

The Underlying Molecular Mechanisms and Prognostic Factors of Rna Binding Protein in Colorectal Cancer: A Study Based on Multiple Online Databases

Qinglian He

Guangdong Medical University

Ziqi Li

Guangdong Medical University

Xue Lei

Guangdong Medical University

Qian Zou

Guangdong Medical University

Haibing Yu

Guangdong Medical University

Yuanlin Ding

Guangdong Medical University

Wei Zhu (✉ zhuwei@gdmu.edu.cn)

Guangdong Medical University <https://orcid.org/0000-0002-1217-1718>

Primary research

Keywords: RNA binding protein, Alternative splicing, Colorectal cancer

Posted Date: April 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-366831/v1>

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

[Read Full License](#)

Abstract

Background: RNA binding protein (RBP) is an active factor involved in the occurrence and development of colorectal cancer (CRC). Therefore, the potential mechanism of RBP in CRC needs to be clarified by dry-lab analyses or wet-lab experiments.

Methods: The differential RBP gene obtained from the GEPIA_2 (Gene Expression Profiling Interactive Analysis 2) were performed functional enrichment analysis. Then, the alternative splicing (AS) events related to survival were acquired by univariate regression analysis, and the correlation between RBP and AS was analyzed by R software. The online databases were conducted to analyze the mutation and methylation of RBPs in CRC. Moreover, 5 key RBP signatures were obtained through univariate and multivariate Cox regression analysis and established as RBP prognosis model. Subsequently, the above model was verified through another randomized group of TCGA CRC cohorts. Finally, multiple online databases and qRT-PCR analysis were carried to further confirm the expression of the above 5 RBP signatures in CRC.

Results: Through a comprehensive bioinformatics analysis, it was revealed that RBPs had genetic and epigenetic changes in CRC. We obtained 300 differentially expressed RBPs in CRC samples. The functional analysis suggested that they mainly participated in spliceosome. Then, a regulatory network for RBP was established to participate in AS and DDX39B was detected to act as a potentially essential factor in the regulation of AS in CRC. Our analysis discovered that 11 differentially expressed RBPs with a mutation frequency higher than 5%. Furthermore, we found that 10 differentially expressed RBPs had methylation sites related to the prognosis of CRC, and a prognostic model was constructed by the 5 RBP signatures. In another randomized group of TCGA CRC cohorts, the prognostic performance of the 5 RBP signatures was verified.

Conclusion: The potential mechanisms that regulate the aberrant expression of RBPs in the development of CRC was explored, a network that regulated AS was established, and the RBP-related prognosis model was constructed and verified, which could improve the individualized prognosis prediction of CRC.

Background

Colorectal cancer (CRC) is one of the most fatal primary digestive tract tumors [1, 2]. Despite some improvements in diagnosis and treatment, global mortality remains high [1]. At present, the field of CRC research is focused on the development of tools for early detection, reliable prognosis and predictive biomarkers, as well as new treatments that can overcome drug resistance [3-5]. As everyone knows, gene regulation in eukaryotes is a multi-step process and new RNA formed after transcription is usually modified, transported, localized and translated. With the emergence of high-throughput technology in genomics, due to genetic and epigenetic mechanisms, the research has been concentrated on the change of transcriptional level [6]. Many studies have often shown that there is a lack of significant correlation between transcripts and protein levels in cells [7]. These observations lead people to believe that other

processes may also play an important role in the cell pool that affects the translation of proteins from their respective transcripts. This paradox can be further explained by identifying post-transcriptional regulatory points, which make a great contribution to the regulation of protein level. These checkpoints are mainly composed of regulation mediated by non-coding RNA (microRNA, circular RNA and long non-coding RNA) and RNA-binding proteins (RBPs) [8, 9].

RBPs bind to newborn RNAs in the whole process of cells [10]. The versatility and wide range of functions of RBPs regulatory targets make them critical post-transcriptional regulators [10]. Therefore, it is necessary to understand the structure and function of these molecules for comprehending the many processes that have changed due to the misregulation of these proteins. Many of these misregulated RBPs have also been shown to contribute to the pathogenesis of cancer [11].

In this study, besides exploring the potential mechanism of regulating the abnormal expression of RBPs, it was also found that using multiple RBP integrated model, RBP may affect the prognosis of CRC, thus improving the prediction accuracy of prognosis. We procured the results of functional analysis of differentially expressed RBPs, in CRC from online database, which prompted the construction of an alternative splicing (AS) network of CRC after acquiring differentially expressed AS events related to the prognosis of RBP. We hypothesized that gene mutation and DNA methylation were the potential mechanisms for regulating abnormal expression of RBP. Hence, a number of online databases were used to analyze and verify that mutation and DNA methylation were involved in the regulation of abnormal expression of RBPs. Moreover, univariate and multivariate proportional hazard regression analysis were applied to further screen prognostic RBP genes from The Cancer Genome Atlas (TCGA) CRC cohorts and establish the optimal risk model, which was verified in randomized trial group. Finally, through the analysis of a variety of publicly available data sets, the expression of RBP genes of the model in CRC was further analyzed.

Materials And Methods

2.1 Acquisition of RBPs

Based on the data reported by Gerstberger *et al.* in 2014 [12], a complete list of 1542 RBPs used in this study was obtained (Supplementary Table 1). This list was used for all the analyses in this study.

2.2 Differential expression analysis and functional enrichment analysis

The abnormally expressed genes in CRC samples were collected on the Gene Expression Profiling Interactive Analysis 2 (GEPIA 2) database (<http://gepia2.cancer-pku.cn/#index>) (condition set to Dataset: COAD or READ, |Log2FC| Cutoff: 1; q-value Cutoff: 0.05; Differential Methods: LIMMA), of which 300 RBP genes were differentially expressed in CRC (Supplementary Table 2). Using R software, the gene name was converted to entrezID by referring to "org.Hs.eg.db" package. Then the R packets "clusterProfiler", "org.Hs.eg.db", "enrichplot" and "ggplot2" were performed to analyze the enrichment of gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) and to visualize the results.

2.3 AS data analysis

The transcriptome and clinical information of CRC were obtained from the TCGA GDC platform (<https://portal.gdc.cancer.gov/>; released before October 27, 2019), and the AS data of TCGA CRC samples were taken from the TCGA SpliceSeq platform (<https://bioinformatics.mdanderson.org/TCGASpliceSeq/PSIdownload.jsp>; released before July, 10, 2020). Univariate Cox analysis with R software was used to find AS events related to survival (Supplementary Table 3). The correlation (correlation coefficient $R > 0.55$, P value < 0.001) between RBP expression and the Percent-spliced-in (PSI) value of survival-related AS was analyzed using the function `cor.test()` in R software. The network diagram was generated by Cytoscape (version 3.7.1).

2.4 Mutation analysis and methylation data analysis of RBP gene

The gene mutation data of 8930 CRC samples were downloaded from COSMIC website (<https://cancer.sanger.ac.uk/cosmic>) (Supplementary Table 4), and the gene mutation frequencies of 1542 RBP were investigated. The mutation of RBP with high mutation frequency in CRC was further studied graphically in cBioPortal (<http://www.cbioportal.org/>). The methylation data of 8930 CRC samples were obtained on the [Catalogue of Somatic Mutations in Cancer](#) (COSMIC) website (Supplementary Table 5) to investigate the methylation of abnormally expressed RBP genes in CRC. In addition, we used the TCGA 450k array downloaded from UCSC (<https://xenabrowser.net/datapages/>; released before July, 20, 2020) (Supplementary Table 5) to analyze the independent prognosis of TCGA CRC methylation sites and found that there were prognostic methylation sites in RBP. Finally, the data of the expression level and the corresponding methylation degree of the interested RBP genes were retrieved on Cancer Cell Line Encyclopedia (CCLE) (<https://portals.broadinstitute.org/ccle>) (Supplementary Table 6), and the Pearson analysis was carried out by using Graphpad 8.0 software.

2.5 Construction and verification of prognostic model

Genes related to CRC survival were taken from the GEPIA 2 database (condition setting: Dataset selection: COAD or READ, Methods: overall or disease free survival; Group Cutoff: Median), of which 96 RBP genes were related to CRC survival (Supplementary Table 2). The transcripts and clinical information of 96 RBPs mentioned above were downloaded from the TCGA GDC platform (<https://portal.gdc.cancer.gov/>; released before October 27, 2019), and the survival-related RBP with significant difference was identified by univariate regression analysis using the 'survival' package of R software. Using R software and data partition function `createDataPartition()`, the patients with TCGA CRC were randomly divided into train group and test group (Supplementary table 7). Then multivariate regression analysis was carried out according to the data of train group and the prognosis model was constructed.

In order to verify the accuracy of the model, the calculation formula of the constructed model was first used: the risk score of the sample = \sum (the Cox regression coefficient X RBP expressed by the log₂ ratio value of a specific RBP). The risk score of the train group was calculated, and the risk was divided into

two groups according to the median value of the risk. After that, R software was used for Kaplan-Meier (KM) survival analysis, univariate and multivariate independent prognostic analysis, and the receiver operator characteristic (ROC) curve drawing. Finally, according to the median value of train group, test group was also divided into high- and low-risk groups. KM survival analysis, univariate and multivariate independent prognostic analysis and the ROC curve were also carried out.

2.6 The five RBP signatures of the model are validated at transcription and protein levels

The expression data of five RBP models in various CRC cell lines were obtained from the CCLE platform (Supplementary Table 8), and their expressions in cell lines from different sources were analyzed. The expression of the five RBP models at the CRC organizational level was searched on The Human Protein Atlas website (<https://www.proteinatlas.org/>).

2.7 Cell lines, RNA isolation and qRT-PCR

CRC cells (HCT 116, SW480, SW620, Caco2, RKO) and normal intestinal epithelial cells (NCM460) grew in 1640 (GIBCO; Thermo Fisher Scientific, Inc.) supplemented with 10% FBS (Sera Gold, Germany) , and 1% penicillin/streptomycin. RNA was isolated using TRI reagent (SIGM-Aldrich, USA) following manufacturer's instructions. Synthesis of cDNA was using the Prime-Script RT reagent kit (Takara Bio, Inc.), 2.0 µg RNA was converted into cDNA. qRT-PCR was performed to quantify the transcript levels under various conditions. GAPDH was used as an internal control, and the $\Delta\Delta C_t$ method was used to calculate gene expression. The primers used in this study were CAPRIN2 (forward): 5'-CAGAGACTCCTGAGGCAGCAATTC-3', CAPRIN2 (reverse): 5'-GAAGCCCTGTTCAGAGCCCTTTG-3', GAPDH (forward): 5'-CTCCTCTGTTCGACAGTCAGC-3', GAPDH (reverse): 5'-CCCAATACGACCAAATCCGTT-3'. Statistical analysis was performed using GraphPad Prism version 8.0 (GraphPad Software, Inc.). The data were expressed as mean \pm standard deviation. The student's t test was used to compare the two groups. P value <0.05 was considered to indicate a statistically significant difference.

Results

3.1 Differentially expressed RBPs in CRC and their enrichment network

We acquired an exhaustive list of 1542 RBPs from the literature published by Gerstberger *et al.* (Supplementary Table 1). The CRC differential genes were downloaded from the GEPIA 2 online database, and it was found that there were 71 down-regulated RBP genes and 229 up-regulated RBP genes in CRC (Supplementary Table 2). In order to understand the potential mechanism of RBP in CRC, KEGG analysis and GO analysis were performed on the above 300 differentially expressed RBP genes (Figure 1).

3.2 RBPs in CRC participate in the network regulation of AS

It could be seen from Figure 1 that RBPs were enriched in the spliceosome pathway, indicating that RBPs were important part of participating in AS. As we all know, AS events are mainly regulated by splicing

factors. Splicing factors bind to pre-mRNA and affect exon selection and splicing site selection [13]. More importantly, AS disorders in the tumor microenvironment may be caused by a limited number of splicing factors [14]. Thus, an important question was whether a large part of these AS events related to the prognosis of CRC (Supplementary Table 3) may be regulated by some key splicing factors. According to the data collected by zhen *et al.* in 2018 [15], most of the splicing factors belong to RBPs. Therefore, the regulatory relationship between RBPs and AS events was worth exploring. For this purpose, the expression of RBPs from the RNA sequencing data of the TCGA CRC cohort was analyzed. Next, in the CRC cohort, through the splicing regulation network established by significant correlation ($|R|>0.55$, t test, $P<0.05$), the correlation between the expression levels of these 1542 RBPs and the PSI value of each AS events related to the prognosis of CRC was analyzed. In the splicing-related network shown in Figure 2, there were 31 AS events related to the prognosis including 24 up-regulated AS events (red dots) and 7 down-regulated AS events (green dots) significantly correlated with 22 RBPs (purple dots). Interestingly, most RBPs (purple dots) were associated with more than one AS event, some of which played an opposite role in the regulation of differential AS events. Moreover, our network revealed that different splicing factors competed for the same binding site (AS event), which at least partly explained why the transcript was able to produce several different splicing isoforms. In addition, we observed from the Figure 2 that DDX39B as RBP had the most AS events, mainly up-regulating AS events. This showed that DDX39B may be a key factor in modulating AS events related to CRC prognosis.

3.3 Potential mechanisms that may be involved in the regulation of differentially expressed RBP genes in CRC

Gene mutations are ubiquitous and occur spontaneously. Not all mutations cause obvious changes in cell functions [16]. However, mutations in key cellular genes cause developmental disorders, which is one of the main ways that proto-oncogenes are transformed into a carcinogenic state [17]. The gradual accumulation of multiple mutations in life will result in cancer, which is also one of the important mechanisms for the occurrence and development of CRC [18]. We speculated that the underlying mechanism of RBP gene differential expression may be caused by mutations, so we used 8,920 CRC samples on the COSMIC platform to investigate the occurrence of differentially expressed RBP gene mutations. We were surprised to discover that all the differentially expressed RBP genes were mutated in CRC samples, and the mutation frequency of 11 RBP genes (PRKDC, RBMS3, SRRM2, HELZ2, MSI2, AFF3, DZIP1, TNRC6A, SND1, QKI, ESRP1) was more than 5% (Supplementary Table 4). Then, in order to further understand the mutations of the first 11 RBP genes, we analyzed 526 TCGA colorectal adenocarcinoma samples (TCGA, pancancer Atlas) on the cBioportal platform, and 180 (34%) samples showed mutations (Figure 3A). The mutation frequency of these 11 differentially expressed RBP genes in CRC was indeed high, and the mutation frequency of PRKDC and HELZ2 was more than 10% in the Figure 3A. These two genes may be key therapeutic targets for CRC.

Additionally, we also noticed the role of epigenetics in CRC. Therefore, we investigated the methylation of differentially expressed RBP genes on 8920 CRC samples on the COSMIC platform and revealed that 21 differentially expressed RBP genes had DNA methylation changes (Supplementary Table 5). Thus, we

used the TCGA 450k array to perform independent prognostic analysis of TCGA CRC methylation sites and determined that 64 RBPs with prognostic methylation sites (Supplementary Table 5) included 10 differentially expressed RBP genes (RPL37, NOL10, CD3EAP, EIF5A, OASL, NHP2, RRS1, NUFIP1, RRP12 and EIF4E). Then we tested the correlation between the degree of methylation of these 10 genes on CRC cell lines and their expression on the CCLE platform. It was observed that the degree of RPL37 methylation was negatively correlated with its mRNA expression level (Figure 3B).

3.4 Construction of RBPs related prognosis model

RBPs play an important role in the occurrence and development of CRC [19]. We collected 1,855 survival-related genes for CRC through the GEPIA 2 database and found that among the 1,542 RBPs that have been cataloged, only 96 RBPs were related to the survival of CRC. Simultaneously, CRC transcripts and clinical information from the TCGA platform were downloaded, and data integration was carried out to obtain the clinical information of 540 CRC patients using R software. Thus, the 96 RBPs obtained above were subjected to univariate cox regression analysis to further screen twenty RBPs related to CRC survival (Figure 4). Next, using R software, the above 540 CRC patients were randomly divided into two groups according to their survival status (to ensure that the number of surviving patients and the number of dead patients is not much different), namely the train group and the test group (Supplementary Table 4). Finally, according to the train group, a multivariate analysis was performed to construct a CRC prognostic prediction model of five genes (CAPRIN2, RPL3L, CCAR2, GSPT1 and MRPS18C) (Table 1).

3.5 Validation of RBPs related prognostic model

By using the risk score formula to combine the effects of each of these five RBP genes, the RBP risk score was calculated for each patient in the train group and the test group. According to the RBP risk score, CRC patients were divided into low-risk and high-risk groups (Figure 5). We conducted a survival analysis on the risk scores of the train group and the test group and determined that the risk scores were both poor prognostic indicators (Figure 6). For further verification, univariate and multivariate independent prognostic analysis involving age, gender, and stage in the train group and the test group determined that the RBP risk score was an independent predictor of patient survival (Figure 6). Lastly, the ROC curve of the train group and the test group was drawn using R. According to its area under the curve (AUC) value, the accuracy of the model was basically at a medium level (Figure 7).

3.6 Genetic alterations of five RBP signatures in CRC

We put the five RBP signatures of the prognostic model on the Oncomine platform (Hong colorectal statistics) and compared the CRC samples with the normal samples. We observed that the transcription level of CAPRIN2, GSPT1 and CCAR2 (CCAR2 is also known as DBC1) showed an upward trend, while the transcription level of RPL3L and MRPS18C decreased in CRC (Figure 8). Subsequently, we performed RT-PCR detection of CAPRIN2 mRNA in six existing CRC cell lines and normal intestinal epithelial cells (Figure 9A), which confirmed that CAPRIN2 was indeed highly expressed in CRC. Besides, we investigated the transcription levels of the five signatures in various CRC cell lines on the CCLE platform. The

expression of RPL3L mRNA in CRC cell lines of patients with TNM stage III was lower than that of patients with TNM stage II (Figure 9B). In addition, except for the LS411N cell line (from poorly differentiated CRC patient), the expression of CCAR2 mRNA in CRC cell lines of patients with Ducks' type B was significantly lower than that of patients with Ducks' type C-D, indicating that CCAR2 may be related to the progression and malignant degree of CRC (Figure 9C). Finally, we analyzed the immunohistochemical results of five signatures in normal colorectal tissue and CRC on the platform of The Human Protein Atlas (<https://www.proteinatlas.org/>). According to the results of immunohistochemistry, the expression level of MRPS18C in CRC was lower than that in normal tissues (Figure 9D and 9E), which was consistent with the result of mRNA level.

Discussion And Conclusion

RBPs are a general term for a class of proteins that accompany RNA to regulate metabolic processes and bind to RNA [11]. Their main role is to mediate the maturation, transport, localization and translation of RNA. One RBP may have multiple target RNAs, and its expression defects can cause multiple diseases [10, 12]. Some scholars believed that RBP is the key to regulate the malignant transformation of CRC [20]. This study provided a comprehensive picture of RBP regulation and its importance in the occurrence and progression of CRC. In our analysis, we revealed that the main proportion of RBP in CRC samples was up-regulated. This was consistent with previously published literature, which showed that most RBPs are up-regulated in various cancers compared to their normal counterparts [20]. This suggests that most of them may have tumor-promoting effects.

After understanding the knowledge that abnormal changes in RBP may lead to CRC, we investigated the main functional pathways of abnormal expression of RBP in CRC. The enriched spliceosome pathway had attracted our attention. Genomic research showed that more than 90% of the genes in the human body have AS events [21]. This process is strictly regulated in different tissues and different physiological stages, and its imbalance leads to a variety of diseases [14]. The *in vivo* regulation of AS is mainly achieved by the recruitment of trans-acting splicing factors by cis-elements in the precursor mRNA [13]. Generally, trans-splicing factors have a modular structure, which includes one or more RNA binding domains and different functional modules [13]. However, the current research on these functional domains is still limited to a few typical splicing factors, such as the serine/arginine protein family and heterogeneous nuclear ribonucleoproteins protein family, but little is known about the functional modules of other RBPs [22]. In-depth understanding of these functional modules will provide a basis for scientists to further study and even synthesize new RNA splicing factors *de novo*. After realizing the importance of RBP in AS, our research constructed its network for regulating AS. We observed that RBP DDX39B could regulate multiple prognostic-related AS events and even control AS events related to different prognosis of the same genes, such as EIF2B1, EXOSC10, which showed that DDX39B may be an essential RBP affecting CRC. *DDX39B* gene encodes a member of the DEAD-box family of RNA-dependent ATPases, which mediates ATP hydrolysis during pre-mRNA splicing [23]. DDX39B protein is an essential splicing factor required for the association of U2 small ribonucleoprotein with pre-mRNA and also functions in the

export of mRNA from the nucleus to the cytoplasm [23]. Tumors related to DDX39B include prostate cancer and melanoma [24, 25], and the situation in CRC has not been reported in the literature.

Cancer is the result of the interaction between environmental factors and cell genetic material, the result of multi-factor, multi-stage, and multi-gene effects, and the result of accumulation of gene mutations [26]. Therefore, cancer is a genetic disease. In the numerous of human genes, proto-oncogenes and tumor suppressor genes are closely related to the occurrence and development of cancer [27]. Mutations in proto-oncogenes and tumor suppressor genes can cause cell canceration [28]. In our research, we found that these aberrantly expressed RBPs had mutations in more or less CRC samples. Among them, the two RBPs with the highest mutation frequency, PRKDC and HELZ2, were searched on the cBioportal platform. The *PRKDC* gene encodes the catalytic subunit of DNA-dependent protein kinase (DNA-PK) [29]. Together with Ku70/Ku80 heterodimeric protein, PRKDC has an effect on DNA double-strand break repair and recombination [30]. PRKDC-related pathways include AKT signaling pathway [31], and PRKDC expression has been shown to be positively correlated with the poor prognosis of CRC, which is a key factor in promoting drug resistance and proliferation of CRC [32, 33].

Epigenetics means that the DNA sequence does not change, but the gene expression has undergone heritable changes [34]. One of the epigenetic factors that we studied here was RBP methylation. Changes in gene methylation may have critical effects on gene expression. As a recognized fact, in most cases, hypermethylation of gene promoters leads to the inhibition of transcription, and vice versa [35]. In our research, we found that methylation participated in regulating the expression of RBP in CRC. It was shown that 21 abnormally expressed RBPs had hypomethylation and/or hypermethylation in CRC. Additionally, we also identified 10 RBPs with methylation sites that were related to prognosis. This may emphasize the importance of abnormally expressed RBPs in CRC, so the abnormal expression of these RBPs in CRC samples was regulated by one or more mechanisms.

Further, we identified five RBPs with prognostic significance in CRC tumor patients. These signatures composed of CAPRIN2, RPL3L, CCAR2, GSPT1 and MRPS18C divided CRC patients into low-risk and high-risk group. The five RBP signatures used to construct the model were closely related to the patient's prognosis in univariate and multivariate analysis and were independent factors of the patient's prognosis. Among them, CAPRIN2 and RPL3L were malignant genes, CCAR2, and GSPT1 and MRPS18C were protective genes. Previous studies have reported that CCAR2 is involved in the regulation of the WNT signaling pathway [36], and the corresponding gene products may also become a therapeutic target for excessive inflammation [37]. GSPT1 is involved in the regulation of mammalian cell growth [38]. It is highly expressed in HCT 116 [39], which is consistent with the results of our analysis. The remaining three RBPs currently have no CRC-related clinical or basic research, and their mechanism of action in CRC is still unclear and needs to be further explored.

The clinical characteristics of CRC, such as pathological type, tissue type, and location of occurrence, are significantly different in different patients, and the prognosis of patients with the same stage is also different [40]. Therefore, more understanding of the prognostic factors of CRC is needed. The model

divided patients into high-risk groups and low-risk groups. The survival rates between the two groups were significantly different. The 1-year, 3-year, and 5-year ROC curve AUC drawn by the train group were all greater than 0.70. The AUC of 1-year, 3-year and 5-year ROC area of test group were greater than 0.63, which further verified the model. Collectively, this prediction model may be used as a useful supplement to TNM staging.

To the best of our knowledge, this study provided the first relatively comprehensive view of the abnormally regulated RBP in CRC and its mechanism, which may result in abnormal regulation. It also provided insights into the regulation of RBP related to AS events and insights that may be related to CRC-associated pathways. We also developed the RBP signatures, which proved to be reliable independent prognostic factors in CRC. This may be clinically helpful, while making treatment-related decisions for CRC patients. However, as an exploratory study, its application value still needs to be further verified by multi-center large sample clinical research.

Abbreviations

RBP RNA binding protein

CRC colorectal cancer

AS alternative splicing

TCGA The Cancer Genome of Atlas

GEPIA 2 Gene Expression Profiling Interactive Analysis 2

GO gene ontology

KEGG Kyoto Encyclopedia of Genes and Genomes

PSI Percent-spliced-in

CCLC Cancer Cell Line Encyclopedia

KM Kaplan-Meier

ROC receiver operator characteristic

AUC area under the curve

DNA-PK DNA-dependent protein kinase

COSMIC Catalogue of Somatic Mutations in Cancer

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All listed authors were actively involved in the study, reviewed and approved the submitted manuscript.

Availability of data and materials

All data and materials could be found in our published paper.

Conflict of Interest

None

Funding

This work was supported by National Natural Science Foundation of China (81472275), Natural Science Foundation of Guangdong Province (2020A151501303, 2014A030313542), Major projects of key platforms for colleges and universities in Guangdong Province (2020KTSCX045, 2019KTSCX046), Research Fund of Guangdong Medical University (GDMUZ2020001).

Authors' contributions

QLH, ZQL and XL conducted data analysis and drafted manuscripts. HBY and YLD were involved in research design and data collection. Finally, WZ revised the manuscript.

Acknowledgements

None

References

1. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol*. 2019; 16(12):713-732.
2. Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M *et al*. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *Eur J Cancer*. 2018; 103:356-387.
3. Ladabaum U, Dornitz JA, Kahi C, Schoen RE. Strategies for Colorectal Cancer Screening. *Gastroenterology*. 2020; 158(2):418-432.
4. Sepulveda AR, Hamilton SR, Allegra CJ, Grody W, Cushman-Vokoun AM, Funkhouser WK *et al*. Molecular Biomarkers for the Evaluation of Colorectal Cancer: Guideline From the American Society

- for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and the American Society of Clinical Oncology. *J Clin Oncol*. 2017; 35(13):1453-1486.
5. Kather JN, Halama N, Jaeger D. Genomics and emerging biomarkers for immunotherapy of colorectal cancer. *Semin Cancer Biol*. 2018; 52(Pt 2):189-197.
 6. Thakore PI, Black JB, Hilton IB, Gersbach CA. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat Methods*. 2016; 13(2):127-137.
 7. Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*. 2016; 165(3):535-550.
 8. Janakiraman H, House RP, Gangaraju VK, Diehl JA, Howe PH, Palanisamy V. The Long (lncRNA) and Short (miRNA) of It: TGF β -Mediated Control of RNA-Binding Proteins and Noncoding RNAs. *Mol Cancer Res*. 2018; 16(4):567-579.
 9. Cai Y, Lei X, Chen Z, Mo Z. The roles of cirRNA in the development of germ cells. *Acta Histochem*. 2020; 122(3):151506.
 10. Gerstberger S, Hafner M, Ascano M, Tuschl T. Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. *Adv Exp Med Biol*. 2014; 825:1-55.
 11. Kim MY, Hur J, Jeong S. Emerging roles of RNA and RNA-binding protein network in cancer cells. *BMB Rep*. 2009; 42(3):125-130.
 12. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nature reviews Genetics*. 2014; 15(12):829-845.
 13. Lee Y, Rio DC. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem*. 2015; 84:291-323.
 14. Urbanski LM, Leclair N, Anczuków O. Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. *Wiley Interdiscip Rev RNA*. 2018; 9(4):e1476.
 15. Zong Z, Li H, Yi C, Ying H, Zhu Z, Wang H. Genome-Wide Profiling of Prognostic Alternative Splicing Signature in Colorectal Cancer. *Frontiers in oncology*. 2018; 8:537.
 16. Simpson AJ. The natural somatic mutation frequency and human carcinogenesis. *Adv Cancer Res*. 1997; 71:209-240.
 17. Anderson MW, Reynolds SH, You M, Maronpot RM. Role of proto-oncogene activation in carcinogenesis. *Environ Health Perspect*. 1992; 98:13-24.
 18. Nguyen LH, Goel A, Chung DC. Pathways of Colorectal Carcinogenesis. *Gastroenterology*. 2020; 158(2):291-302.
 19. Fagoonee S, Picco G, Orso F, Arrigoni A, Longo DL, Forni M *et al*. The RNA-binding protein ESRP1 promotes human colorectal cancer progression. *Oncotarget*. 2017; 8(6):10007-10024.
 20. Chatterji P, Rustgi AK. RNA Binding Proteins in Intestinal Epithelial Biology and Colorectal Cancer. *Trends Mol Med*. 2018; 24(5):490-506.

21. Zaharieva E, Chipman JK, Soller M. Alternative splicing interference by xenobiotics. *Toxicology*. 2012; 296(1-3):1-12.
22. Kędzierska H, Piekietko-Witkowska A. Splicing factors of SR and hnRNP families as regulators of apoptosis in cancer. *Cancer Lett*. 2017; 396:53-65.
23. Shen H. UAP56- a key player with surprisingly diverse roles in pre-mRNA splicing and nuclear export. *BMB Rep*. 2009; 42(4):185-188.
24. Nakata D, Nakao S, Nakayama K, Araki S, Nakayama Y, Aparicio S *et al*. The RNA helicase DDX39B and its paralog DDX39A regulate androgen receptor splice variant AR-V7 generation. *Biochem Biophys Res Commun*. 2017; 483(1):271-276.
25. Awasthi S, Chakrapani B, Mahesh A, Chavali PL. DDX39B promotes translation through regulation of pre-ribosomal RNA levels. 2018; 15(9):1157-1166.
26. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009; 458(7239):719-724.
27. Spandidos DA. Oncogenes and tumor suppressor genes as paradigms in oncogenesis. *J buon*. 2007; 12 Suppl 1:S9-12.
28. Spandidos DA, Liloglou T, Field JK. Prognostic significance of oncogenes and tumor suppressor genes in human malignancy. *Stem Cells*. 1993; 11(3):194-198.
29. Yin Y, He Q, Li Y, Long J, Lei X, Li Z *et al*. Emerging functions of PRKDC in the initiation and progression of cancer. 2020:300891620950472.
30. Shrivastav M, De Haro LP, Nickoloff JA. Regulation of DNA double-strand break repair pathway choice. *Cell Res*. 2008; 18(1):134-147.
31. Li C, Liu X, Liu Y, Liu X, Wang R, Liao J *et al*. Keratin 80 promotes migration and invasion of colorectal carcinoma by interacting with PRKDC via activating the AKT pathway. *Cell Death Dis*. 2018; 9(10):1009.
32. Sun S, Cheng S, Zhu Y, Zhang P, Liu N, Xu T *et al*. Identification of PRKDC (Protein Kinase, DNA-Activated, Catalytic Polypeptide) as an essential gene for colorectal cancer (CRCs) cells. *Gene*. 2016; 584(1):90-96.
33. Wu L, Zhang J, Wu H, Han E. DNA-PKcs interference sensitizes colorectal cancer cells to a mTOR kinase inhibitor WAY-600. *Biochem Biophys Res Commun*. 2015; 466(3):547-553.
34. Harvey ZH, Chen Y, Jarosz DF. Protein-Based Inheritance: Epigenetics beyond the Chromosome. *Mol Cell*. 2018; 69(2):195-202.
35. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature reviews Genetics*. 2007; 8(4):286-298.
36. Yu EJ, Kim SH, Kim HJ, Heo K, Ou CY, Stallcup MR *et al*. Positive regulation of β -catenin-PROX1 signaling axis by DBC1 in colon cancer progression. *Oncogene*. 2016; 35(26):3410-3418.
37. Escande C, Chini CC, Nin V, Dykhouse KM, Novak CM, Levine J *et al*. Deleted in breast cancer-1 regulates SIRT1 activity and contributes to high-fat diet-induced liver steatosis in mice. *J Clin Invest*. 2010; 120(2):545-558.

38. Hoshino S, Miyazawa H, Enomoto T, Hanaoka F, Kikuchi Y, Kikuchi A *et al.* A human homologue of the yeast GST1 gene codes for a GTP-binding protein and is expressed in a proliferation-dependent manner in mammalian cells. *Embo j.* 1989; 8(12):3807-3814.
39. Xiao R, Li C, Chai B. miRNA-144 suppresses proliferation and migration of colorectal cancer cells through GSPT1. *Biomed Pharmacother.* 2015; 74:138-144.
40. Gallois C, Pernot S, Zaanani A, Taieb J. Colorectal Cancer: Why Does Side Matter? *Drugs.* 2018; 78(8):789-798.

Figures

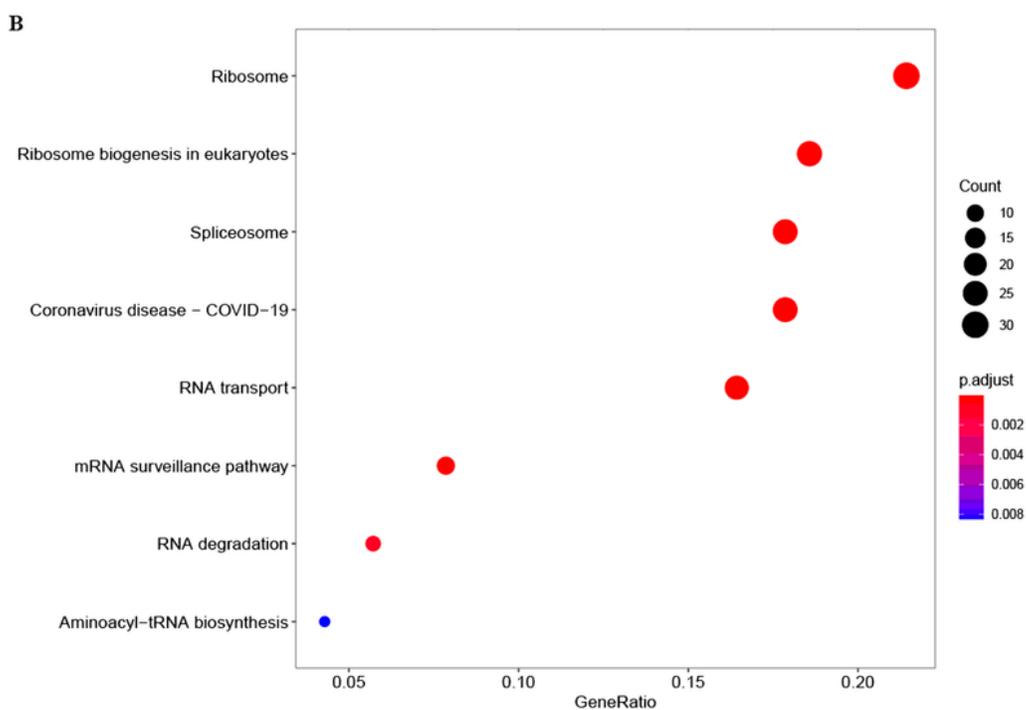
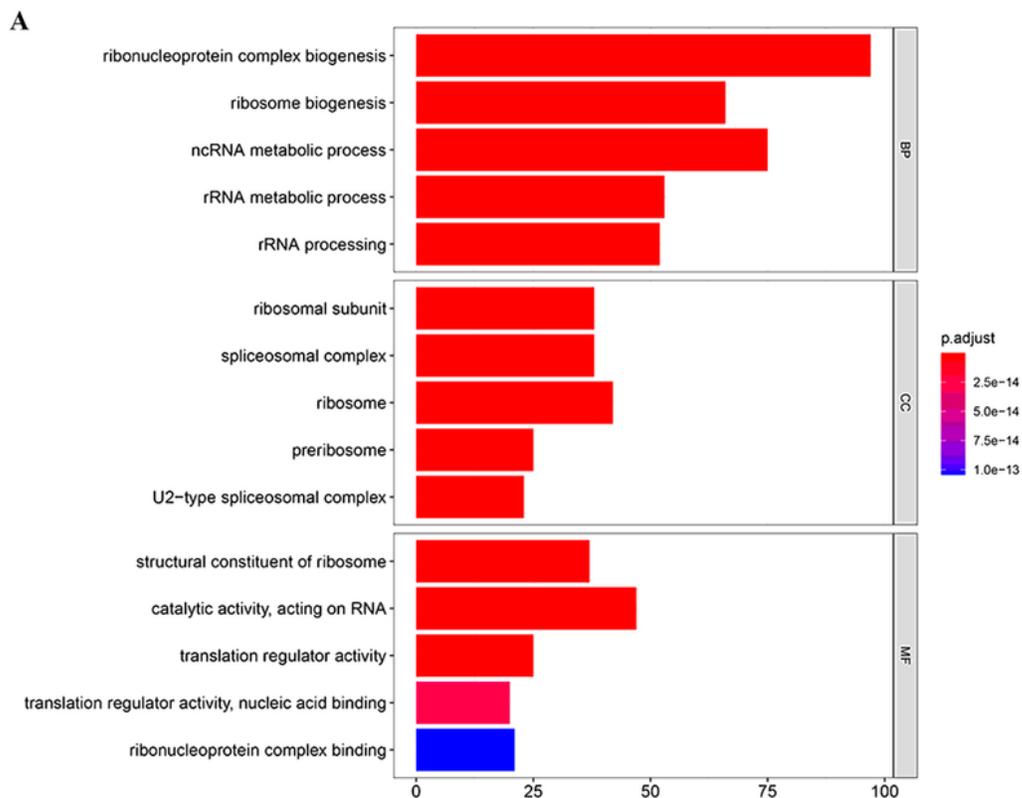


Figure 1

Functional enrichment analysis of RBP gene abnormally expressed in CRC. (A) is the results of biological process (BP), cellular component (CC) and molecular function (MF) enrichment in GO analysis of abnormally expressed RBP genes in CRC. (B) is the result of enriched signal pathway in KEGG analysis.

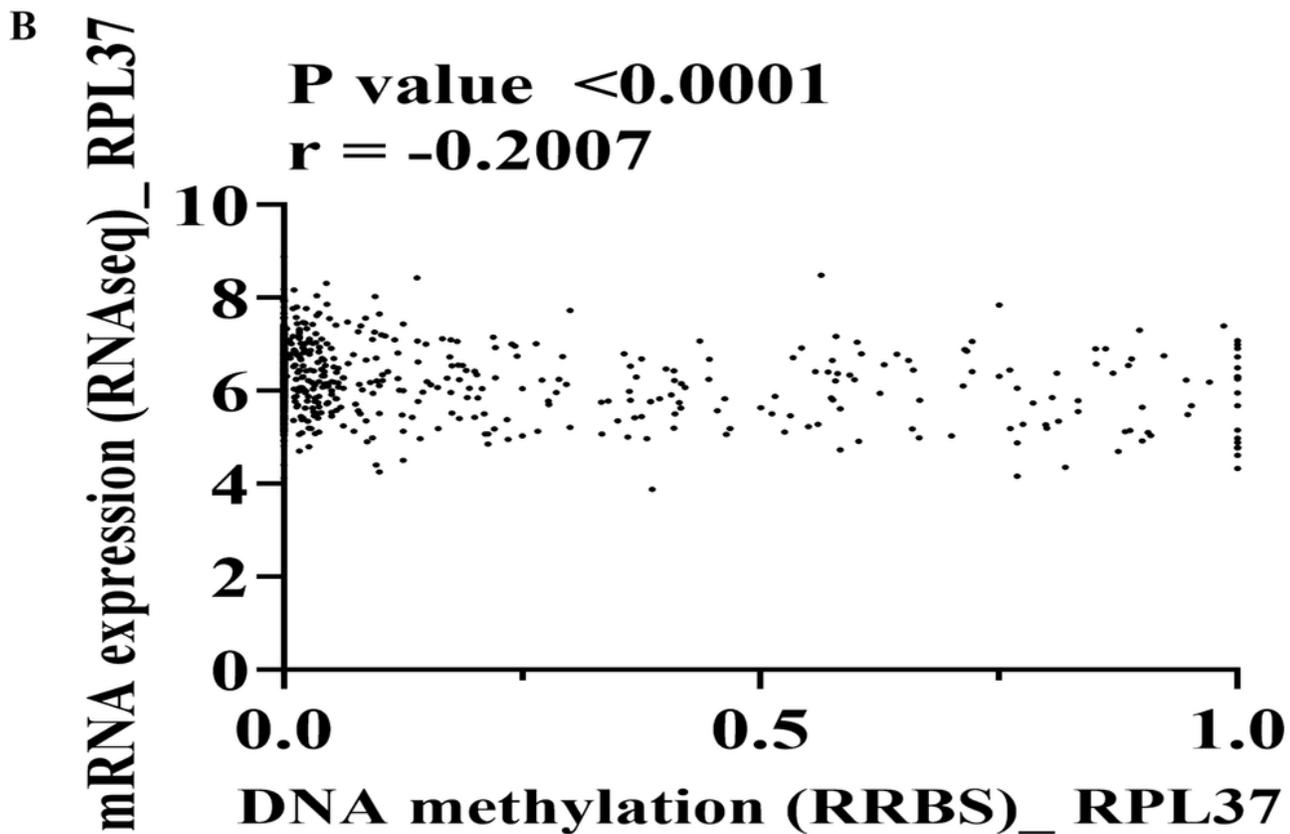
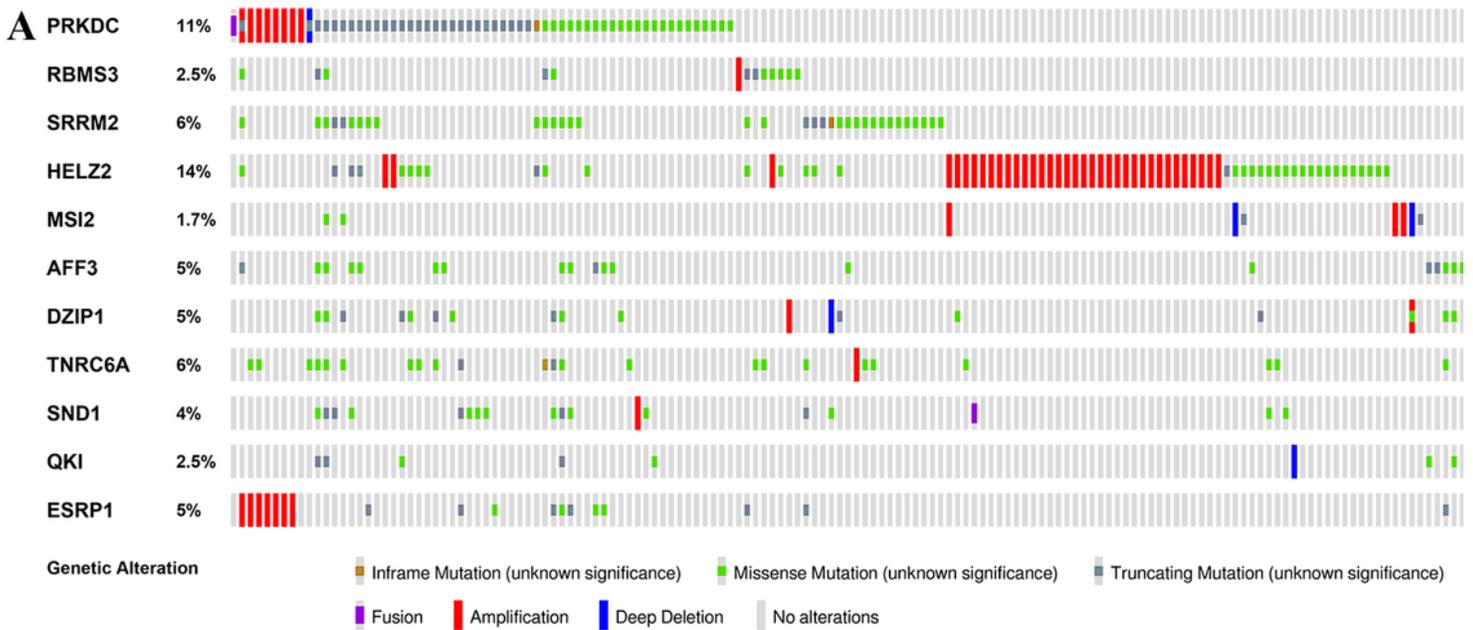


Figure 3

(A) Mutations of 11 interest genes in CRC. The graph shows CRC samples with gene mutation. Each gray bar represents a CRC sample. The brown stripes denote samples with inframe mutations in particular RBPs, the green stripes indicate the samples with missense mutation, while the black stripes indicate that truncating mutations were identified in RBP of CRC samples. (B) Analysis of the correlation between methylation degree and expression of RPL37 in CRC cell lines.

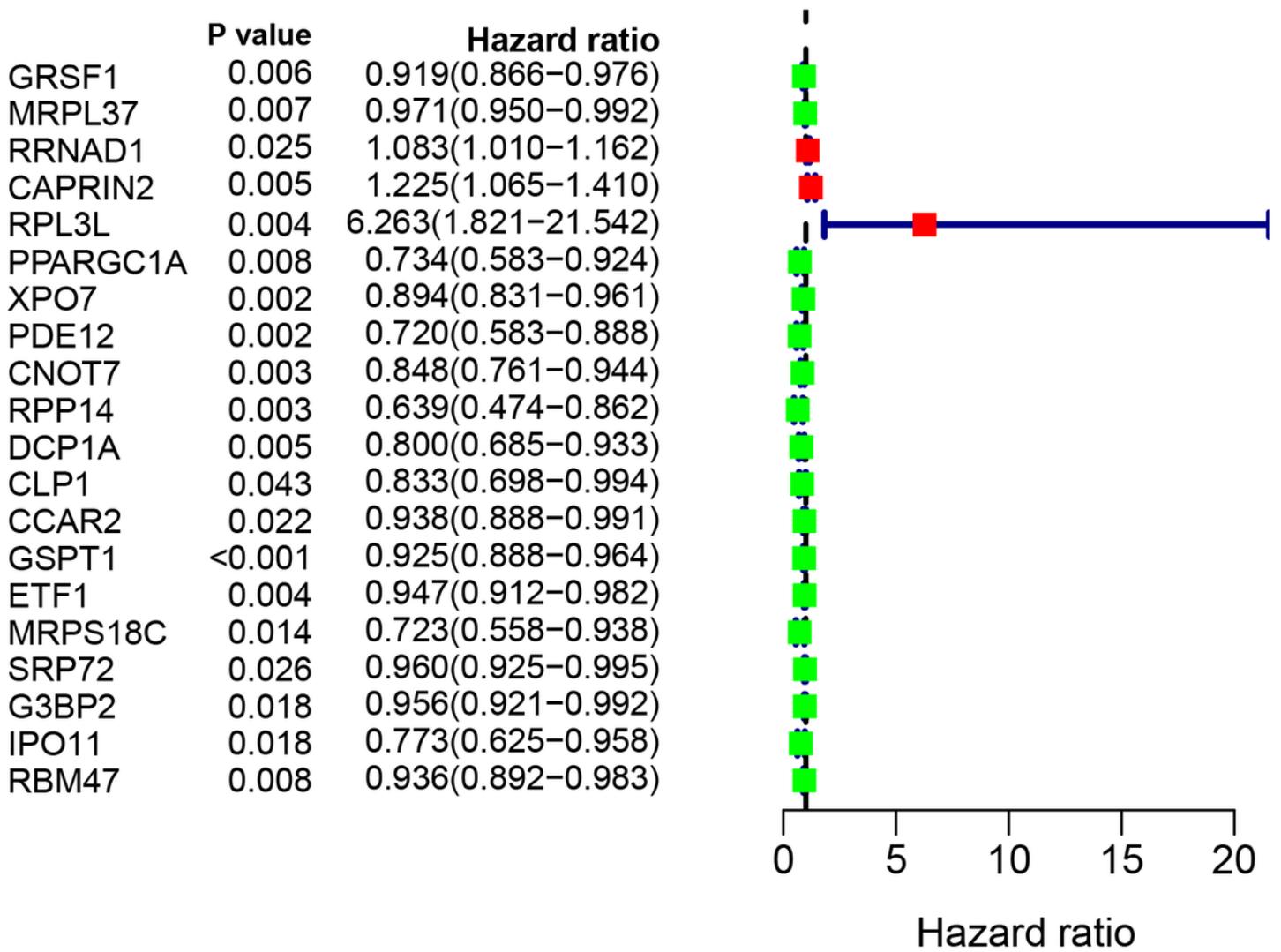


Figure 4

Multivariate regression analysis identified RBP genes associated with CRC prognosis.

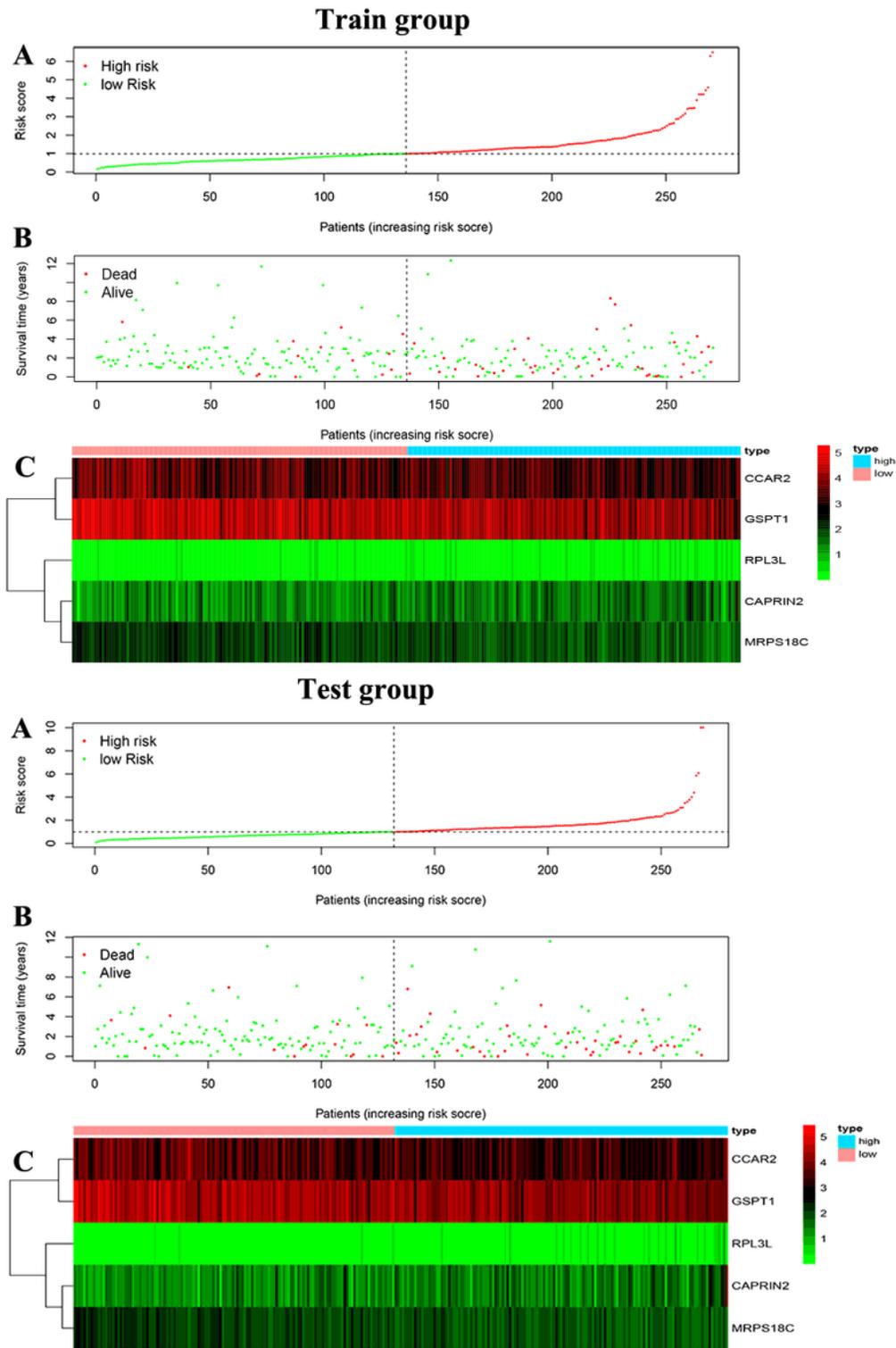


Figure 5

Characteristics of the prognostic gene signatures. (A-B) The distribution of risk score and patient's survival time, as well as status for Train group (A) and Test group (B). (A) The black dotted line is the optimum cutoff dividing patients into low-risk and high-risk groups. (C-D) Heatmap of the RBP gene expression profiles in prognostic signature for Train group (C) and Test group (D).

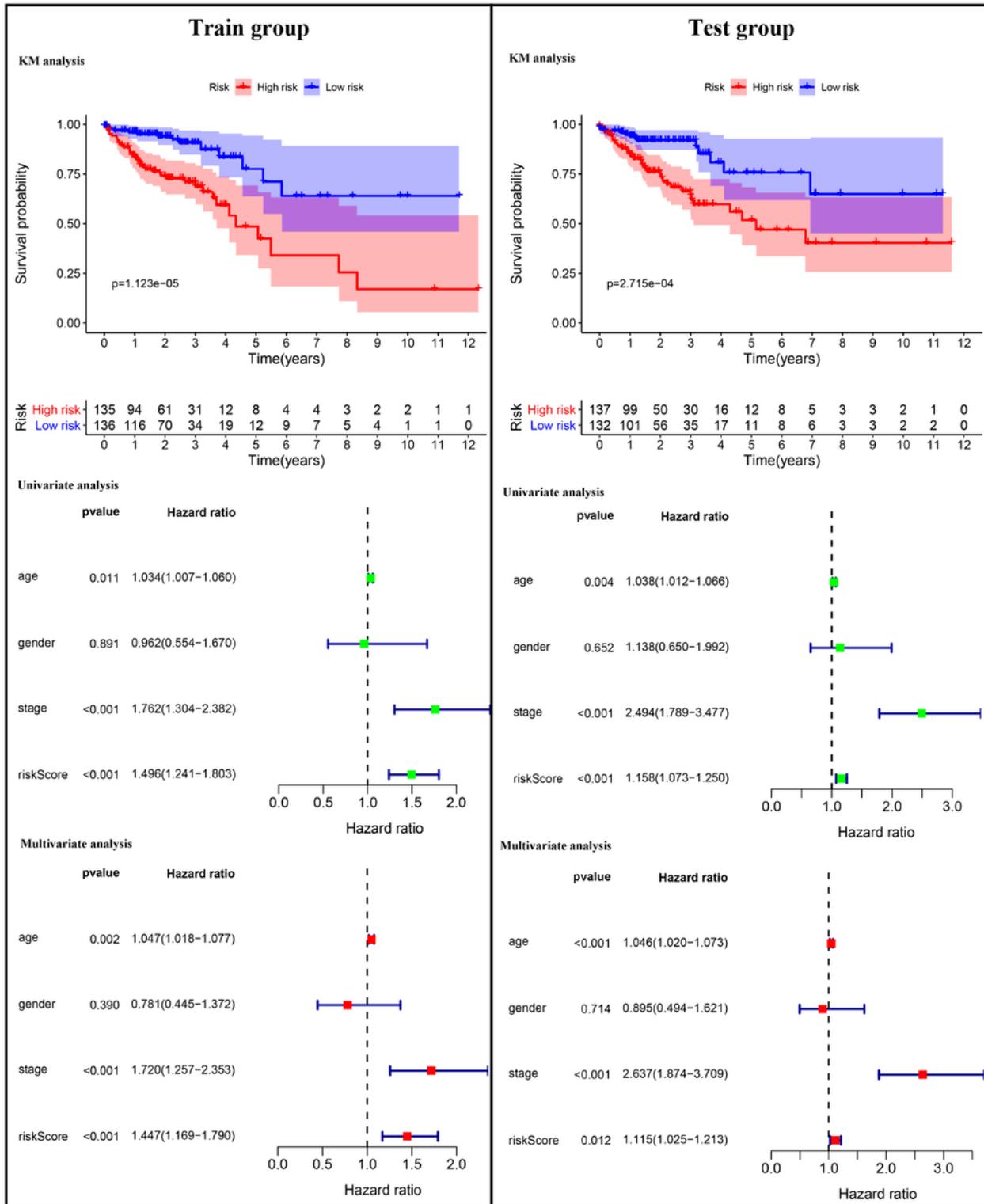


Figure 6

The KM survival curve, univariate independent prognostic analysis and multivariate independent prognostic analysis of the Train group (A) and the Test group (B) in the prediction model.

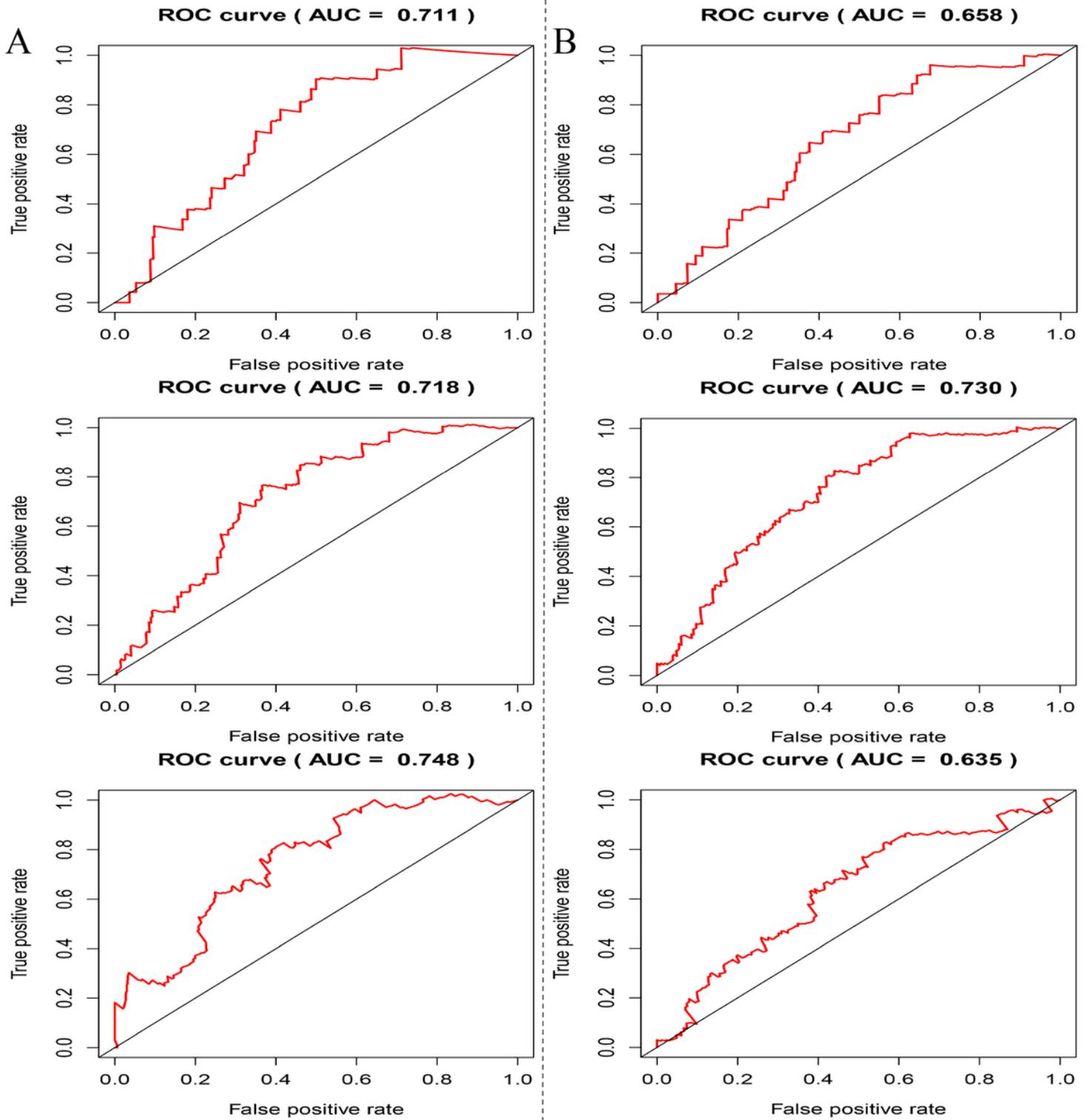


Figure 7

The 1-year, 3-year and 5-year ROC curves of the Train (A) group and Test group (B) in the prediction model.

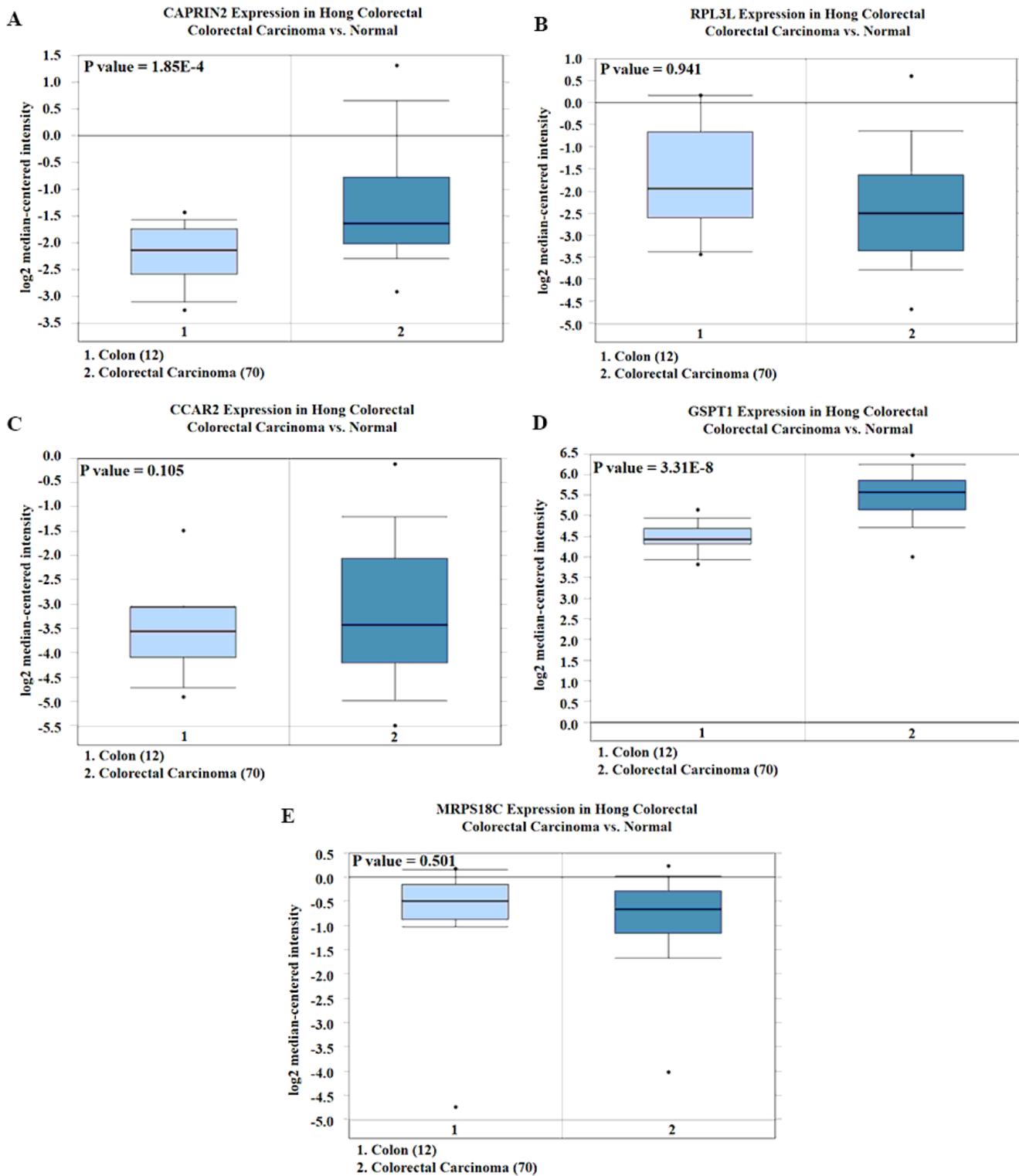


Figure 8

The expression of RBP genes in prognostic signatures for oncomine platform (A-E).

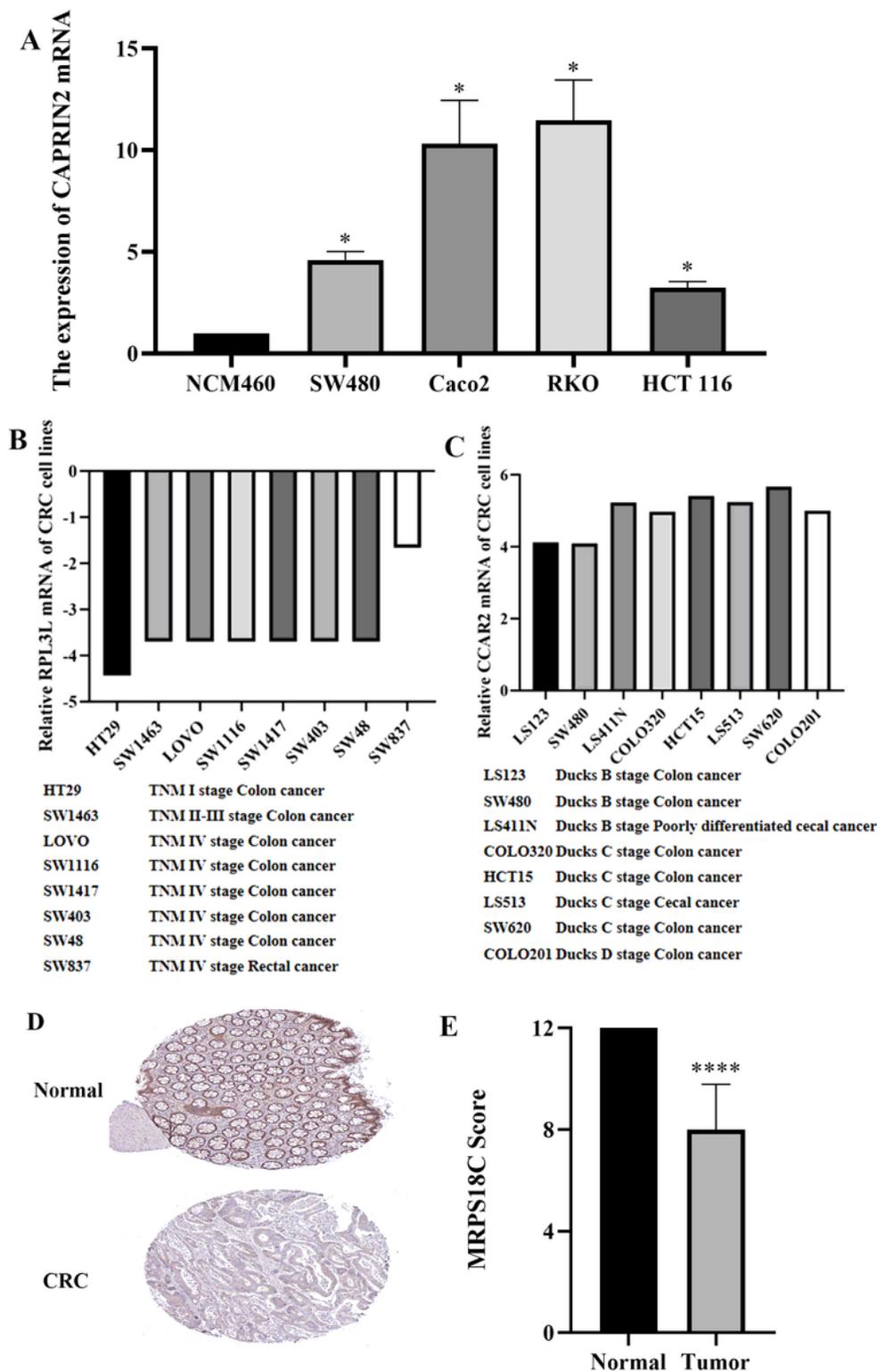


Figure 9

Verification of CAPRN2 expression in CRC at RNA and protein levels (A). Relative expression of RPL3L mRNA in CRC cell lines from different TNM stages (B). Relative expression of CCAR2 mRNA in CRC cell lines from different Ducks' stages (C). Immunohistochemical results of MRPS18C in normal and CRC tissues (D-E). (* $p < 0.05$)

Supplementary Files

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

- [SupplementaryTable1.xls](#)
- [SupplementaryTable2.xls](#)
- [SupplementaryTable3.xls](#)
- [SupplementaryTable4.xls](#)
- [SupplementaryTable5.xls](#)
- [SupplementaryTable6.xls](#)
- [SupplementaryTable7.xls](#)
- [SupplementaryTable8.xls](#)