highly with 1 another. Finally, Applications of multivariate Cox regression were conducted to develop a model of prognostic risk. Based on the regression coefficient from multivariate Cox regression analysis and mRNA expression level, the prognostic risk model was shown as risk score = (CoefficientRBP1 $\times$ expression of RBP1) + (CoefficientRBP2 $\times$ expression of RBP2) + (CoefficientRBPn $\times$ expression of RBPn).[21,22] The optimal cut-off values for risk scores were identified using the Survminer R package to classify patients into high- and low-risk catalogues. The predicting accuracy of this model was assessed by Kaplan-Meier survival curves and the approach of the time-dependent receiver operating characteristic curve (ROC) analysis. The prognostic risk model was further evaluated by using the distribution of risk scores, the scatter plot of survival status and the expression heat map. Besides, the model of prognostic risk was verified in the testing, TCGA, GSE14520, and ICGC cohort.

2.5. Independent prognostic role of the prognostic risk model

Uni- and multi-variate Cox regression analysess had been adopted to investigated whether the prognostic risk model was independent of other clinical data (age, sex, histological grade, and pathological stage and risk score) for HCC patients. With the clinical characteristics as the independent variable and OS as the dependent variable, the ratio of hazard and 95 per cent confidence interval were determined. *P* value less than .05 was considered to be significant.

2.6. Building and validating a predictive nomogram

The present research constructed a nomogram to predict the 1-year, 3-year, and 5-year OS probability for HCC patients by utilizing independent prognostic factors that were selected by multivariate Cox regression analysis. The accuracy of the nomogram was validated by comparing the prediction probability of the nomogram with the actual observation probability through the calibration curves. The better coincidence with the reference line indicated the higher accuracy of nomogram prediction. ROC curves were also used to evaluate the predictive accuracy of the nomogram.

2.7. Statistical analysis

R software (version 3.6.1) and Perl (version 5.26.3) were used to analyze the RNA expression spectrum and respective clinical information data of HCC patients. The rank correlation between

Table 1

Patients' characteristics in the GEO, TCGA, and ICGC cohorts.

	TCGA (n, %)	GSE14520 (n, %)	ICGC (n, %)
Age			
<=65	209 (65.5%)	200 (90.5%)	88 (38.4%)
>65	110 (34.5%)	21 (9.5%)	141 (61.6%)
Gender			
Female	100 (31.3%)	30 (13.6)	61 (26.6%)
Male	219 (68.7%)	191 (86.4)	168 (73.4%)
TNM stage			
Stage I	160 (50.2%)	93 (42.1%)	36 (15.7%)
Stage II	76 (23.8%)	77 (34.8%)	105 (45.9%)
Stage III	80 (25.1%)	49 (22.2%)	69 (30.1%)
Stage IV	3 (0.9%)	2 (0.9%)	19 (8.3%)
Histological grade			
G1	44 (13.8%)		
G2	154 (48.3%)		
G3	109 (34.2%)		
G4	12 (3.8%)		
Survival status			
Alive	212 (66.5%)	136 (61.5%)	189 (82.5%)
Dead	107 (33.5%)	85 (38.5%)	40 (17.5%)
Median follow-up time (yr)[a]	1.67 (1.00–3.32)	4.36 (1.48–4.82)	2.14 (1.40–3.04)
Risk			
Low	232 (72.7%)	85 (38.5%)	145 (63.3%)
High	87 (27.3%)	136 (61.5%)	84 (36.7%)

GEO = gene expression omnibus, TCGA = the Cancer Genome Atlas.
[a] median values are shown with 25th–75th percentiles in parenthesis.

the risk score and level of immune infiltration was assessed with the Pearson correlation coefficient test, the independent *t*-test was utilized to evaluate the differences between the variables. Statistical significance was identified by $P < .05$.

3. Results

3.1. Screening of differentially expressed RBPs in HCC patients

Patients' characteristics in the GEO, TCGA, and ICGC cohorts were showed in the Table 1, we choose these 3 cohorts because they contain a large number of patients. The expression of 1542 RBPs in HCC patients from TCGA dataset were analyzed. As shown in the heat map and volcano map (Fig. 1A, B), 133 differentially expressed RBPs could be identified in HCC tissues versus normal tissues, including 111 differentially expressed RBPs that were upregulated and 22 that were downregulated.

3.2. Functional enrichment analysis and PPI network establishment of differentially expressed RBPs

In order to reveal possible biological functions of differentially expressed RBPs, we performed GO term and KEGG pathway analysis. There were 247 pathways considering to be significantly enriched, including 152 biological process (BP) terms, 49 cellular component terms, 46 molecular function terms. The most significant BP, cellular component, molecular function concentrate on the regulation of mRNA metabolic process, cytoplasmic ribonucleoprotein granule, catalytic activity and acting on RNA, respectively (Fig. 1C). Meanwhile, KEGG pathway analysis identified 6 significantly enriched pathways: mRNA surveillance pathway, RNA degradation, Spliceosome, RNA transport, Ribosome and Ribosome biogenesis in eukaryotes (Fig. 1D). To further investigate the potential interactions between differentially expressed RBPs, we developed a PPI network based on data from the STRING databank by utilizing software Cytoscape (Fig. 1E). Figure 1F showed the top ten hub genes of the PPI network. The top ten hub genes were PABPC1, PIWIL1, ELAVL2, GSPT2, LIN28A, SNRPE, BOP1, DDX39A, DDX39B, DDX4.

3.3. Identify the prognostic RBPs included in the risk model in the training cohort

Patients with OS time of under 1 month and incomplete clinical data in the entire TCGA cohort were excluded, resulting in the inclusion of 319 patients in model construction. The entire TCGA cohort was split into a testing cohort (n=159) and a training cohort (n=160) randomly. After removing RBPs not in GSE14520, a total of 81 differentially expressed RBPs remained for further study. For the purpose of identifying the prognostic relevance of RBPs, approach of univariate Cox regression was applied to evaluate the expression of these RBPs in the training cohort, yielding 23 RBPs that were associated with prognosis (Fig. 2A). Lasso regression analysis was used to eliminate genes that lead to overfitting of the model and to pick out 10 candidate prognostic-associated RBPs (Fig. 2B). Subsequently, all the candidate RBPs were measured by the approach of multivariate Cox regression assay. In the end, 5 RBPs were identified to establish the prognostic risk model. The 5 RBPs were ribosomal protein L10-like (RPL10L), EZH2, PPARGC1A, zinc finger protein 239 (ZNF239) and interferon-induced protein with tetratricopeptide repeats 1 (IFIT1; Fig. 2C).

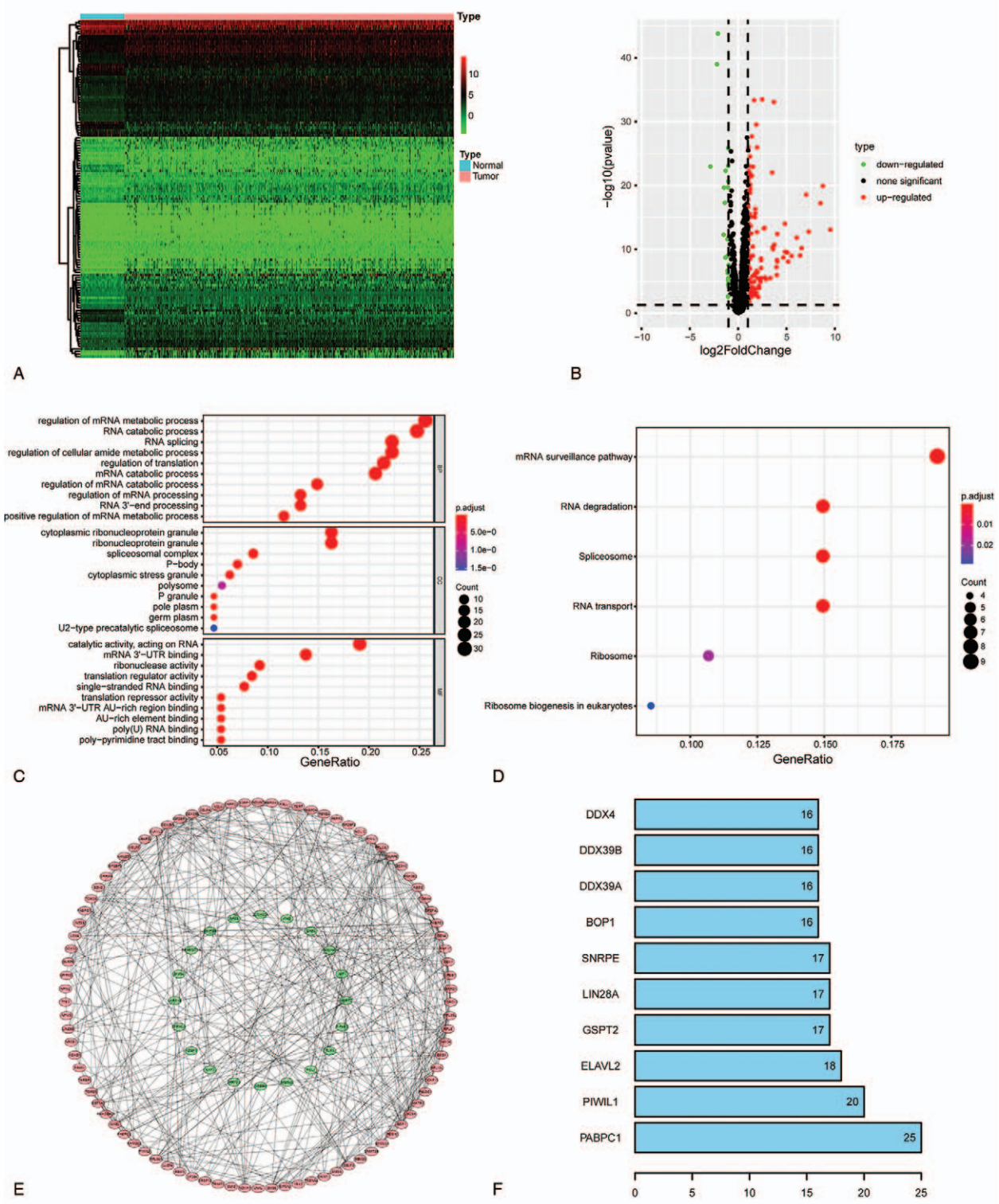

Figure 1. Differentially expressed RBPs for functional enrichment profiling and protein-protein interaction (PPI) network formation. The heat map and volcano plot of differentially expressed RBPs (A, B). GO and KEGG analyses of differentially expressed RBPs (C, D). PPI network of differentially expressed RBPs, the pink nodes represent upregulated differentially expressed RBPs, the green nodes represent downregulated differentially expressed RBPs (E). The top ten hub RBPs (F).

3.4. Establishment of the model for the prognostic risk in the training cohort

With the purpose of investigating the capacity of these 5 prospective RBPs in predicting the outcomes of patients with HCC, these 5 RBPs have been adopted to develop a prognostic model. The risk score was determined by the following method: Risk score = (0.1400 × RPL10L expression) + (0.4536 × EZH2 expression) + (-0.1195 × PPARGC1A expression) + (0.1537 × ZNF239 expression) +

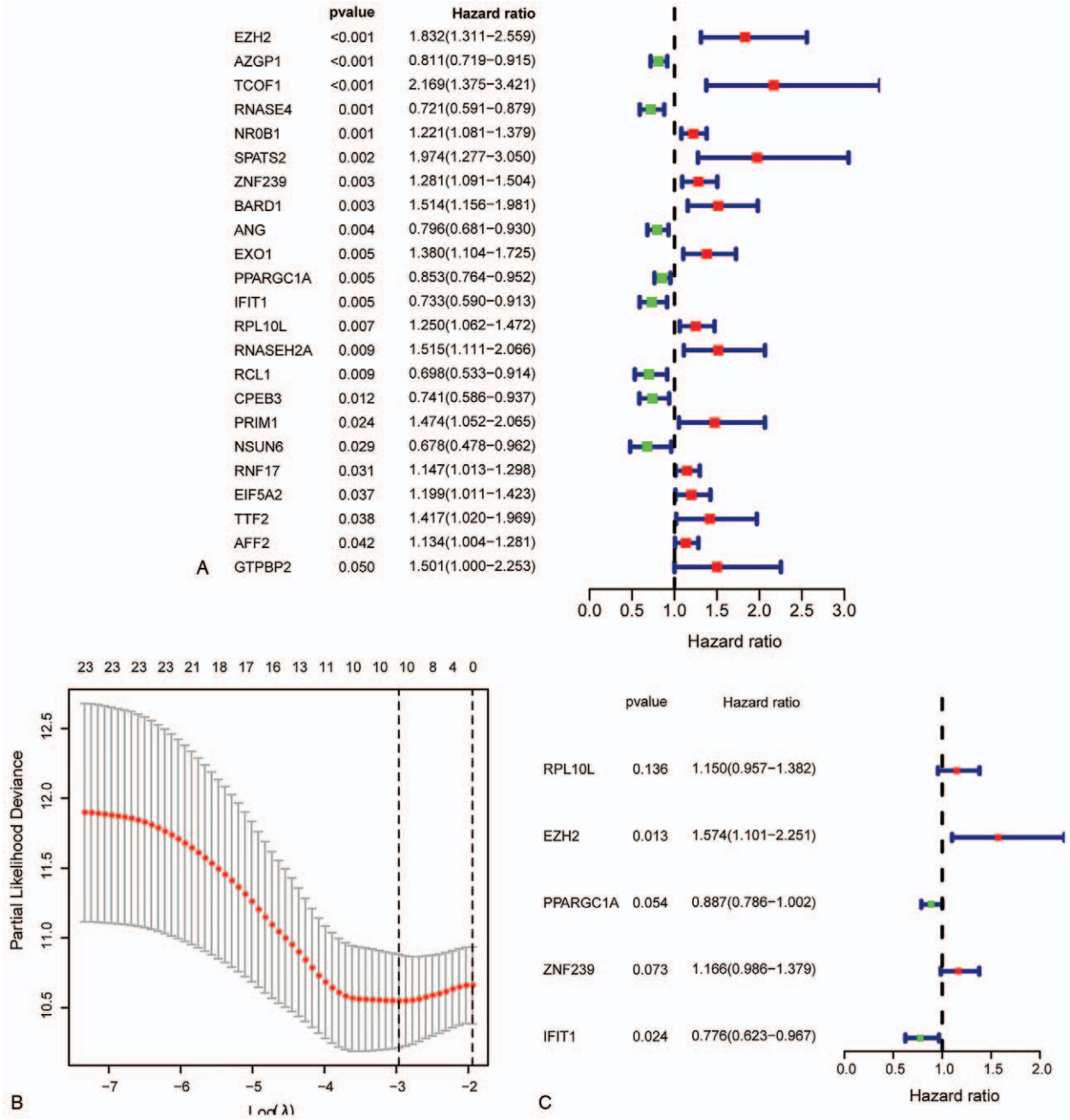

Figure 2. Identification of risk RBPs in the prognostic risk model. Identification of 23 prognostic-associated RBPs through univariate Cox regression analysis (A). Further analysis through Lasso regression analysis (B). Risk gene in the prognostic risk model (C).

(-0.2530 × IFIT1 expression). The optimal cut-off value of -0.142 for the risk score was determined utilizing the Survminer R package. Patients in the training cohort were divided to high-risk (n = 41) and low-risk (n = 119) groups in accordance with cut-off values. In accordance of Kaplan-Meier analysis, the high-risk group had remarkable poorer OS than the low-risk group ($P < .001$; Fig. 3A). The prognostic value of the 5 RBPs characteristics was further assessed with time-dependent ROC analysis. For 1-year survival, 3-year survival, and 5-year survival, the area under the ROC curve (AUC) was 0.763, 0.763, and 0.731, respectively. (Fig. 3B). The risk score analysis for the prognostic risk model in the training cohort was depicted in Figure 3C, which indicated that prognosis for the high-risk group has been shown to be unfavorable in comparison to the low risk group.

3.5. Validation of the prognostic risk model

With the aim of evaluating whether the 5-RBPs related prognostic risk model had parallel predictive value in cohorts of other HCC patients, the same formula was used to determine risk score for the testing cohort, the TCGA full cohort, the GSE14520 cohort and ICGC cohort separately. Individuals in the testing and the entire TCGA cohort were sorted to high- and low-risk groups in accordance with the optimal cut-off value for the training cohort. Through the Survminer R package as the

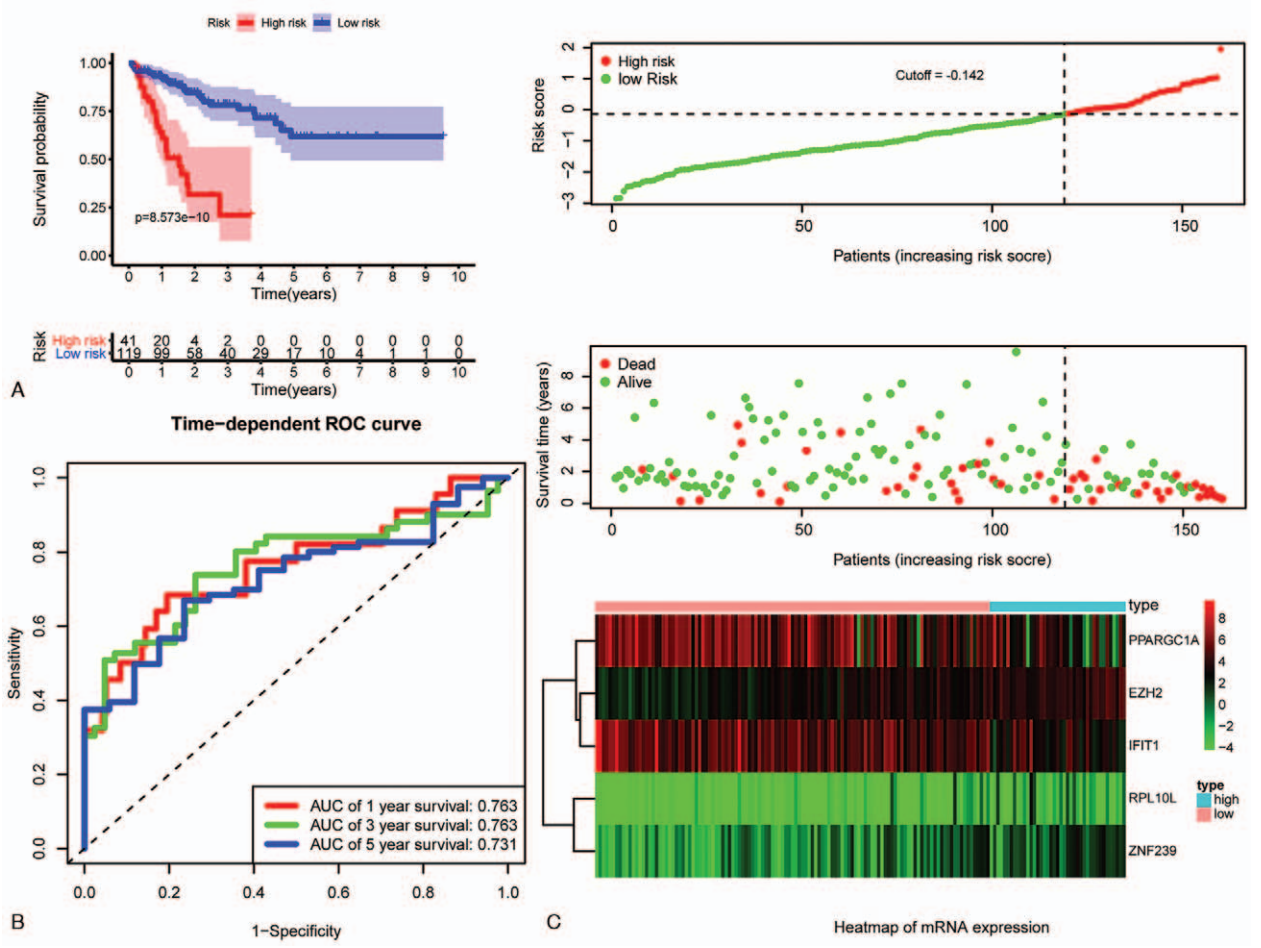

Figure 3. Validation of the risk model in the training cohort. Kaplan–Meier analysis (A), time-dependent ROC analysis (B) and risk score analysis (C) for the prognostic risk model in the training cohort. ROC = receiver operating characteristic.

training cohort, we obtained the optimum cutoff value of 1.111 in the GSE14520 cohort and -0.683 in the ICGC cohort. The 1-, 3-, and 5-year AUC of the testing cohort, TCGA cohort, GSE14520 cohort and ICGC cohort were showed in Figure 4, respectively. Our findings showed that in the testing cohort, TCGA cohort, GSE14520 cohort as well as ICGC cohort, Prognosis for the high-risk group has shown to be unfavorable in comparison to the low risk set, which conforms the results of the training set. The expression levels of EZH2, RPL10L, ZNF239 in the high-risk category were higher compared to the low-risk category, while PPARGC1A, IFIT1 level was reduced compared to the low-risk category (Fig. 4). Taking all into consideration, these findings suggested that our model of prognostic risk could reliably predict the OS of HCC patients.

3.6. Independent prognostic role of the prognostic risk model in the whole TCGA cohort

We evaluated the prognostic value of different clinical variables in the TCGA cohort of HCC patients by uni- and multi-variate Cox regression assay. Both uni- and multivariate Cox regression assays indicated that risk score and pathological stage were independent prognostic role. Among the parameters of age, gender, histological grade and pathological stage and risk score, 1-, 3-, and 5-year AUC of risk score were the largest and the hazard rate of the risk score was also the largest (Fig. 5), which indicated that risk score predicted OS at 1, 3, and 5 years more accurately than other clinical characteristics.

3.7. Constructing a predictive nomogram in the whole TCGA cohort

A nomogram created using 2 independent prognostic factors including pathological stage and risk score was used to predict OS at 1, 3, and 5 years in the whole TCGA cohort (Fig. 6A). ROC curve was used to evaluate the prediction accuracy of nomogram. The area under the ROC curve for 1-,3-, and 5-year was 0.793, 0.786, 0.767 (Fig. 6B), which manifested that the nomogram had very good prediction accuracy. Calibration curve analysis demonstrated that the 1-, 3-, and 5-year survival rates that were predicted by the nomogram corresponded well with the observed survival rates. (Fig. 6C, D, E). Time-dependent ROC analysis was used to assess the accuracy of different nomograms constructed from risk score and clinical features (Fig. 6F, G, H). The predictive ability of the nomogram constructed by age, gender, grade, stage, and risk score and the nomogram constructed by stage and risk score was better than other nomograms, and their predictive abilities were similar. In order to facilitate the evaluation by clinicians, we finally used the nomogram constructed by stage and risk score.

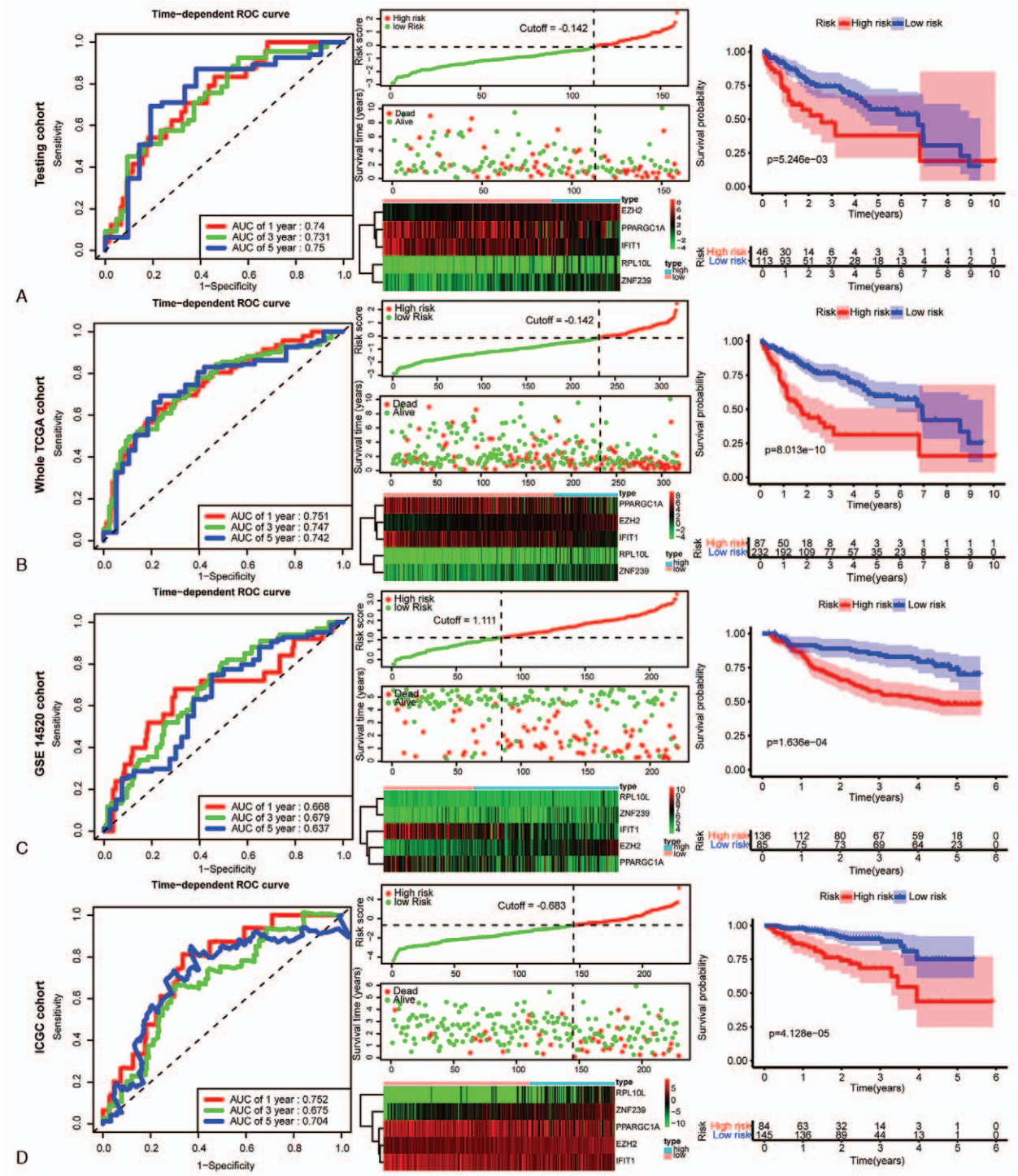

Figure 4. Validation of the risk model in the testing cohort, whole TCGA cohort, GSE14520 cohort and ICGC cohort. Time-dependent ROC analysis, risk score analysis and Kaplan–Meier analysis in the testing cohort (A). Time-dependent ROC analysis, risk score analysis and Kaplan–Meier analysis in the whole TCGA cohort (B). Time-dependent ROC analysis, risk score analysis and Kaplan–Meier analysis in the GSE14520 cohort (C). Time-dependent ROC analysis, risk score analysis and Kaplan–Meier analysis in the ICGC cohort (D). ROC = receiver operating characteristic, TCGA = the Cancer Genome Atlas.

3.8. Clinical utility of the prognostic risk model in the whole TCGA cohort

For the purposes of assessing the clinical utility of the predictive model, the correlation between risk factors (risk score and risk genes) derived from the model and the clinical properties of the entire TCGA cohort was assessed. As shown in Figure 7, compared with low histological grades, high histological grades had significant correlations with higher expression of EZH2, ZNF239, and risk score, and lower expression of IFIT1 and PPARGC1A (all $P < .05$). Compared with low pathological stage,

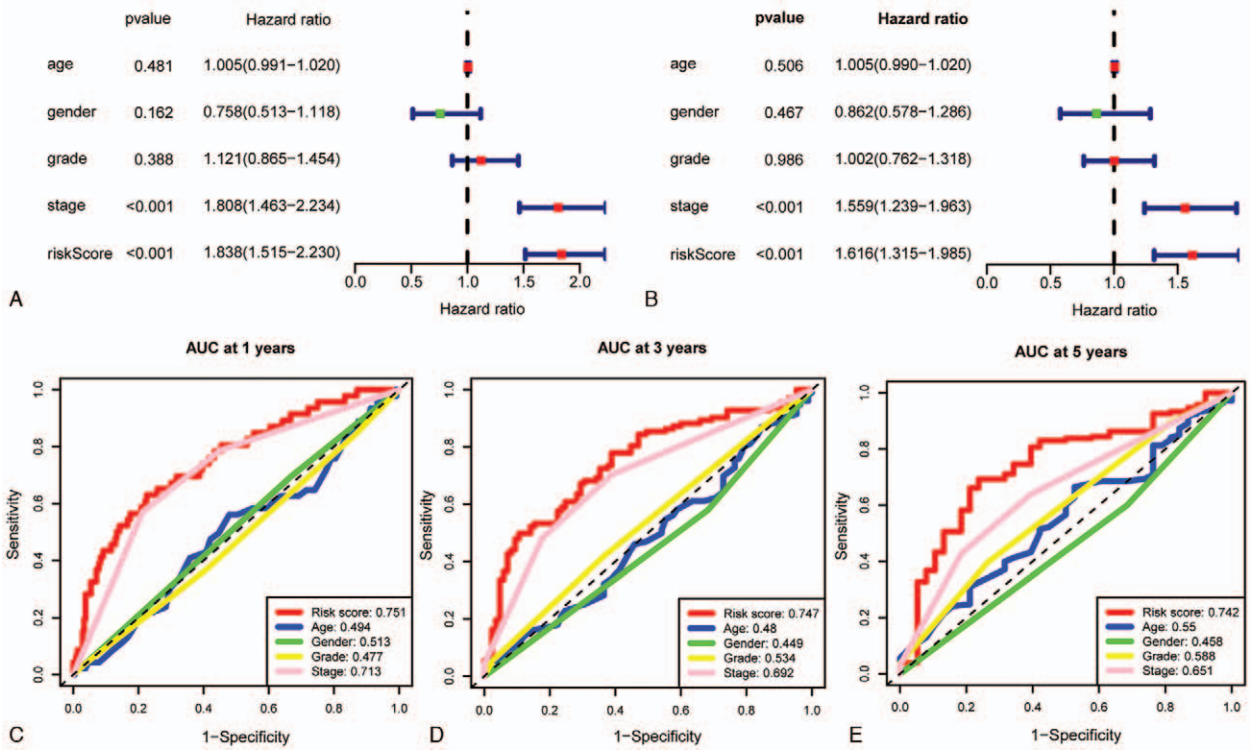

Figure 5. Independent prognostic role of the model of prognostic risk in the whole TCGA cohort. Uni- (A) and multi-variate Cox regression assay (B) of the whole TCGA cohort. Time-dependent ROC analysis of different clinical parameters in the whole TCGA cohort at 1, 3, and 5 years (C, D, E). TCGA = the Cancer Genome Atlas.

high pathological stage had higher values of EZH2 and risk score, and lower values of IFIT1 and PPARGC1A (all $P < 0.05$). The value of IFIT1 for male was higher than that for female, and the value of ZNF239 was lower than female. The expression of RPL10L in elderly patients was greater than that of young patients. These findings demonstrated that these RBP genes were intimately associated with the development of HCC.

Tumor-infiltrating immune cells played a key function in the genesis and progress of tumors.[21] Therefore, the potential association between risk score and level of immune infiltration in the whole TCGA set was analyzed. As shown in Figure 8, risk score had a positive correlation with neutrophil, B cell, CD4/CD8 + T cells, dendritic, and macrophage (all $P < .05$). These findings revealed that the score obtained by our model might reflect the status of the tumor immune cell infiltration in HCC patients.

4. Discussion

Various researches have established that RBPs aberrantly express themselves in diverse human diseases, including those of human malignancies.[13,14] Many RBPs have been verified as key molecules in the occurrence and development of cancer. Abnormal expression of RBPs has also been strongly associated with the outcome of cancer victims. Therefore, RBPs may be essential for the progression and prognosis of HCC.

In our study, a total of 133 differentially expressed RBPs were obtained by comparing liver cancer tissues with normal tissues based on the data from TCGA-LIHC. To investigate the possible biological functions of differentially expressed RBPs, GO term and KEGG pathway analyses were performed. Moreover, to further research the underlying interaction of differentially expressed RBPs, we constructed the PPI network and identified ten hub RBPs. These results will lay a foundation for the future study of the mechanism of HCC occurrence and progression. Next, the 23 prognostic-associated RBPs in the training cohort were obtained by applying univariate Cox regression assay. Lasso- and multi-variate Cox regression analyses were used to screen out paramount RBPs (RPL10L, EZH2, PPARGC1A, ZNF239, IFIT1) from the prognostic-associated RBPs. Subsequently, a prognostic model of RNA-binding proteins was developed in accordance with the regression coefficients from multivariate Cox regression analysis and the mRNA expression status of the 5 RBPs. The Kaplan-Meier survival curves and time-dependent ROC curve analyses revealed that this model possessed superior performance in diagnosis and could screen out the patients with poor prognosis. Besides, we verified the stability and reliability of the model in the testing cohort, whole TCGA-LIHC cohort, GSE14520 cohort and ICGC cohort. It was shown by uni-and multivariate Cox regression assays that our model, as well as pathological stage, could serve as independent predictors of prognosis in HCC patients. Research has shown that pathological stage could independently predict the prognosis of liver cancer patients,[23] which is consistent with our results. Further research found that our model predicted OS at 1-, 3-, and 5 years in liver cancer patients more accurately than other clinical parameters. We also found that risk factors (risk score and risk

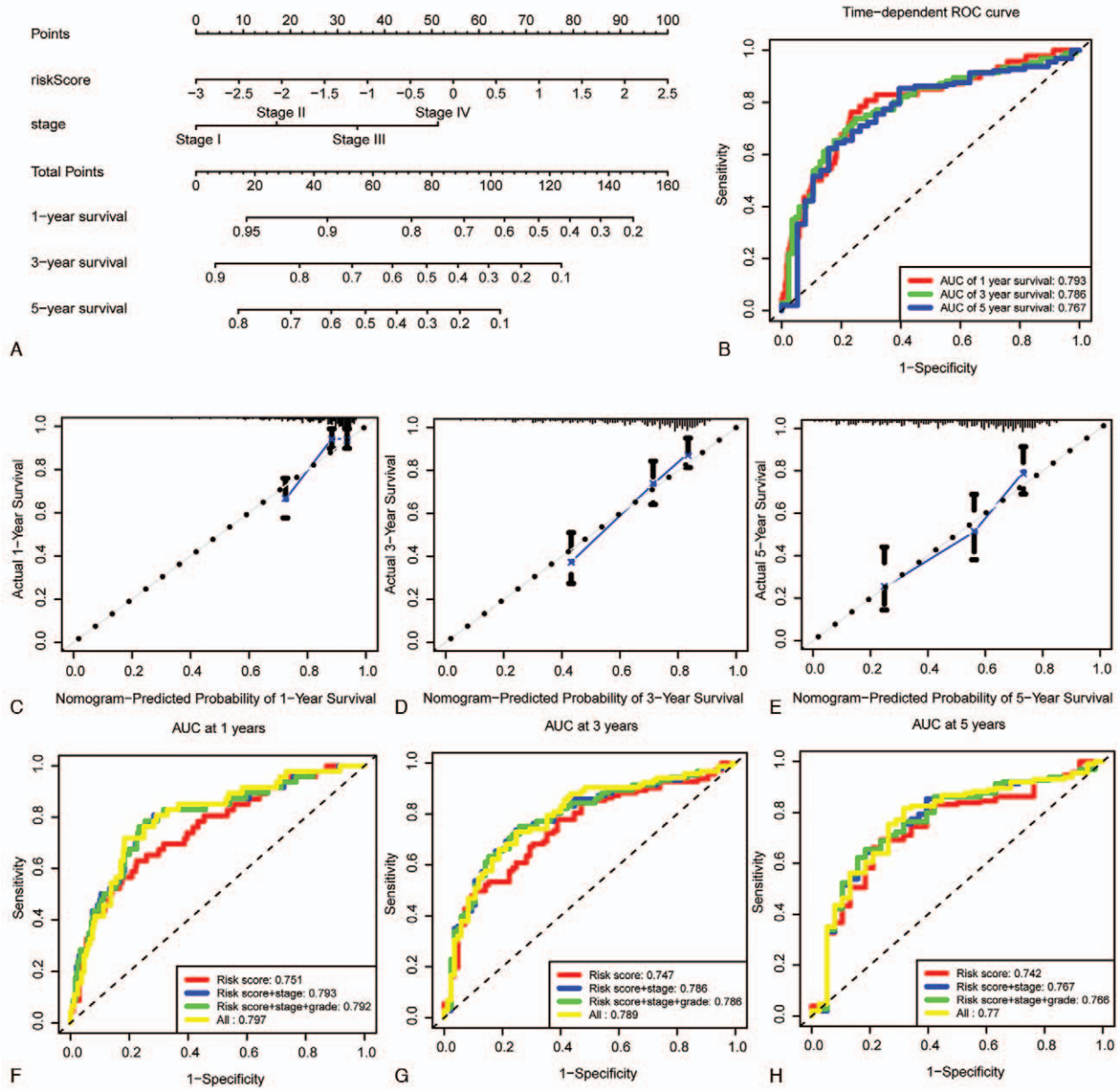

Figure 6. Establishment of a nomogram for predicting overall survival for HCC patients. Nomogram combining risk score with pathological stage (A). Time-dependent ROC analysis of the nomogram (B). The calibration plot for validation of the nomogram (C, D, E). Time-dependent ROC analysis appraises the accuracy of the nomograms. The red, blue, green or yellow line represents the nomogram, the yellow line represents the nomogram constructed by age, gender, grade, stage and risk score (F, G, H).

gene) obtained in the model were closely related to the progress of HCC, which, on the other hand, confirmed that our model had good prediction performance. These findings suggest that our prognostic risk model could reliably predict the OS of HCC patients.

In the current study, among these 5 RBPs, EZH2, PPARGC1A and IFIT1 have been reported to be essential to the progression and prognosis of cancer. Song et al demonstrated increased levels of H3K27me3 by overexpression of EZH2 and silenced the Wnt signaling inhibitor expression, leading to initiation of Wnt/β-Catenin signaling and subsequent induction of cell proliferation and tumor development.[24] Kido et al. conformed that PPARGC1A was beneficial to the survival of individuals with liver malignancies and had a negative correlation with the expression of the testis specific protein Y.[25] Zhang et al identified elevated expression of IFIT1 as a positive outcome indicator of progression-free survival as well as the period of OS in glioblastoma.[26] In our study, we found that PPARGC1A and IFIT1 were favorable factors for prognosis of HCC, while EZH2 was a risk factor for prognosis of HCC, which indicated that they had the prospect of becoming a new molecular target for liver cancer treatment.

Numerous researches and clinical studies have confirmed the significance of immune infiltration in solid neoplasms.[27] Wei

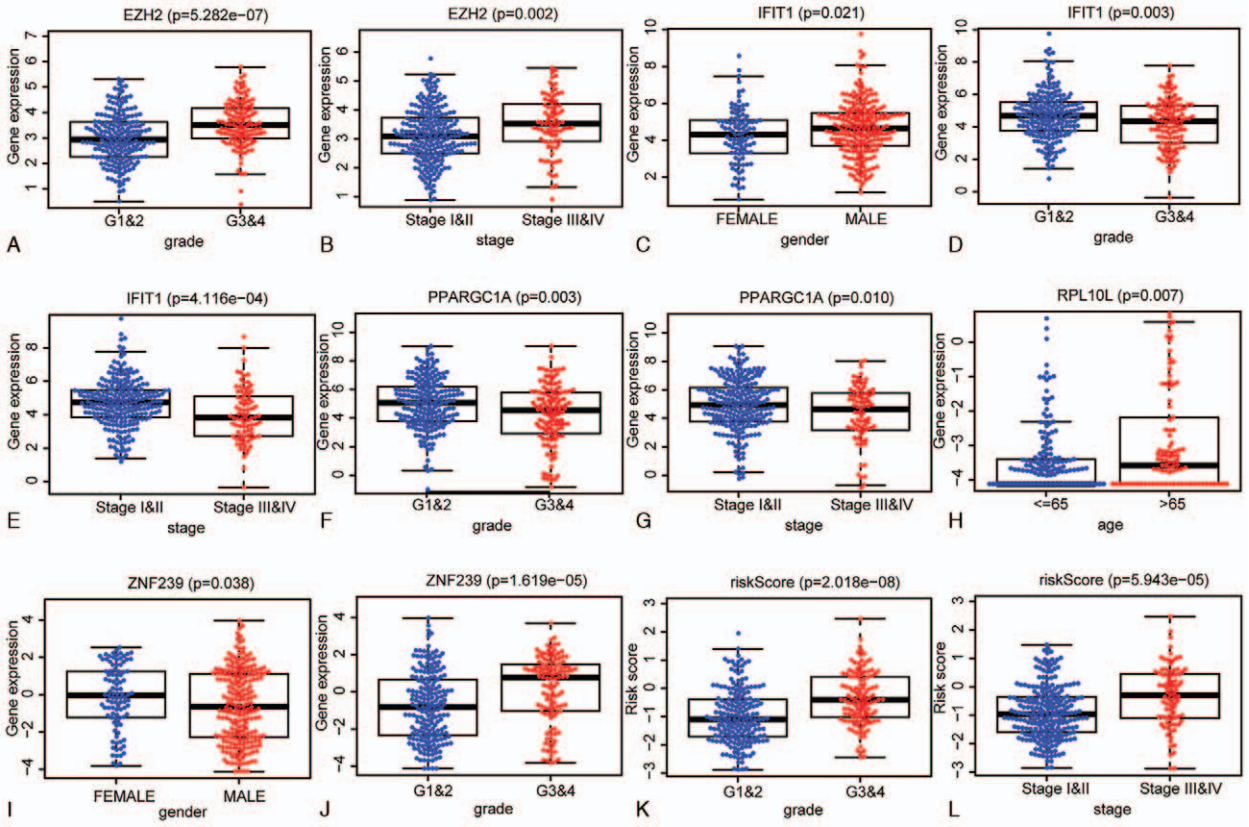

Figure 7. Association of variables in the modelling with clinical attributes of patients in entire TCGA cohort. EZH2 expression and histological grade (A). EZH2 expression and pathological stage (B). IFIT1 expression and gender (C). IFIT1 expression and histological grade (D). IFIT1 expression and pathological stage (E). PPARGC1A expression and histological grade (F). PPARGC1A expression and pathological stage (G). RPL10L expression and age (H). ZNF239 expression and gender (I). ZNF239 expression and histological grade (J). Risk score and histological grade (K). Risk score and pathological stage (L).

et al reported that sperm-associated antigen 5 was a marker of adverse outcome, and its expression was positively associated with the infiltration of neutrophils, CD8+ T cells, B cells, dendritic and macrophages.[28] For the sake of evaluating whether our model had an impact on immune cell infiltration in HCC, we performed an analysis of the correlation between risk score and exposure to immune infiltration levels. We found that risk score had a positive correlation with B cell, CD4+ T cell, CD8 + T cell, dendritic, macrophage and neutrophil, which also indicated that our model had excellent predictive performance.

Recently, there are a lot of studies on the construction of predictive models in HCC. For example, a 4-gene based prognostic model (CENPA, SPP1, MAGEB6, and HOXD9) with the area under the ROC curve of 0.767, 0.737, and 0.692 for 1-year, 3-year, and 5-year was established.[22] Besides, a 9-mRNA signature (RGCC, CDH15, XRN2, RAB3IL1, THEM4, PIF1, MANBA, FKTN, and GABARAPL) with the area under the ROC curve of 0.781, 0.707, and 0.704 for 1-year, 3-year, and 5-year was constructed.[29] In the present study, we developed a reliable model with the area under the ROC curve for 1-, 3-, and 5-year was 0.793, 0.786, 0.767 to further improve the prediction ability of the prediction model. Importantly, our model was superior to previous models in predicting immune cell infiltration and HCC progression. However, there are also some limitations in our research. Firstly, the current research results rely on gene mining methods and a prospective cohort study is required to validate the results. Additionally, the potential mechanisms of how the RBPs-associated gene affects the progression of liver cancer needs further study.

5. Conclusions

We developed and validated 5 RBPs-related models for prognosis to reliably predict OS in HCC patients. Scoring of risk prediction facilitates the screening of patients at high risk of mortality, thereby optimizing decision-making for individualized treatment.

Author contributions

Conceptualization: Ye Liu, Xiaohong Liu, Haofeng Lu.
Data curation: Yang Gu.
Formal analysis: Ye Liu, Yang Gu.
Investigation: Ye Liu, Yang Gu.
Methodology: Ye Liu, Yang Gu.
Project administration: Ye Liu, Xiaohong Liu.
Resources: Haofeng Lu.

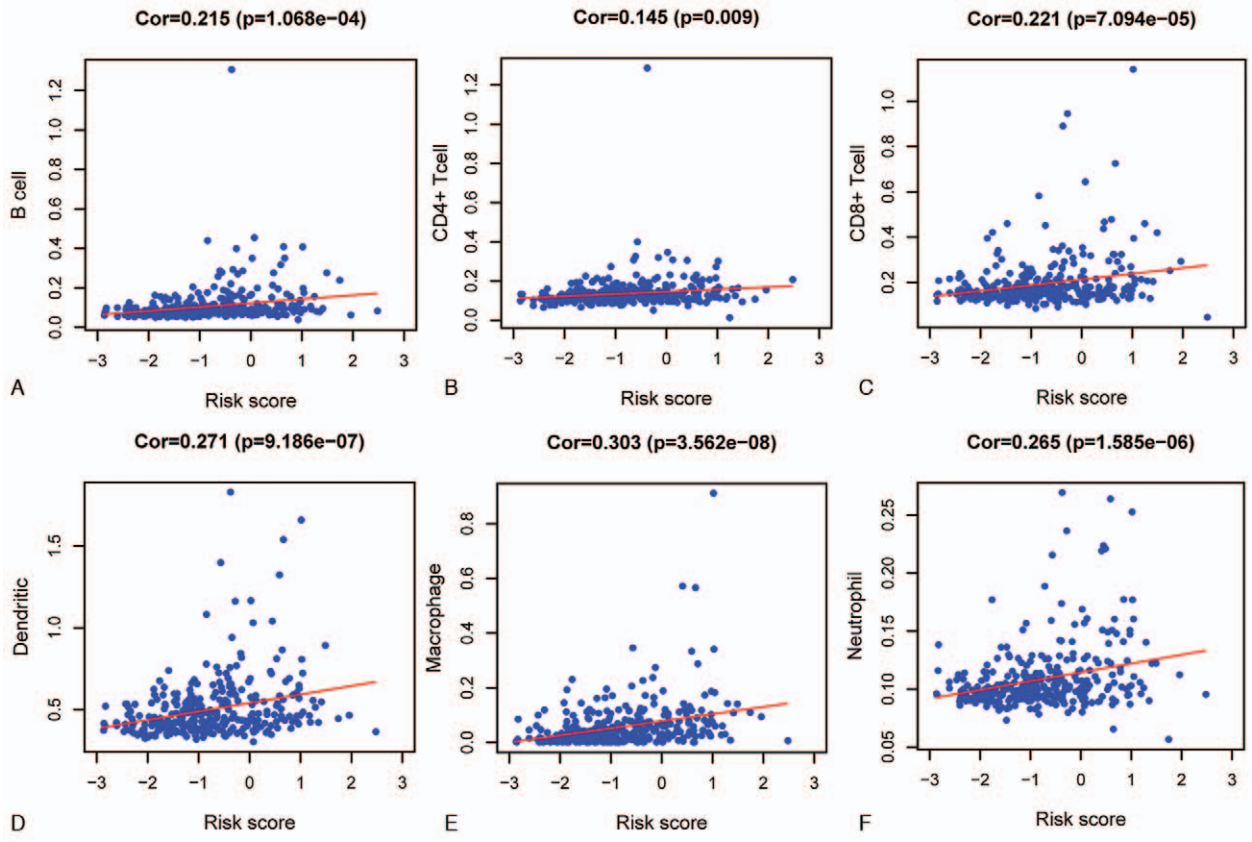

Figure 8. The association on risk score and level of immune infiltration in the entire TCGA cohort. (A). CD4+ T cell (B). CD8+ T cell (C). Dendritic (D). Macrophage (E). Neutrophil (F).

Software: Ye Liu, Yang Gu.
Supervision: Haofeng Lu.
Writing – original draft: Ye Liu, Xiaohong Liu.
Writing – review & editing: Ye Liu, Xiaohong Liu, Haofeng Lu.

References

[1] Villanueva A. Hepatocellular carcinoma. N Engl J Med 2019;380:1450–62.
[2] Llovet JM, Zucman-Rossi J, Pikarsky E, et al. Hepatocellular carcinoma. Nature Rev Dis Primers 2016;2:16018.
[3] Long J, Bai Y, Yang X, et al. Construction and comprehensive analysis of a ceRNA network to reveal potential prognostic biomarkers for hepatocellular carcinoma. Cancer Cell International 2019;19:90.
[4] Sun XY, Yu SZ, Zhang HP, Li J, Guo WZ, Zhang SJ. A signature of 33 immune-related gene pairs predicts clinical outcome in hepatocellular carcinoma. Cancer Med 2020;9:2868–78.
[5] Zheng R, Qu C, Zhang S, et al. Liver cancer incidence and mortality in China: temporal trends and projections to 2030. Chin J Cancer Res 2018;30:571–9.
[6] Craig AJ, von Felden J, Garcia-Lezana T, Sarcognato S, Villanueva A. Tumour evolution in hepatocellular carcinoma. Nature Rev Gastroenterol Hepatol 2020;17:139–52.
[7] Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. Nature Rev Mol Cell Biol 2018;19:327–41.
[8] Treiber T, Treiber N, Plessmann U, et al. A compendium of RNA-binding proteins that regulate MicroRNA biogenesis. Molecular cell 2017;66:270–84.
[9] Coppin L, Leclerc J, Vincent A, Porchet N, Pigny P. Messenger RNA life-cycle in cancer cells: emerging role of conventional and non-conventional RNA-binding proteins? Int J Mol Sci 2018;19:
[10] Pereira B, Billaud M, Almeida R. RNA-binding proteins in cancer: old players and new actors. Trends Cancer 2017;3:506–28.
[11] Kechavarzi B, Janga SC. Dissecting the expression landscape of RNA-binding proteins in human cancers. Genome biology 2014;15:R14.
[12] Bisogno LS, Keene JD. RNA regulons in cancer and inflammation. Curr Opin Genet Dev 2018;48:97–103.
[13] Liang Y, Li E, Min J, et al. miR-29a suppresses the growth and metastasis of hepatocellular carcinoma through IFITM3. Oncol Rep 2018;40:3261–72.
[14] Dong W, Dai ZH, Liu FC, et al. The RNA-binding protein RBM3 promotes cell proliferation in hepatocellular carcinoma by regulating circular RNA SCD-circRNA 2 production. EBioMedicine 2019;45:155–67.
[15] Zheng R, Wan C, Mei S, et al. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res 2019;47(D1):D729–d735.
[16] Liu T, Ortiz JA, Taing L, et al. Cistrome: an integrative platform for transcriptional regulation studies. Genome biology 2011;12:R83.
[17] Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. Nat Rev Genet 2014;15:829–45.
[18] Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 2010;13:R25.
[19] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[20] Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research 2019;47(D1):D607–d613.
[21] She Y, Kong X, Ge Y, et al. Immune-related gene signature for predicting the prognosis of head and neck squamous cell carcinoma. Cancer Cell International 2020;20–2.

[22] Long J, Zhang L, Wan X, et al. A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. J Cell Mol Med 2018;22:5928–38.
[23] Li W, Lu J, Ma Z, Zhao J, Liu J. An integrated model based on a six-gene signature predicts overall survival in patients with hepatocellular carcinoma. Front Genet 2020;10–213.
[24] Song H, Yu Z, Sun X, et al. Androgen receptor drives hepatocellular carcinogenesis by activating enhancer of zeste homolog 2-mediated Wnt/β-catenin signaling. EBioMedicine 2018;35:155–66.
[25] Kido T, Lau YC. The Y-linked proto-oncogene TSPY contributes to poor prognosis of the male hepatocellular carcinoma patients by promoting the pro-oncogenic and suppressing the anti-oncogenic gene expression. Cell Biosci 2019;9–12.
[26] Zhang JF, Chen Y, Lin GS, et al. High IFIT1 expression predicts improved clinical outcome, and IFIT1 along with MGMT more accurately predicts prognosis in newly diagnosed glioblastoma. Hum Pathol 2016;52:136–44.
[27] Shao S, Cao H, Wang Z, et al. CHD4/NuRD complex regulates complement gene expression and correlates with CD8 T cell infiltration in human hepatocellular carcinoma. Clinical Epigenetics 2020;12–31.
[28] Chen W, Chen X, Li S, Ren B. Expression, immune infiltration and clinical significance of SPAG5 in hepatocellular carcinoma: a gene expression-based study. J Gene Med 2020;22:
[29] Ni FB, Lin Z, Fan XH, et al. A novel genomic-clinicopathologic nomogram to improve prognosis prediction of hepatocellular carcinoma. Clin Chim Acta 2020;504:88–97.