


RNA Binding Protein-Based Model for Prognostic Prediction of Colorectal Cancer

Technology in Cancer Research & Treatment
Volume 20: 1-14
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15330338211019504
journals.sagepub.com/home/tct


Ting Li, PhD¹ , Wenjia Hui, MD¹, Halina Halike, MD¹, and Feng Gao, PhD¹

Abstract

Background: Colorectal cancer (CRC) is a kind of gastrointestinal tumor with serious high morbidity and mortality. Several reports have implicated the disorder of RNA-binding proteins (RBPs) in plenty of tumors, associating it to tumorigenesis and disease progression. The study is intended to construct novel prognostic biomarkers associated with CRC patients. **Methods:** Data of gene expression was acquired from the TCGA database, prognosis-related genes were selected. Besides, we analyzed GO and KEGG pathways. Univariate and multivariate Cox analyses were performed to generate a prognostic-related gene signature, which was evaluated by the Kaplan-Meier (K-M) and the Receiver Operating Characteristic (ROC) curve. The independent prognostic factor was established by survival analysis. GSE38832 dataset was used to validate the signature. Finally, expression of 8 genes was further confirmed by qRT-PCR in SW480 and SW620 cell lines. **Results:** We obtained 224 differentially expressed RBPs in total, of which 78 were downregulated and 146 were upregulated. Univariate COX analysis was conducted in the TCGA cohort to select 13 RBPs with $P < 0.005$, stepwise multivariate COX regression analysis was used to construct an 8-RBP signature (TERT, PPARGC1A, BRCA1, CELF4, TDRD7, LUZP4, PNLDC1, ZC3H12C). Based on the model, systematic analysis illustrated that a high risk score was obviously connected to a poor prognosis. The prognostic value of the risk score was validated in GSE38832 dataset, indicating that the risk model was accurate and effective. The prognostic signature-based risk score was identified as an independent prognostic indicator for CRC. The expression results of qRT-PCR were consistent with the results of differential expression analysis. **Conclusions:** The eight-RBP signature can predict the survival of CRC patients and potentially act as CRC prognostic biomarker.

Keywords

RNA binding protein, colorectal cancer, prognostic signature, survival analysis

Abbreviations

CRC, colorectal cancer; GEO, Gene Expression Omnibus; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; OS, Overall survival; PPI, Protein-protein interaction; RBPs, RNA binding proteins; ROC, Receiver operating characteristic; TCGA, The Cancer Genome Atlas

Received: August 21, 2020; Revised: March 30, 2021; Accepted: April 28, 2021.

Introduction

Colorectal cancer (CRC) is the leading cause of cancer mortality and morbidity. Individuals with colorectal cancer generally have a survival rate of fewer than 5 years because of early metastasis. Even though the rapid development of treatments such as radiotherapy, surgery, targeted therapies and chemotherapy, the poor prognosis and high recurrence rate remain a concern.¹

¹ Department of Gastroenterology, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, Xinjiang Province, China

Corresponding Author:

Feng Gao, PhD, Department of Gastroenterology, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi 830000, Xinjiang Province, China.
Email: xjgf@sina.com



RNA binding proteins function in conjunction with several different RNA types, involving tRNAs, mRNAs, rRNAs, miRNAs, ncRNAs, snoRNAs and snRNAs. They either directly interact with RNA or act as a constituent of a ribonucleoprotein complex indirectly associated with RNA. Regulating the epithelial homeostasis, injury response and malignant transformation of intestinal epithelial cells by RNA binding protein (RBPS) is an emerging research hotspot.² At present, there are over 1500 experimentally validated RBP-coding genes, accounted for a large proportion of all protein-coding genes.³ RBP is essential for regulating many basic cellular processes, including RNA splicing, modification, degradation, stability, localization, modification, translation, and transport. RBPs can bind a particular target RNA to form ribonucleoprotein (RNP) complexes and regulating gene expression after transcription.⁴ RBP is crucial in post-transcriptional modulation that subsequently participates in several pathways, including apoptosis, differentiation, angiogenesis, proliferation and migration.⁵ Recent evidence suggests that RBPs participate in the pathogenesis of cardiovascular disease as key regulators, such as HUR, SRSF1, MUR, and Quaking.^{6,7}

Recent evidence about RBPs and colorectal cancer suggests about 30% of colorectal cancer patients over-express Lin28b.^{8,9} Human colorectal cancers highly express IMP1,¹⁰ which is linked to lymph node metastasis, poor prognosis, TNM stage, and tumor size.¹¹ Overexpression of MSI can activate mTORC1 complex with inhibition of PTEN to transform intestinal epithelial cells and form tumors.^{12,13} Moreover, there is high HUR protein expression in colon cancer cell cytoplasm and nucleus¹⁴ and is underexpressed in healthy colon cells.¹⁵ By up-regulating the expression of zinc finger E-box binding homeobox 1 (ZEB1) mRNA, forkhead box K2 protein (FOXK2) endorses CRC cells migration.¹⁶ In conclusion, RBP might play a crucial role in the regulation of cancer. Nevertheless, most roles of RBP have not been studied. A comprehensive analysis of RBPs will contribute to our understanding of their role in tumors.

Here, we obtained the RNA-SEQ data and clinicopathological data from the TCGA database to identify prognosis-related RBPs. Survival-related RBPs were screened by univariate and multivariate COX regression, and their potential function and clinical significance were systematically explored. The study was designed and conducted in strict conformed to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement and the Reporting Recommendations for Tumor Marker Prognostic Studies.^{17,18} The findings of the present study might be possibly beneficial in the development of prognosis biomarkers. The established signature could be regarded as a novel independent prognostic factor that has a pivotal role in predicting the prognostic of CRC patients.

Materials and Methods

Data Source and Preprocessing

The source of gene expression profile and relevant clinical data were obtained from the Cancer Genome Atlas (TCGA) dataset.

Table 1. Primer Sequences for 5 Hub Genes.

Gene		Primer sequence
TERT	Forward:	TTGGTGGATGATTTCTTGTG
	Reverse:	GGTGAGACTGGCTCTGATGG
PPARGC1A	Forward:	GAGCAATAAAGCGAAGAG
	Reverse:	GTGTTGTGACTGCGACTG
BRCA1	Forward:	GCTGTGCTCATACTACTG
	Reverse:	TTTGTGACCCTTTCTGT
CELF4	Forward:	CTTTCCTCAGCCGCTCCA
	Reverse:	TGCATCAGCTCAGCGTCCC
TDRD7	Forward:	AGCAACCCTCAGACAACC
	Reverse:	GCATCAGGCTTAACTCCA
LUZP4	Forward:	TTTCGGAAGCTAACGCTTTCT
	Reverse:	CCGATGGCGATGTCTATGAGC
PNLDC1	Forward:	TTGAATCCCACCAAGAAT
	Reverse:	GAGGAAGGCATCATACGC
ZC3H12C	Forward:	CCACGAGAATAGACAGCATC
	Reverse:	AGTTATCGGGCAAGGAAT
U6	Forward:	CTCGTTCGGCACA
	Reverse:	AACGCTTCACGAATTGT

To ensure high-quality analyses, patients with missing or incomplete data or survival time less than 14 days were excluded. Limma package was used to analyze all the original information and genes that had an average count value < 1 were excluded, Wilcox test was utilized to test samples. Genes were identified for follow-up analyses which logFC > 0.5 and FDR < 0.05. “pheatmap” R package was used to visualize the various RBPs expression patterns.

Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Oncology (GO) Analyses

To determine the function of differently expressed RBPs, GO and KEGG analyses were completed. The org.Hs.eg.db (version 3.7.0) and clusterProfiler (version 3.10.1)¹⁹ were used to carry out all the KEGG pathway and GO analyses.

PPI Network Construction and Module Selection

STRING website (<http://string-db.org>) was utilized for analyzing the protein-protein interaction (PPI) among the differently expressed RBPs. Cytoscape 3.7.0 software was used in network construction and visualization. After that, key modules with node counts > 5 and scores > 7 were chosen via MCODE (Molecular Complex Detection) plug-in in PPI network. *P*-values < 0.05 were treated as statistically significant.

Assembly and Verification of the 8-RBP Signature

The evaluation of key modules survival-associated genes was completed via Univariate Cox regression. Based on the aforementioned primary selected survival-related genes, the TCGA cohort was subjected to multivariate Cox for optimal model

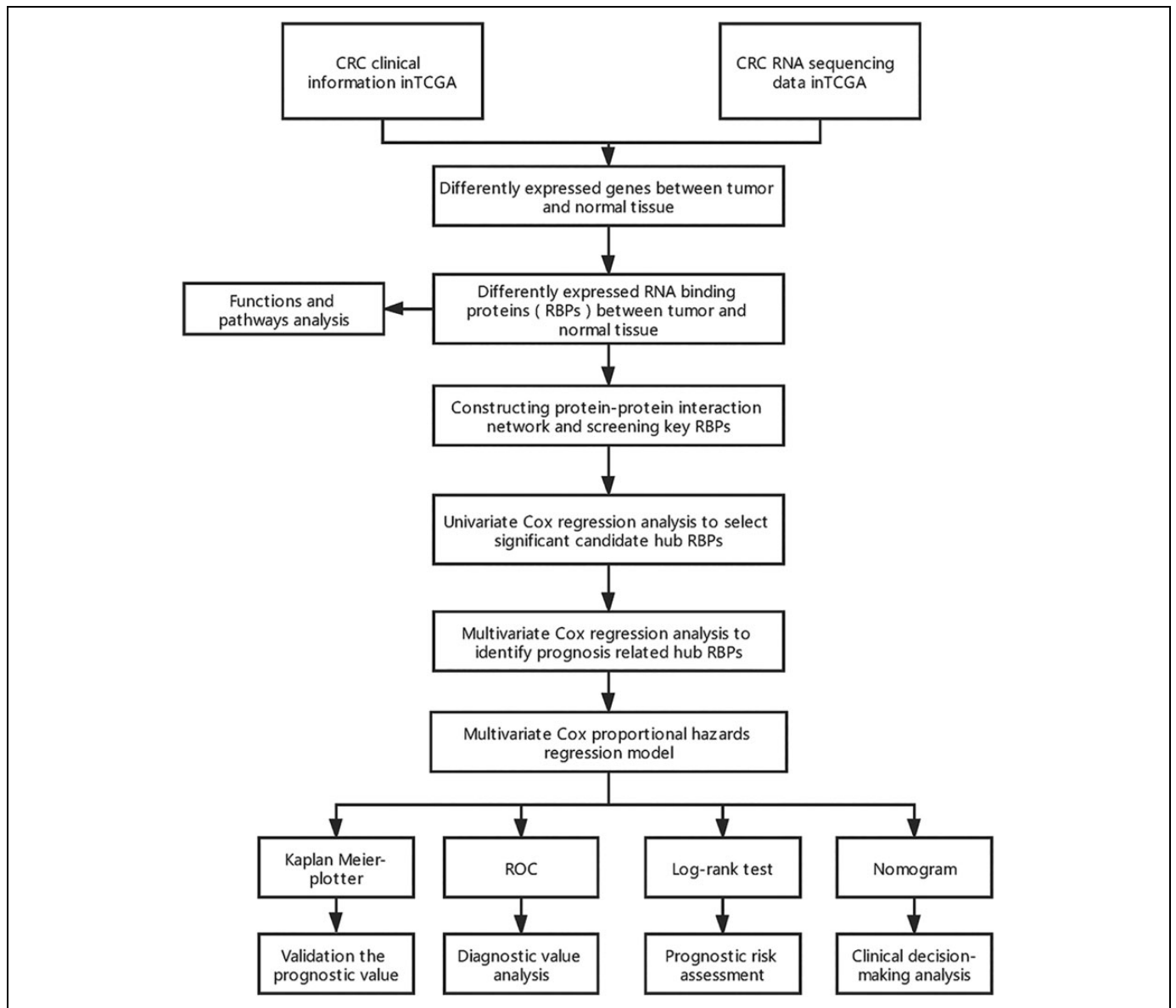


Figure 1. Framework for analyzing the RBPs in CRC.

determination. A risk score for every patient was computed as the sum of each gene's score as follows:

$$Riskscore = \sum_{i=1}^n Exp_i \beta_i$$

β refers to the coefficient value; Exp serves as the level of expressed genes. The CRC patients were grouped into low-risk and high-risk based on median risk score. Next, the difference in OS among the subgroups was determined by Kaplan-Meier (KM) and log-rank methods through the “survival” R package. Moreover, model predictive power was evaluated by calculating the AUC of 1-year, 3-year, and 5-year time-dependent ROC curve using “survivalROC” package.²⁰ Afterward, The GSE38832 dataset was used to further validate the prognostic value of the risk signature.

Determination of the Independent Prognostic Capacity of the Multi-Gene Signature

The Univariate and multivariate Cox regressions were utilized to identify whether the risk score and the respective clinicopathological properties were independent prognostic aspects. $P < 0.05$ being considered meaningfully.

Development of the Nomogram

A nomogram was used to predict CRC prognosis. The nomogram was established by “rms” R package and included all feature genes that had a significant association with OS. After that, we plot Calibration to assess the difference in actual versus nomogram predicted OS.

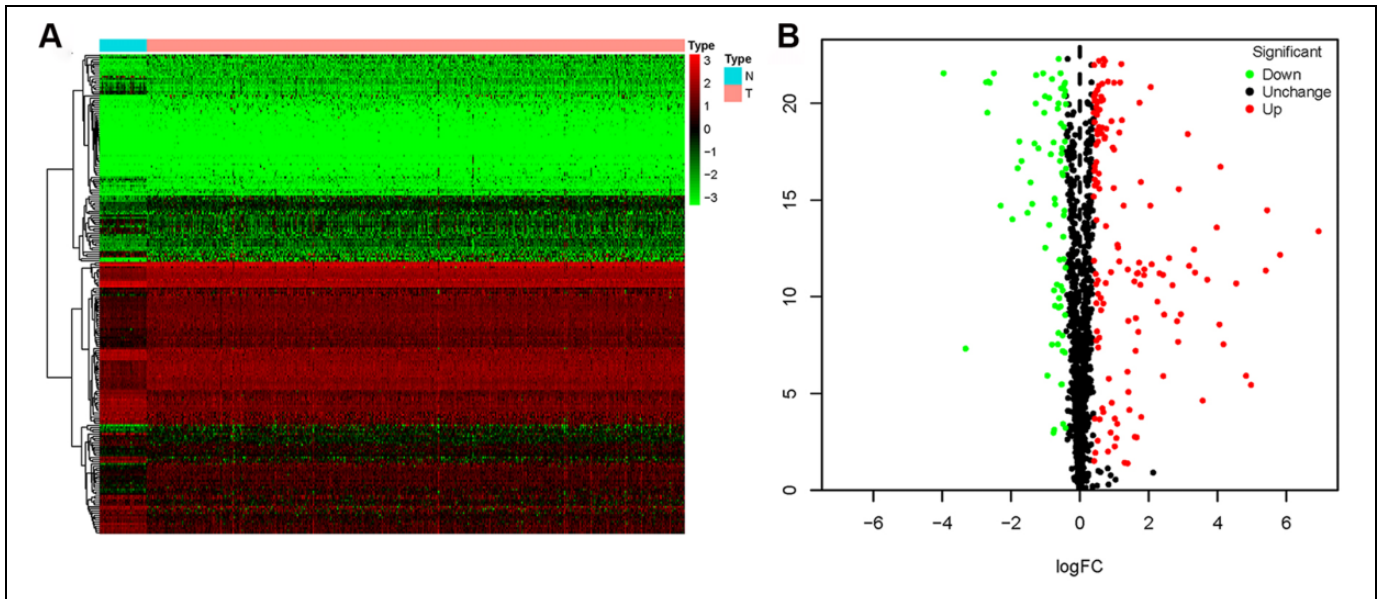


Figure 2. Heat map and volcano plot of differentially expressed RBPs. (A) Heat map; (B) Volcano plot.

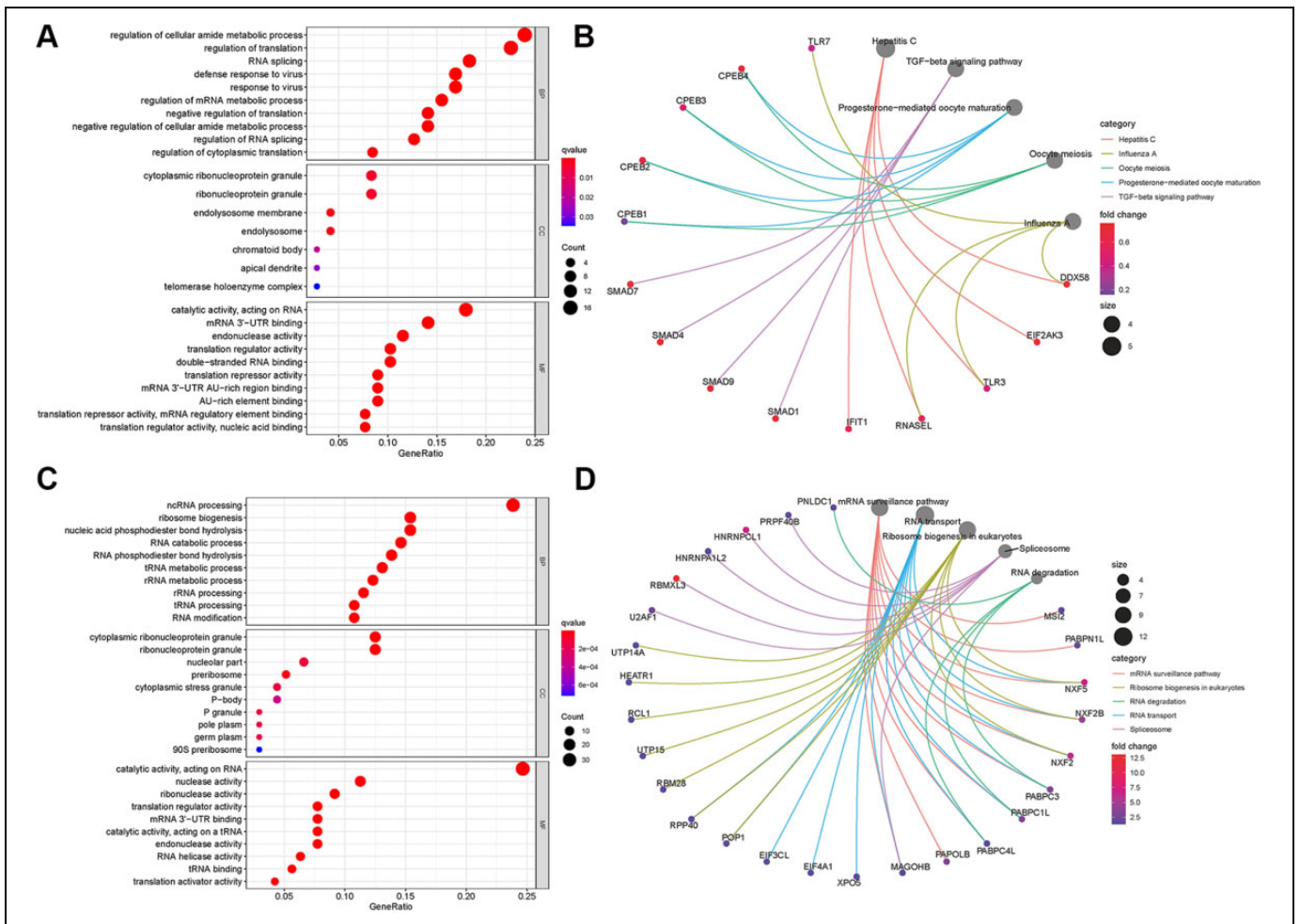


Figure 3. The GO and KEGG analysis of differentially expressed RBPs. A, GO analysis for downregulated RBPs. B, KEGG pathway analysis for downregulated RBPs. C, GO analysis for upregulated RBPs. D, KEGG pathway analysis for upregulated RBPs.

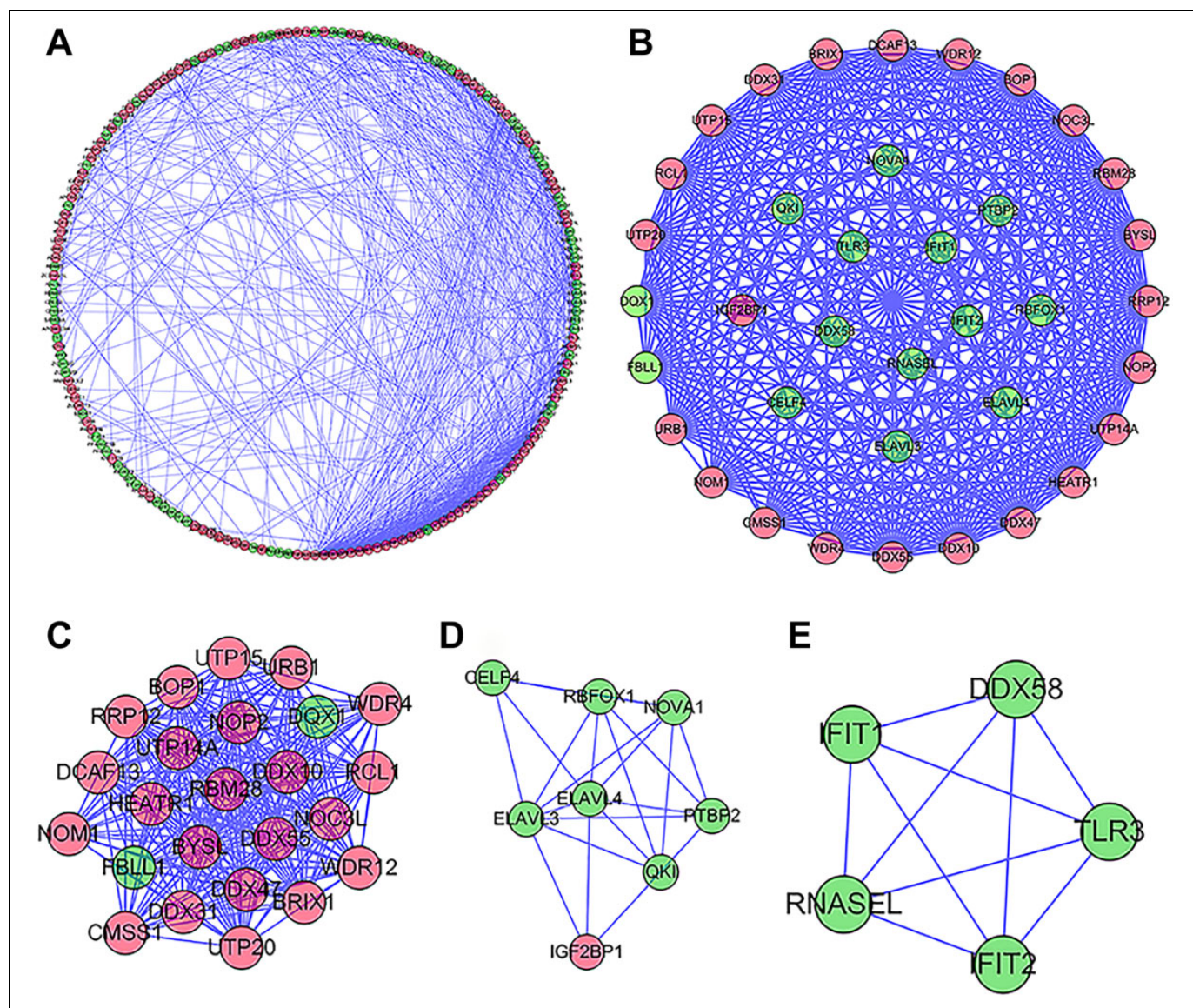


Figure 4. The PPI network and their modules. (A) PPI network of differentially expressed RBPs; (B) The key PPI network module. Green circles: down-regulation at a fold change above 2; red circles: up-regulation at a fold change above 2. (C) Key module 1 in PPI network. (D) Key module 2 in PPI network. (E) Key module 3 in PPI network.

Cell Culture and qRT-PCR

Human CRC cell lines with high metastatic potential (SW620) and low metastatic potential (SW480) were obtained from an American-type culture collection (Manassas, VA, USA). DMEM medium with 10% FBS (Invitrogen) in a humidified incubator at 37 °C and 5% CO₂. Total RNA from cell lines was extracted by TRIzol reagent. cDNA was synthesized using cDNA Synthesis Kits (Toyobo, FSQ-301). qRT-PCR was carried out on the iQ5 Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA). Primer sequences were listed in Table 1. U6 was used as the internal reference. The qRT-PCR relative quantitative method was used to analyze the experimental results. Statistical analyses were performed with GraphPad Prism 8.0.1 software. The expression levels of 8

genes were analyzed by unpaired t-test, and $P < 0.05$ indicated that the difference was statistically significant.

Results

The RBPs Differential Expression in CRC

Herein, a comprehensive investigation on significant functions and prognostic features of RBPs was undertaken. The study procedure was represented in Figure 1. Based on Wilcoxon test, we identified 469 and 41 tumor and normal colon tissue samples, respectively. Regarding RBPs, 78 were considerably upregulated, whereas 146 were downregulated (Figure 2).

Table 2. KEGG and GO Analysis of Sub-Pathway.

Module 1	ID	Description	<i>P</i> adjust	q value	Count
BP	GO:0042254	Ribosome biogenesis	8.03E-26	3.42E-26	17
BP	GO:0006364	rRNA processing	2.00E-21	8.53E-22	14
BP	GO:0016072	rRNA metabolic process	1.47E-20	6.26E-21	14
CC	GO:0030684	Preribosome	2.81E-13	1.48E-13	8
CC	GO:0032040	Small-subunit processome	5.97E-09	3.14E-09	5
CC	GO:0030686	90 S preribosome	1.99E-07	1.05E-07	4
MF	GO:0140098	RNA catalysis	2.79E-08	1.47E-08	9
MF	GO:0003724	RNA helicase activity	8.14E-07	4.28E-07	5
MF	GO:0004386	Helicase activity	2.15E-05	1.13E-05	5
KEGG	hsa03008	Ribosome biogenesis in eukaryotes	6.38E-12	NA	6
Module 2					
MF	GO:0003730	mRNA 3'-UTR binding	7.95E-05	1.52E-05	3
MF	GO:0035925	mRNA 3'-UTR AU-rich region binding	0.000246	4.72E-05	2
MF	GO:0017091	AU-rich element binding	0.000246	4.72E-05	2
BP	GO:0008380	RNA splicing	5.10E-05	2.33E-05	5
BP	GO:0000377	Regulation of RNA splicing	0.00028	0.000128	4
BP	GO:0000398	Dysregulation of RNA splicing	0.00028	0.000128	4
CC	GO:0030426	Growth cone	0.009286	0.006109	2
CC	GO:0030427	Site of polarized growth	0.009286	0.006109	2
CC	GO:0150034	Distal axon	0.016029	0.010546	2
KEGG	hsa05206	MicroRNAs in cancer	0.038552	NA	1
Module 3					
BP	GO:0051607	Defense response to virus	7.81E-08	2.07E-08	5
BP	GO:0009615	Response to virus	1.91E-07	5.06E-08	5
BP	GO:0071360	Cellular response to exogenous dsRNA	5.05E-07	1.34E-07	3
KEGG	hsa05160	Hepatitis C	1.86E-06	1.12E-06	4
KEGG	hsa05164	Influenza A	0.000256	0.000154	3
KEGG	hsa05168	Herpes simplex virus 1 infection	0.004033	0.002426	3
KEGG	hsa05161	Hepatitis B	0.008249	0.004962	2

Abbreviations: BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Differently Expressed RBPs GO and KEGG Pathway Enrichment

To annotate mechanisms and function of screened RBPs, differently expressed RBPs were sorted into 2 sub-sets (the down-regulated and upregulated groups) and then subjected to GO and KEGG pathway analyses. The outcomes demonstrated that the genes in the downregulated group were markedly enriched in the biological process (BP) and correlated with regulation of cellular amide metabolic process, regulation of translation, RNA splicing (Figure 3A). The upregulated genes were notably enhanced in terms of ncRNA production, nucleic acid phosphodiester bond hydrolysis, and ribosome biogenesis (Figure 3C). Regarding the analysis of cellular component (CC), there was abundance of downregulated RBPs in the cytoplasmic ribonucleoprotein granule, endolysosome membrane, and ribonucleoprotein granule (Figure 3A). However, elevated RBPs were predominant in the cytoplasmic ribonucleoprotein granule, nucleolar part, and the ribonucleoprotein granule (Figure 3C). The molecular function (MF) revealed downregulated RBPs were high in the catalytic activity, acting on RNA,

mRNA 3'-UTR binding, endonuclease activity (Figure 3A). In contrast, the elevated RBPs were mostly high in catalytic activity, acting on RNA, nuclease activity, ribonuclease activity (Figure 3C).

Overall, the KEGG pathways exhibited 6 effects from downregulated differently expressed RBPs (Figure 3B). These included enrichment in the Hepatitis C, TGF-beta signaling pathway, Progesterone-mediated oocyte maturation, Oocyte meiosis. The elevated differently expressed RBPs were associated with mRNA surveillance pathway, RNA transport, Ribosome biogenesis in eukaryotes (Figure 3D).

Protein-Protein Interaction (PPI) System Assembly and the Key Modules

The mutual interaction among the differently expressed RBPs in CRC was investigated thoroughly via the PPI linkages of the chosen genes. The linkages were constructed from the String database, using Cytoscape software to form the PPI network (Figure 4A). The co-expression network was performed to find the possible key modules via the MODE tool (Figure 4B).

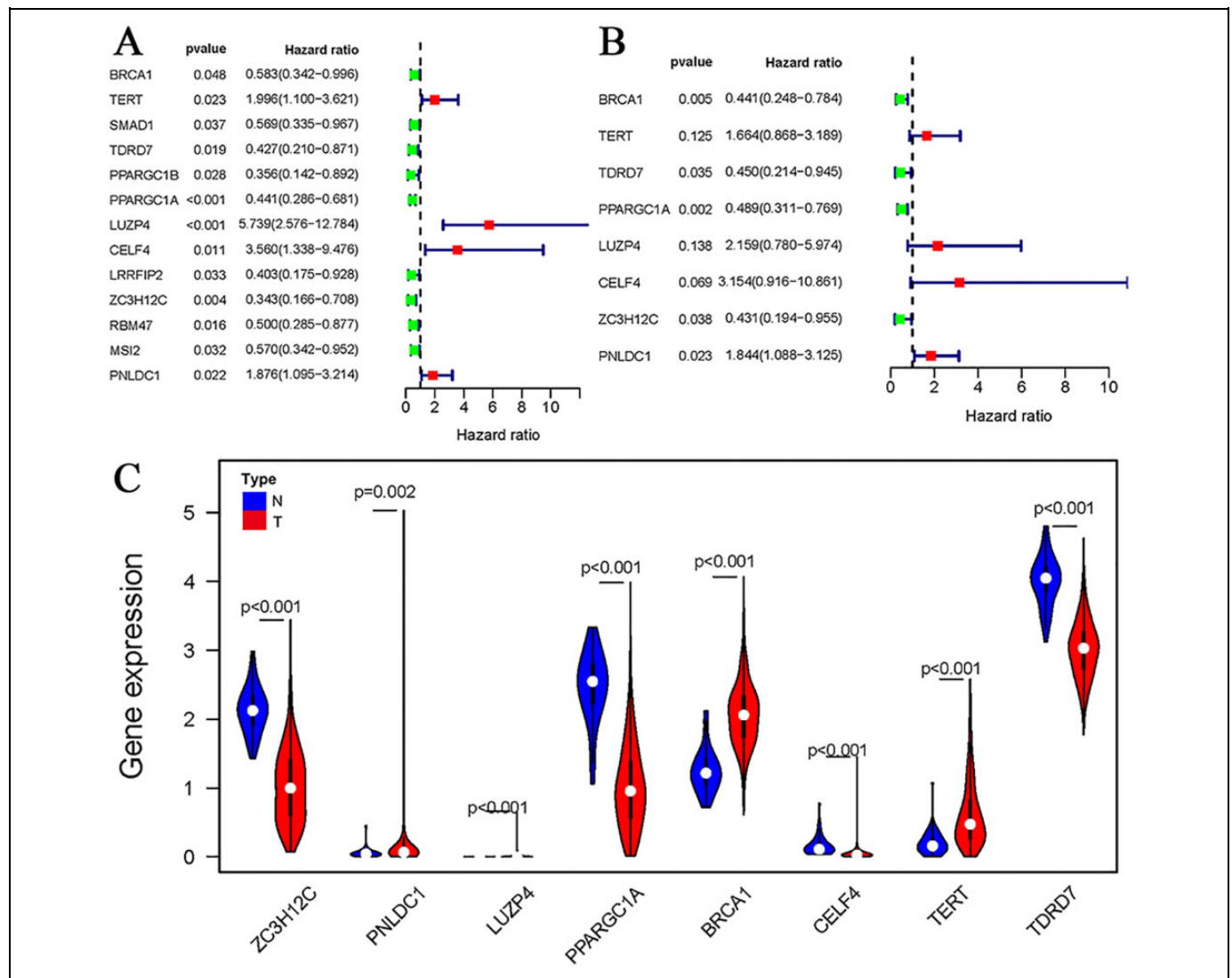


Figure 5. Construction of prognostic risk signature with feature RBPs. (A) Univariate Cox regression; (B) Multivariate Cox regression; (C) The expression array among the tumor and normal samples' 8 RBPs.

Module 1 contained 24 nodes and 254 edges (Figure 4C), and module 2 involved 8 nodes and 23 edges (Figure 4D). Module 3 included 5 nodes and 10 edges (Figure 4E). According to the GO and pathway evaluations, RBPs in the key module 1 were primarily enriched in rRNA processing, ribosome biogenesis, and rRNA metabolism. On the other hand, the module 2 RBPs were linked to RNA splicing, mRNA 3'-UTR AU-rich region binding, and mRNA 3'-UTR binding. Module 3 was highly enhanced in defensive reaction to virus, response to virus, and cellular response to exogenous dsRNA (Table 2).

The Selection of Survival-Related Genes

Overall, there were 13 survival-related RBPs confirmed from the TCGA cohort via univariate Cox regressions (Figure 5A). Whereafter, Multiple stepwise Cox regression analysis was utilized to test these candidate RBPs. 8 feature RBPs (BRCA1,

Table 3. The 8-Prognosis Hub RBPs.

RBP name	Coef	HR	Lower 95% CI	Upper 95% CI	P value
BRCA1	-0.8181	0.4413	0.2484	0.7839	0.0053
TERT	0.5093	1.6641	0.8682	3.1894	0.1250
TDRD7	-0.7989	0.4498	0.2142	0.9447	0.0348
PPARGC1A	-0.7152	0.4891	0.3112	0.7685	0.0019
LUZP4	0.7698	2.1593	0.7805	5.9741	0.1382
CELF4	1.1488	3.1544	0.9162	10.8606	0.0686
ZC3H12C	-0.8421	0.4308	0.1943	0.9554	0.0382
PNLDC1	0.6121	1.8443	1.0885	3.1250	0.0229

Abbreviations: CI, confidence interval; Coef, coefficient; HR, hazard ratio.

TERT, TDRD7, PPARGC1A, LUZP4, CELF4, ZC3H12C, and PNLDC1) were discovered (Figure 5B, Table 3). Figure 5C shows the expression of these 8 RBPs among the tumor and normal samples.

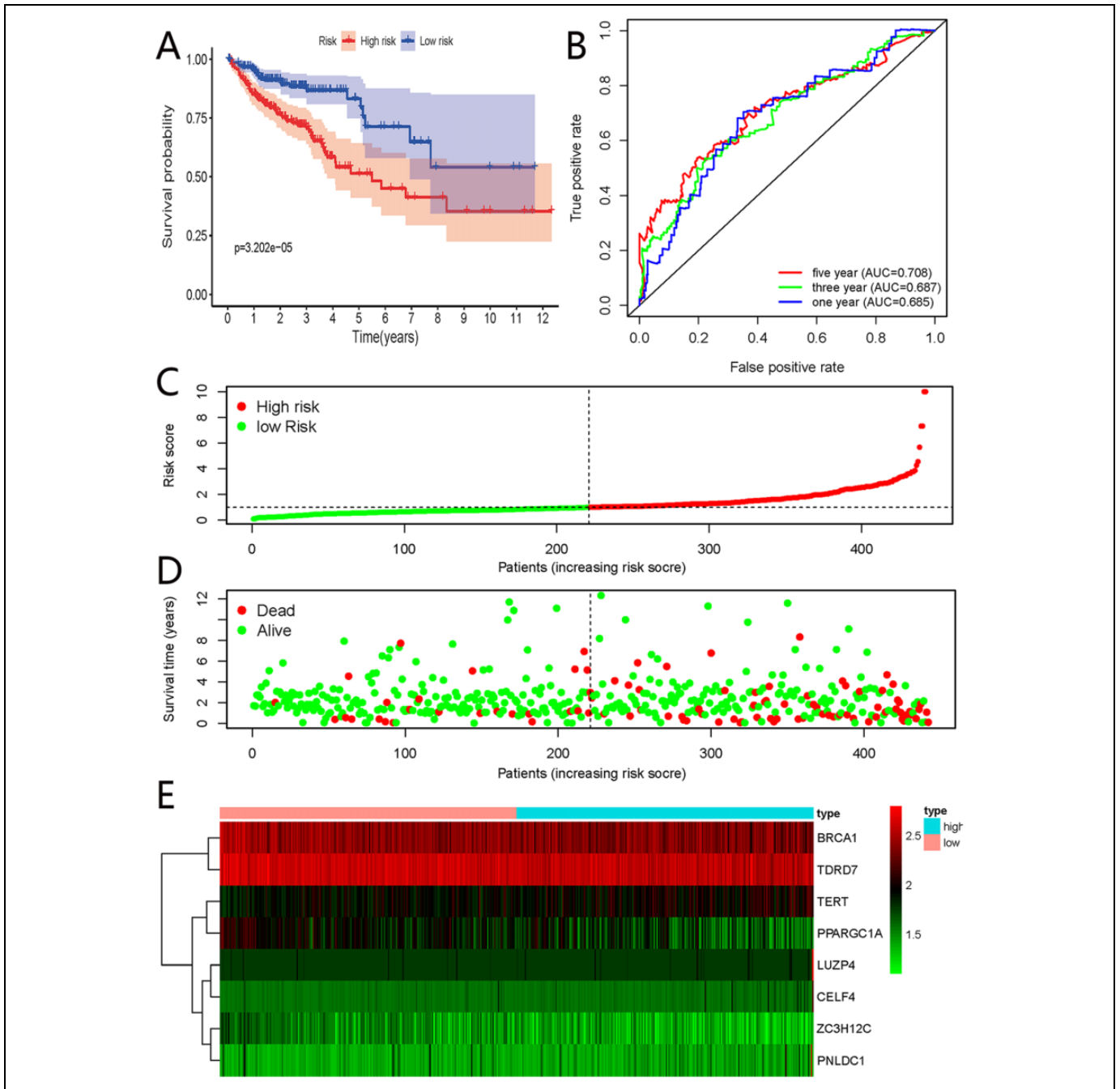


Figure 6. Risk score analysis of the eight-RBP prognostic model in the TCGA cohort. (A) The survival plot for the low- and high-risk subgroups; (B) OS predictive ROC plots as per risk score; (C) The risk score curve; (D) Survival status; (E) Expression heat map.

Construct the Predictive Model in the TCGA Cohort

The predictive model was constructed with the 8 hub RBPs. Here, LUZP4, TERT, PNLDC1 and CELF4 served as high-risk RBPs ($HR > 1$). On the other hand, the remaining 4 (BRCA1, TDRD7, PPARGC1A and ZC3H12C) were confirmed to be low-risk RBPs ($HR < 1$). The formula below was applied in computing each patient's risk score:

$$\begin{aligned} \text{Risk score} = & (-0.8181 \times \text{BRCA1}) + (0.5093 \times \text{TERT}) \\ & + (-0.799 \times \text{TDRD7}) + (-0.7152 \\ & \times \text{PPARGC1A}) + (0.7699 \times \text{LUZP4}) \\ & + (1.1488 \times \text{CELF4}) + (-0.8421 \times \text{ZC3H12C}) \\ & + (0.6121 \times \text{PNLDC1}) \end{aligned}$$

In terms of the median risk score, CRC patients were separated into 2 subsets: the high and low-risk group.

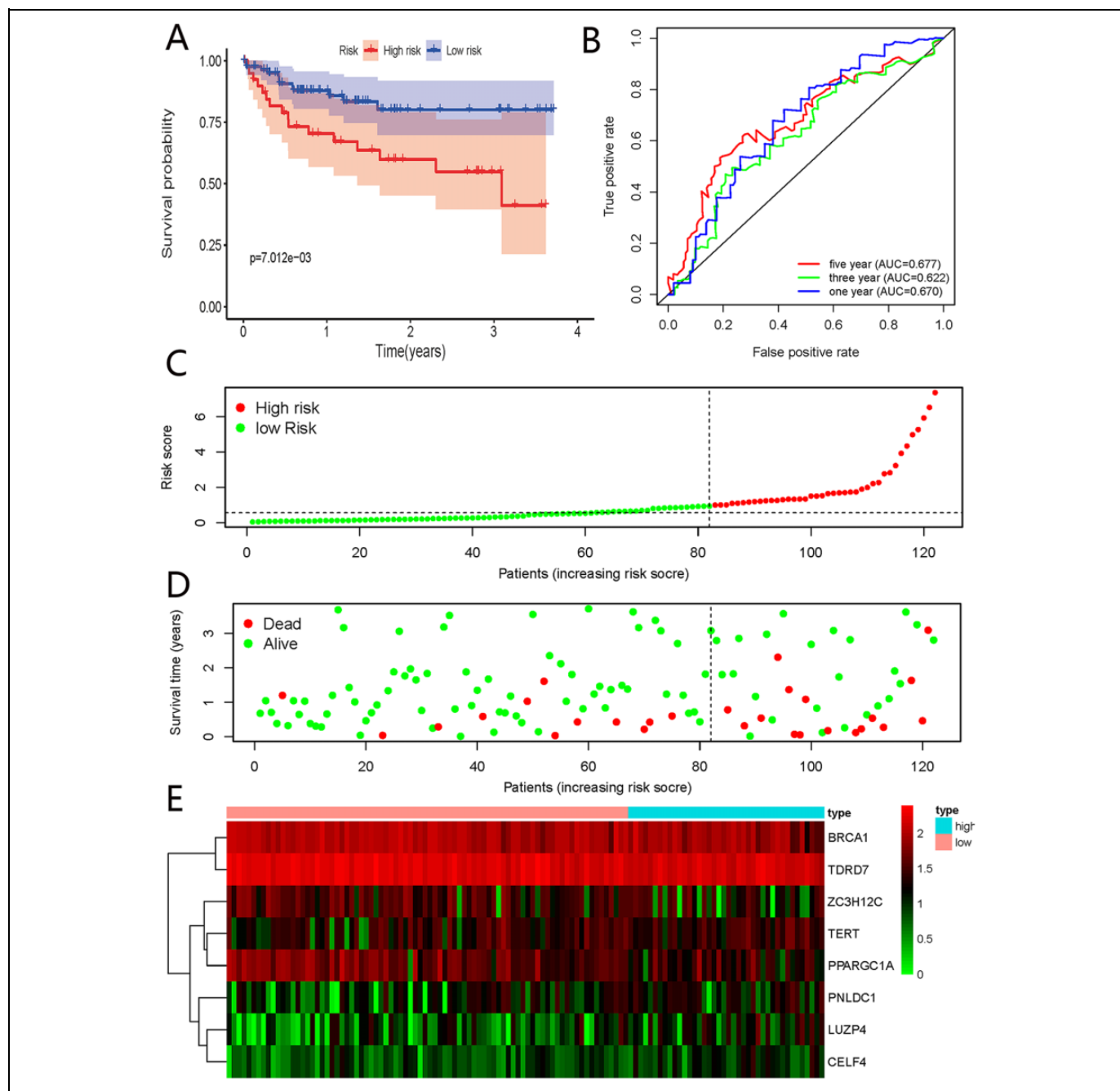


Figure 7. Validation of the prognostic signature in the GSE38832 dataset. (A) Low- and high-risk subsets survival plots; (B) OS predictive ROC plots based on risk score; (C) The risk score plot.; (D)Survival status; (E) Expression heat map.

Furthermore, to comprehend possible impact on OS of the patient from risk score, Kaplan-Meier method was used to analyze the high and low-risk groups. The analysis revealed that the OS in low-risk was higher than in the high-risk group (Figure 6A). Furthermore, we analyzed ROC based on 1, 3, and 5 years to evaluate the specificity and sensitivity of prognostic signature (Figure 6B). The corresponding AUCs values for 1 year, 3 years and 5 years survival were 0.685, 0.687 and 0.708. The results suggested that the signature has

high prognostic accuracy. The distributions of the risk scores, OS and OS status were shown in Figure 6C-E. From the graph, we can see that CRC mortality rises highly as the risk score increases. Besides, the expression levels of 8 RBPs were visualized in heatmaps. ZC3H12C, BRCA1, TDRD7 and PPARGC1A had a higher expression in the low-risk group than in the high-risk group, while TERT, CELF4, LUZP4 and PNLDC1 was highly expressed in the high-risk group.

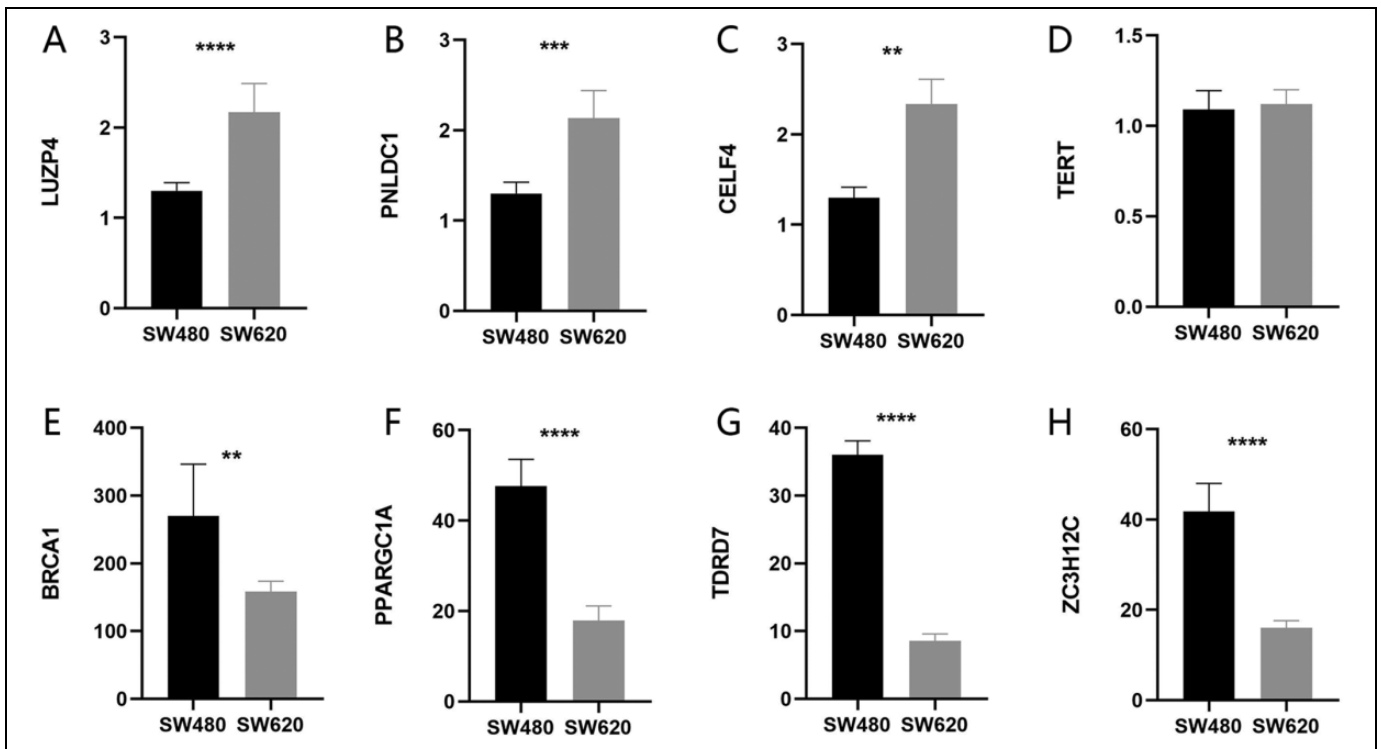


Figure 8. Expression levels of 8 genes in SW480 and SW620 cell lines. (A) LUZP4; (B) PNLDC1; (C) CELF4; (D) TERT; (E) BRCA1; (F) PPARGC1A; (G) TDRD7; (H) ZC3H12C.

Validation of the Signature in the GSE38832 Dataset

The GSE38832 GEO dataset was used to further validate the prognostic value of the risk signature. The outcomes from KM analysis in the GEO dataset (Figure 7A) showed survival of the low-risk patients was higher than the high-risk ones. Additionally, according to ROC values (Figure 7B), the corresponding values of AUC in the GEO dataset were 0.670, 0.622 and 0.677 for 1 year, 3 years and 5 years survival. This indicated a good specificity and sensitivity of the prognostic model. The TCGA cohort exhibited equal risk score, survival period distribution and patients' state (Figure 7C-E). Altogether, these outcomes provide important insights into the eight-RBP signatures, which have a high selectivity of high-risk CRC patients with severe prognoses. The result was consistent with the TCGA results, indicating that the risk model was accurate and effective.

SW480 and SW620 have been shown to exhibit several phenotypic differences including metastatic potential.^{21,22} Highly metastatic SW620 cell lines could be considered high-risk patients, while poorly metastatic SW480 cell lines act as low-risk patients. The expression levels of 8 genes as presented in Figure 8. Expression levels of genes are consistent with our results except for TERT.

The Signature-Based Risk Score Acted as Independent Prognostic CRC Parameter

The univariate and multivariate Cox regression analyses were confirmed in the TCGA cohort to determine if the eight-RBP

risk signature were independent prognostic factors. The results obtained from the univariate Cox model suggested that the risk score obtained from the signature was correlated to OS worsening (HR = 1.024, $P < 0.001$, 95% CI [1.013-1.035]) (Figure 9A). In the meantime, stage (HR = 2.597, $P < 0.001$, 95% CI [1.992-3.386]) was certified to be highly related to OS (Figure 9A). Afterward, the entire variables were analyzed via multivariate Cox regression. Further statistical tests revealed that the signature from the risk score maintained a reduced OS risk factor (HR = 1.021, $P < 0.001$, 95% CI [1.010-1.032]) (Figure 9B). In summary, the results indicate that a risk score derived from signature could be regarded as an independent prognostic parameter in individuals with CRC.

Generate a Nomogram

Next, a nomogram incorporating the signatures was built to predict probable OS in 1 year, 3 years, and 5 years (Figure 9C). The calibration plot showed good agreement between nomogram prediction and actual observations. (Figure 9D-E). Together these results provide important insights into the risk score, which perfectly matches the projections and experimental outcomes.

Discussion

Colorectal cancer exhibits high malignancy and is highly associated with liver and lung metastasis, which have a strong impact on the survival prognosis.²³ Hence, discovering a highly

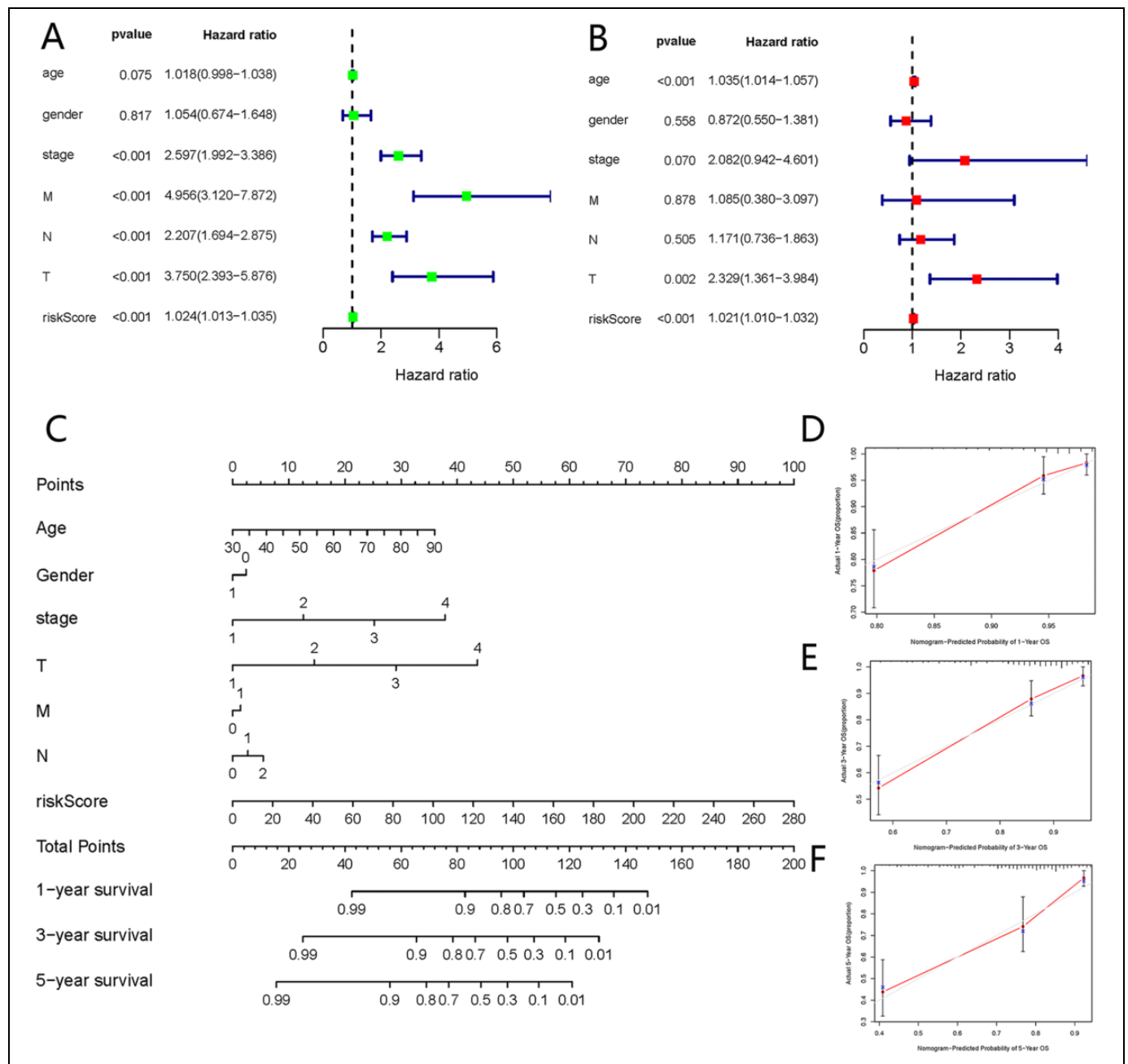


Figure 9. The independent prognostic factors and assembly of gene-based prognostic model. (A) Univariate Cox regression; (B) Multivariate Cox regression; (C) Prediction nomogram for 1-, 3-, and 5-year OS in CRC patients; (D) Calibration plot for 1-year; (E) Calibration plot for 3-year; (F) Calibration plot for 5-year.

sensitive and specific prognostic biomarker is important. Prior studies have noted the importance of RNA binding proteins, which is important for tumorigenicity and progression.²⁴⁻²⁶ However, very few RBPs have been deeply investigated, and some of them may be associated with genesis and development of carcinoma.²⁷

Prior studies have noted that post-transcriptional regulation is important in RNA, especially in genetic expressions. Some reports have confirmed that RBPs are key regulators of post-transcription processes such as differentiation, apoptosis,

migration, angiogenesis as well as cell proliferation. HuR protein is one of the hot spots now, which belongs to ELAV family of RBPs. HuR can identify and bind to target genes, thus stabilizes the target genes, suppresses the mRNA degradation, and regulates multifarious processes such as proliferation, tumor growth, transcription and translation, differentiation, apoptosis, angiogenesis and protein transport.²⁸ The insulin-like growth factor-2 mRNA binding proteins (IGF2BPs or IMPs) are preserved RBPs subsets. IMP1 plays a key role in cell proliferation and growth by combining and shielding several mRNAs.²⁹ It

has previously been observed that over 80% of CRC overexpress IMP1,¹⁰ a regulator of cell cycle migration and progression.³⁰ IMP1 is also linked to lymph node metastasis and invasion.^{11,31,32} Additionally, protein synthesis occurs in ribonucleoprotein granule. The mutation of it affects translation and plays an important regulatory role in carcinogenesis.³³ The whole-genome sequence analysis of cancer cell confirmed ribonucleoprotein regulates gene mutation regulations in colorectal carcinomas, endometrial cancer, chronic lymphocytic leukemia (CLL), T-cell acute high-grade gliomas, and lymphoblastic leukemia (T-ALL) cancer cells.³⁴⁻³⁷ In brief, RBPs affect the development of several diseases by regulating transcription. Our results are consistent with previous studies. The enrichment analysis in this study confirms that RBPs dysregulation was highly connected with control of translation, such as RNA splicing, catalytic activity, ribonucleoprotein granule, cytoplasmic ribonucleoprotein granule, mRNA 3'-UTR binding and acting on RNA. The KEGG pathway results reflect that the abnormal expressed RBPs might take part in the mediation of carcinoma progression by controlling mRNA surveillance pathway, TGF- β signaling pathway and RNA transport. All these results showed that most of the dysregulation RBPs link with RNA modification, and the processing and modification of RNA are connected with tumorigenesis.

The current study selected the feature RBPs using univariate Cox and multiple Cox regression analysis. Cumulatively, 8 genes related to the survival in CRC patient was confirmed. This included TERT, BRCA1, TDRD7, LUZP4, PPARGC1A, CELF4, PNLDC1 and ZC3H12C. Among the RBPs genes, a small number of them are associated with colorectal cancer. The BRCA1 is mainly linked with hereditary breast cancer and meta-analysis confirmed that BRCA1 mutation carriers increase the risk of colorectal cancer.³⁸ TERT is involved in the early stages of colorectal cancer development, initially affecting the tumor's stromal microenvironment by inducing COX-2 expression.³⁹ CUG binding protein 4 (CUBP4) or CELF4 has multiple functions, and the main function is related processes like splicing and translation. Previous research has shown that reduces CELF2 expression may be related to tumor promotion and development by regulating the transcription process.⁴⁰ PPARGC1A is a tumor inhibitor in ovarian and colorectal carcinomas, and it can also act as a negative prognostic marker for CRC.^{41,42} Several RBPs are associated with other tumors, the relationship between most RBPs and CRC is still not clear. LUZP as a regulatory protein, mutation and Excessive expression of LUZP result in tumorigenesis. Solid tumors like pancreatic, breast and cervical carcinomas have exhibited abnormal expression of LUZP.⁴³⁻⁴⁵ ZC3H12A, a novel RNA-binding protein (RBP) called MCP1P1 or Regnase-1, links with immune homeostasis and post-transcriptional regulation.^{46,47} Prior studies in lung carcinoma found that ZC3H12A activates macrophages.⁴⁸ PNLDC1, a PARN-like 3'-to-5' exonuclease located at the membrane of the mitochondria in a mouse, is related to a mature piRNA development.⁴⁹ The TUDOR domain-containing proteins (TDRDs) play major roles in identifying methyl-lysine/

arginine residue. The dysregulation of TDRD could initiate tumorigenesis.⁵⁰

In this study, to predict the survival of CRC patients, a risk score formula based on the 8-gene signature was made. In CRC patients, high expression of LUZP4, TERT, PNLDC1 and CELF4 related to a more unfavorable clinical outcome, whereas high expression of TDRD7, BRCA1, ZC3H12C and PPARGC1A were connected with a good prognosis. One of the important findings in this study was that the 8-RBPs risk signature was constructed and demonstrated a robust prognostic prediction in CRC. The ROC analysis of the model demonstrated a higher precision for both the TCGA cohort (1-year AUC = 0.685; 3-year AUC = 0.687; 5-year AUC = 0.708) and GEO dataset (1-year AUC = 0.670; 3-year AUC = 0.622; 5-year AUC = 0.677). The results of cellular experience provide strong support for our finding. Expression levels of genes are consistent with ours except for TERT. This might be due to the differing behavior of cells *in vitro* or even to technical reasons. It should be validated in future large sample clinical studies. In conclusion, the above results suggest that the model has a significant prognostic value for CRC.

In the present study, a novel prognosis predictive model of RBPs signature was unveiled, and the predictive capability of the model was effectively assessed. This signature could serve as promising biomarkers for supervising the development of colorectal cancer. The present study outlines important CRC pathogenesis and provides information on therapy and colorectal cancer prognosis.

Acknowledgments

The results of this study are derived from the TCGA and GEO databases. We are grateful to the authors whose data is used in this study.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Ting Li, PhD  <https://orcid.org/0000-0001-9470-0451>

References

1. Al Bandar MH, Kim NK. Current status and future perspectives on treatment of liver metastasis in colorectal cancer (Review). *Oncol Rep.* 2017;37(5):2553-2564. doi:10.3892/or.2017.5531
2. Andersson-Rolf A, Zilbauer M, Koo BK, Clevers H. Stem cells in repair of gastrointestinal epithelia. *Physiology (Bethesda, Md).* 2017;32(4):278-289. doi:10.1152/physiol.00005.2017
3. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Gen.* 2014;15(12):829-845. doi:10.1038/nrg3813

4. Dreyfuss G, Kim VN, Kataoka N. Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Bio.* 2002; 3(3):195-205. doi:10.1038/nrm760
5. Wang ZL, Li B, Luo YX, et al. Comprehensive genomic characterization of RNA-binding proteins across human cancers. *Cell Rep.* 2018;22(1):286-298. doi:10.1016/j.celrep.2017.12.035
6. de Bruin RG, Rabelink TJ, van Zonneveld AJ, van der Veer EP. Emerging roles for RNA-binding proteins as effectors and regulators of cardiovascular disease. *Eur Heart J.* 2017;38(18): 1380-1388. doi:10.1093/eurheartj/ehw567
7. Sonnenschein K, Fiedler J, Pfanne A, et al. Therapeutic modulation of RNA-binding protein Rbm38 facilitates re-endothelialization after arterial injury. *Cardiovasc Res.* 2019;115(12):1804-1810. doi:10.1093/cvr/cvz063
8. King CE, Cuatrecasas M, Castells A, Sepulveda AR, Lee JS, Rustgi AK. LIN28B promotes colon cancer progression and metastasis. *Cancer Res.* 2011;71(12):4260-4268. doi:10.1158/0008-5472.can-10-4637
9. King CE, Wang L, Winograd R, et al. LIN28B fosters colon cancer migration, invasion and transformation through let-7-dependent and -independent mechanisms. *Oncogene.* 2011; 30(40):4185-4193. doi:10.1038/onc.2011.131
10. Ross J, Lemm I, Berberet B. Overexpression of an mRNA-binding protein in human colorectal cancer. *Oncogene.* 2001; 20(45):6544-6550. doi:10.1038/sj.onc.1204838
11. Dimitriadis E, Trangas T, Milatos S, et al. Expression of oncofetal RNA-binding protein CRD-BP/IMP1 predicts clinical outcome in colon cancer. *Int J Cancer.* 2007;121(3):486-494. doi:10.1002/ijc.22716
12. Li N, Yousefi M, Nakauka-Ddamba A, et al. The Msi family of RNA-binding proteins function redundantly as intestinal oncoproteins. *Cell Rep.* 2015;13(11):2440-2455. doi:10.1016/j.celrep.2015.11.022
13. Wang S, Li N, Yousefi M, et al. Transformation of the intestinal epithelium by the MSI2 RNA-binding protein. *Nat Commun.* 2015;6:6517. doi:10.1038/ncomms7517
14. Denkert C, Koch I, von Keyserlingk N, et al. Expression of the ELAV-like protein HuR in human colon cancer: association with tumor stage and cyclooxygenase-2. *Mod Pathol.* 2006;19(9): 1261-1269. doi:10.1038/modpathol.3800645
15. Young LE, Sanduja S, Bemis-Standoli K, Pena EA, Price RL, Dixon DA. The mRNA binding proteins HuR and tristetraprolin regulate cyclooxygenase 2 expression during colon carcinogenesis. *Gastroenterol.* 2009;136(5):1669-1679. doi:10.1053/j.gastro.2009.01.010
16. Du F, Qiao C, Li X, et al. Forkhead box K2 promotes human colorectal cancer metastasis by upregulating ZEB1 and EGFR. *Theranostics.* 2019;9(13):3879-3902. doi:10.7150/thno.31716
17. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC Med.* 2012;10: 51. doi:10.1186/1741-7015-10-51
18. Moons KG, Altman DG, Reitsma JB, Collins GS. New guideline for the reporting of studies developing, validating, or updating a multivariable clinical prediction model: The TRIPOD statement. *Adv Anat Pathol.* 2015;22(5):303-305. doi:10.1097/pap.0000000000000072
19. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284-287. doi:10.1089/omi.2011.0118
20. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics.* 2000;56(2):337-344. doi:10.1111/j.0006-341x.2000.00337.x
21. Leibovitz A, Stinson JC, McCombs WB III, McCoy CE, Mazur KC, Mabry ND. Classification of human colorectal adenocarcinoma cell lines. *Cancer Res.* 1976;36(12):4562-4569.
22. Provenzani A, Fronza R, Loreni F, Pascale A, Amadio M, Quattrone A. Global alterations in mRNA polysomal recruitment in a cell model of colorectal cancer progression to metastasis. *Carcinogenesis.* 2006;27(7):1323-1333. doi:10.1093/carcin/bgi377
23. Adam R, de Gramont A, Figueras J, et al. Managing synchronous liver metastases from colorectal cancer: a multidisciplinary international consensus. *Cancer Treat Rev.* 2015;41(9):729-741. doi: 10.1016/j.ctrv.2015.06.006
24. Pereira B, Billaud M, Almeida R. RNA-binding proteins in cancer: old players and new actors. *Trend Cancer.* 2017;3(7):506-528. doi:10.1016/j.trecan.2017.05.003
25. Wu Y, Chen H, Chen Y, et al. HPV shapes tumor transcriptome by globally modifying the pool of RNA binding protein-binding motif. *Aging.* 2019;11(8):2430-2446. doi:10.18632/aging.101927
26. Lujan DA, Ochoa JL, Hartley RS. Cold-inducible RNA binding protein in cancer and inflammation. *Wiley Interdiscip Rev RNA.* 2018;9(2). doi:10.1002/wrna.1462
27. 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 (Clifton, NJ).* 2019;1882:239-252. doi:10.1007/978-1-4939-8879-2_22
28. Brody JR, Dixon DA. Complex HuR function in pancreatic cancer cells. *Wiley Interdiscip Rev RNA.* 2018;9(3):e1469. doi:10.1002/wrna.1469
29. Noubissi FK, Elcheva I, Bhatia N, et al. CRD-BP mediates stabilization of betaTrCP1 and c-myc mRNA in response to beta-catenin signalling. *Nature.* 2006;441(7095):898-901. doi:10.1038/nature04839
30. Boyerinas B, Park SM, Shomron N, et al. Identification of let-7-regulated oncofetal genes. *Cancer Res.* 2008;68(8):2587-2591. doi:10.1158/0008-5472.can-08-0264
31. Hamilton KE, Noubissi FK, Katti PS, et al. IMP1 promotes tumor growth, dissemination and a tumor-initiating cell phenotype in colorectal cancer cell xenografts. *Carcinogenesis.* 2013;34(11): 2647-2654. doi:10.1093/carcin/bgt217
32. Madison BB, Liu Q, Zhong X, et al. LIN28B promotes growth and tumorigenesis of the intestinal epithelium via Let-7. *Genes Develop.* 2013;27(20):2233-2245. doi:10.1101/gad.224659.113
33. Goudarzi KM, Lindström MS. Role of ribosomal protein mutations in tumor development (Review). *Int J Oncol.* 2016;48(4): 1313-1324. doi:10.3892/ijo.2016.3387
34. De Keersmaecker K, Atak ZK, Li N, et al. Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and

- RPL10 in T-cell acute lymphoblastic leukemia. *Nat Genet.* 2013; 45(2):186-190. doi:10.1038/ng.2508
35. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* 2014;505(7484):495-501. doi:10.1038/nature12912
36. Nieminen TT, O'Donohue MF, Wu Y, et al. Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. *Gastroenterol.* 2014;147(3):595-598. e5. doi:10.1053/j.gastro.2014.06.009
37. Novetsky AP, Zigelboim I, Thompson DM Jr, Powell MA, Mutch DG, Goodfellow PJ. Frequent mutations in the RPL22 gene and its clinical and functional implications. *Gynecol Oncol.* 2013;128(3):470-474. doi:10.1016/j.ygyno.2012.10.026
38. Oh M, McBride A, Yun S, et al. BRCA1 and BRCA2 gene mutations and colorectal cancer risk: systematic review and meta-analysis. *J Natl Cancer Inst.* 2018;110(11):1178-1189. doi:10.1093/jnci/djy148
39. Ayiomamitis GD, Notas G, Vasilakaki T, et al. Understanding the interplay between COX-2 and hTERT in colorectal cancer using a multi-omics analysis. *Cancers (Basel).* 2019;11(10):1536. doi:10.3390/cancers11101536
40. Ramalingam S, Ramamoorthy P, Subramaniam D, Anant S. Reduced expression of RNA binding protein CELF2, a putative tumor suppressor gene in colon cancer. *Immuno-Gastroenterol.* 2012;1(1):27-33. doi:10.7178/ig.1.1.7
41. Feilchenfeldt J, Bründler MA, Soravia C, Tötsch M, Meier CA. Peroxisome proliferator-activated receptors (PPARs) and associated transcription factors in colon cancer: reduced expression of PPARgamma-coactivator 1 (PGC-1). *Cancer Lett.* 2004;203(1): 25-33.
42. Zhang Y, Ba Y, Liu C, et al. PGC-1alpha induces apoptosis in human epithelial ovarian cancer cells through a PPARgamma-dependent pathway. *Cell Res.* 2007;17(4):363-373. doi:10.1038/cr.2007.11
43. Jang SY, Jang SW, Ko J. Regulation of ADP-ribosylation factor 4 expression by small leucine zipper protein and involvement in breast cancer cell migration. *Cancer Lett.* 2012;314(2):185-197. doi:10.1016/j.canlet.2011.09.028
44. Peng Y, Clark C, Luong R, et al. The leucine zipper putative tumor suppressor 2 protein LZTS2 regulates kidney development. *The J Biol Chem.* 2011;286(46):40331-40342. doi:10.1074/jbc.M111.302059
45. Kang H, Jang SW, Ko J. Human leucine zipper protein sLZIP induces migration and invasion of cervical cancer cells via expression of matrix metalloproteinase-9. *J Biol Chem.* 2011; 286(49):42072-42081. doi:10.1074/jbc.M111.272302
46. Neubauer H, Chen R, Schneck H, et al. New insight on a possible mechanism of progestogens in terms of breast cancer risk. *Horm Mol Biol Clin Investig.* 2011;6(1):185-192. doi:10.1515/hmbci.2010.082
47. Ahmed IS, Chamberlain C, Craven RJ. S2R(Pgrmc1): the cytochrome-related sigma-2 receptor that regulates lipid and drug metabolism and hormone signaling. *Expert Opin Drug Metab Toxicol.* 2012;8(3):361-370. doi:10.1517/17425255.2012.658367
48. Zhou Z, Jiang R, Yang X, et al. circRNA mediates silica-induced macrophage activation via HECTD1/ZC3H12A-dependent ubiquitination. *Theranostics.* 2018;8(2):575-592. doi:10.7150/thno.21648
49. Izumi N, Shoji K, Suzuki Y, Katsuma S, Tomari Y. Zucchini consensus motifs determine the mechanism of pre-piRNA production. *Nature.* 2020;578(7794):311-316. doi:10.1038/s41586-020-1966-9
50. Hou Y, Liu W, Yi X. PHF20L1 as a H3K27me2 reader coordinates with transcriptional repressors to promote breast tumorigenesis. *Sci Adv.* 2020;6(16): eaaz0356. doi:10.1126/sciadv.aaz0356