Six RNA binding proteins (RBPs) related prognostic model predicts overall survival for clear cell renal cell carcinoma and is associated with immune infiltration

Qianwei Xing[1#], Jiaochen Luan[2#], Shouyong Liu[2#], Limin Ma[1*], Yi Wang[1*]

ABSTRACT

The aim of this article was to construct an accurate prognostic model using RNA-binding proteins (RBPs) to predict overall survival (OS) for patients with clear cell renal cell carcinoma (ccRCC) as well as to reveal its associations with immune infiltration. Expression profiles based on RBPs and clinical follow-up parameters were obtained from the Cancer Genome Atlas (TCGA) and the ArrayExpress databases. Through univariate COX and LASSO regression analyses, the RBPs based signature was developed. A total of six RBPs (CLK2, IGF2BP2, RNASE2, EZH2, PABPC1L, and RPL22L1) were eventually used to establish a prognostic signature. Based on this signature, ccRCC patients were classified into high-risk and low-risk subgroups and significant OS was obtained in both the internal and external datasets ($p < 0.05$). AUCs of its ROC curve were all above 0.70 and this signature was an independent prognostic factor of OS for ccRCC ($p < 0.05$). Nomograms were also constructed to visualize the relationships among individual predictors and 1-, 3-, and 5-year OS for ccRCC. Furthermore, the established RBPs based signature was strongly related to critical clinicopathologic characteristics such as grade ($p = 8.921e-12$), stage ($p = 1.421e-11$), M ($p = 1.662e-05$), and T stage ($p = 7.907e-10$). Moreover, 12 kinds of tumor-infiltrating immune cells were significantly linked to high-risk and low-risk groups classified by our constructed model (all $p < 0.05$). Our study successfully identified six RBPs as a robust prognostic signature in ccRCC by both external and internal verifications. Besides, our established model displayed significant associations with immune infiltration. In addition to original clinical parameters, our findings may further help clinicians in predicting patients' survival status and creating individualized treatment plans.

KEYWORDS: RNA-binding proteins; overall survival; clear cell renal cell carcinoma; model; prognosis

INTRODUCTION

Renal cell carcinoma (RCC) is estimated to have caused approximately 76,080 new cases as well as 13,780 new deaths in the United States in 2021 alone [1]. Accounting for approximately 70-80% of RCC, clear cell renal cell carcinoma (ccRCC) comprises the majority of cancer deaths [2,3]. Due to the resistance to radiotherapy and chemotherapy, the surgical resection is recommended as the primary therapy for ccRCC by the clinical guidelines [4,5]. Despite the tremendous progress in novel diagnostic tools and early surgical treatment, the cancer metastasis of ccRCC is still extremely common and 2 years survival rate of metastatic patients is <20% [6]. Therefore, new diagnostic markers and therapeutic targets are urgently required to understand the potential molecular mechanism and predict the disease occurrence, progression, and metastasis for these cases.

RNA-binding proteins (RBPs), also known as proteins interacting with different types of RNAs (ncRNAs, rRNAs, miRNAs, snRNAs, tRNAs, mRNAs, and snoRNAs), are recommended as pivotal post-transcriptional regulators not only regulating the spatiotemporal expression of genes but also modulating the disease pathogenesis [7]. Accounting for 7.5% of about 20,500 protein-coding genes in humans [8], RBPs blind to RNA or constitute crucial components of ribonucleoprotein (RNPs) to participate in RNA metabolism [9]. Till now, genome-wide screening of the human genome has identified more than 1500 RBPs that played an essential role in biogenesis, surveillance, transport, localization, and degradation of RNA in line with the genetic and biochemical studies [8,10-12]. According to target RNA categorization, it is found that 50% of RBPs involve pathways of mRNA metabolism, 11% of them establish ribosomal proteins, and the others associate with different kinds of non-coding RNA metabolism [8,13].

[1]Department of Urology, Affiliated Hospital of Nantong University, Nantong, Jiangsu Province, China
[2]Department of Urology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu Province, China
#These authors contributed equally to this work

*Corresponding authors: Yi Wang, Department of Urology, Affiliated Hospital of Nantong University, No. 20 West Temple Road, Nantong, Jiangsu Province, China. E-mail: wangyi_urology@ntu.edu.cn
Limin Ma, Department of Urology, Affiliated Hospital of Nantong University, No. 20 West Temple Road, Nantong, Jiangsu Province, China. E-mail: ntmalimin@163.com

With the discovery of non-coding RNAs and increasing understanding of post-transcriptional regulation in tumors, cancer-related RBPs are employed to construct highly intricate regulatory networks. Moreover, the disturbance of these networks is likely to have a relationship with primary carcinogenic hits, increasing aggressiveness, and accelerating progression [14]. Accumulating data have underscored that RBPs primarily alter various cancer-associated downstream targets to exert influence on carcinogenesis and development. Researches as UNR in melanoma [15], LARP1 in ovarian cancer (OC) [16], IMP3 in leukemia [17], QKI in kidney cancer [18], LIN28B as well as MSI in colon cancer [19], and IMP2 in glioma [20] have been applied to disclose the RBP cancer-specific post-transcriptional networks. However, the roles of most RBPs have not yet been found in tumors, and the functions of RBPs in the progress of cancer remain relatively unexplored. In the current research, our efforts were made to establish a RBPs based signature to predict overall survival (OS) for ccRCC. Our results were anticipated to help clinician predict patients' survival status and to promote the specific individualized treatment than original clinical parameters.

MATERIALS AND METHODS

Identification of expression profiles and differentially expressed RBPs (DERBPs) from public databases

Expression profiles based on RBPs together with clinical follow-up parameters were obtained from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/). As detailed in Supplementary Table S1, the census of human RBPs was obtained from the article of "A census of human RNA-binding proteins" in Nature Reviews Genetics [8]. The raw data were pre-handled by the "limma" package of R software and standardized by log2 transformation. In addition, to screen DERBPs between ccRCC tumor tissues and adjacent normal kidney tissues, log2|fold change (FC)| ≥1 and false discovery rate (FDR)<0.05 were set as the cutoff criterion. Moreover, the external validation cohort (E-MTAB-1980 dataset) was obtained from the ArrayExpress dataset (https://www.ebi.ac.uk/arrayexpress/).

Establishment and validation of the risk score model

To explore the associations between OS and DERBPs in the training database and testing database, the univariate Cox regression analysis was applied and $p = 0.05$ was set as the cut-off value. Then, LASSO Cox regression method was utilized to establish the model based on prognostic RBPs. LASSO regression could fit the generalized linear model, contributing to variable selection and regularization. Subsequently, the risk score algorithm was constructed as following:

$$Riskscore = \sum_{i=1}^{n} expi * \beta i$$

Therein, βi represented the regression coefficient of each gene, and expi represented the expression level of each gene.

Nomogram construction and validation

According to the independent prognostic clinical parameters and our established RBPs signature, a novel nomogram was conducted to forecast the likelihood of OS for ccRCC. To estimate the accuracy of the nomogram, ROC curves as well as the area under the ROC curve (AUC) was plotted. Moreover, calibration curves were generated from the "rms" package of R software, and applied to compare the observed and predicted results of this nomogram. Similarly, the methods were employed in the external validation set to verify these outcomes.

Verification of the mRNA expression and prognosis utilizing the ICGC, GEO datasets, and Kaplan–Meier plotter website

International Cancer Genome Consortium (ICGC) dataset cohort (http://dcc.icgc.org) and two Gene Expression Omnibus (GEO) datasets (https://www.ncbi.nlm.nih.gov/geo/; GSE14994 and GSE6344) were utilized to verify the mRNA expression of these six hub RBPs. Kaplan–Meier plotter online tool (http://kmplot.com/analysis/) was employed to evaluate the prognosis of hub RBPs in ccRCC cohorts [21].

Quantitative real-time PCR (qRT-PCR)

We followed the manufacturer's instructions and qRT-PCR was utilized to verify the mRNA expression of these six hub RBPs in four pairs of ccRCC tumor and adjacent normal kidney tissues acquired from Affiliated Hospital of Nantong University by means of StepOne Plus RT-PCR system (Applied Biosystems, Foster City, CA, USA). Our used primers were displayed as following: Actin (F: 5'-ATGACTTAGTTGCGTTACACC-3', R: 5'-GACTTCCTGTAACAACGCATC-3');

CLK2 (F: 5'-GGGGAGTTACCGTGAACACTA-3', R: 5'-CGTGTCCGGTCACTACTACTTG-3');

EZH2 (F: 5'-GTACACGGGGATAGAGAATGTGG-3', R: 5'-GGTGGGCGGCTTTCTTTATCA-3');

IGF2BP2 (F: 5'-AGCCTGTCACCATCCATGC-3', R: 5'-CTTCGGCTAGTTTGGTCTCATC-3');

PABPC1L (F: 5'-AACATCTACGTGAAGAACCTCCC-3', R: 5'-CACTCAGCATTTTCCCAAACTG -3');

RNASE2 (F: 5'-TGTGGTAACCCAAATATGACCTG-3', R: 5'-GGTCTCGTCGTTGATCTCTGT-3');

RPL22L1 (F: 5'-GCAATTTCTACGGGAGAAGGTT-3', R: 5'- ACTCGAAGCCAATCACGAAGA-3');

Validation of the protein expression utilizing the Human Protein Atlas (HPA) database and CPTAP analysis

Using the HPA online database (http://www.proteinatlas.org/), the protein expression of the hub RBPs in ccRCC was validated by immunohistochemical (IHC) staining. As detailed in HPA database, CLK2 was stained by HPA055366 antibody in IHC; EZH2 was stained by CAB009589 antibody in IHC; IGF2BP2 was stained by HPA035145 antibody in IHC; RNASE2 was stained by HPA044983 antibody in IHC; RPL22L1 was stained by HPA056207 antibody in IHC; whereas PABPC1L immunohistochemistry outcomes had not been provided yet. Scale bar for each IHC picture was 200 um. We also utilized the UALCAN website (http://ualcan.path.uab.edu/analysis-prot.html) to validate the protein of the hub RBPs expression between the primary ccRCC tumor and normal tissues by Clinical Proteomic Tumor Analysis Consortium (CPTAC) dataset analysis [22].

Tumor-infiltrating immune cells (TIICs) estimation

The expressions of TIICs in every ccRCC sample from the TCGA dataset were calculated as previously described [23]. Based on our established model, ccRCC patients were divided into a high- and low-risk group. Through R packages, we could explore whether or not TIICs were linked to these two groups, under the threshold of $p < 0.05$.

Statistical analysis

Statistical analysis was accomplished by utilizing the R software 3.6.3. Student's t-test or Wilcoxon rank-sum test was utilized for continuous variables and Chi-squared test or Fisher's exact test was employed for categorical variables. Kaplan–Meier survival curves as well as the log-rank test were implemented utilizing the "survival" package of R software. The ROC curves were plotted using the R package "survival ROC." For the whole statistical analyses, p values were two sided and its values below 0.05 were regarded to be significantly different.

Availability of data and material

RNA sequencing data of ccRCC, together with clinical follow-up parameters, were got from the Cancer Genome Atlas (TCGA) database and the ArrayExpress database (E-MTAB-1980). ICGC dataset and two GEO datasets (GSE14994 and GSE6344) were utilized to verify the mRNA expression of these six hub RBPs.

RESULTS

RBPs based expression profiles and differently expressed RBPs identification

The whole workflow of this study is shown in Supplementary Figure S1. RNA sequencing data of ccRCC and the clinicopathological characteristics including 72 adjacent normal renal tissues and 539 ccRCC tumor samples, were obtained from the TCGA cohort. When selecting |log2(FC)|>1 and FDR<0.05 as the threshold, 125 differently expressed RBPs were screened out from a list of 1542 RBPs, including 38 downregulated and 87 upregulated RBPs (Supplement Table S2). The expression heatmap of 125 differently expressed RBPs and the volcano plot of all RBPs are demonstrated in Figure 1A and 1B.

Prognostic model (risk score) construction

Based on the univariate Cox regression analysis, 54 candidate RBPs were identified (Figure 2A). Then, the LASSO Cox regression model was performed and six vital prognostic RBPs including CLK2, IGF2BP2, RNASE2, EZH2, PABPC1L, and RPL22L1 were finally selected (Figure 2B-C and Table 1). As a result, a six prognostic RBPs signature was constructed and the risk score of each sample was calculated: Risk score = (0.01812 × ExpCLK2) + (0.02605 × ExpIGF2BP2) + (0.04622 × ExpRNASE2) + (0.05813 × ExpEZH2) + (0.01903 × ExpPABPC1L) + (0.01191 × ExpRPL22L1).

Evaluation, external and internal verification of six RBPs based signature (risk score)

Based on six RBPs established signature (risk score), patients with ccRCC were classified into two groups (high- and

TABLE 1. Coefficients of these six key prognostic RBPs

Gene	Coefficients
CLK2	0.018122682
IGF2BP2	0.026048548
RNASE2	0.046223445
EZH2	0.058125193
PABPC1L	0.019031592
RPL22L1	0.011911132

RBPs: RNA-binding proteins

TABLE 2. External and internal verification datasets of 1-year, 3-year, and 5-year ROC

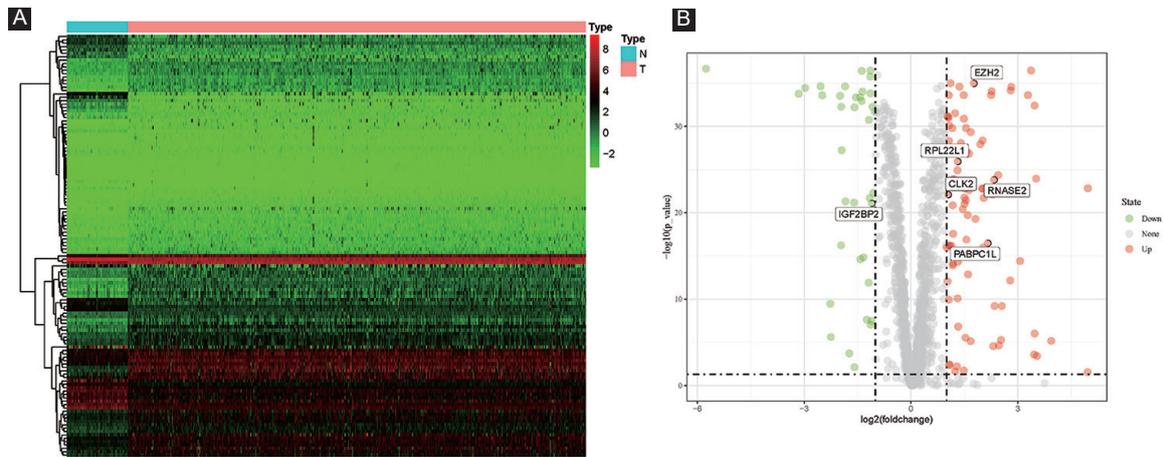| Datasets	1-year ROC	3-year ROC	5-year ROC
The whole training dataset (TCGA)	0.724	0.716	0.741
The external validation dataset (ArrayExpress)	0.790	0.830	0.862
The internal validation dataset 1 (test 1)	0.705	0.723	0.745
The internal validation dataset 2 (test 2)	0.742	0.710	0.738

FIGURE 1. One hundred and twenty-five differentially expressed RNA-binding proteins get from TCGA ccRCC cohort; (A) heatmap; N=adjacent normal renal tissues; T=ccRCC tumor tissues; scale=FPKM values of gene expression; (B) volcano plot.
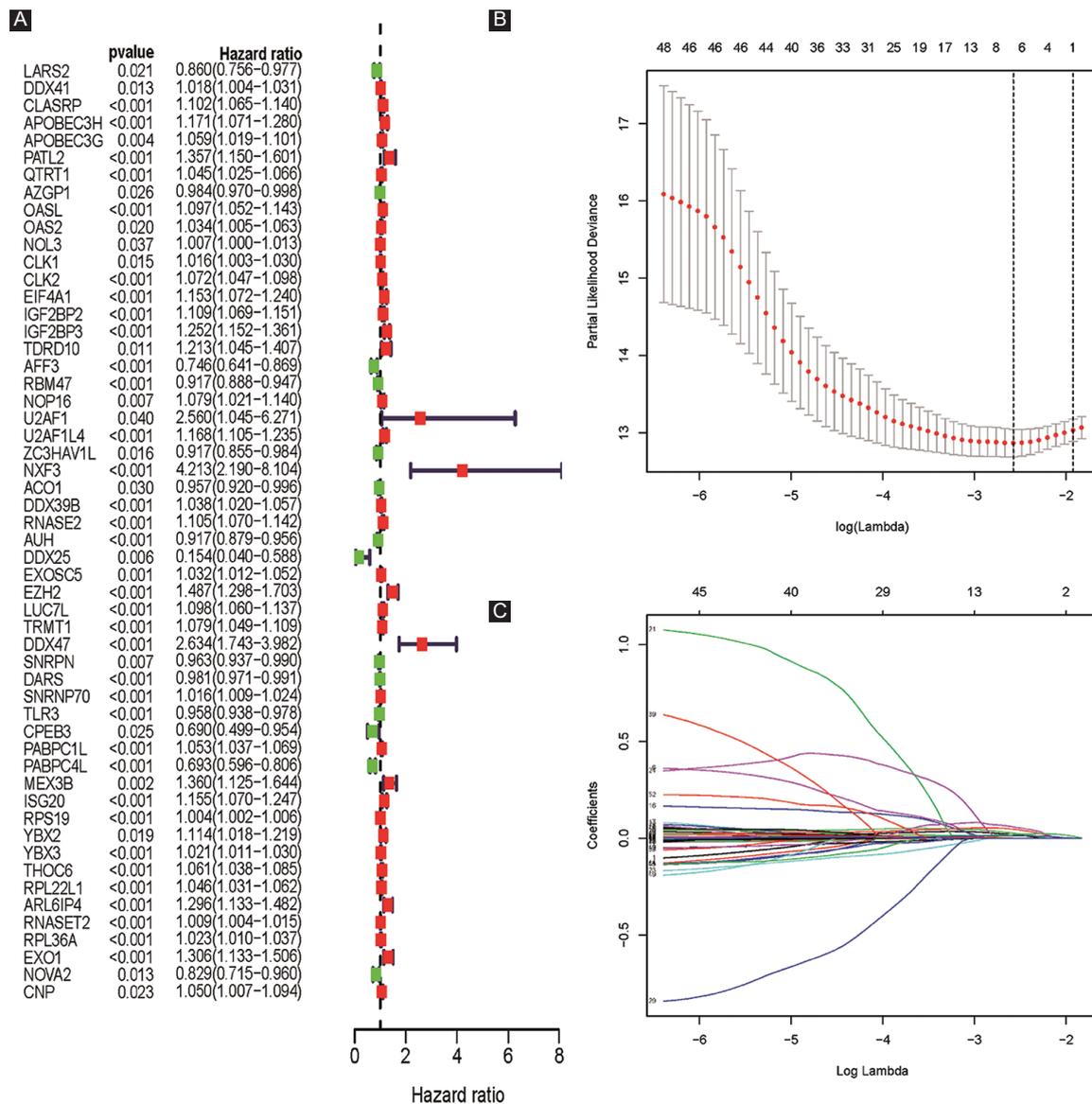

FIGURE 2. Prognostic model index (risk score) construction based on univariate Cox regression analysis and LASSO analysis; (A) the forest plot of 54 differentially expressed RNA-binding proteins (RBPs) screened out by univariate Cox regression; (B-C) LASSO coefficient profiles of the prognostic RBPs.
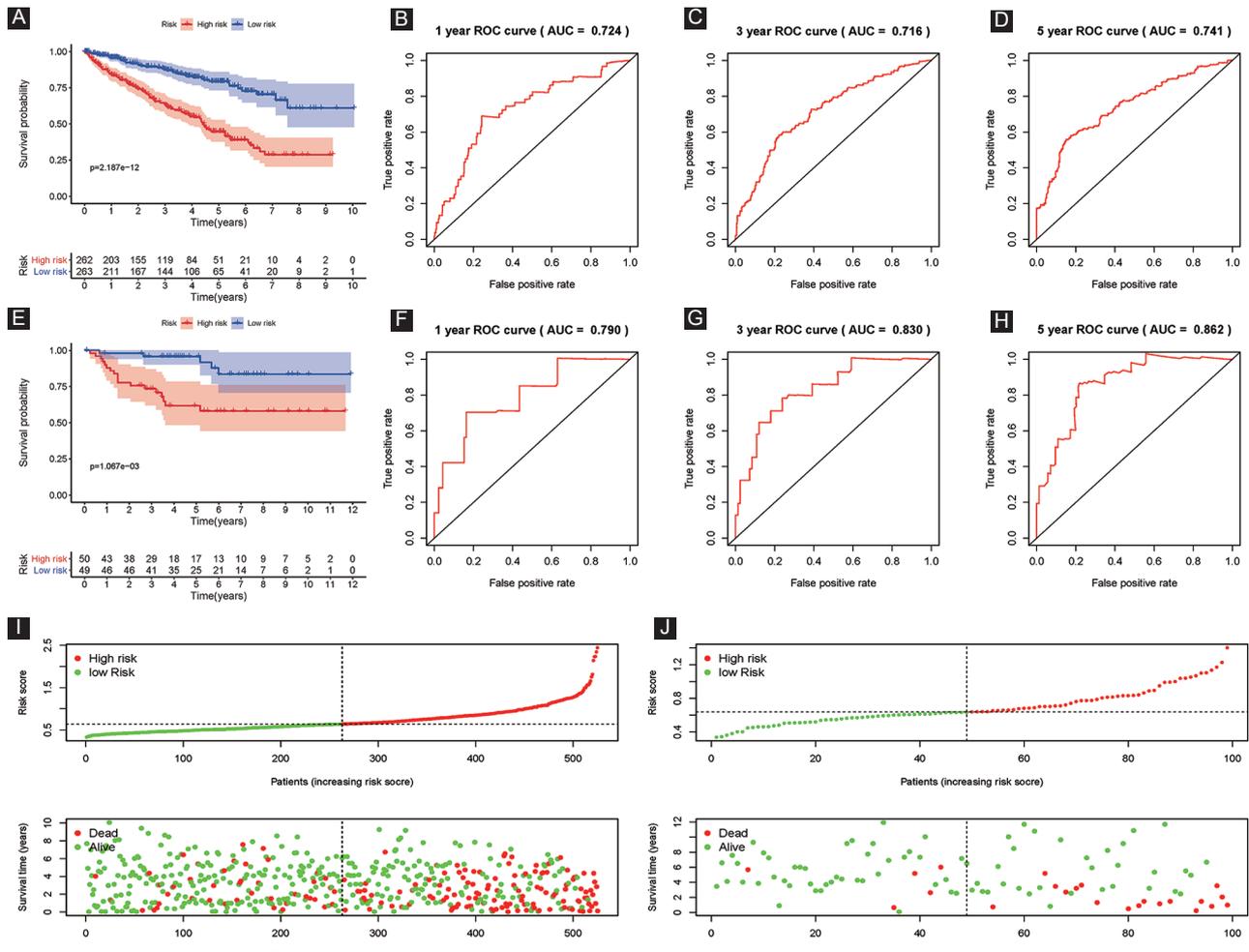
FIGURE 3. Evaluation and external verification of our established signature; (A) Kaplan–Meier survival curves of OS in the whole training dataset (TCGA); (B) 1-year ROC in the whole training dataset (TCGA); (C) 3-year ROC; (D) 5-year ROC; (E) Kaplan–Meier survival curves of OS in the external validation dataset (ArrayExpress); (F) 1-year ROC in the external validation dataset (ArrayExpress); (G) 3-year ROC; (H) 5-year ROC; (I-J) the distribution of risk scores for each sample and patients' survival status in the whole training dataset (TCGA) and the external validation dataset (ArrayExpress); therein, A-D and I represented the whole training dataset (TCGA); E-H and J represented the external validation dataset (ArrayExpress).

low-risk groups), and KM survival analysis shed light on that patients with low risks had a much better OS than those with high risks ($p = 2.187e–12$, Figure 3A). To better evaluate our established model, the ROC curve and its AUC were further analyzed. Our results indicated that 1-, 3-, and 5-year AUC values were 0.724, 0.716, and 0.741, separately, showing superior predictive veracity of patients' survival results (Figure 3B-D and Table 2). In addition, when the risk score increased, patients would have more dead events (Figure 3I).

The external validation database (E-MTAB-1980, n = 99), the internal validation dataset 1 (n = 264), and the internal validation dataset 2 (n = 261) were utilized as validation databases to verify our signature. In terms of Kaplan–Meier survival analysis, all three validation sets showed similar outcomes (the external validation dataset: $p = 1.067e–03$; the internal validation dataset 1: $p = 5.082e–06$; and the internal validation dataset 2: $p = 3.507e–07$; Figure 3E, Figure 4A and E). ROC analysis displayed that the AUC for 1-year, 3-year, and 5-year OS of these three databases were all above 0.70 (Figure 3F-H, Figure 4B-D,

Figure 4F-H, Table 2). Figure 3J and Figure 4I-J displayed that patients would have more dead events, when the risk score increased. In a word, our established model possessed superior sensitivity and specificity in predicting OS for ccRCC.

Our established six RBPs based signature could serve as an independent prognostic parameter for OS

In the univariate Cox analysis, the high-risk groups revealed a 5.411-fold, 4.453-fold, 6.776-fold, and 57.907-fold increased risk of death than those in the low-risk groups (the whole training dataset (TCGA): 95% CI 3.897-7.512; the internal validation dataset 1: 95% CI 2.824-7.023; the internal validation dataset 2: 95% CI 4.142-11.084; and the external validation dataset: 95% CI 10.449-320.911, all $p < 0.001$; respectively). By means of multivariate Cox analysis, the six RBPs based model was strongly associated with OS for patients with ccRCC in the whole training database: HR = 3.417, 95% CI 2.259-5.168;

TABLE 3. Univariate and multivariate Cox regression analysis of external and internal verification datasets for overall survival (OS)

The whole training dataset (TCGA)	Univariate Cox regression analysis				Multivariate Cox regression analysis			
ID	HR	HR.95L	HR.95H	*p*-value	HR	HR.95L	HR.95H	*p*-value
Age	1.030913	1.017301	1.044707	7.14E-06	1.035101	1.020556	1.049854	1.77E-06
Gender	0.939895	0.682968	1.293475	0.703593641	1.033268	0.738757	1.445188	0.848385584
Race	1.175718	0.705799	1.958508	0.534112006	1.275535	0.766027	2.123932	0.349558444
Grade	1.971561	1.640971	2.368751	4.20E-13	1.28167	1.017062	1.615121	0.03543498
Stage	1.880446	1.664529	2.12437	3.38E-24	1.803986	1.245201	2.613526	0.00181205
T	2.043146	1.72421	2.421079	1.57E-16	1.002533	0.757815	1.326277	0.985863735
M	2.135679	1.688608	2.701115	2.42E-10	0.742863	0.374737	1.472621	0.394560117
N	0.862088	0.737405	1.007852	0.062626764	0.869547	0.737792	1.024831	0.095437453
Risk score	5.410606	3.897003	7.512095	6.50E-24	3.417165	2.259309	5.168402	5.85E-09

The internal validation dataset 1 (test 1)	Univariate Cox regression analysis				Multivariate Cox regression analysis			
ID	HR	HR.95L	HR.95H	*p*-value	HR	HR.95L	HR.95H	*p*-value
Age	1.042564	1.021728	1.063825	5.19E-05	1.047857	1.025271	1.07094	2.61E-05
Gender	1.010839	0.637407	1.603052	0.963450866	1.216074	0.73948	1.999832	0.440825876
Race	1.149275	0.577583	2.286829	0.691857074	1.329803	0.669684	2.640615	0.41542614
Grade	1.978801	1.495423	2.618424	1.79E-06	1.347296	0.936298	1.938707	0.108390345
Stage	1.736677	1.454669	2.073355	1.03E-09	1.653178	0.979126	2.791262	0.059968007
T	1.902481	1.497071	2.417677	1.44E-07	0.979773	0.648727	1.479753	0.922618285
M	1.883329	1.319557	2.687969	0.00048717	0.631105	0.239803	1.66092	0.351185025
N	0.814787	0.65018	1.021068	0.075257409	0.811419	0.634599	1.037506	0.095641512
Risk score	4.453297	2.823815	7.023073	1.31E-10	3.376678	1.896403	6.012412	3.56E-05

The internal validation dataset 2 (test 2)	Univariate Cox regression analysis				Multivariate Cox regression analysis			
ID	HR	HR.95L	HR.95H	*p*-value	HR	HR.95L	HR.95H	*p*-value
Age	1.021395	1.00335	1.039765	0.01992815	1.032345	1.011747	1.053361	0.001963071
Gender	0.901166	0.578116	1.404734	0.645895533	0.920735	0.574228	1.476335	0.731736537
Race	1.194868	0.558591	2.55591	0.646298532	1.151368	0.508761	2.605642	0.735174194
Grade	1.980887	1.558245	2.518162	2.37E-08	1.312002	0.963009	1.78747	0.085237027
Stage	2.015703	1.700396	2.38948	6.66E-16	2.147403	1.169713	3.942286	0.013674269
T	2.152136	1.692616	2.736408	3.99E-10	0.99652	0.659971	1.504691	0.986769909
M	2.348955	1.713978	3.219172	1.09E-07	0.684919	0.225055	2.084439	0.505106802
N	0.908011	0.729234	1.130616	0.388358571	0.891494	0.708698	1.121439	0.326580501
Risk score	6.776	4.142479	11.08374	2.52E-14	3.584984	1.7738	7.245523	0.000375966

The external validation dataset (ArrayExpress)	Univariate Cox regression analysis				Multivariate Cox regression analysis			
ID	HR	HR.95L	HR.95H	P value	HR	HR.95L	HR.95H	pvalue
Age	1.043727	1.001971	1.087222	0.03992667	1.0259	0.981825	1.071953	0.253746248
Gender	2.265277	0.672896	7.62596	0.186735984	2.612452	0.665756	10.25136	0.168601041
Grade	2.981514	1.670809	5.320429	0.000218	1.498271	0.670429	3.348328	0.324411222
T	2.564016	1.675159	3.92451	1.45E-05	1.615248	0.984168	2.650997	0.057850092
N	3.866228	2.31875	6.446455	2.17E-07	2.385944	1.245581	4.570339	0.00873782
M	6.113103	2.570636	14.53727	4.20E-05	3.682659	1.382799	9.807631	0.0090944
Risk score	57.90725	10.44915	320.9113	3.39E-06	2.054828	0.14946	28.25042	0.590182717

in the internal validation dataset 1: HR = 3.377, 95% CI 1.896-6.012; and in the internal validation dataset 2: HR = 3.585, 95% CI 1.774-7.246, all *p* < 0.001; except for the external validation dataset: HR = 2.055, 95% CI 0.149-28.250; *p* = 0.590) (Table 3).

Construction of the novel nomogram on the basis of clinical characteristics and the signature

To provide a quantitative method to predict the ccRCC patients' prognosis in clinical trials, we established a compound nomogram in both the TCGA and ArrayExpress databases. Our

outcomes presented that this novel nomogram could better predict OS of patients (Figure 5A) and its 1-year, 3-year, and 5-year AUC values and C-index in the TCGA dataset were 0.842, 0.806, 0.788, and 0.79, respectively, showing an excellent prognostic ability (Table 4 and Supplement Figure S2). In the calibration curve, the diagonal line represented the most ideal outcome; the closer the predictive values were to the diagonal line, the more consistent they were with the actual situation. Calibration plots of this nomogram revealed that the predictive values were significantly similar to the ideal ones (Figure 5C). We also built another prognostic
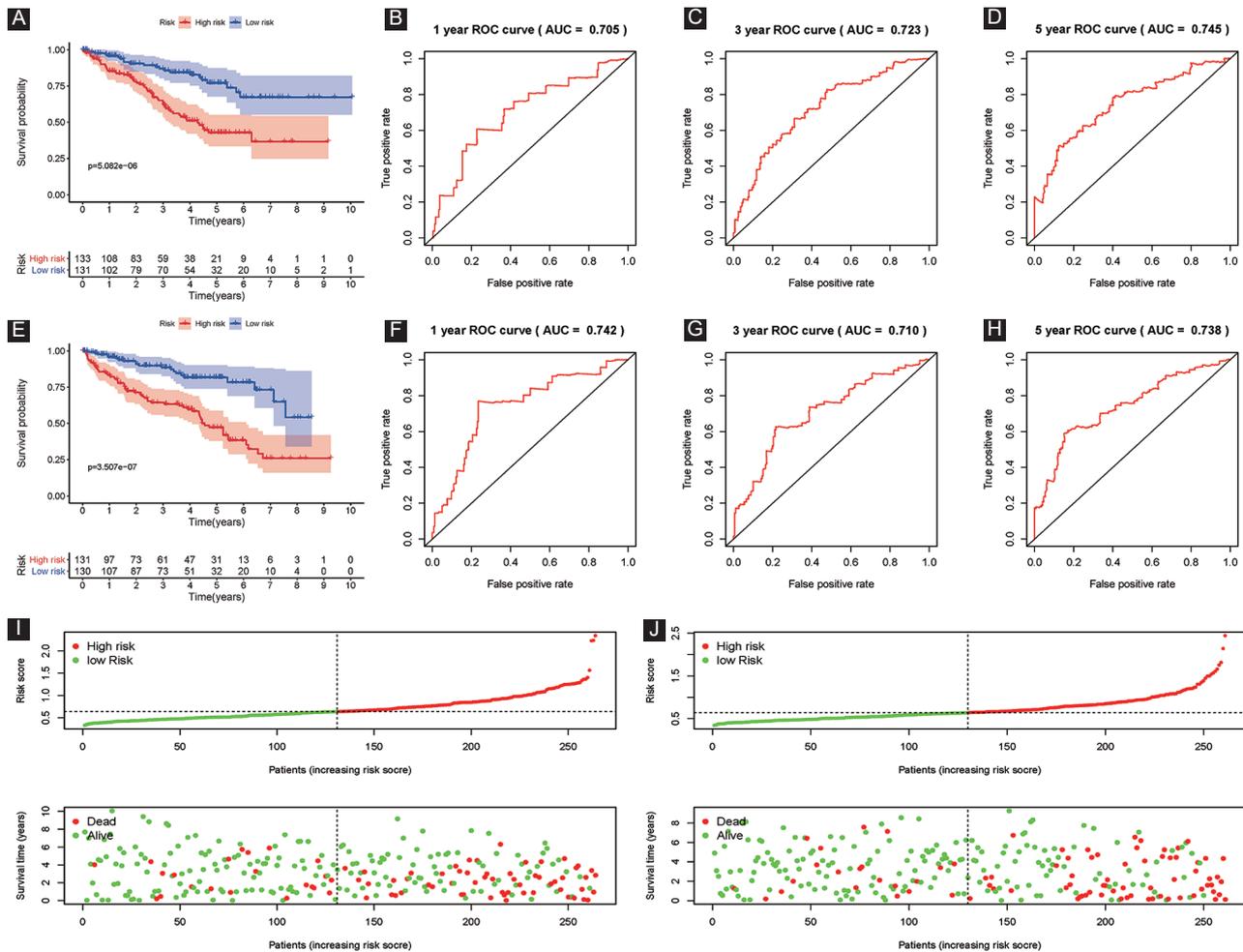
FIGURE 4. Internal verification of six RNA-binding proteins established signature; (A) Kaplan–Meier survival curves of OS in the internal validation dataset 1 (test 1); (B) 1-year ROC in the internal validation dataset 1 (test 1); (C) 3-year ROC; (D) 5-year ROC; (E) Kaplan–Meier survival curves of OS in the internal validation dataset 2 (test 2); (F) 1-year ROC in the internal validation dataset 2 (test 2); (G) 3-year ROC; (H) 5-year ROC; (I-J) the distribution of risk scores for each sample and patients' survival status in the internal validation dataset 1 (test 1) and in the internal validation dataset 2 (test 2); therein, A-D and I represented the internal validation dataset 1 (test 1); E-H and J represented the internal validation dataset 2 (test 2).

TABLE 4. 1-year, 3-year, and 5-year ROC and C-index of nomogram for in TCGA and ArrayExpress datasets

ROC	1-year ROC	3-year ROC	5-year ROC	C-index
TCGA cohort	0.842	0.806	0.788	0.790
ArrayExpress cohort	0.895	0.897	0.861	0.872

nomogram in the ArrayExpress dataset (E-MTAB-1980) as an external validation set to verify the previous results (Figure 5B). Its 1-, 3-, and 5-year AUC values and C-index in the ArrayExpress dataset were 0.895, 0.897, 0.861, and 0.872, separately, showing a better predictive accuracy in OS (Table 4 and Supplement Figure S2). Calibration plot also displayed the satisfactory conformity between the predicted and actual values (Figure 5D).

Association between these six prognostic RBPs, risk score, and clinicopathologic characteristics

The relationships between clinicopathologic characteristics, risk score, and six prognostic RBPs were explored. Our results revealed that the six RBPs based signature (risk score) was firmly related to grade ($p = 8.921e{-}12$), tumor stage ($p = 1.421e{-}11$), M stage ($p = 1.662e{-}05$), and T stage ($p = 7.907e{-}10$) (Supplement Figure S3). In addition, the correlation between six hub RBPs and clinical features was also analyzed (Table 5).

Validation of the mRNA expression and the prognosis of six RBPs in ccRCC

ICGC dataset (http://dcc.icgc.org), containing 45 normal renal and 91 tumor samples, was applied to verify the mRNA expression of these six RBPs (CLK2, EZH2, IGF2BP2, PABPC1L, RNASE2, and RPL22L1). As displayed in Figure 6A-F, they were differentially expressed in tumors compared with normal tissues (all $p < 0.001$). Results from GSE14994 and GSE6344 datasets showed that EZH2, IGF2BP2, and RNASE2 had significant expressions in tumors compared with normal tissues, while the others did not (all $p < 0.01$; Figure 6G-L). KM plotter displayed that six RBPs were remarkably related to OS in ccRCC patients
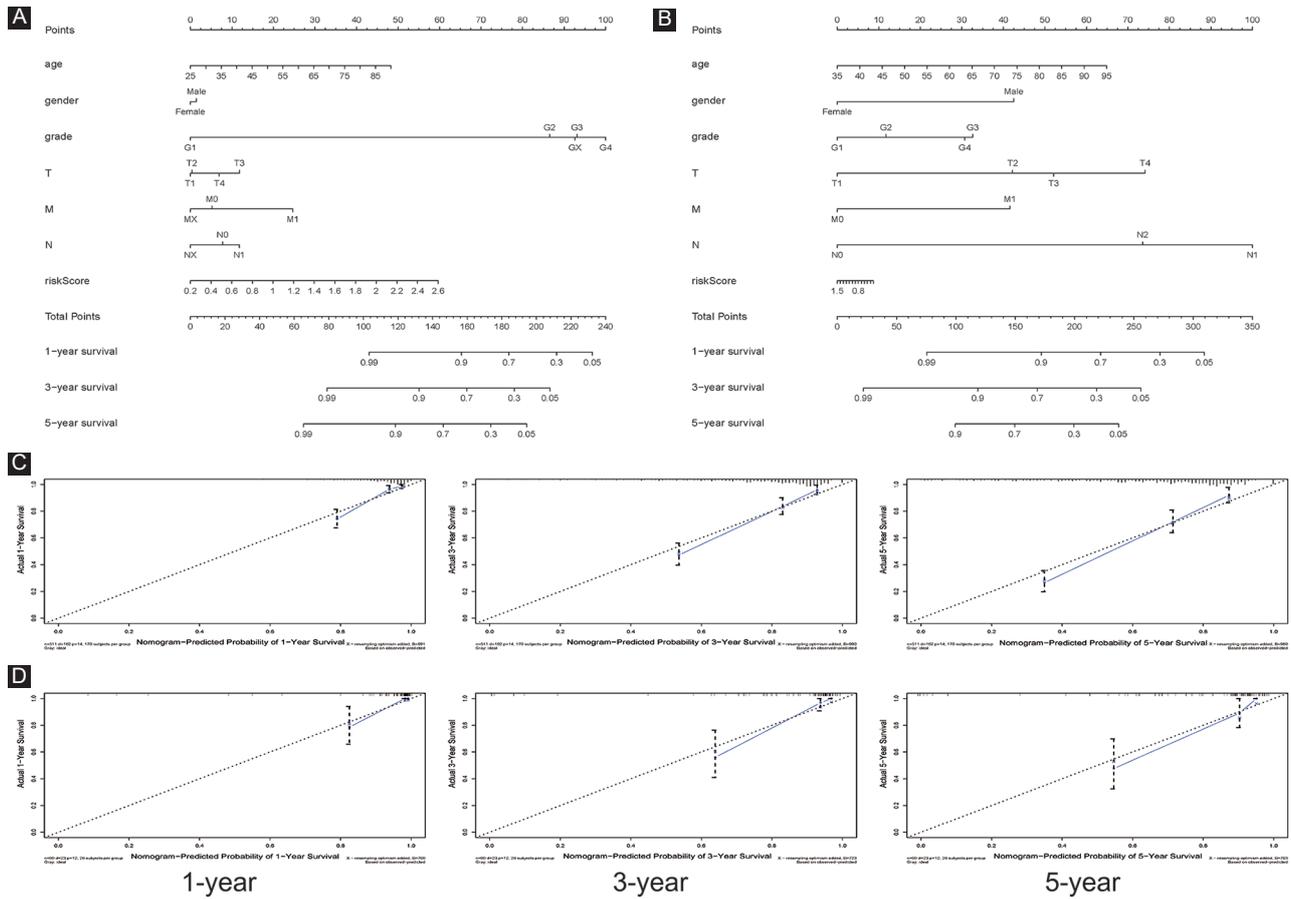
FIGURE 5. Nomogram and calibration plots in both TCGA and ArrayExpress databases; (A-B) nomogram in the TCGA and ArrayExpress databases, respectively; (C-D) calibration plot of 1-, 3-, and 5-year OS prediction in the TCGA and ArrayExpress databases, separately.

TABLE 5. Clinical correlation analysis between these six prognostic RBPs, our established risk score and clinical features

ID	Age	Gender	Race	Grade	Stage	T	M	N
CLK2	−0.752 (0.453)	1.485 (0.138)	6.75 (0.034)	−1.75 (0.081)	−3.094 (0.002)	−2.643 (0.009)	−2.764 (0.006)	0.077 (0.939)
IGF2BP2	−0.828 (0.409)	−0.696 (0.487)	6.183 (0.045)	−3.653 (2.911e−04)	−2.895 (0.004)	−3.077 (0.002)	−2.657 (0.009)	0.41 (0.682)
RNASE2	0.455 (0.649)	−1.429 (0.154)	9.985 (0.007)	−4.582 (6.12e−06)	−4.023 (7.549e−05)	−3.758 (2.187e−04)	−1.928 (0.056)	−0.386 (0.700)
EZH2	−0.157 (0.875)	−0.904 (0.366)	5.34 (0.069)	−5.535 (5.193e−08)	−5.348 (1.709e−07)	−4.874 (1.783e−06)	−3.901 (1.534e−04)	1.27 (0.205)
PABPC1L	−1.316 (0.189)	1.084 (0.279)	6.654 (0.036)	−4.553 (6.87e−06)	−4.048 (6.62e−05)	−3.817 (1.699e−04)	−3.142 (0.002)	−0.08 (0.936)
RPL22L1	−0.63 (0.529)	−1.81 (0.071)	1.338 (0.512)	−4.974 (9.155e−07)	−5.45 (1.035e−07)	−4.479 (1.107e−05)	−3.259 (0.001)	0.623 (0.534)
Risk score	−0.927 (0.355)	−0.431 (0.667)	2.07 (0.355)	−7.014 (8.921e−12)	−7.028 (1.421e−11)	−6.381 (7.907e−10)	−4.48 (1.662e−05)	0.349 (0.728)

The data in the brackets represent the *p*-value. The data outside the brackets represent the correlation coefficient. RBPs: RNA-binding proteins

(all $p < 0.001$; Figure 6M-R). qRT-PCR was employed to validate the mRNA expression of these six hub RBPs in four pairs of ccRCC tumor samples and adjacent normal kidney tissues. Based on our results, only RPL22L1 showed significant results ($p = 0.0067$). *p* values of CLK2, EZH2, IGF2BP2, PABPC1L, and RNASE2 were all above 0.054. This might be due to the relatively small number of samples (Figure 6S-X).

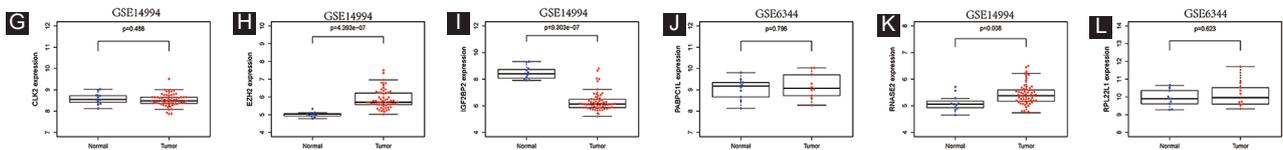## Verification of the protein expression of the critical RBPs in ccRCC

Due to the absence of PABPC1L protein in CPTAC and HPA datasets, only five proteins of CLK2, EZH2, IGF2BP2,

RNASE2, and RPL22L1 were analyzed. As presented in Figure 7A-E, these RBPs were differently expressed in ccRCC tumor samples compared with adjacent normal kidney tissues (all $p < 0.001$), except for EZH2 ($p = 0.962$). Besides, immunohistochemistry outcomes were used to validate the protein expression of these hub RBPs (Figure 7F-J). Antibody HPA055366 staining for CLK2 in normal kidney tissue was medium, whereas it was low in tumor tissue. Antibody CAB009589 staining for EZH2 in normal kidney tissue was not detected, whereas it was low in tumor tissue. Antibody HPA035145 staining for IGF2BP2 in normal kidney tissue was low, whereas it was medium in tumor tissue. Antibody
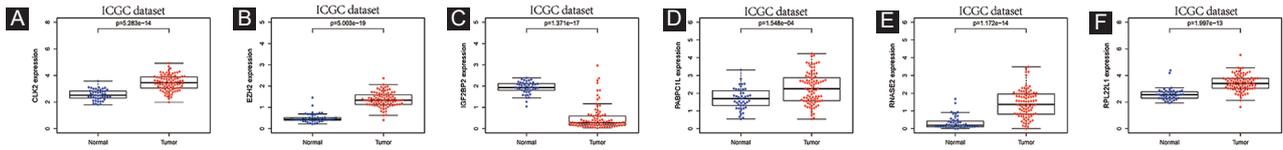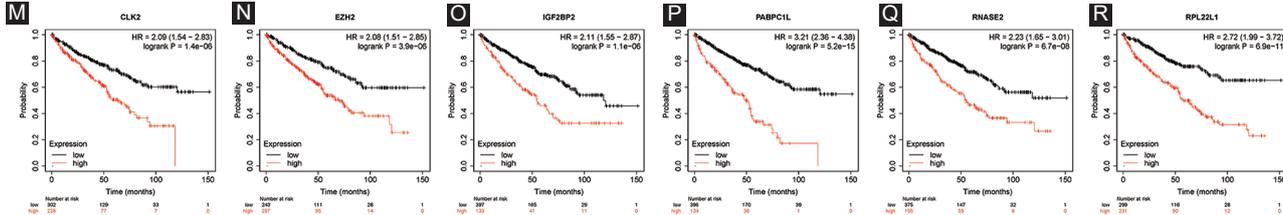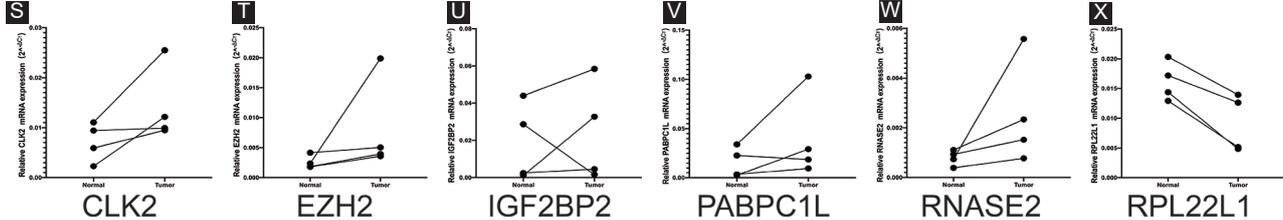
FIGURE 6. Validation of the mRNA expression and prognostic value of six critical RNA-binding proteins (RBPs) in ccRCC; boxplot of six critical RBPs mRNA expression in ICGC dataset (N=45; T=91); (A) CLK2; (B) EZH2; (C) IGF2BP2; (D) PABPC1L; (E) RNASE2; (F) RPL22L1; boxplot of six critical RBPs mRNA expression in GSE14994 dataset (N=11; T=59) and GSE6344 dataset (N=10; T=10); (G) CLK2, (H) EZH2, (I) IGF2BP2, (J) PABPC1L, (K) RNASE2, (L) RPL22L1; Scale=Gene count values; Kaplan–Meier plotter of (M) CLK2, (N) EZH2, (O) IGF2BP2, (P) PABPC1L, (Q) RNASE2, (R) RPL22L1; quantitative real-time PCR validation of (S) CLK2, (T) EZH2, (U) IGF2BP2, (V) PABPC1L, (W) RNASE2, (X) RPL22L1 mRNA expressions in clinical ccRCC samples (N=4; T=4).

HPA044983 staining for RNASE2 in normal kidney tissue was medium, whereas it was not detected in tumor tissue. Antibody HPA056207 staining for RPL22L1 in normal kidney tissue was medium, whereas it was low in tumor tissue. Scale bar for each IHC picture was 200 um.

Clinical factors stratified by our established signature for OS

According to our established model, ccRCC patients were further divided into subgroups for five clinical factors (stage, grade, N, T, and M). Our results shed light on that except for N1 ($p = 0.809$), our signature was able to predict OS in Grade 1-2, Grade 3-4, Stage III-IV, Stage I-II, T3-4 stage, T1-2 stage, N0, M0, and M1 ccRCC patients (all $p < 0.05$; Figure 8).

## TIICs stratified by our established model

As detailed in Figure 9A-L, 12 out of 21 TIICs (dendritic cells resting, B cells naive, macrophages M2, macrophages M0, monocytes, mast cells resting, T cells CD4 memory activated, plasma cells, T cells CD8, T cells CD4 memory resting,

T cells regulatory (Tregs), and T cells follicular helper) were all significantly stratified by our established model (all $p < 0.05$). Figure 9M summarizes all of the 21 TIICs in high-risk and low-risk groups stratified by our established model by radar chart.

DISCUSSION

In line with the latest cancer statistics reported from the World Health Organization, the occurrence rate of RCC dramatically increased over the past few decades and it was estimated to have over 140,000 ccRCC-related death events per year [24]. Therefore, it was important to identify reliable biomarkers for better predicting ccRCC patients' survival. With the development of cancer precision medicine, a variety of signatures or biomarkers have been established for predicting prognosis and therapeutic benefits. Studies illustrated that under the guidance of biomarkers, response rates seen with targeted agents have reached approximately 30%, which were much higher than that of chemotherapy [25]. As reported, RBPs dysregulation happens in the genesis and development

CPTAC protein validation
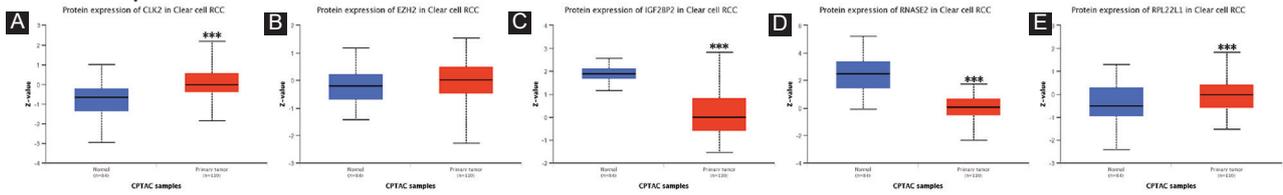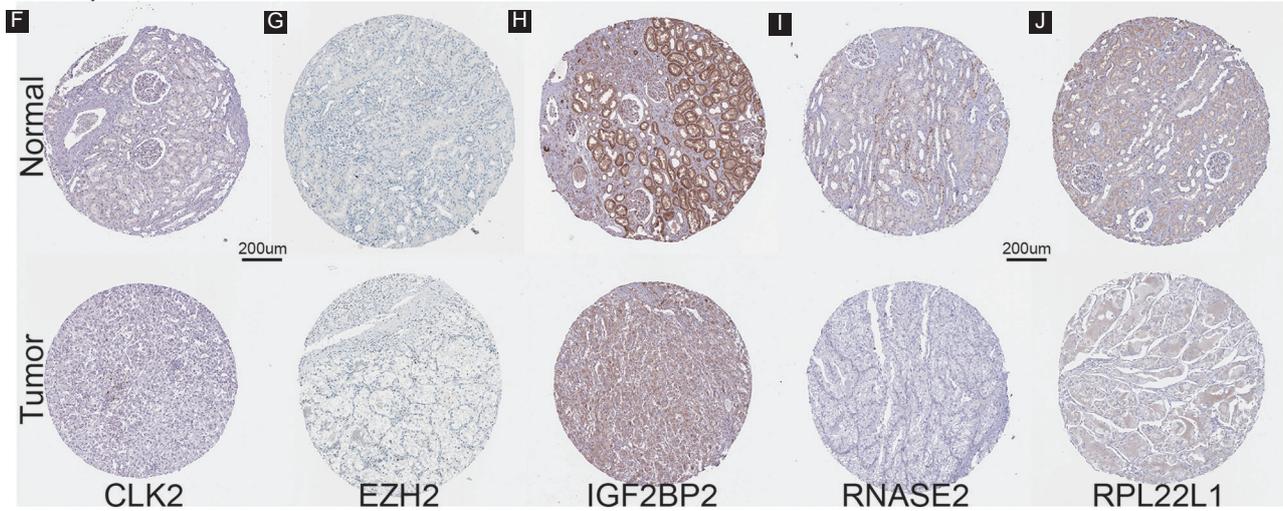

HPA protein validation

FIGURE 7. Verification of the protein expression of the critical RNA-binding proteins (RBPs) in ccRCC; boxplot of six critical RBPs protein expression by Clinical Proteomic Tumor Analysis Consortium analysis (A) CLK2; (B) EZH2; (C) IGF2BP2; (D) RNASE2; (E) RPL22L1; immunohistochemistry outcomes from Human Protein Atlas database (F) CLK2; (G) EZH2; (H) IGF2BP2; (I) RNASE2; (J) RPL22L1; scale bar=200 um; ***$p < 0.001$.
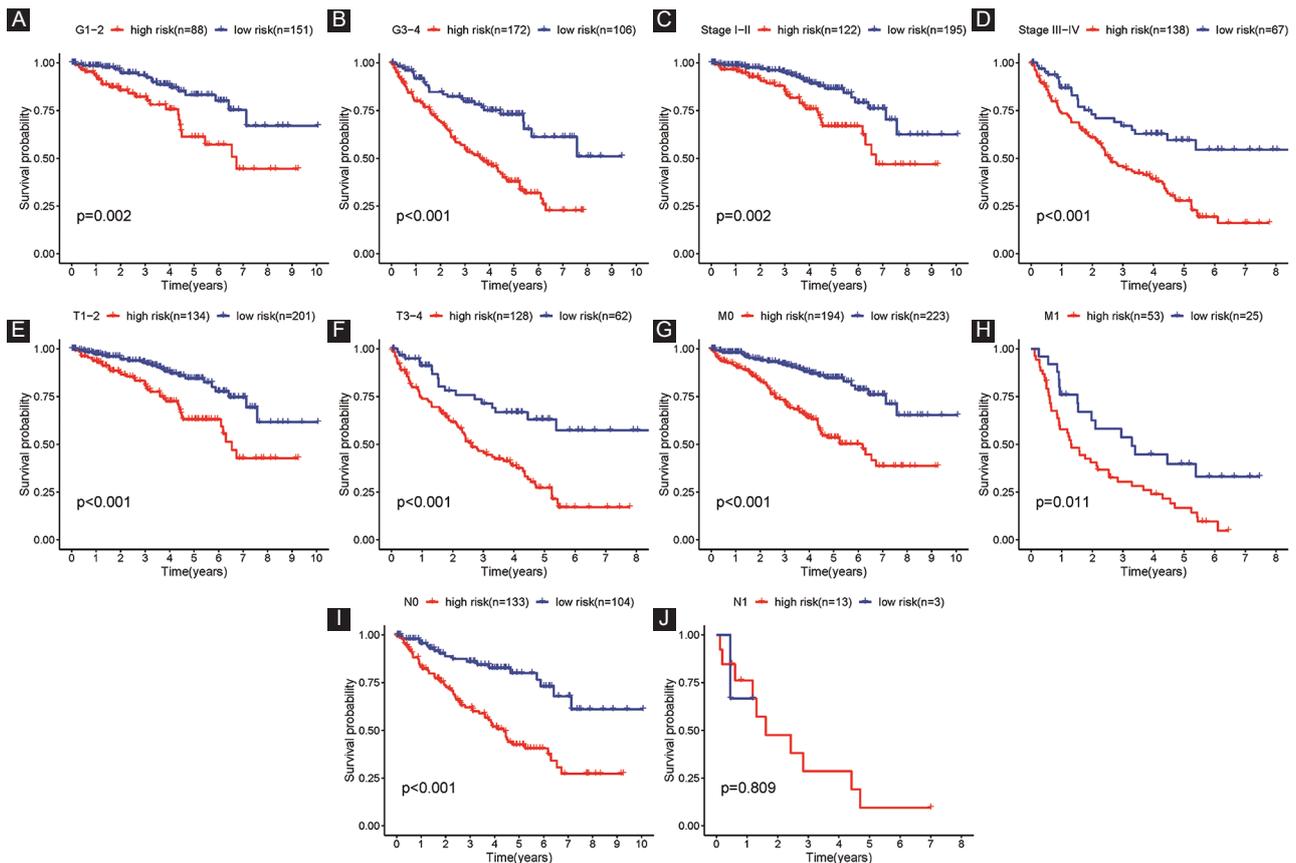

FIGURE 8. Clinicopathological parameters stratified by risk score for OS; (A) Grade 1-2 stratified by risk score for OS; (B) Grade 3-4; (C) Stage I-II; (D) Stage III-IV; (E) T1-2 stage; (F) T3-4 stage; (G) M0; (H) M1; (I) N0; (J) N1.
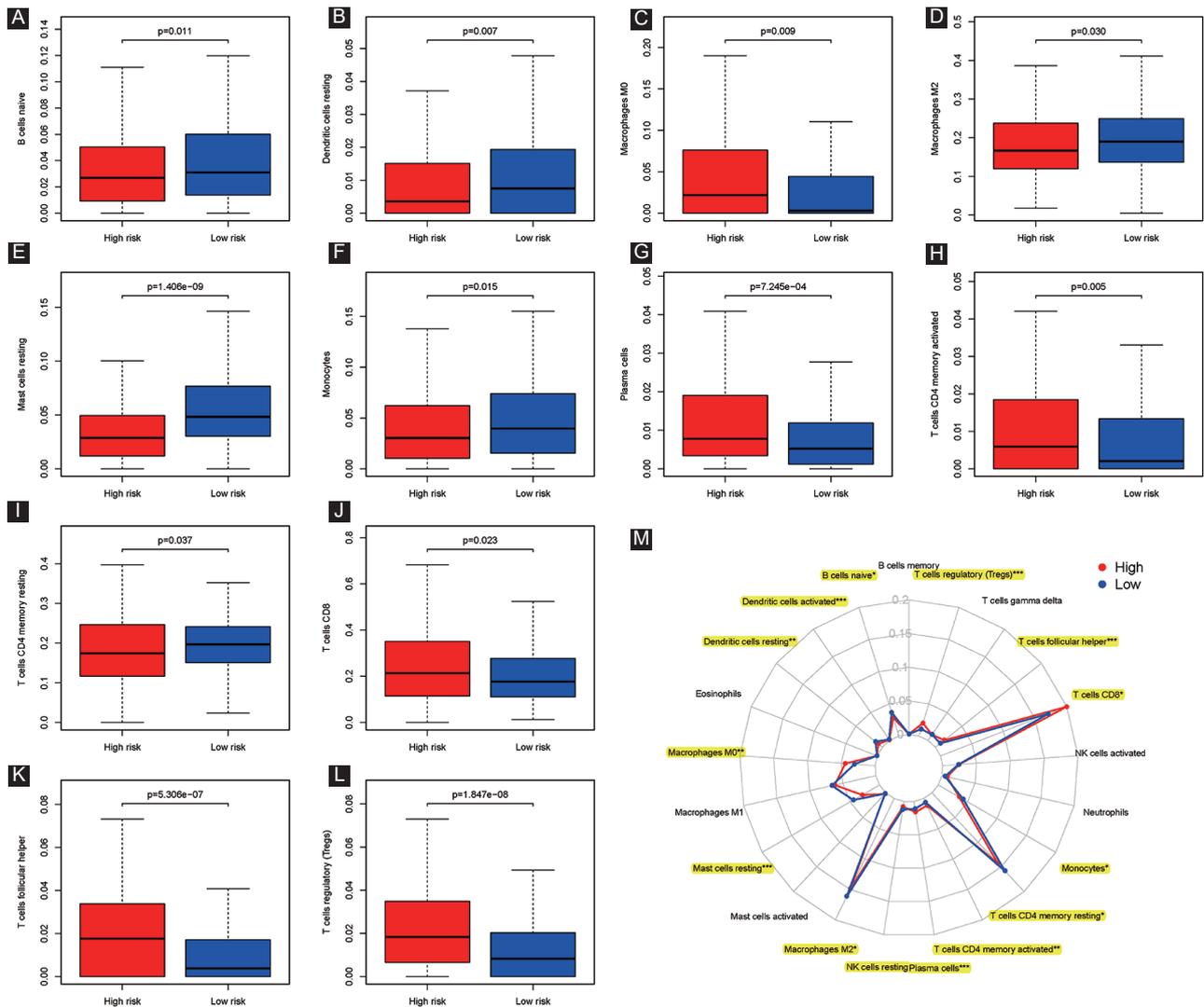
FIGURE 9. TIICs stratified by our established model; (A) B cells naive distribution; (B) dendritic cells resting; (C) macrophages M0; (D) macrophages M2; (E) mast cells resting; (F) monocytes; (G) plasma cells; (H) T cells CD4 memory activated; (I) T cells CD4 memory resting; (J) T cells CD8; (K) T cells follicular helper; (L) T cells regulatory (Tregs); (M) radar chart.

of various malignant tumors [14,26,27]. Nevertheless, pivotal functional roles of most RBPs in human cancer remain unclear [8,14]. Moreover, few studies focused on the roles of RBPs in ccRCC's survival prediction. Based on the ccRCC data from the TCGA set, 125 differently expressed RBPs between kidney and ccRCC tissues were selected. In addition, we adopted univariate COX and LASSO regression analysis to identify hub RBP genes and to build a prognostic signature. This work might contribute to identifying new effective biomarkers for the prognosis of ccRCC.

By means of univariate Cox and LASSO regression analysis, we singled out six hub RBPs (CLK2, IGF2BP2, RNASE2, EZH2, PABPC1L, and RPL22L1). Several studies demonstrated these RBPs played important roles in tumorigenesis and development, even in kidney cancer [14,28,29]. CLK2 could serve as an oncogene in breast cancer, whereas downregulation of CLK2 could suppress tumor growth [30]. Furthermore, CLK2 acted a pivotal part in the control of cell cycle and prognosis

of glioblastoma by regulating FOXO3a/p27 pathways [31]. As for IGF2BP2, it was reported that IGF2BP2 was differentially expressed in pancreatic cancer, and its upregulation promotes cancer cells' growth through stimulating the PI3K/Akt pathway [32]. Wan et al. figured out that RNASE2 was identified as the valuable prognostic predictor in ccRCC patients and utilized to explore the occurrence mechanisms of renal carcinoma and to design individualized treatments for patients [33]. EZH2, namely, the enzymatic subunit of polycomb repressive complex 2 [34,35], has been found to be of great importance in various cancers, including bladder cancer, breast cancer, prostate cancer, and so on [36-38]. Due to the ability of suppressing its enzymatic function, EZH2 became an anti-cancer therapy target and might bring a substantial therapeutic advance in clinical trials [39,40]. Regarding PABPC1L, Wu et al. demonstrated that PABPC1L suppressed migration and cell proliferation in colorectal cancer (CRC), and the expression of PABPC1L in CRC was highly associated with age,

pathologic node, pathologic metastasis, pathologic stage, and death [41]. With regard to RPL22L1, Wu et al. demonstrated that RPL22L1 induced epithelial-to-mesenchymal transition in OC and was critical in triggering cell metastasis and maintaining the aggressive phenotype of OC [42].

Next, we established the six RBPs signatures (risk score). The survival curves revealed that ccRCC patients in high-risk groups were associated with poor OS in the training cohort (TCGA) and three validation cohorts (E-MTAB-1980; internal validation cohort dataset 1; and internal validation cohort dataset 2). ROC analysis displayed that our six RBP genes based signature had a moderate performance for predicting ccRCC patients' OS. Moreover, our established signature could be an independent prognostic parameter of OS for ccRCC. It was worth noting that the *p*-value in the external validation dataset was more than 0.05 in the multivariate Cox regression analysis for the possible reason that its sample sizes were not large enough. Thus, it was acknowledged that the risk score had superior specificity and sensitivity in predicting ccRCC patients' OS.

In addition, a novel nomogram was constructed that integrated the risk score and several clinical factors (grade, gender, age, T, M, and N) to predict the ccRCC patients' OS. As for the calibration plot, very excellent outcomes were found in the TCGA cohort between the predicted and actual values. Similarly, a satisfactory agreement was also observed in the external validation dataset. In all, the novel nomogram might better help clinician predict ccRCC patients' survival status, improve risk stratification, and provide the individualized treatment than before.

In the study, we comprehensively explored the relationship between six RBPs, risk score, and different clinical factors in ccRCC. The results presented that the six RBPs based signature (risk score) was strongly associated with grade, tumor stage, M and T stage, and we also found that risk score would rise when clinicopathological factors (grade, tumor stage, M, and T stage,) increased. Taken together, RBPs might likely have malignant pathological implications in ccRCC, and these discoveries could provide novel insights into the underlying mechanism of RBPs in the progression of cancer. What's more, through characterizing RBP expression in ccRCC or risk score, novel therapeutic targets could be developed and survival could be predicted for ccRCC patients. A growing number of studies indicated that tumor immune infiltration played key roles in tumorigenesis and tumor progression, having an effect on immunotherapy [43-45]. Hence, in this article, we aimed to identify the associations between our established signature and tumor immune infiltration. The outcomes of us found that 12 out of 21 TIICs (dendritic cells resting, B cells naive, macrophages M2, macrophages M0, monocytes, mast cells resting, T cells CD4 memory activated, T cells regulatory

(Tregs), plasma cells, T cells CD8, T cells CD4 memory resting, and T cells follicular helper) were all significantly linked to high-risk and low-risk groups, indicating that immune infiltration was significantly related to our established model for ccRCC patients.

The strength of the article was that our established RBPs related signature was successfully established and evaluated in the other three validation sets (ArrayExpress cohort, the internal validation dataset 1, and the internal validation dataset 2). Moreover, six hub RBPs mRNA or protein expression were also validated by the ICGC, GEO, CPTAP, HPA datasets, and qRT-PCR verification, making our results more persuasive. However, several limitations should also be mentioned. First, our research was retrospective, thus the veracity and availability of six RBPs based signatures should be tested in other public databases, even in the prospective research. Second, two datasets (TCGA and E-MTAB-1980) were screened out of the study with no complete clinical parameters, which might decrease the statistical reliability of multivariate Cox regression analysis. Third, more clinico-pathological characteristics were required to be fetched into the novel prognostic nomogram, and additional biomarkers were needed to explore and identify. Fourth, in regard to the expression of IGF2BP2 mRNA, there are many reasons for the difference between the public database and our experiment. For example, the patient samples in the TCGA database are White, Asian, and Black or African-American, and most of them are white. However, in our study, all of the patients are Asian. In addition, the transcriptome profiling data from the TCGA cohort were the RNA-seq data, while the transcriptome profiling data from the GEO cohort were the microarray data which were produced with Illumina HumanHT-12 V4.0 Array, and in our study, we validated the expression level of IGF2BP2 mRNA by qRT-PCR. Besides, the tumor samples consisted of many mixed components. Therefore, there may be a discrepancy about the expression levels of several genes. Moreover, the study revealed that our six RBPs prognostic signature was significantly associated with the OS in ccRCC. However, it was merely analyzed and evaluated by data mining. Hence, more attention should be paid to uncover their roles in ccRCC by functional experiments. Despite the deficiencies described above, the predictive value of the signature in ccRCC patients could not be ignored. In the future, well-designed and multi-institutional studies were still required to verify our findings.

CONCLUSION

Taken together, our results successfully singled out six critical RBPs (CLK2, IGF2BP2, RNASE2, EZH2, PABPC1L, and RPL22L1) as a robust prognostic signature in ccRCC by

external and internal verification, helping clinician predict patients' survival status. Moreover, this signature could also be an independent prognostic factor for ccRCC. Besides, our established model displayed significant associations with immune infiltration. Further prospective studies were required to verify our established signature and to understand the roles of these six RBPs.

REFERENCES

[1] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. CA Cancer J Clin 2021;71(1):7-33.
https://doi.org/10.3322/caac.21654.

[2] Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M, et al. Renal cell carcinoma. Nat Rev Dis Primers 2017;3:17009.
https://doi.org/10.1038/nrdp.2017.9.

[[3] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA Cancer J Clin 2017;67(1):7-30.

[4] Yin L, Li W, Wang G, Shi H, Wang K, Yang H, et al. NR1B2 suppress kidney renal clear cell carcinoma (KIRC) progression by regulation of LATS 1/2-YAP signaling. J Exp Clin Cancer Res 2019;38(1):343.
https://doi.org/10.1186/s13046-019-1344-3.

[5] Li QK, Pavlovich CP, Zhang H, Kinsinger CR, Chan DW. Challenges and opportunities in the proteomic characterization of clear cell renal cell carcinoma (ccRCC): A critical step towards the personalized care of renal cancers. Semin Cancer Biol 2019;55:8-15.
https://doi.org/10.1016/j.semcancer.2018.06.004.

[6] Mickisch GH. Principles of nephrectomy for malignant disease. BJU Int 2002;89(5):488-95.
https://doi.org/10.1046/j.1464-410x.2002.02654.x

[7] Jadhav S, Ajay AK, Trivedi P, Seematti J, Pellegrini K, Craciun F, et al. RNA-binding protein Musashi homologue 1 regulates kidney fibrosis by translational inhibition of p21 and numb mRNA. J Biol Chem 2016;291(27):14085-94.
https://doi.org/10.1074/jbc.m115.713289.

[8] Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. Nat Rev Genet 2014;15(12):829-45.
https://doi.org/10.1038/nrg3813.

[9] Masuda K, 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.
https://doi.org/10.1002/wrna.1520.

[10] Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett 2008;582(14):1977-86.
https://doi.org/10.1016/j.febslet.2008.03.004.

[11] Torun A, Enayat S, Sheraj I, Tuncer S, Ulgen DH, Banerjee S. Butyrate mediated regulation of RNA binding proteins in the post-transcriptional regulation of inflammatory gene expression. Cell Signal 2019;64:109410.
https://doi.org/10.1016/j.cellsig.2019.109410.

[12] Jain A, Brown SZ, Thomsett HL, Londin E, Brody JR. 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.
https://doi.org/10.1007/978-1-4939-8879-2_22.

[13] Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. Nat Rev Mol Cell Biol 2018;19(5):327-41.
https://doi.org/10.1038/nrm.2017.130.

[14] Pereira B, Billaud M, Almeida R. RNA-binding proteins in cancer: Old players and new actors. Trends Cancer 2017;3(7):506-28.
https://doi.org/10.1016/j.trecan.2017.05.003.

[15] Wurth L, Papasaikas P, Olmeda D, Bley N, Calvo GT, Guerrero S, et al. UNR/CSDE1 drives a post-transcriptional program to promote melanoma invasion and metastasis. Cancer Cell 2019;36(5):337.

[16] https://doi.org/10.1016/j.ccell.2019.08.013.

[16] Hopkins TG, Mura M, Al-Ashtal HA, Lahr RM, Abd-Latip N, Sweeney K, 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.
https://doi.org/10.1093/nar/gkv1515.

[17] Palanichamy JK, Tran TM, Howard JM, Contreras JR, Fernando TR, Sterne-Weiler T, et al. RNA-binding protein IGF2BP3 targeting of oncogenic transcripts promotes hematopoietic progenitor proliferation. J Clin Invest 2016;126(4):1495-511.
https://doi.org/10.1172/jci80046.

[18] Wang Y, Chen SX, Rao X, Liu Y. Modulator-dependent RBPs changes alternative splicing outcomes in kidney cancer. Front Genet 2020;11:265.
https://doi.org/10.3389/fgene.2020.00265.

[19] Wang S, Li N, Yousefi M, Nakauka-Ddamba A, Li F, Parada K, et al. Transformation of the intestinal epithelium by the MSI2 RNA-binding protein. Nat Commun 2015;6:6517.
https://doi.org/10.1038/ncomms7517

[20] Degrauwe N, Schlumpf TB, Janiszewska M, Martin P, Cauderay A, Provero P, et al. The RNA binding protein IMP2 preserves glioblastoma stem cells by preventing let-7 target gene silencing. Cell Rep 2016;15:1634-47.
https://doi.org/10.1016/j.celrep.2016.04.086.

[21] Nagy Á, Lánczky A, Menyhárt O, Győrffy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. Sci Rep 2018;8:9227.
https://doi.org/10.1038/s41598-018-29514-3.

[22] Chen F, Chandrashekar DS, Varambally S, Creighton CJ. Pan-cancer molecular subtypes revealed by mass-spectrometry-based proteomic characterization of more than 500 human cancers. Nat Commun 2019;10(1):5679.
https://doi.org/10.1038/s41467-019-13528-0.

[23] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12(5):453-7.
https://doi.org/10.1038/nmeth.3337.

[24] Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136(5):E359-86.
https://doi.org/10.1002/ijc.29210.

[25] Schwaederle M, Zhao M, Lee JJ, Eggermont AM, Schilsky RL, Mendelsohn J, et al. Impact of precision medicine in diverse cancers: A meta-analysis of phase II clinical trials. J Clin Oncol 2015;33(32):3817-25.
https://doi.org/10.1200/jco.2015.61.5997.

[26] Lujan DA, Ochoa JL, Hartley RS. Cold-inducible RNA binding protein in cancer and inflammation. Wiley Interdiscip Rev RNA 2018;9(2):1462.
https://doi.org/10.1002/wrna.1462.

[27] Wu Y, Chen H, Chen Y, Qu L, Zhang E, Wang Z, et al. HPV shapes tumor transcriptome by globally modifying the pool of RNA binding protein-binding motif. Aging (Albany NY) 2019;11(8):2430-46.
https://doi.org/10.18632/aging.101927.

[28] Wang Z, Tang W, Yuan J, Qiang B, Han W, Peng X. Integrated Analysis of RNA-Binding Proteins in Glioma. Cancers 2020;12(4):892.
https://doi.org/10.3390/cancers12040892.

[29] Chénard CA, Richard S. New implications for the QUAKING RNA binding protein in human disease. J Neurosci Res 2008;86(2):233-42.

[30] Yoshida T, Kim JH, Carver K, Su Y, Weremowicz S, Mulvey L, et al. CLK2 is an oncogenic kinase and splicing regulator in breast cancer. Cancer Res 2015;75(7):1516-26.

[31] Park SY, Piao Y, Thomas C, Fuller GN, de Groot JF. Cdc2-like kinase 2 is a key regulator of the cell cycle via FOXO3a/p27 in glioblastoma. Oncotarget 2016;7(18):26793-805.
https://doi.org/10.18632/oncotarget.8471.

[32] Xu X, Yu Y, Zong K, Lv P, Gu Y. Up-regulation of IGF2BP2 by multiple mechanisms in pancreatic cancer promotes cancer proliferation

by activating the PI3K/Akt signaling pathway. J Exp Clin Cancer Res 2019;38:497.
https://doi.org/10.1186/s13046-019-1470-y.

[33] Wan B, Liu B, Huang Y, Yu G, Lv C. Prognostic value of immune-related genes in clear cell renal cell carcinoma. Aging (Albany NY) 2019;11(23):11474-89.
https://doi.org/10.18632/aging.10548.

[34] Müller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, et al. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell 2002;111(2):197-208.

[35] Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta A. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 2002;111:185-96.
https://doi.org/10.1016/s0092-8674(02)00975-3.

[36] Gui Y, Guo G, Huang Y, Hu X, Tang A, Gao S, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. Nat Genet 2011;43(9):875-8.
https://doi.org/10.1038/ng.907.

[37] Pugh TJ, Weeraratne SD, Archer TC, Pomeranz Krummel DA, Auclair D, Bochicchio J, et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature 2012;488(7409):106-10.
https://doi.org/10.1038/nature11329.

[38] Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 2015;518(7540):495-501.
https://doi.org/10.1038/nature14169.

[39] Gibaja V, Shen F, Harari J, Korn J, Ruddy D, Saenz-Vash V, et al. Development of secondary mutations in wild-type and mutant EZH2 alleles cooperates to confer resistance to EZH2 inhibitors. Oncogene 2016;35(5):558-66.
https://doi.org/10.1038/onc.2015.114.

[40] Kim KH, Roberts CW. Targeting EZH2 in cancer. Nat Med 2016;22(2):128-34.

[41] Wu YQ, Ju CL, Wang BJ, Wang RG. PABPC1L depletion inhibits proliferation and migration via blockage of AKT pathway in human colorectal cancer cells. Oncol Lett 2019;17(3):3439-45.
https://doi.org/10.3892/ol.2019.9999.

[42] Wu N, Wei J, Wang Y, Yan J, Qin Y, Tong D, et al. Ribosomal L22-like1 (RPL22L1) promotes ovarian cancer metastasis by inducing epithelial-to-mesenchymal transition. PLoS One 2015;10(11):e0143659.
https://doi.org/10.1371/journal.pone.0143659.

[43] Wang Y, Chen Y, Zhu B, Ma L, Xing Q. A novel nine apoptosis-related genes signature predicting overall survival for kidney renal clear cell carcinoma and its associations with immune infiltration. Front Mol Biosci 2021;8:567730.
https://doi.org/10.3389/fmolb.2021.567730.

[44] Nishikawa H, Koyama S. Mechanisms of regulatory T cell infiltration in tumors: Implications for innovative immune precision therapies. J Immunother Cancer 2021;9(7):e002591.
https://doi.org/10.1136/jitc-2021-002591.

[45] Cheng W, Xu B, Zhang H, Fang S. Lung adenocarcinoma patients with KEAP1 mutation harboring low immune cell infiltration and low activity of immune environment. Thorac Cancer 2021
https://doi.org/10.1111/1759-7714.14089.

Related articles published in BJBMS

1. The prognostic significance of different proportion of signet-ring cells of colorectal carcinoma
   Wei Chen et al., BJBMS, 2020

2. Development and validation of a ferroptosis-related lncRNAs prognosis signature in colon cancer
   Hua-jun Cai et al., BJBMS, 2020

3. Development and validation a novel preoperative comprehensive prognostic score based on inflammatory and nutritional score, coagulation indicator and tumor marker in esophageal squamous cell carcinoma
   Jifeng Feng et al., BJBMS, 2021

SUPPLEMENTAL DATA
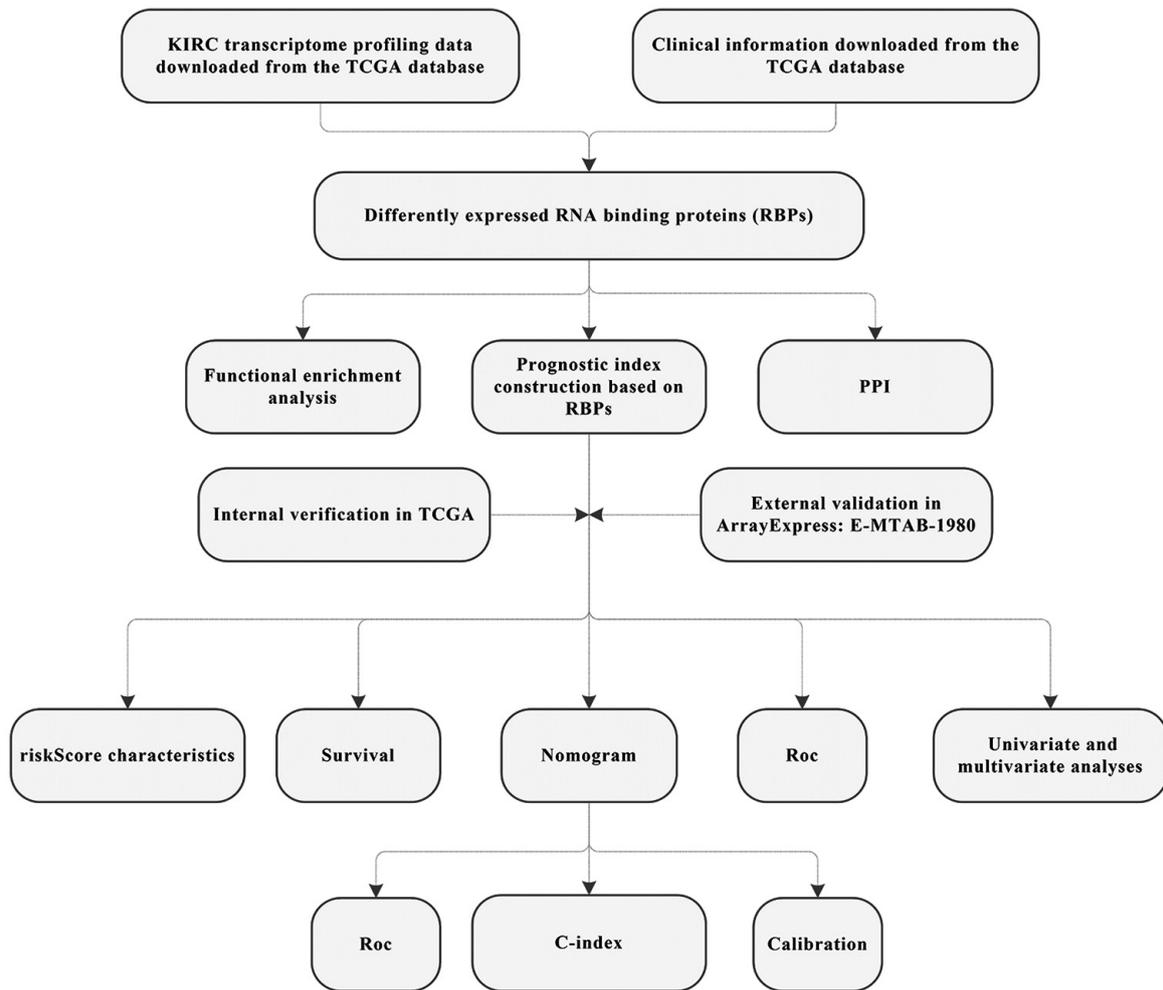
SUPPLEMENT FIGURE S1. The whole workflow chart of this study.
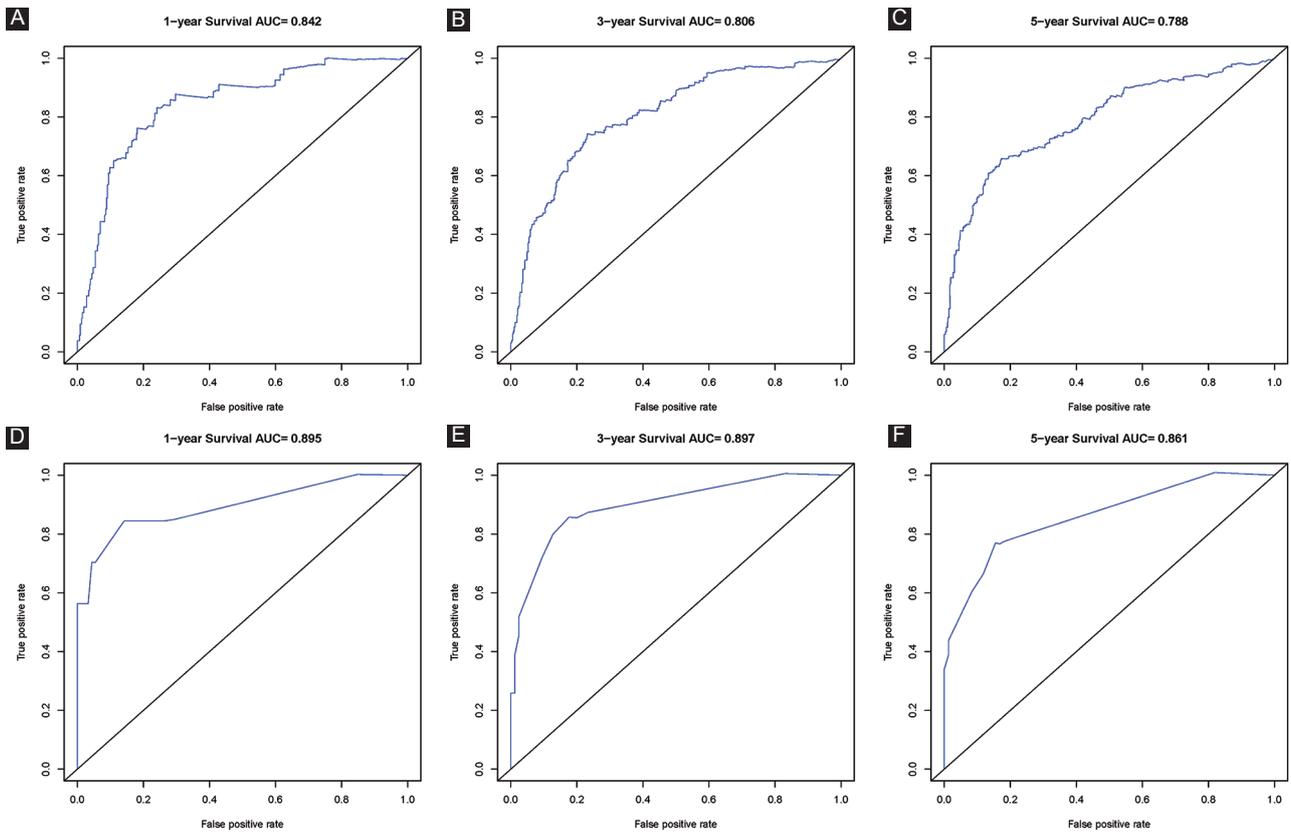
SUPPLEMENT FIGURE S2. One-, 3-, and 5-year ROC analyses of nomogram in both TCGA and ArrayExpress databases (E-MTAB-1980); (A-C) 1-, 3-, and 5-year in TCGA dataset; (D-F) 1-, 3-, and 5-year in ArrayExpress dataset (E-MTAB-1980).
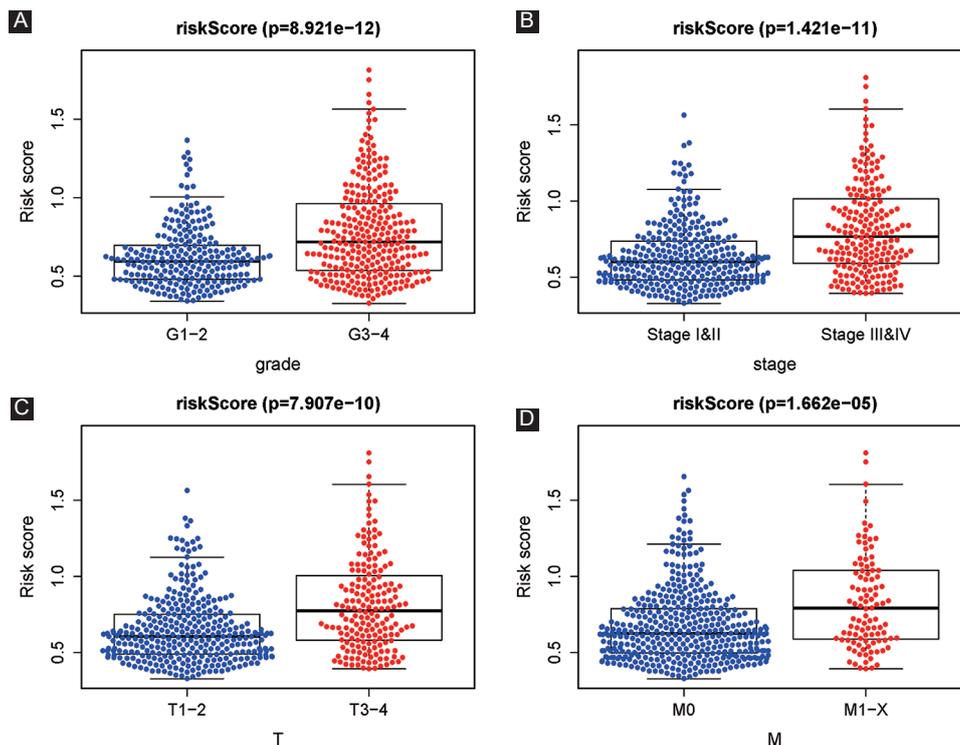
SUPPLEMENT FIGURE S3. Relationships between clinicopathologic characteristics and our established risk score; distribution of risk scores in (A) grade; (B) stage; (C) T stage; (D) M stage.

SUPPLEMENT TABLE 1. GO results of differently expressed RBPs;

GO results	ID	Description	GeneRatio	pvalue	p.adjust	qvalue	Count
BP	GO: 0008380	RNA splicing	31/114	4.69E-24	6.97E-21	6.08E-21	31
BP	GO: 0006397	mRNA processing	32/114	1.93E-23	1.43E-20	1.25E-20	32
BP	GO: 0006401	RNA catabolic process	22/114	4.28E-15	2.12E-12	1.85E-12	22
BP	GO: 1903311	regulation of mRNA metabolic process	19/114	3.24E-14	1.20E-11	1.05E-11	19
BP	GO: 0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	19/114	2.83E-13	7.01E-11	6.12E-11	19
BP	GO: 0000398	mRNA splicing, via spliceosome	19/114	2.83E-13	7.01E-11	6.12E-11	19
BP	GO: 0000375	RNA splicing, via transesterification reactions	19/114	3.33E-13	7.07E-11	6.17E-11	19
BP	GO: 0043484	regulation of RNA splicing	13/114	8.29E-13	1.54E-10	1.35E-10	13
BP	GO: 0051607	defense response to virus	16/114	1.13E-12	1.87E-10	1.64E-10	16
BP	GO: 0009615	response to virus	17/114	1.48E-11	2.21E-09	1.93E-09	17
CC	GO: 0036464	cytoplasmic ribonucleoprotein granule	14/116	1.36E-11	1.87E-09	1.50E-09	14
CC	GO: 0035770	ribonucleoprotein granule	14/116	2.53E-11	1.87E-09	1.50E-09	14
CC	GO: 0043186	P granule	5/116	4.91E-08	1.45E-06	1.17E-06	5
CC	GO: 0045495	pole plasm	5/116	4.91E-08	1.45E-06	1.17E-06	5
CC	GO: 0060293	germ plasm	5/116	4.91E-08	1.45E-06	1.17E-06	5
CC	GO: 0005681	spliceosomal complex	10/116	2.97E-07	7.34E-06	5.90E-06	10
CC	GO: 0033391	chromatoid body	4/116	6.70E-07	1.42E-05	1.14E-05	4
CC	GO: 0000932	P-body	7/116	9.00E-07	1.66E-05	1.34E-05	7
CC	GO: 0036020	endolysosome membrane	3/116	8.06E-05	0.001324915	0.001064832	3
CC	GO: 0005844	polysome	5/116	9.65E-05	0.001428211	0.001147851	5
MF	GO: 0140098	catalytic activity, acting on RNA	29/120	1.19E-22	2.37E-20	1.78E-20	29
MF	GO: 0004540	ribonuclease activity	13/120	4.96E-13	4.93E-11	3.71E-11	13
MF	GO: 0004518	nuclease activity	15/120	1.84E-11	1.22E-09	9.15E-10	15
MF	GO: 0003729	mRNA binding	19/120	3.05E-11	1.52E-09	1.14E-09	19
MF	GO: 0004004	ATP-dependent RNA helicase activity	9/120	7.40E-10	2.41E-08	1.81E-08	9
MF	GO: 0003724	RNA helicase activity	9/120	8.48E-10	2.41E-08	1.81E-08	9
MF	GO: 0008186	RNA-dependent ATPase activity	9/120	8.48E-10	2.41E-08	1.81E-08	9
MF	GO: 0003730	mRNA 3'-UTR binding	8/120	1.56E-08	3.88E-07	2.91E-07	8
MF	GO: 0008026	ATP-dependent helicase activity	9/120	2.28E-08	4.54E-07	3.41E-07	9
MF	GO: 0070035	purine NTP-dependent helicase activity	9/120	2.28E-08	4.54E-07	3.41E-07	9

SUPPLEMENT TABLE 2. KEGG results of differently expressed RBPs;

ID	Description	GeneRatio	pvalue	p.adjust	qvalue	Count
hsa03015	mRNA surveillance pathway	7/47	8.22E-07	4.28E-05	3.55E-05	7
hsa05164	Influenza A	7/47	5.16E-05	0.001286	0.001068	7
hsa03013	RNA transport	7/47	7.42E-05	0.001286	0.001068	7
hsa03010	Ribosome	6/47	0.000289	0.003763	0.003123	6
hsa05134	Legionellosis	3/47	0.004397	0.045729	0.037953	3