

Construction of an RNA Binding Protein-based Model for Prognosis Prediction of Colorectal Cancer

Ting li

People's Hospital of Xinjiang Uygur Autonomous Region <https://orcid.org/0000-0001-9470-0451>

Wenjia Hui

People's Hospital of Xinjiang Uygur Autonomous Region

Halina Halike

People's Hospital of Xinjiang Uygur Autonomous Region

Feng Gao (✉ xjgf@sina.com)

<https://orcid.org/0000-0001-7355-2672>

Primary research

Keywords: Colorectal cancer, RNA binding protein, Prognostic signature, Survival analysis

Posted Date: June 16th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-35100/v1>

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

[Read Full License](#)

Abstract

Background: Colorectal cancer (CRC) is a prevalent gastrointestinal tumor with high incidence and mortality. Dysregulation of RNA binding proteins (RBPs) has been found in a variety of cancers and is related to oncogenesis and progression. This study aimed to develop and validate new biomarkers related to CRC prognosis by a series of bioinformatics analysis.

Methods: We mined the gene expression data of 510 CRC samples from The Cancer Genome Atlas (TCGA) database, differentially expressed genes were screened and prognosis-related genes were identified. Furthermore, gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were carried out. A prognosis-related gene signature was constructed by univariate and multivariate Cox analysis. Kaplan–Meier curves and time-dependent receiver operating characteristic (ROC) curves were utilized to evaluate the signature. The test set was used to validate the RBPs risk score model. Survival analysis was carried out to determine the independent prognostic significance of the signature. A nomogram combined with the gene signature was constructed.

Results: A total of 224 aberrantly expressed RBPs were obtained, comprising 78 downregulated and 146 upregulated RBPs. 13 RBPs with $p < 0.005$ were revealed in univariate Cox regression analysis of train group, then stepwise multivariate Cox regression was applied for constituting an eight-RBP (BRCA1, TERT, TDRD7, PPARGC1A, LUZP4, CELF4, ZC3H12C, PNLDC1) signature prognostic biomarkers. Further analysis demonstrated that high risk score for patients was significantly related to poor overall survival according to the model. The area under the time-dependent receiver operator characteristic curve of the prognostic model was 0.730 at 5 years. The signature-based risk score was an independent prognostic factor in CRC patients. We also established a nomogram based on eight RBPs and internal validation in the train set, which displayed a favorable discriminating ability for Colorectal cancer.

Conclusions: The established eight-RBP signature may serve as a novel independent prognostic factor that could be an important tool to predict the prognostic outcome of CRC patients. However, the specific biological mechanism needs further verification.

Background

Colorectal cancer (CRC) is a major cause of cancer mortality and morbidity worldwide. CRC patients usually show a survival rate of < 5 years due to early metastasis. Although treatments (such as surgery, radiotherapy, chemotherapy, and targeted therapy) have been developed fleetly, high recurrence and poor prognosis remain troubling issues[1]. The molecular pathogenesis of colorectal cancer (CRC) is an elaborate multistep process. RBPs are a class of proteins that interact with a variety of types of RNAs involve in rRNAs, ncRNAs, snRNAs, miRNAs, mRNAs, tRNAs and snoRNAs. It can directly interact with RNA or can be part of ribonucleoprotein complexes without direct contact with RNA. An emerging node of regulation of intestinal epithelial homeostasis, response to injury and malignant transformation is

through RNA binding proteins (RBPs) [2]. Currently, there are more than 1500 experimentally validated RBP coding genes in the human genome, accounting for about 7.5% of all protein-coding genes [3].

RBPs are vital for the regulation of several essential cellular processes, such as RNA splicing, modifications, transport, localization, stability, degradation, and translation. RBPs function by binding to their target RNA, forming ribonucleoprotein (RNP) complexes [4] and regulating gene expression post-transcriptionally in a plethora of ways. As a crucial participant in post-transcriptional regulation, RBPs participates in most processes of post-transcriptional regulation, and eventually engages in a series of processes including cell proliferation, differentiation, migration, apoptosis and angiogenesis by regulating the expression of related proteins, and its abnormal regulation is one of the important mechanisms for development of cancer [5]. Previous studies have indicated that RBPs such as SRSF1, Quaking, Muscleblind and HuR as pivotal moderators to regulate the occurrence and progression of cardiovascular diseases by mediating a wide range of post-transcriptional events [6, 7]. There have been some previous studies on RBPs and Colorectal cancer. Several studies have shown that LIN28B is overexpressed in about 30% of colorectal tumors[8, 9]. IMP1 is overexpressed in more than 80% of human CRC [10] and linked to TNM stage, tumor size, lymph node metastasis, and poor prognosis[11]. The overexpression of MSI is sufficient to transform the intestinal epithelium and form tumors via activation of the mTORC1 complex with inhibition of Pten [12, 13]. High HuR protein expression is found in both the nucleus and cytoplasm of human colon cancers [14]. While low HuR protein expression is observed in the normal colon[15]. The forkhead box K2 protein (FOXK2) promotes colorectal cancer metastasis by upregulating mRNA expression of zinc finger E-box binding homeobox 1 (ZEB1) [16]. Therefore, *RBPs* might play a critical role in regulating cancer, however, the functions of most RBPs have not yet been determined in cancers. A systematic functional analysis of RBPs will help us thoroughly investigate its role in tumors.

In the present study, RNA-seq data and clinic pathological data were accessed from the TCGA database for the identification of prognosis-related RBPs. Univariate Cox and multivariate Cox regression were performed to select survival-related RBPs, and systematically explored them to analyze the potential functional and clinical significance. The study was designed and carried out strictly according to the Reporting Recommendations for Tumor Marker Prognostic Studies[17] and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement[18]. Our results might be potentially helpful for developing prognosis biomarkers. It will enhance our understanding of molecular mechanisms of the model in CRC.

Methods

Data source and preprocessing

The Cancer Genome Atlas (TCGA) database (TCGA, <https://portal.gdc.cancer.gov/>) was utilized to access CRC mRNA data with corresponding clinical information (510 samples). All raw data was preprocessed by Limma package and excluded genes with an average count value less than 1. Wilcox Test was performed to test CRC samples. Genes with $\log_{2}FC > 0.5$ and $FDR < 0.05$ were selected for follow-up

analyses. Heatmaps and volcano plot were utilized to visualize the different expression patterns of RBPs through “pheatmap” R package.

KEGG pathway and GO enrichment analysis

These differently expressed RBPs were evaluated by functional enrichment analysis to explore the functions using the clusterProfiler (version 3.10.1) and org.Hs.eg.db (version 3.7.0) package[19]. Significantly enriched Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways with a P value < 0.05 and q value < 0.05 were visualized using the R (version 4.0.0) software.

PPI network construction and module screening

The differently expressed RBPs were submitted to the STRING database (<http://www.string-db.org/>)[20] to identify protein-protein interaction information. The network was constructed with the Cytoscape 3.7.0 software. Subsequently, the key modules were screened from the PPI network with scores > 7 and node counts > 5 by using the MCODE (Molecular Complex Detection) plug-in in Cytoscape. All $P \leq 0.05$ were considered as significant difference.

Construction and validation of the eight-RBP signature.

Univariate Cox regression analysis was conducted to screen survival-related genes in the key modules of TCGA dataset using survival R package. The “caret” package was used to randomly divide the samples with complete survival information into two subgroups: A training set (n = 222) and a test set (n = 222). Based on the above preliminary screened significant candidate genes. A series of multivariate Cox was therefore used in the training set based on the feature genes selected before using the survival R package [21, 22]. Constructing an optimal model to calculate a risk score to predict the prognosis of CRC patients, consequently identifying the optimal feature genes. The risk score formula for each sample was as follows:

$$Risk\ score = \sum_{i=1}^n Exp^i \beta_i$$

where β represents the coefficient value, and Exp represented the gene expression level. According to the median risk score survival analysis, CRC patients were divided into low-risk and high-risk groups. Then, Kaplan-Meier (KM) and log-rank methods were used to test the difference in the two groups by using the ‘survival’ R package (version 2.43; <https://cran.r-project.org/web/packages/survival/index.html>). After that, the model was validated in the test set. Additionally, a receiver operating characteristic (ROC) curve analysis was implemented in both the training set and test set using the SurvivalROC package to evaluate the prognostic capability of the above model[23]. And the AUC values of the 1-year, 3-year and 5-year survival rates were calculated.

Independent prognostic ability of the multi-gene signature

Univariate and multivariate Cox regression analyses were conducted both in training and test set to determine whether risk score and corresponding clinicopathologic features were independent prognostic factors for CRC patients.

Development of the nomogram.

In order to provide clinicians with a quantitative tool to predict the individual probability of survival time, the nomogram was conducted using rms R package to forecast the likelihood of OS. $P < 0.05$ was considered to be a significant difference.

Results

Identification of differently expressed RBPs in CRC patients

In this study, we conducted a systematic analysis of critical roles and prognostic values of RBPs in CRC by several advanced computational methods. The study design was illustrated in Additional file 1. A total of 469 tumor samples and 41 normal colon tissue samples were obtained and were screened by Wilcox Test. 224 RBPs identified as significantly different between tumor and normal tissues, of which 78 RBPs were upregulated and 146 were downregulated. The expression distribution of these differently expressed RBPs was displayed in Fig. 1.

GO and KEGG pathway enrichment analysis of the differently expressed RBPs

To investigate the function and mechanisms of the identified RBPs, we divided these differently expressed RBPs into two groups: upregulated or downregulated expression. Then we uploaded these differently expressed RBPs for functional enrichment analysis. The results indicated that downregulated differently expressed RBPs were significantly enriched in the biological process related to regulation of cellular amide metabolic process, regulation of translation, RNA splicing (Fig. 2a). The upregulated differently expressed RBPs were significantly enriched in ncRNA processing, RNA phosphodiester bond hydrolysis, nucleic acid phosphodiester bond hydrolysis. (Fig. 2d). In terms of molecular function, the downregulated RBPs were notably enriched in catalytic activity, acting on RNA, mRNA 3'-UTR binding, endonuclease activity. (Fig. 2a), while the upregulated differently expressed RBPs were significantly enriched in catalytic activity, acting on RNA, nuclease activity, ribonuclease activity (Fig. 2d). Through the cellular component (CC) analysis, we found that the decreased differently expressed RBPs were enriched in cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, endolysosome membrane (Fig. 2a), and upregulated differently expressed RBPs were mainly enriched in cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, preribosome. (Fig. 2d).

The results of KEGG pathways about the downregulated differently expressed RBPs are 6 (Fig. 2b-c), of which were mainly enriched in the Hepatitis C, TGF-beta signaling pathway, Progesterone – mediated oocyte maturation, Oocyte meiosis. The outcome of KEGG pathways about the upregulated differently

expressed RBPs is 6, enriched in the mRNA surveillance pathway, RNA transport, Ribosome biogenesis in eukaryotes (Fig. 2e-f).

Protein-protein interaction (PPI) network construction and key modules selecting

To further investigated the roles of differently expressed RNA binding proteins in CRC, we created the PPI network using Cytoscape software, which incorporated 190 nodes and 895 edges based on the data from STRING database (Fig. 3a). The co-expression network was processed using the MODE tool to identify possible key modules(Fig. 3b). Module 1 included 24 nodes and 254 edges (Fig. 3c), and module 2 consisted of 8 nodes and 23 edges (Fig. 3d). Module 3 contained 5 nodes and 10 edges (Fig. 3e), The GO and pathway analyses showed that the RBPs in the key module 1 greatly abounded in ribosome biogenesis,rRNA processing,rRNA metabolic process. While the RBPs in module 2 were associated with mRNA 3'-UTR binding, RNA splicing, mRNA 3'-UTR AU-rich region binding. Module 3 were mainly enriched in defense response to virus, response to virus, cellular response to exogenous dsRNA(Table 1).

Table 1
KEGG and GO analysis of subnetwork.

Module 1	ID	Description	p.adjust	qvalue	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	90S preribosome	1.99E-07	1.05E-07	4
MF	GO:0140098	catalytic activity, acting on RNA	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	negative regulation 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

Module 1	ID	Description	p.adjust	qvalue	Count
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

Prognosis related RBPs selecting

A total of 224 key differently expressed RBPs were identified from the PPI network. To investigate the prognostic significance of these RBPs, 13 survival-related feature genes were identified in the training set as being associated with prognosis following univariate Cox proportional hazards regression analysis (Fig. 4a). Subsequently, these 13 candidate RBPs were analyzed by multiple stepwise Cox regression to investigate their impact on patient survival time and clinical outcomes, eight hub RBPs were found, including BRCA1, TERT, TDRD7, PPARGC1A, LUZP4, CELF4, ZC3H12C and PNLDC1 (Fig. 4b, Table 2). Expression pattern of 8 selected RBPs between tumor and normal colon tissue (Fig. 4c).

Table 2
Eight prognosis-associated hub RBPs identified by multivariate Cox regression analysis.

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
<i>CI</i> confidence interval, <i>HR</i> hazard ratio					

Signature development in the training set

Eight feature RBPs were selected by the multivariate Cox regression model and their survival risk was evaluated to comprehensively investigate the relationship between the 8 RBPs and prognosis of CRC. The eight hub RBPs were used to construct the predictive model. Therein, TERT, LUZP4, CELF4 and PNLDC1

were observed to be high-risk RBPs (Hazard ratio > 1), while the other four were low-risk RBPs (Hazard ratio < 1). The risk score of each patient was calculated according to the following formula:

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

According to the median risk score, patients were divided into high- and low-risk groups. Moreover, to know better about the role of the risk score in the patient's survival, Kaplan–Meier analysis was performed on both the high and low-risk samples. The results demonstrated that the patients in the high-risk subgroup were with poor OS compared to those in the low-risk subgroup (Additional file 2a). Besides, 1-, 3- and 5-year time-dependent ROC analyses were performed to evaluate the signature prognostic sensitivity and specificity (Additional file 2b). The AUCs of the signature was 0.670, 0.675 and 0.678 at 1-, 3- and 5-year survival times, respectively, suggesting that this signature may have high prognostic accuracy.

According to the formula, the risk score of patients was calculated and arranged from low to high. As shown in Additional file 2c-e, we could observe that the risk score of the patients with high survival risk was over 1, while those of low-risk patients were all less than 1. Then, scatter diagrams of patients' survival time were plotted based on the risk score. It was found that with the increase of the risk score, the mortality of CRC patients was increased. However, the survival rate of patients no significant downward trend. Moreover, the expression levels of the 8 RBPs were analyzed in heatmaps. TERT exhibited high expression in the high-risk group, ZC3H12C, and PPARGC1A showed low expression in the high-risk group relative to the low-risk group.

Validation of the signature in the test set.

By using the established cut-off point, patients were assigned to a low- or high-risk group in the test sets. The patients' risk score distribution and survival status were ranked by the risk score in the test set. The results from KM analysis for overall survival suggested that patients in the low-risk group had significantly better survival than those in the high-risk group ($p < 0.05$). (Fig. 5a). Furthermore, ROC analysis was conducted to determine the sensitivity and specificity of the risk score for survival prediction. As plotted in Fig. 5b, the AUC values of 1-year, 3-year and 5-year survival rates in training set were 0.701, 0.709, 0.730, respectively (Fig. 5b). These results suggested that the prognostic model has better sensitivity and specificity. The risk score, survival time distribution and patients' status are the same as the training set (Fig. 5c-e). Taken together, these results indicated that this eight-RBP signature could effectively screen out high-risk CRC patients with relatively worse clinical outcomes.

The signature-based risk score was an independent prognostic factor in CRC patients

In order to determine whether the eight-RBP risk signature acts as an independent prognostic indicator, we performed univariate and multivariate Cox regression analyses of signature-based risk score in test set.

The univariate Cox analysis indicated that signature-based risk score was significantly associated with worse OS with HR = 1.759 ($p = 0.003$, 95% CI [1.212–2.554]) (Fig. 6a). Meanwhile, age (HR = 1.031, 95% CI [1.001–1.063], $p = 0.045$) were also proved to be significantly associated with the OS (Fig. 6a). After that, all the variables were enlisted into the multivariate Cox analysis. Notably, the signature-based risk score still exerted as a risk factor for lower overall survival (HR = 1.635, 95% CI [1.083–2.469], $p = 0.019$) of CRC patients (Fig. 6b). Hence, these data demonstrated that the signature-based risk score was an independent prognostic factor in CRC patients.

Construction of a nomogram based on the eight hub RBPs

Then, we constructed a nomogram that integrated eight-RBP signature to develop a quantitative method for CRC prognosis (Fig. 6c). Based on the multivariate Cox analysis, points were assigned to individual variables by using the point scale in the nomogram. We draw a horizontal line to determine the point of each variable and calculate the total points for each patient by summing the points of all variables, and normalize it to a distribution of 0 to 100. We can calculate the estimated survival rates for CRC patients.

Discussion

Colorectal cancer is a highly malignant tumor, which is particularly prone to liver and lung metastasis, seriously affecting the survival prognosis of patients[24]. Therefore, finding a prognostic marker with high specificity and sensitivity is the advisable choice applied in the clinical practice. Extensive evidence displayed RBPs play a pivotal role in the development and progression of various human tumors[25–27]. However, only a small part of RBPs have been studied in-depth and partially confirmed that they contributed to occurrence and development of cancers[28].

In the current study, we download mRNA expression profiles and corresponding patients' clinical information of CRC from TCGA database. 224 RBPs identified as significantly different by using the Wilcox Test for the differential analysis, consisting of 146 up-regulated RBPs and 78 down-regulated RBPs. Then, we systematically explored the potential functional pathways and constructed a PPI network of these differently expressed RBPs. Moreover, univariate Cox regression analysis, survival analysis, multivariate stepwise Cox regression analysis and ROC analysis were performed on Hub RBPs to explore its biological function and clinical significance further. These findings may contribute to developing novel biomarkers for prognosis of CRC patients.

The function enrichment analysis demonstrated that the differently expressed RBPs were significantly related to regulation of translation, RNA splicing, cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, catalytic activity, acting on RNA, mRNA 3'-UTR binding. The KEGG pathways analysis showed that the aberrantly expressed RBPs regulate the tumorigenesis and progression of CRC by affecting mRNA surveillance pathway, RNA transport, TGF – beta signaling pathway. The above information suggests that most of the enrichment pathways relate to RNA modification, which also indicates that the modification and processing of RNA is related to tumorigenesis. RNA metabolism and

RNA processing have been increasingly recognized in various diseases[29, 30], Post-transcriptional regulation of RNA stability is an important procedure in gene expression processing. Previous studies have proved that RBPs as a key participant in post-transcriptional regulation, and are involved in most processes such as cell proliferation, differentiation, migration, apoptosis and angiogenesis by regulating the expression of related proteins. HuR, a member of ELA V family of RBPs, High HuR protein expression is found in both the nucleus and cytoplasm of human colorectal cancer [14]. HuR protein can recognize and combined messenger RNA of target genes, thus stable target genes to inhibit the degradation of messenger RNA, participate in the post-transcriptional regulation of target genes, and regulate a variety of physiological and biochemical processes such as tumor cell growth, proliferation, differentiation, signal transduction, transcription and translation, angiogenesis, protein transport, cell apoptosis and its metabolism[31]. The insulin-like growth factor-2 mRNA binding proteins (IGF2BPs or IMPs) belong to a conserved subfamily of RBPs. The IMPs have been studied for their roles in regulation of post-transcriptional processes such as mRNA localization, turnover, and translational control [32]. IMP1 plays a functional role in the RNA stability by binding and shielding several mRNAs that play critical roles in cell growth and proliferation from proteolytic degradation [33]. IMP1 is overexpressed in more than 80% of human CRC[10], which regulates cell cycle progression and migration [34], and correlates with invasion, lymph node metastasis, and worse prognosis[11, 35, 36]. Moreover, ribonucleoprotein granule is an important area that implements protein synthesis. The mutation of ribonucleoprotein regulates the translation process and associated with tumor development [37]. By applying whole-exome sequencing, RNA seq, or whole-genome sequencing, ribonucleoprotein regulates gene mutations have been detected in the genome of cancer cells, including from endometrial cancer, T-cell acute lymphoblastic leukemia (T-ALL), chronic lymphocytic leukemia (CLL), colorectal carcinomas, and high-grade gliomas [38–41]. In brief, RBPs can bind to mRNAs and regulate their expression, which play an important role in the development of many diseases.

Furthermore, the hub RBPs were selected based on univariate Cox regression analysis, and multiple Cox regression analysis. 8 feature genes associated with CRC patient's survival were identified, including BRCA1, TERT, TDRD7, PPARGC1A, LUZP4, CELF4, ZC3H12C, and PNLDC1. Among these genes, several RBPs have been reported to be closely associated with colorectal cancer. BRCA1 associated hereditary breast cancer, and recently, meta-analysis results provide clinicians and healthcare regulatory agencies with evidence of the increased risk of colorectal cancer in BRCA1 mutation carriers[42]. The human telomerase reverse transcriptase (TERT) also plays an important role in colorectal cancer growth, initially affecting the tumor's stromal microenvironment[43]. CUG binding protein 4 (CUBP4) or CELF4 is a multifunctional RBP studied primarily for its role in RNA metabolism related processes like decay, translation and splicing. Previous studies have shown that CELF2 expression is consistently reduced during neoplastic transformation, suggesting that it might play a crucial role in tumor initiation and progression[44]. It is consistent with our findings. PPARGC1A was indicated as a tumor suppressor in colorectal cancer[45] and ovarian cancer[46], as well as a negative prognostic biomarker for CRC. Although the connection between most of the differently expressed RBPs and CRC remains unclear, some RBPs have been reported to be associated with other tumors. LUZP is a regulatory protein that contains

20 amino acids, Excessive expression or mutation of LUZP may lead to the occurrence of malignant tumors, and abnormal expression of LUZP can be detected in solid tumors such as cervical cancer, breast cancer and pancreatic cancer[47–49]. ZC3H12A, also known as MCPIP1 or Regnase-1, is a novel RNA-binding protein (RBP) that plays a key role in post-transcriptional regulation and immune homeostasis[50, 51], a study shows that ZC3H12A is involved in circHECTD1/HECTD1-mediated macrophage activation in lung cancer[52]. PNLDC1 is a PARN-like 3'-to-5' exonuclease localized on the mitochondrial surface in mouse, Associated with the emergence of the mature piRNA[53]. TUDOR domain-containing proteins (TDRDs) are chiefly responsible for recognizing methyl-lysine/arginine residue. TDRD dysregulation contributes to breast tumorigenesis[54].

Then, an 8-gene signature-based risk score formula was developed for survival analysis in the training and test set. High expression of BRCA1, TDRD7, PPARGC1A, and ZC3H12C was associated with a good prognosis in patients with CRC, whereas that of TERT, LUZP4, CELF4, and PNLDC1 were related to poor prognosis. Next, the eight RBPs were used to construct a risk model by multiple stepwise Cox regression analysis to predict prognosis in CRC patients. ROC analysis was performed and showed the higher accuracy of the risk score model both for the training set (1-year AUC = 0.670; 3-year AUC = 0.675; 5-year AUC = 0.678) and test set (1-year AUC = 0.701; 3-year AUC = 0.709; 5-year AUC = 0.730). According to our analysis, the signature-based risk score was an independent prognostic factor in CRC patients. A nomogram was constructed to enable practitioners to predict 1-year, 3-year, and 5-year OS of CRC patients.

Nevertheless, our study had several limitations. Firstly, our prognostic model was only based on the data from TCGA database, this study was a pure bioinformatics study, and the scientific hypothesis was not proved by biological experiments. Secondly, our study was designed based on a retrospective analysis, and a prospective study should be performed to verify the results. Thirdly, the lack of some clinical characteristics in the datasets from TCGA may have decreased the statistical effectiveness and reliability of the multivariate stepwise Cox regression analysis.

In summary, our study not only constructed a new predictive model of RBPs signature prognosis but also by grouping to verify and evaluate the predictive ability of the model. The most important thing is can be used as an independent prognostic factor in CRC. As far as we know, this is the first report of developing a RBPs-associated prognostic model for CRC. Our results would greatly contribute to show the pathogenesis of CRC and enhance our comprehension to treatment and progression of colorectal cancer.

Conclusions

We constructed an eight-RBP signature to predict prognosis for patients with colorectal cancer and validated the model in training and test sets. ROC analysis showed a higher accuracy of the risk score model. The eight-RBP signature was an independent prognostic factor for overall survival. Prospective studies are needed to further test its analytical accuracy for estimating prognosis and to validate its clinical utility in individualized management of CRC patients

Abbreviations

CRC

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

Declarations

Acknowledgements

The results of this study are based on the data from TCGA. We thank the authors who provided the data for this study.

Authors' contributions

TL conceived and designed the study, and wrote the paper. WJ.H and HH acquired part of the data. FG contributed to design, revising it critically for important intellectual content, and final approval of the version to be published. All authors read and approved the final manuscript.

Funding

None.

Availability of data and materials

The datasets generated and analyzed during the current study are available in The Cancer Genome Atlas (TCGA), <https://portal.gdc.cancer.gov/>.

Ethics approval and consent to participate

Not applicable.

Consent for publication:

All listed authors have actively participated in the study and have read and approved the submitted manuscript.

Competing interests

The authors declare that they have no competing interests.

References

1. Al Bandar MH,et al. Current status and future perspectives on treatment of liver metastasis in colorectal cancer (Review). *Oncol Rep.* 2017;37(5):2553–64.
2. Andersson-Rolf A,et al. Stem Cells in Repair of Gastrointestinal Epithelia. *Physiology (Bethesda Md).* 2017;32(4):278–89.
3. Gerstberger S,et al. A census of human RNA-binding proteins. *Nature reviews Genetics.* 2014;15(12):829–45.
4. Dreyfuss G,et al. Messenger-RNA-binding proteins and the messages they carry. *Nature reviews Molecular cell biology.* 2002;3(3):195–205.
5. Wang ZL,et al. Comprehensive Genomic Characterization of RNA-Binding Proteins across Human Cancers. *Cell reports.* 2018;22(1):286–98.
6. de Bruin RG,et al. Emerging roles for RNA-binding proteins as effectors and regulators of cardiovascular disease. *European heart journal.* 2017;38(18):1380–8.
7. Sonnenschein K,et al. Therapeutic modulation of RNA-binding protein Rbm38 facilitates re-endothelialization after arterial injury. *Cardiovascular research.* 2019;115(12):1804–10.
8. King CE,et al. LIN28B promotes colon cancer progression and metastasis. *Cancer research.* 2011;71(12):4260–8.
9. King CE,et al. LIN28B fosters colon cancer migration, invasion and transformation through let-7-dependent and -independent mechanisms. *Oncogene.* 2011;30(40):4185–93.
10. Ross J,et al. Overexpression of an mRNA-binding protein in human colorectal cancer. *Oncogene.* 2001;20(45):6544–50.
11. Dimitriadis E,et al. Expression of oncofetal RNA-binding protein CRD-BP/IMP1 predicts clinical outcome in colon cancer. *International journal of cancer.* 2007;121(3):486–94.
12. Li N,et al. The Msi Family of RNA-Binding Proteins Function Redundantly as Intestinal Oncoproteins. *Cell reports.* 2015;13(11):2440–55.
13. Wang S,et al. Transformation of the intestinal epithelium by the MSI2 RNA-binding protein. *Nature communications.* 2015;6:6517.
14. Denkert C,et al. Expression of the ELAV-like protein HuR in human colon cancer: association with tumor stage and cyclooxygenase-2. *Modern pathology: an official journal of the United States Canadian Academy of Pathology Inc.* 2006;19(9):1261–9.
15. Young LE,et al. The mRNA binding proteins HuR and tristetraprolin regulate cyclooxygenase 2 expression during colon carcinogenesis. *Gastroenterology.* 2009;136(5):1669–79.
16. Du F,et al. Forkhead box K2 promotes human colorectal cancer metastasis by upregulating ZEB1 and EGFR. *Theranostics.* 2019, 9(13):3879–3902.
17. Altman DG,et al. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC Med.* 2012;10:51.
18. Moons KG,et al. New Guideline for the Reporting of Studies Developing, Validating, or Updating a Multivariable Clinical Prediction Model: The TRIPOD Statement. *Advances in anatomic pathology*

- 2015, 22(5):303–305.
19. Yu G,et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*. 2012;16(5):284–7.
 20. Szklarczyk D,et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*. 2019;47(D1):D607-d613.
 21. O'Quigley J,et al. Cox's regression model: computing a goodness of fit statistic. *Comput Methods Programs Biomed*. 1986;22(3):253–6.
 22. Spruance SL,et al. Hazard ratio in clinical trials. *Antimicrob Agents Chemother*. 2004;48(8):2787–92.
 23. Heagerty PJ,et al. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000;56(2):337–44.
 24. Adam R,et al. Managing synchronous liver metastases from colorectal cancer: a multidisciplinary international consensus. *Cancer treatment reviews*. 2015;41(9):729–41.
 25. Pereira B,et al. RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends in cancer*. 2017;3(7):506–28.
 26. Wu Y,et al. HPV shapes tumor transcriptome by globally modifying the pool of RNA binding protein-binding motif. *Aging*. 2019;11(8):2430–46.
 27. Lujan DA,et al. Cold-inducible RNA binding protein in cancer and inflammation. *Wiley interdisciplinary reviews RNA* 2018, 9(2).
 28. Jain A,et al. Evaluation of Post-transcriptional Gene Regulation in Pancreatic Cancer Cells: Studying RNA Binding Proteins and Their mRNA Targets. *Methods in molecular biology*. (Clifton NJ). 2019;1882:239–52.
 29. Martínez-Terroba E,et al. The oncogenic RNA-binding protein SRSF1 regulates LIG1 in non-small cell lung cancer. 2018, 98(12):1562–1574.
 30. Hinkle ER,et al. RNA processing in skeletal muscle biology and disease. 2019, 10(1):1–20.
 31. Brody JR,et al. Complex HuR function in pancreatic cancer cells. *Wiley interdisciplinary reviews RNA*. 2018;9(3):e1469.
 32. Yaniv K,et al. The involvement of a conserved family of RNA binding proteins in embryonic development and carcinogenesis. *Gene*. 2002;287(1–2):49–54.
 33. Noubissi FK,et al. CRD-BP mediates stabilization of betaTrCP1 and c-myc mRNA in response to beta-catenin signalling. *Nature*. 2006;441(7095):898–901.
 34. Boyerinas B,et al. Identification of let-7-regulated oncofetal genes. *Cancer research*. 2008;68(8):2587–91.
 35. Hamilton KE,et al. .IMP1 promotes tumor growth, dissemination and a tumor-initiating cell phenotype in colorectal cancer cell xenografts. *Carcinogenesis*. 2013;34(11):2647–54.
 36. Madison BB,et al. LIN28B promotes growth and tumorigenesis of the intestinal epithelium via Let-7. *Genes Dev*. 2013;27(20):2233–45.

37. Goudarzi KM,et al. Role of ribosomal protein mutations in tumor development (Review). *Int J Oncol*. 2016;48(4):1313–24.
38. De Keersmaecker K,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):186–90.
39. Lawrence MS,et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014;505(7484):495–501.
40. Nieminen TT,et al. Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. *Gastroenterology*. 2014;147(3):595–8.e595.
41. Novetsky AP,et al. Frequent mutations in the RPL22 gene and its clinical and functional implications. *Gynecol Oncol*. 2013;128(3):470–4.
42. Oh M. et al. BRCA1 and BRCA2 Gene Mutations and Colorectal Cancer Risk: Systematic Review and Meta-analysis. *J Natl Cancer Inst*. 2018;110(11):1178–89.
43. Ayiomamitis GD,et al. Understanding the Interplay between COX-2 and hTERT in Colorectal Cancer Using a Multi-Omics Analysis. 2019, 11(10).
44. Ramalingam S,et al. Reduced Expression of RNA Binding Protein CELF2, a Putative Tumor Suppressor Gene in Colon Cancer. *Immuno-gastroenterology*. 2012;1(1):27–33.
45. Feilchenfeldt J,et al. Peroxisome proliferator-activated receptors (PPARs) and associated transcription factors in colon cancer: reduced expression of PPARgamma-coactivator 1 (PGC-1). *Cancer letters*. 2004;203(1):25–33.
46. Zhang Y,et al. .PGC-1alpha induces apoptosis in human epithelial ovarian cancer cells through a PPARgamma-dependent pathway. *Cell research*. 2007;17(4):363–73.
47. Jang SY,et al. Regulation of ADP-ribosylation factor 4 expression by small leucine zipper protein and involvement in breast cancer cell migration. *Cancer letters*. 2012;314(2):185–97.
48. Peng Y,et al. The leucine zipper putative tumor suppressor 2 protein LZTS2 regulates kidney development. *J Biol Chem*. 2011;286(46):40331–42.
49. Kang H,et al. 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):42072–81.
50. Neubauer 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):185–92.
51. Ahmed IS,et al. 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):361–70.
52. Zhou Z,et al. circRNA Mediates Silica-Induced Macrophage Activation Via HECTD1/ZC3H12A-Dependent Ubiquitination. *Theranostics*. 2018;8(2):575–92.
53. Izumi N,et al. Zucchini consensus motifs determine the mechanism of pre-piRNA production. *Nature*. 2020;578(7794):311–6.

54. Hou Y,et al.PHF20L1 as a H3K27me2 reader coordinates with transcriptional repressors to promote breast tumorigenesis. 2020, 6(16):eaaz0356.

Figures

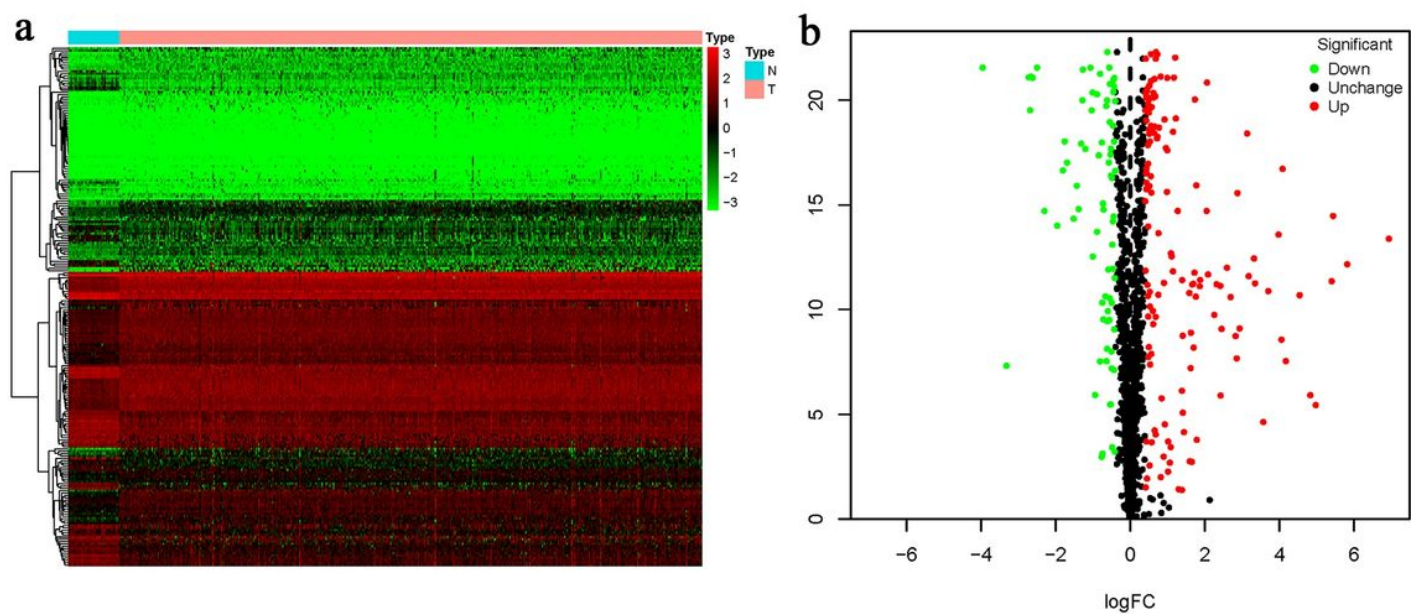


Figure 1
Differentially expressed RBPs in colorectal cancer. (a) Heat map; (b) Volcano plot.

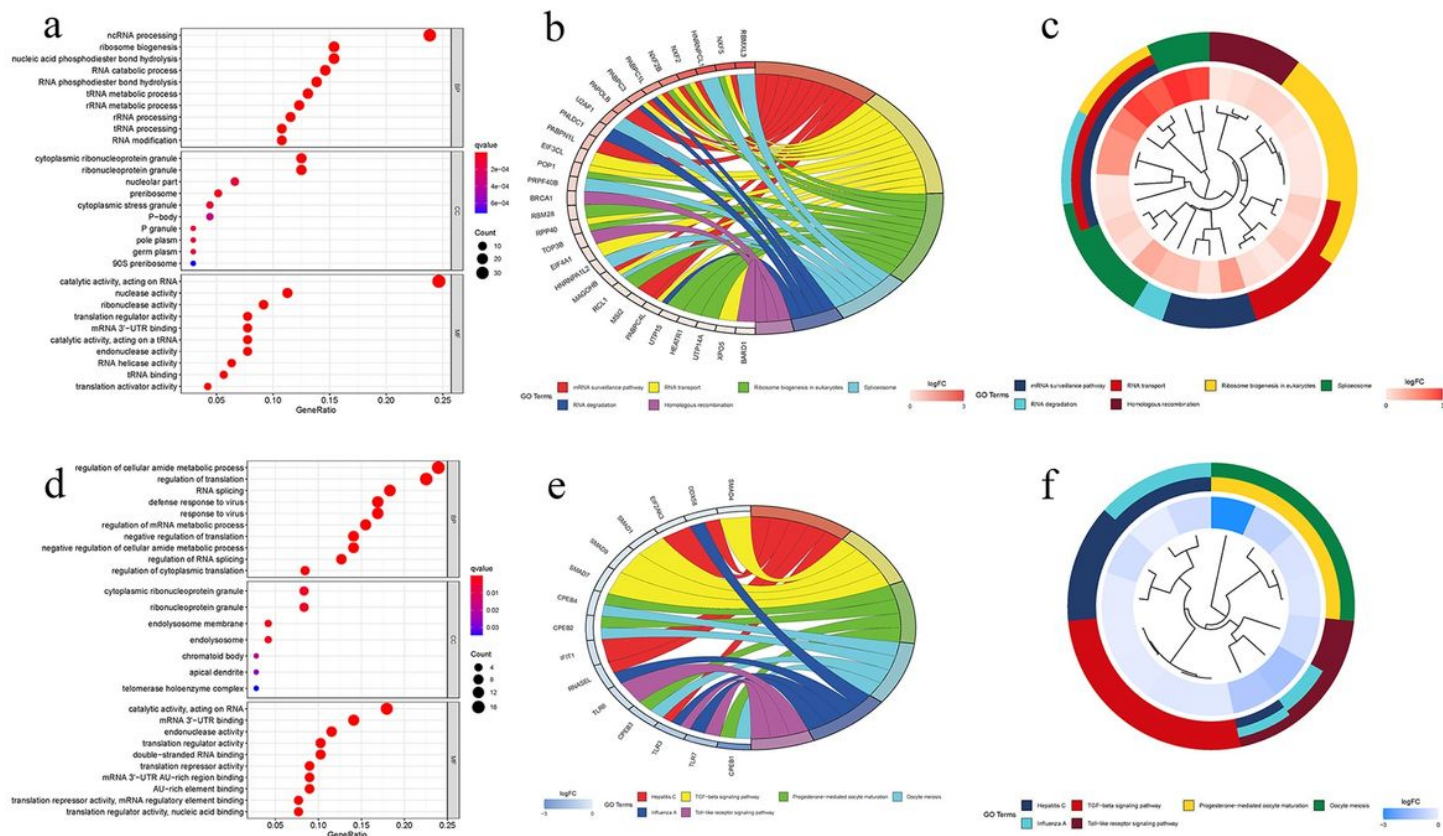


Figure 2

KEGG and GO analysis of differentially expressed RBPs. (a) GO enrichment analysis for downregulated RBPs. (b-c) KEGG pathway analysis for downregulated RBPs. (d) GO enrichment analysis for upregulated RBPs. (e-f) KEGG pathway analysis for upregulated RBPs.

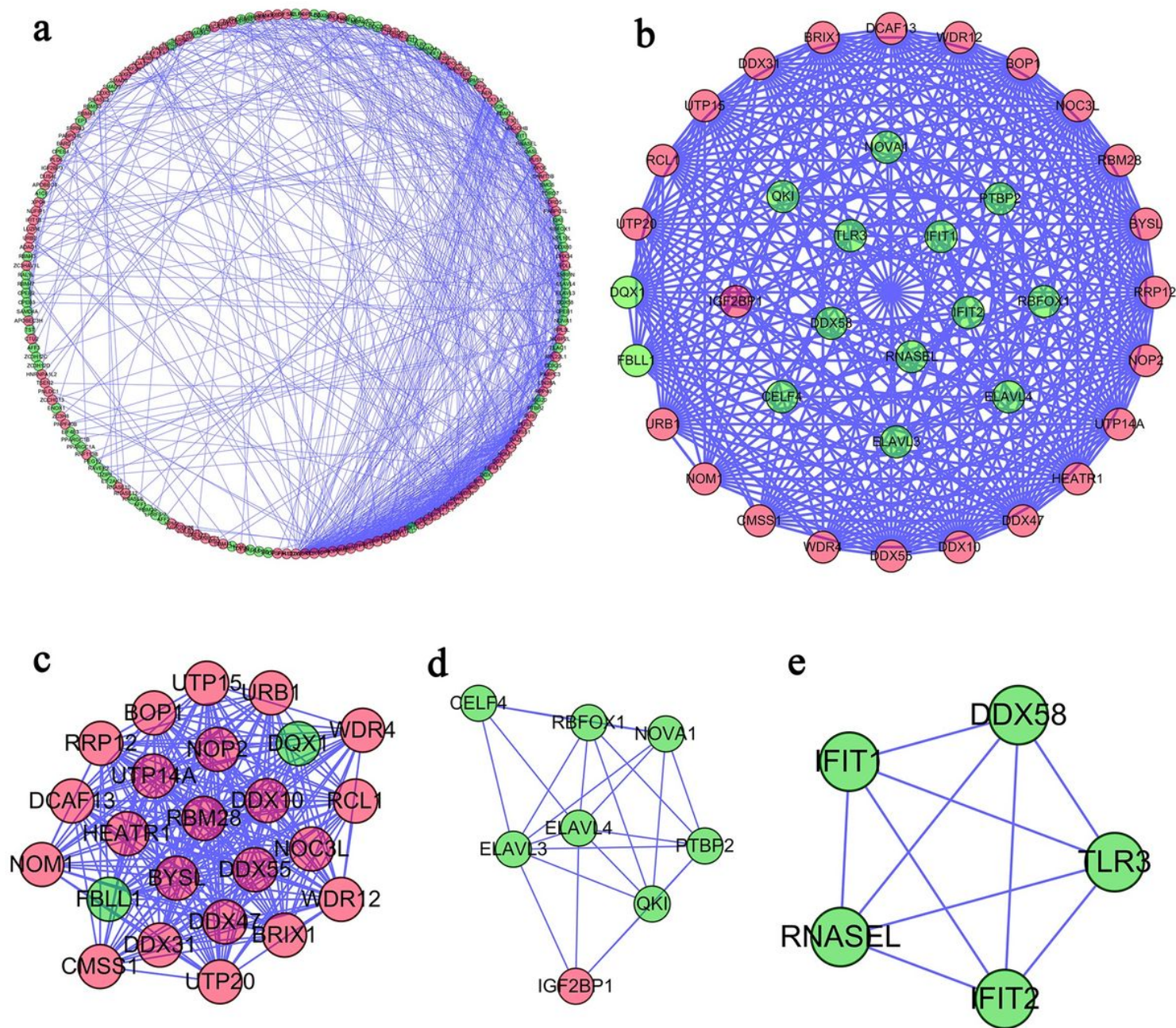


Figure 3

PPI network and module analysis. (a) Protein-protein interaction network of differentially expressed RBPs; (b) critical module from PPI network. Green circles: down-regulation with a fold change of more than 2; red circles: up-regulation with fold change of more than 2. (c) critical module 1 in PPI network. (d) critical module 21 in PPI network. (e) critical module 3 in PPI network.

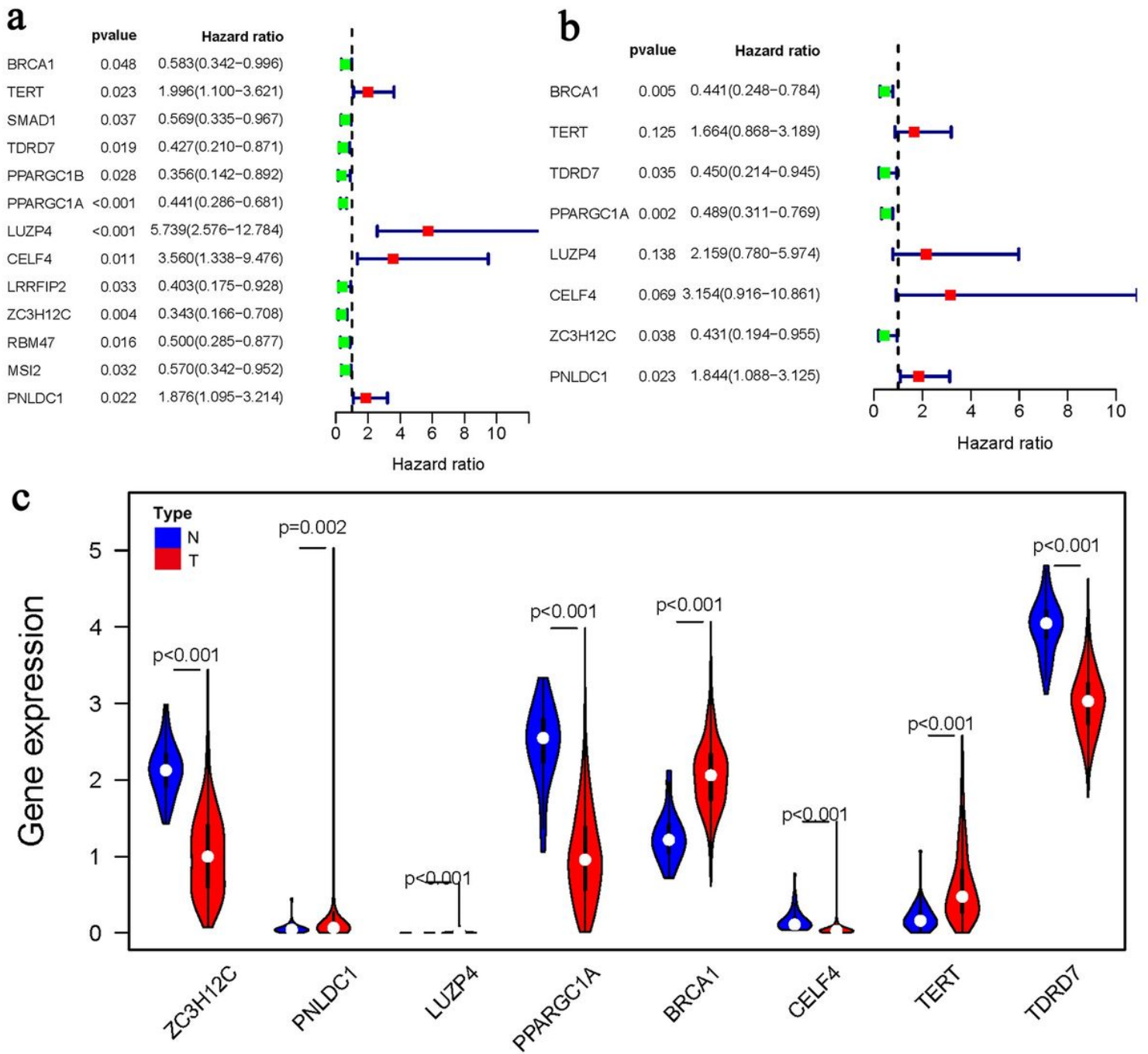


Figure 4

Univariate and multivariate Cox regression analysis for identification of hub RBPs. (a) Univariate Cox regression;(b)Multivariate Cox regression;(c) Expression pattern of 8 selected RBPs between tumor and normal colon tissue

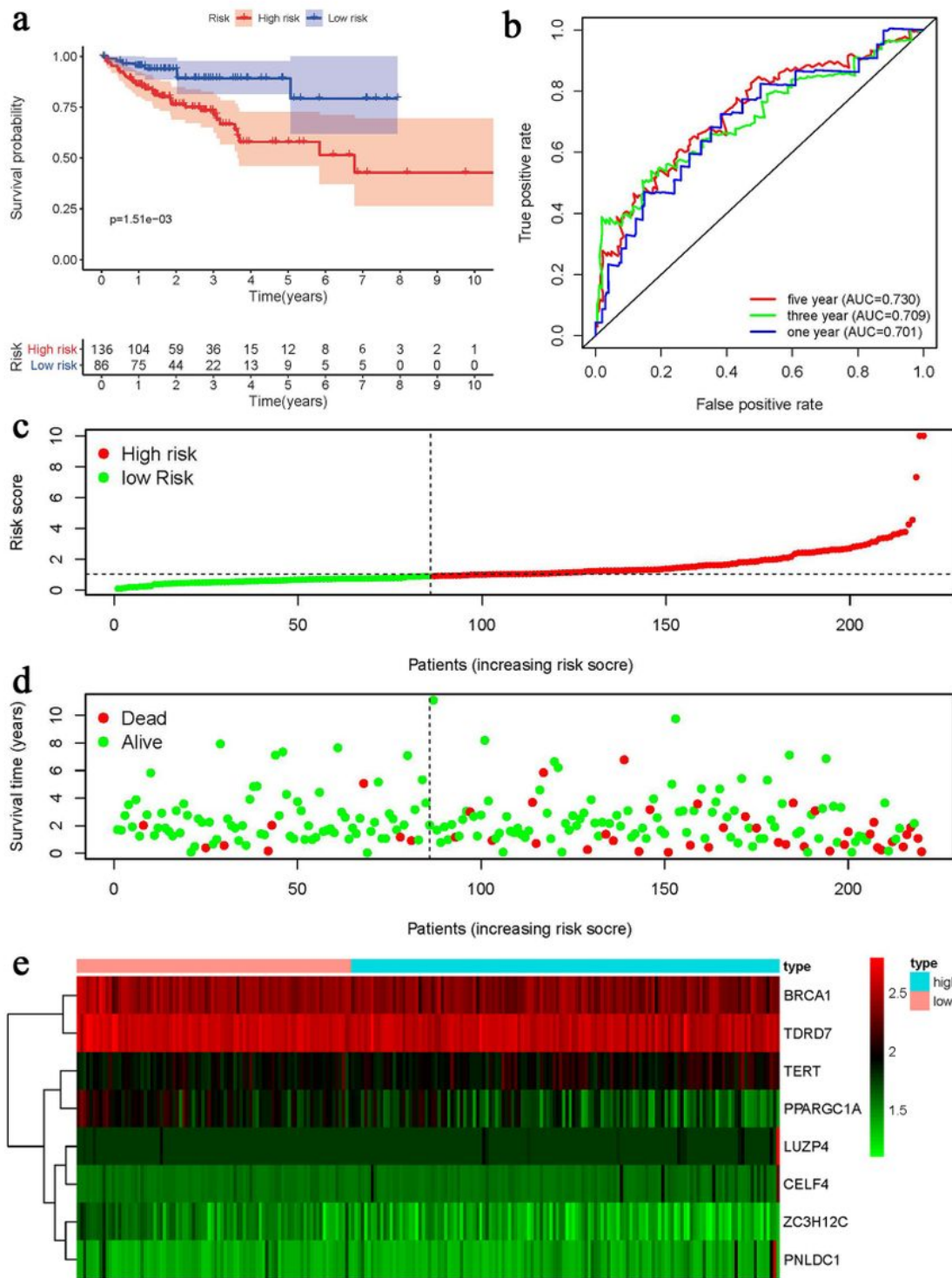


Figure 5

Risk score analysis of the eight-RBP prognostic model in test set. (a) Survival curve for low- and high-risk subgroups; (b) ROC curves for forecasting OS based on risk score; (c) The curve of risk score.; (d) Survival status; (e) Expression heat map,

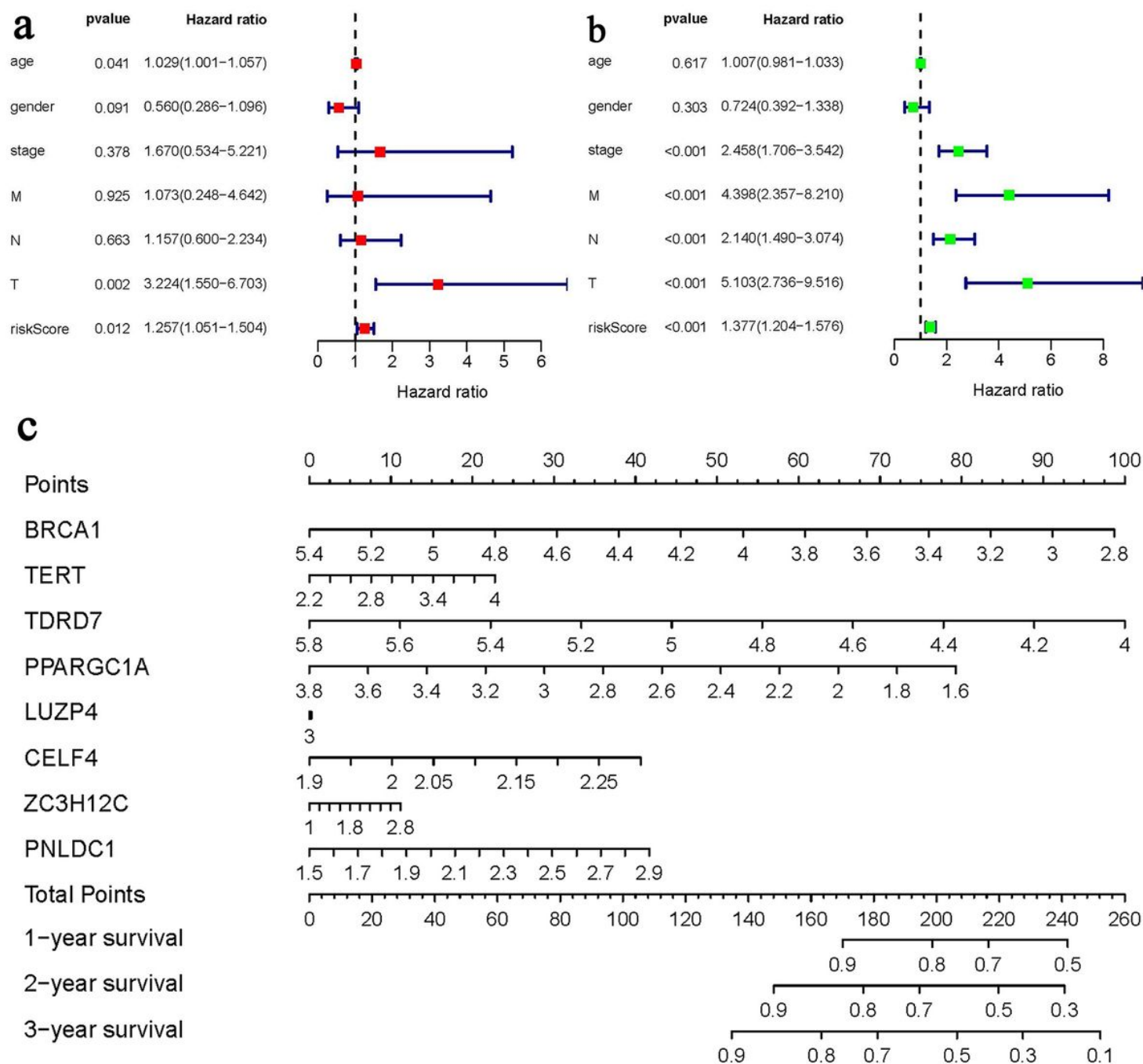


Figure 6

Identifying the independent prognostic parameters and construction of gene-based prognostic model. (a) Univariate Cox regression in test set; (b) Multivariate Cox regression in test set; (c) Nomogram for predicting 1-, 3-, and 5-year OS of CRC patients.

Supplementary Files

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

- [Additionalfile1.tif](#)

- [Additionalfile2.tif](#)