

# Integrated analysis of the functions and prognostic values of RNA binding proteins in colon adenocarcinoma

**Xinhong Liu**

Chongqing University of Education

**Fang Tan**

Our Lady of Fatima University

**Xingyao Long**

Chongqing University of Education

**Ruokun Yi**

Chongqing University of Education

**Dingyi Yang**

Chongqing University

**Xin Zhao** (✉ [zhaoxin@cque.edu.cn](mailto:zhaoxin@cque.edu.cn))

Chongqing University of Education

---

## Research article

**Keywords:** colon adenocarcinoma, RNA-binding protein, survival, prognosis, bioinformatics

**Posted Date:** June 2nd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-29330/v1>

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

[Read Full License](#)

---

# Abstract

## Background

RNA binding proteins (RBPs) play an important role in a variety of cancers. However, the role of RBPs in colorectal adenocarcinoma (COAD) has not been studied. Integrated analysis of RBPs will provide a better understanding of disease genesis and new insights into COAD treatment.

## Methods

The gene expression data and corresponding clinical information for COAD were downloaded from The Cancer Genome Atlas (TCGA) database. Univariate Cox regression analysis was used to screen for RBPs associated with COAD recurrence, and multivariate Cox proportional hazards regression analyses were used to identify genes that were associated with COAD recurrence. A nomogram was constructed to predict the recurrence of COAD, and a receiver operating characteristic (ROC) curve analysis was performed to determine the accuracy of the prediction models. The Human Protein Atlas database was used in prediction models to confirm the expression of key genes in COAD patients.

## Result

A total of 177 differentially expressed RBPs was obtained, comprising 123 upregulated and 54 downregulated. GO and KEGG enrichment analysis showed that the differentially expressed RBPs were mainly related to mRNA metabolism, RNA processing and translation regulation. Seven RBP genes (*TDRD6*, *POP1*, *TDRD7*, *PPARGC1A*, *LIN28B*, *LRRFIP2* and *PNLDC1*) were identified as prognosis-associated genes and were used to construct the prognostic model.

## Conclusion

We constructed a COAD prognostic model through bioinformatics analysis, which indicated that prognostic model RBPs have a potential role in the diagnosis and prognosis of COAD. Moreover, the nomogram can effectively predict the 1-year, 3-year, and 5-year survival rate for COAD patients.

## Background

Colorectal cancer (CRC) is the third most frequently diagnosed malignancy and one of the leading causes of cancer-related mortality worldwide [1]. Colon adenocarcinoma (COAD) is a common type of CRC [2]. Even with the significant progress that has been attained in recent COAD research, the data show that the morbidity and mortality rates of COAD are increasing [3]. Findings have demonstrated that COAD can be successfully treated when identified at an early stage [4]. In the pathogenesis of COAD, accumulation of genetic and epigenetic alterations transforms normal colonic epithelial cells to adenocarcinoma cells [5].

Therefore, it is required to systematically study the genetic and epigenetic alterations in COAD to identify potential diagnostic markers and therapeutic targets.

RNA-binding protein (RBPs) regulates gene expression at the post-transcriptional level mainly through interaction with target RNA [6–8]. A large number of studies have shown that RBPs participate in RNA metabolism and play an important role in the regulation of RNA stability, modification, localization, translation and alternative splicing [9, 10]. Moreover, a recent study found that RBPs can interact directly with chromatin to regulate gene expression at the epigenetic level [11].

Considering the extensive and important role of RBP in post-transcriptional regulation, changes in RBP expression will inevitably lead to the occurrence of various diseases. A total of 1542 RBPs have been identified by high throughput screening in human cells, however, the role of these RBPs in the occurrence and development of COAD has not been examined [12].

In this study, we obtained COAD expression profile data through TCGA database. Through preprocessing and differential analysis of the data, we obtained RBPs that were differentially expressed in COAD patients and normal samples. Then, the Cox analysis was performed to determine the RBPs that are significantly related to COAD survival, and a COAD risk model was then constructed. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the risk model genes were then carried out to reveal the potential functional mechanism of RBPs in COAD. Finally, the expression level of the risk model genes was identified, and it was determined that some genes were appropriate for use as potential prognostic biomarkers.

## Methods

### Identification of differentially expressed RBPs

The transcriptome data and corresponding clinical information for 473 COAD samples and 41 normal samples were downloaded from TCGA. The transcriptome data were preprocessed using the Bioconductor-limma R package [13]. False discovery rate  $< 0.05$  and  $|\log_2$  fold change (FC)|  $\geq 1$  were used as the criteria for differential RBPs (DE-RPBs) screening. The average count value of all DE-RPBs was greater than 1.

### GO And KEGG Functional Enrichment Analyses

The biological functions of these DE-RPBs including molecular functions, biological processes and cellular components, were systematically studied by GO enrichment [14]. The potential biological pathways of DE-RPBs were detected using KEGG database [15]. All GO and KEGG pathway enrichment analyses were carried out using the Bioconductor – clusterProfiler and enrichplot packages [16] with a *P*-value less than 0.05.

# PPI Network Construction And Module Screening

START (<https://string-db.org/>) [17] was used to detect the protein-protein interaction (PPIs) among all DE-RPBs, and the network was constructed using Cytoscape 3.8.0 software. Then the subnetwork were selected from PPI network by using MCODE (Molecular Complex Detection) plugin and Cytoscape.  $P$ -value  $< 0.05$  was considered significant.

## Prognostic Model Construction And Survival Analysis

Univariate Cox regression analysis was used to evaluate the prognostic value of DE-RPBs. DE-RPBs related to overall survival time ( $p < 0.05$ ) was selected DE-RPBs related to overall survival time ( $p < 0.05$ ) was selected for subsequent analysis. A risk signature was formulated according to the multivariate Cox proportional hazards regression analyses. The risk-score for each COAD patient was calculated as follows:

$$\text{Risk-score} = \text{beta-gene1} \times \text{exp-gene1} + \text{beta-gene2} \times \text{exp-gene2} + \text{beta-gene} \times \text{exp-gene}$$

The samples were randomly divided into the train group and the test group, and the train group and the test group were divided into the high-risk group and the low-risk group according to the middle risk value. Beta-gene represents the regression coefficient that derived from the multivariate Cox regression analysis. Exp-gene represents the expression of genes.

## Independent prognostic analysis

The patient's age, gender, cancer stage, and other information was preprocessed and then integrated with the risk value. Then, single-factor and multi-factor independent prognostic analyses of the train group and test group were performed by utilizing the "survival" package in R (version 3.6.3).

## Nomogram construction

A nomogram for individualized prediction of overall survival was generated based on the results of the multivariate analysis (Cox model) to predict 1-, 3-, and 5-year overall survival. The nomogram plots were generated using the "rms" package of R software (version 3.6.3). The total points were calculated, and the 1- year, 3- year, and 5-year patient survival rates were predicted based on the total points.

## Prognostic model gene expression validation

The expression levels of prognostic model genes in COAD and normal samples were verified by the Human Protein Atlas (HPA) database. This is a web-based database that provides protein expression in

normal samples, cancer samples, cells, and blood.

## Statistical analysis

Most of the statistical analyses were performed using the bioinformatic tools mentioned above. When we conducted differential expression analysis, only RBPs with  $|\log_2 \text{FC}| \geq 1$  and  $P < 0.05$  were considered as statistically significant. The univariate and multivariate Cox regression analyses were performed utilizing the “survival” package in R (version 3.6.3). Cox  $P < 0.05$  was regarded as statistically significant for survival analysis. Boxplot generation were conducted using the R packages “ggplot2”, “ggpubr,” and “ggsignif”. A survival curve created using the R packages “survival” and “survminer” was utilized to estimate the differences in the overall survival between high-risk and low-risk COAD.

## Results

### Identification of differentially expressed RBPs

The transcriptome data and corresponding clinical information for 473 COAD samples and 41 normal samples were downloaded from TCGA. After preprocessing and difference analysis of the raw data of the 1542 RBPs [12], we found that there were 123 upregulated and 54 downregulated RBPs (Fig. 1A, B).

### GO And KEGG Enrichment Analysis Of Differentially Expressed RBPs

In order to explore the function and molecular mechanism of these differentially expressed RBPs, we performed GO and KEGG enrichment analysis on upregulated and downregulated RBPs, respectively. Upregulated differentially expressed RBPs (UP-DE-RBPs) were significantly enriched in biological processes associated with ncRNA processing, RNA phosphodiester bond hydrolysis, and ribosome biogenesis. In the cellular component enrichment analysis, UP-DE-RBPs were significantly enriched in cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, and nucleolar part. The molecular function analysis showed that UP-DE-RBPs were significantly enriched in catalytic activity, acting on RNA, ribonuclease activity, and translation regulator activity (Fig. 2A, B). Downregulated differentially expressed RBPs (DOWN-DE-RBPs) were significantly enriched in biological processes associated with defense response to virus, regulation of cytoplasmic translation, and regulation of translation. In cellular component enrichment analysis, DOWN-DE-RBPs were significantly enriched in endolysosome membrane, apical dendrite, and mitochondrial matrix. The molecular function analysis showed that DOWN-DE-RBPs were significantly enriched in mRNA 3'-UTR AU-rich region binding, AU-rich element binding and mRNA 3'-UTR binding and (Fig. 2C, D). In regard to the KEGG enrichment analysis, the UP-DE-RBPs were mainly enriched in ribosome biogenesis in eukaryotes, mRNA surveillance pathway, and RNA

transport (Fig. 2E, F). The DOWN-DE-RBPs were mainly enriched in progesterone-mediated oocyte maturation, oocyte meiosis, and hepatitis C (Fig. 2G, H).

## PPI Network Construction

In order to more clearly understand the mechanism of these DE-RBPs, we used the STRING database and Cytoscape software to construct a PPI network (Fig. 3A). This PPI network contained a total of 145 nodes and 596 edges. Next, we used the plug-in MODE in Cytoscape to analyze the PPI network to identify the potential key subnetwork. Finally, we determined the first three important subnetworks. Subnetwork 1 included 21 nodes and 195 edges, subnetwork 2 included 13 nodes and 38 edges, and subnetwork 3 included 5 nodes and 10 edges. An overall analysis was conducted by integrating these three key subnetworks into a network (Fig. 3B). The GO and KEGG analyses showed that the subnetwork 1 genes were mainly enriched in ncRNA processing, preribosome, snoRNA binding, and ribosome biogenesis in eukaryotes. The subnetwork 2 genes were significantly enriched in DNA alkylation, cytoplasmic ribonucleoprotein granule, and catalytic activity acting on RNA. The subnetwork 3 genes were significantly enriched in RNA splicing via transesterification reactions with bulged adenosine as nucleophile, perikaryon, and the mRNA 3'-UTR AU-rich region binding.

## Prognosis-related RBP Screening

We conducted a univariate Cox regression analysis of 177 differentially expressed RBPs and found that TDRD6, POP1, TDRD7, LUZP4, PPARGC1A, LIN28B, PABPC1L, LRRFIP2, ZC3H12C, RBM47, CELF4, and PNLDC1 were significantly related to survival (Fig. 4A).

## Prognosis-related Genetic Risk Score Model Construction And Validation

To further determine the RBPs with the greatest potential prognosis ability, multivariate Cox regression analyses were used to identify seven RBPs: TDRD6, POP1, TDRD7, PPARGC1A, LIN28B, LRRFIP2, and PNLDC1. The seven RBPs were analyzed and used to construct a predictive model (Fig. 4B). The formula for calculating the risk-score of COAD patient is as follows:

$$\text{Risk-score} = (-1.9076 \times \text{ExpTDRD6}) + (-0.5828 \times \text{ExpPOP1}) + (-0.6433 \times \text{ExpTDRD7}) + (-0.5888 \times \text{ExpPPARGC1A}) + (1.2823 \times \text{ExpLIN28B}) + (-0.8568 \times \text{ExpLRRFIP2}) + (0.3925 \times \text{ExpPNLDC1})$$

In order to evaluate the predictive ability of the model, we randomly divided 446 samples of COAD patients with detailed survival information into two groups, namely the train group and the test group. At the same time, we divided the train group and the test group into high-risk and low-risk groups according to the median risk score for survival analysis. In the train and test groups, patients in the high-risk group

exhibited a significantly lower overall survival rate than low-risk group (Fig. 4C, D). Then, we performed a time-dependent receiver operating characteristic (ROC) analysis to further evaluate the prognostic accuracy of this model. The AUC of the ROC curve for overall survival in the train and test groups was 0.715 and 0.681, respectively (Fig. 4E, F). The risk-score, survival status, and expression heatmap of patients in the train and test groups are shown in Fig. 5.

## Independent Prognostic Analysis

In order to further explore the independent prognostic factors of COAD, we integrated the patient's age, gender, cancer stage, risk value, and other information, and conducted single-factor and multi-factor independent prognostic analyses on the train group and the test group, respectively. The results are shown in Fig. 6A, B. In the single-factor independent prognostic analysis of the train group and the test group, the P-value of age and cancer stage was less than 0.05, indicating that these two factors are significantly related to the survival of COAD patients. In the multi-factor independent prognostic analysis, the P-value of age and cancer stage was less than 0.05, indicating that these two factors can be used as independent prognostic factors for COAD patients.

## Nomogram Construction

In order to more accurately predict the survival rate of COAD patients, we constructed a nomogram using the expression levels and points of model genes (Fig. 7). By corresponding to the expression level and score of the Cox model genes, the total points of the patient can be calculated. Based on the total points, the patient's predicted 1-year, 3-year, and 5-year survival rates are intuitively displayed.

## Prognostic Model Gene Expression Validation

In order to further confirm the expression of key genes in COAD patients in the risk model, we used the Human Protein Atlas database to search for the expression of these seven genes. It was found that TDRD6, TDRD7, LIN28B, LRRFIP2, and PNLDC1 were significantly increased in COAD patients (Fig. 8). At present, the database does not contain data related to POP1 or PPARGC1A protein expression.

## Discussion

COAD is a noncutaneous carcinoma that is common worldwide with high morbidity and mortality. Effective prevention methods are clearly underutilized. Both diagnostic and prognostic models that can precisely and accurately predict the status and survival time of COAD are urgently needed. In the field of COAD, there have been limited studies on molecular biomarkers that can predict disease status and prognosis [18]. Through Cox analysis of TCGA COAD expression profile data, we constructed a risk model

for seven genes, conducted a series of bioinformatics analyses on the basis of the risk model, and finally identified new diagnostic and prognostic markers for COAD.

The GO and KEGG enrichment analyses of these DE-RBPs showed that the UP-DE-RBPs were significantly enriched in RNA phosphodiester bond hydrolysis, RNA catabolic process, RNA stabilization, and translation regulator activity. The DOWN-DE-RBPs were significantly enriched for regulation of mRNA processing, regulation of mRNA metabolic process, RNA splicing, and mRNA 3'-UTR binding. Previous experiments have shown that RNA splicing, processing, localization, transport, and stability are directly related to the occurrence of COAD [19, 20]. Circular RNAs (circRNAs) and micro RNAs (miRNAs) are the most common products of RNA splicing, and a large number of studies have also shown that they play an important role in the occurrence of COAD [21–24]. Previous experiments have shown that circRNA can specifically bind to RBP, which in turn affects the occurrence of disease [25, 26]. Therefore, we predict that these differentially expressed RBPs may intentionally bind with circRNA, thereby affecting the occurrence of COAD. The interaction between risk model genes and circRNA and its mechanism of action in COAD will be studied in follow-up work.

Subsequently, we used univariate Cox regression analysis to identify twelve RBPs that were significantly associated with COAD survival. Then, these were gradually reduced to seven using multivariate Cox regression analysis, and were subsequently used to construct a COAD risk model. The *P*-values of the survival curves of the train group and the test group are both less than 0.01, indicating that there is a significant survival difference between the high-risk group and the low-risk group. This indicates that our risk model can be used to predict the prognosis of COAD. The ROC curve of the risk model shows that our model has better prediction accuracy (train group AUC = 0.715 and test group AUC = 0.681).

The COAD nomogram we constructed can predict 1-year, 3-year, and 5-year overall survival of COAD patients. According to the results of the model, the higher the score mean the lower overall survival, which provides important warning information for the treatment of patients. The genes in the nomogram can be used as the best target for our further study of the pathogenesis of COAD.

Finally, we conducted an immunohistochemical prediction of the expression levels of seven genes in the risk model and found that TDRD6, TDRD7, LIN28B, LRRFIP2, and PNLDC1 were significantly increased in COAD patients compared with normal colon tissue. Because the database has not yet included the immunohistochemical data of POP1 and PPARGC1A, we cannot ascertain the difference in expression between the two in patients and normal samples. The heat maps of the expression of these seven genes in the high-risk and low-risk groups of the training and test groups showed that high expression of TDRD6, TDRD7, LRRFIP2, PNLDC1, POP1, and PPARGC1A was associated with a good prognosis in patients with COAD, whereas that of LIN28B was related to poor prognosis.

LIN28B is highly expressed in primary tumors and various cancer cell lines, and numerous experiments have shown that LIN28B plays an important role in the occurrence of cancer [27–31]. Studies have shown that high expression of LIN28B can promote the occurrence of COAD, cancer cell migration, and drug resistance [32, 33]. TDRD6, TDRD7, LRRFIP2, PNLDC1, POP1, and PPARGC1A are involved in RNA

splicing and stability, but no previous studies have proved their relationship with COAD. Our research provides a risk prediction model for COAD and also provides a new direction for in-depth study of the pathogenesis of COAD.

Although we performed detailed bioinformatics analyses on the expression data of TCGA COAD patients and normal samples, provided new prognostic indicators, and generated a COAD-RBPs-risk model, there are still deficiencies associated with the current study. First, the data are all from the TCGA database, and the sample data still requires further expansion. Second, we only analyzed the expression data, and after synthesizing other types of data, different situations may occur. Finally, we have not yet carried out experimental verification of the research results, which will be the focus of our follow-up research.

## Conclusions

Overall, our comprehensive bioinformatics analysis examined the key RBP modules specifically associated with the overall survival of COAD patients. The constructed prognostic model and nomogram exhibited good predictive accuracy with regard to survival of COAD and these RBP genes could be used in clinical adjuvant treatments.

## Abbreviations

CRC: colorectal cancer; COAD: colon adenocarcinoma; RBP: RNA-binding protein; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; TCGA: The Cancer Genome Atlas; DE-RPB: differential RPB; UP-DE-RBP: upregulated differentially expressed RBP; DOWN-DE-RBP: downregulated differentially expressed RBP; PPI: protein-protein interaction; ROC: receiver operating characteristic; AUC: area under curve; HPA: Human Protein Atlas database;

## Declarations

### Ethics approval and consent to participate

The research didn't involve animal experiments and human specimens, no ethics related issues.

### Consent for publication

Not Applicable.

### Availability of data and materials

The data of this study are from TCGA database.

### Competing interests

The authors declare that they have no competing interests.

## Funding

This research was funded by the Program for Innovation Team Building at Institutions of Higher Education in Chongqing (CXTDX201601040). The funding body had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

## Authors' contributions

XHL and XZ contributed to the conception of the study.

FT performed the data analyses.

XYL and DYY contributed significantly to process data.

RKY wrote the manuscript.

All of the authors read and approved the final manuscript.

## Acknowledgements

N/A

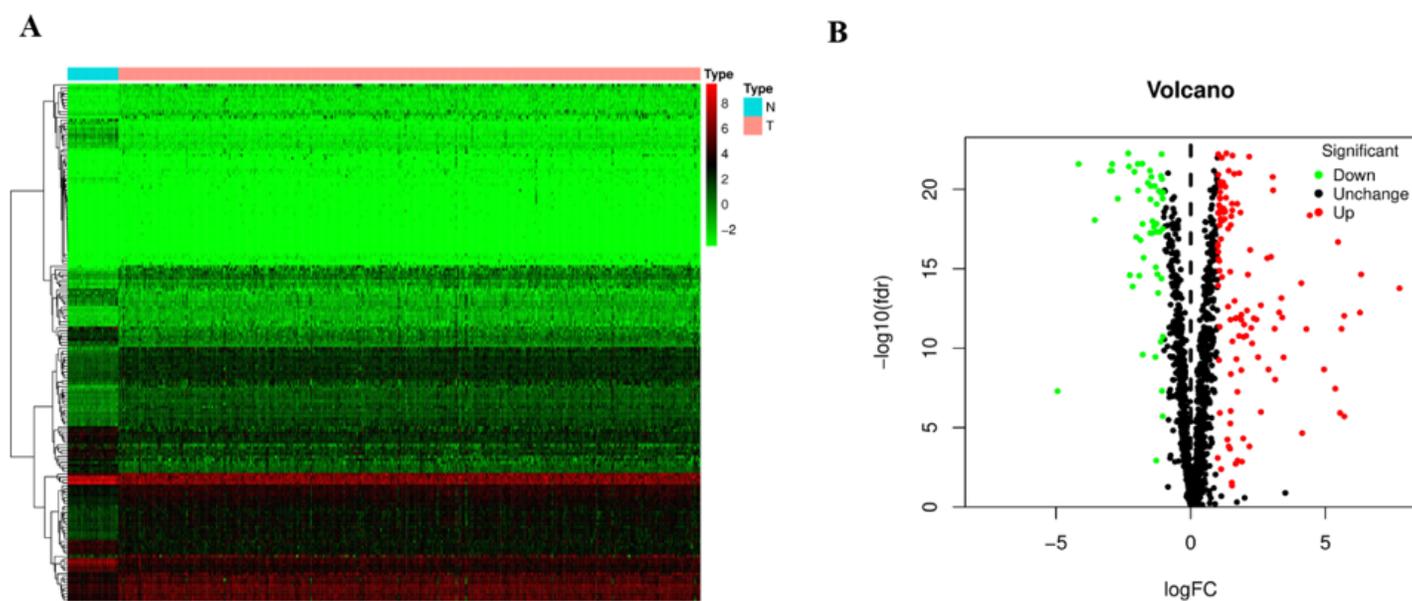
## References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics. 2019. *CA Cancer J Clin.* 2019;69:7–34.
2. Barret M, Chaussade S, Coriat R. Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med.* 2014;370(14):1298–306.
3. Cai L, Bennedsen ALB, Qvortrup C, Gögenur I. Increasing incidence of colorectal cancer in young patients. *Ugeskr Laeger.* 2019;30(1):V09190524. 182(.
4. Roncucci L, Mariani F. Prevention of colorectal cancer: How many tools do we have in our basket? *Eur J Intern Med.* 2015;26(10):752–6.
5. Khare S, Verma M. Epigenetics of colon cancer. *Methods Mol Biol.* 2012;863:177–85.
6. Calabretta S, Richard S. Emerging Roles of Disordered Sequences in RNA-Binding Proteins. *Trends Biochem Sci.* 2015;40(11):662–72.
7. Velasco MX, Kostı A, Penalva LOF, Hernández G. The Diverse Roles of RNA-Binding Proteins in Glioma Development. *Adv Exp Med Biol.* 2019;1157:29–39.
8. Cooke A, Schwarzl T, Huppertz I, Kramer G, Mantas P, Alleaume AM, Huber W, Krijgsveld J, Hentze MW. The RNA-Binding Protein YBX3 Controls Amino Acid Levels by Regulating SLC mRNA Abundance. *Cell Rep.* 2019;11(11):3097–106. 27(.
9. Frisone P, Pradella D, Di Matteo A, Belloni E, Ghigna C, Paronetto MP. SAM68: Signal Transduction and RNA Metabolism in Human Cancer. *Biomed Res Int.* 2015;2015:528954.

10. Perron G, Jandaghi P, Solanki S, Safisamghabadi M, Storoz C, Karimzadeh M, Papadakis AI, Arseneault M, Scelo G, Banks RE, Tost J, Lathrop M, Tanguay S, Brazma A, Huang S, Brimo F, Najafabadi HS, Riazalhosseini Y. A General Framework for Interrogation of mRNA Stability Programs Identifies RNA-Binding Proteins that Govern Cancer Transcriptomes. *Cell Rep.* 2018;23(6):1639–50.
11. Xiao R, Chen JY, Liang Z, Luo D, Chen G, Lu ZJ, Chen Y, Zhou B, Li H, Du X, Yang Y, San M, Wei X, Liu W, Lécuyer E, Graveley BR, Yeo GW, Burge CB, Zhang MQ, Zhou Y, Fu XD. Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription. *Cell.* 2019;178(1):107–21.
12. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet.* 2014;15(12):829–45.
13. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
14. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25(1):25–9.
15. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 2004;32(Database issue):D277–80.
16. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–7.
17. Von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 2005;33(Database issue):D433–7.
18. Liu Z, Bai Y, Xie F, Miao F, Du F. Comprehensive Analysis for Identifying Diagnostic and Prognostic Biomarkers in Colon Adenocarcinoma. *DNA Cell Biol.* 2020;39(4):599–614.
19. Hollander D, Donyo M, Atias N, Mekahel K, Melamed Z, Yannai S, Lev-Maor G, Shilo A, Schwartz S, Barshack I, Sharan R, Ast G. A network-based analysis of colon cancer splicing changes reveals a tumorigenesis-favoring regulatory pathway emanating from ELK1. *Genome Res.* 2016;26(4):541–453.
20. Amirkhah R, Naderi-Meshkin H, Shah JS, Dunne PD, Schmitz U. The Intricate Interplay between Epigenetic Events, Alternative Splicing and Noncoding RNA Deregulation in Colorectal Cancer. *Cells.* 2019;8(8):929.
21. Ruan J. A robust circular RNA-based prognostic signature for postoperative recurrence in stage II/III colon cancer. *AIMS Genet.* 2019;6(4):67–9.
22. He JH, Han ZP, Luo JG, Jiang JW, Zhou JB, Chen WM, Lv YB, He ML, Zheng L, Li YG, Zuo JD. Hsa\_Circ\_0007843 Acts as a miR-518c-5p Sponge to Regulate the Migration and Invasion of Colon Cancer SW480 Cells. *Front Genet.* 2020;25:11:9.

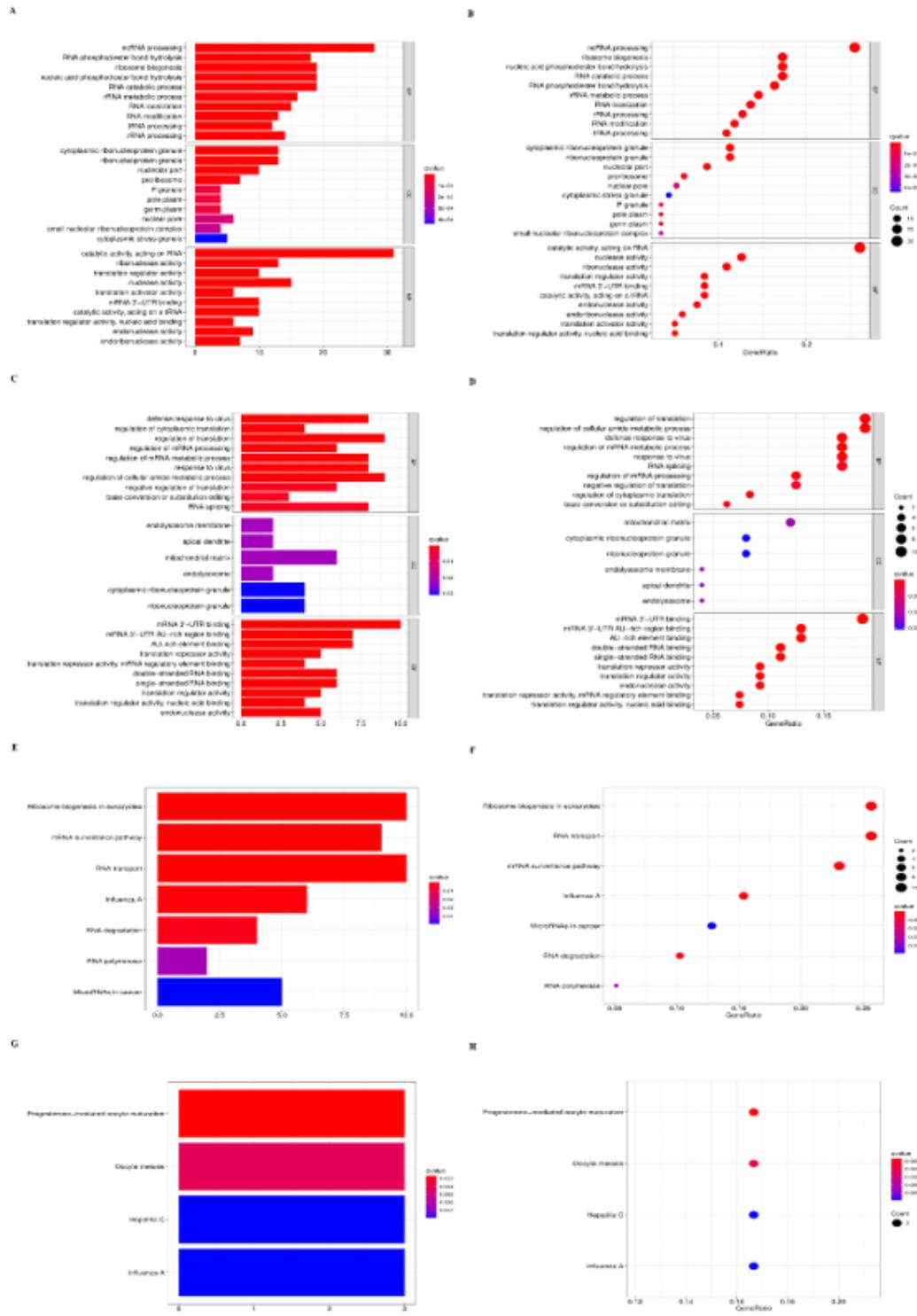
23. Huang G, Ma J, Zhang L. Integrin Subunit Alpha 5 (ITGA5) Gene Circular RNA Sponges microRNA-107 in Colorectal Carcinoma Cells and Tissues and Regulates the Expression of the Forkhead Box J3 (FOXJ3) Gene. *Med Sci Monit.* 2020;26:e920623.
24. He JH, Li YG, Han ZP, Zhou JB, Chen WM, Lv YB, He ML, Zuo JD, Zheng L. The CircRNA-ACAP2/Hsa-miR-21-5p/ Tiam1 Regulatory Feedback Circuit Affects the Proliferation, Migration, and Invasion of Colon Cancer SW480 Cells. *Cell Physiol Biochem.* 2018;49(4):1539–50.
25. Zang J, Lu D, Xu A. The interaction of circRNAs and RNA binding proteins: An important part of circRNA maintenance and function. *J Neurosci Res.* 2020;98(1):87–97.
26. Zhu YJ, Zheng B, Luo GJ, Ma XK, Lu XY, Lin XM, Yang S, Zhao Q, Wu T, Li ZX, Liu XL, Wu R, Liu JF, Ge Y, Yang L, Wang HY, Chen L. Circular RNAs negatively regulate cancer stem cells by physically binding FMRP against CCAR1 complex in hepatocellular carcinoma. *Theranostics.* 2019;26(12):3526–40. 9(.
27. Kugel S, Sebastián C, Fitamant J, Ross KN, Saha SK, Jain E, Gladden A, Arora KS, Kato Y, Rivera MN, Ramaswamy S, Sadreyev RI, Goren A, Deshpande V, Bardeesy N, Mostoslavsky R. SIRT6 Suppresses Pancreatic Cancer through Control of Lin28b. *Cell.* 2016;165(6):1401–15.
28. Ma L, Zhao Q, Chen W, Zhang Y. Oncogene Lin28B increases chemosensitivity of colon cancer cells in a let-7-independent manner. *Oncol Lett.* 2018;15(5):6975–81.
29. Ottaviani S, Stebbing J, Frampton AE, Zagorac S, Krell J, de Giorgio A, Trabulo SM, Nguyen VTM, Magnani L, Feng H, Giovannetti E, Funel N, Gress TM, Jiao LR, Lombardo Y, Lemoine NR, Heeschen C, Castellano L. TGF- $\beta$  induces miR-100 and miR-125b but blocks let-7a through LIN28B controlling PDAC progression. *Nat Commun.* 2018;9(1):1845.
30. Qu TT, Chen F, Wang J, Zhang YJ, Cheng MB, Sun WZ, Shen YF, Zhang Y. PCAF-mediated acetylation of Lin28B increases let-7 biogenesis in lung adenocarcinoma H1299 cells. *BMC Cancer.* 2018;18(1):27.
31. Ji W, Diao YL, Qiu YR, Ge J, Cao XC, Yu Y. LINC00665 promotes breast cancer progression through regulation of the miR-379-5p/LIN28B axis. *Cell Death Dis.* 2020;11(1):16.
32. Madison BB, Liu Q, Zhong X, Hahn CM, Lin N, Emmett MJ, Stanger BZ, Lee JS, Rustgi AK. LIN28B promotes growth and tumorigenesis of the intestinal epithelium via Let-7. *Genes Dev.* 2013;27(20):2233–45.
33. King CE, Cuatrecasas M, Castells A, Sepulveda AR, Lee JS, Rustgi AK. LIN28B promotes colon cancer progression and metastasis. *Cancer Res.* 2011;71(12):4260–8.

## Figures



**Figure 1**

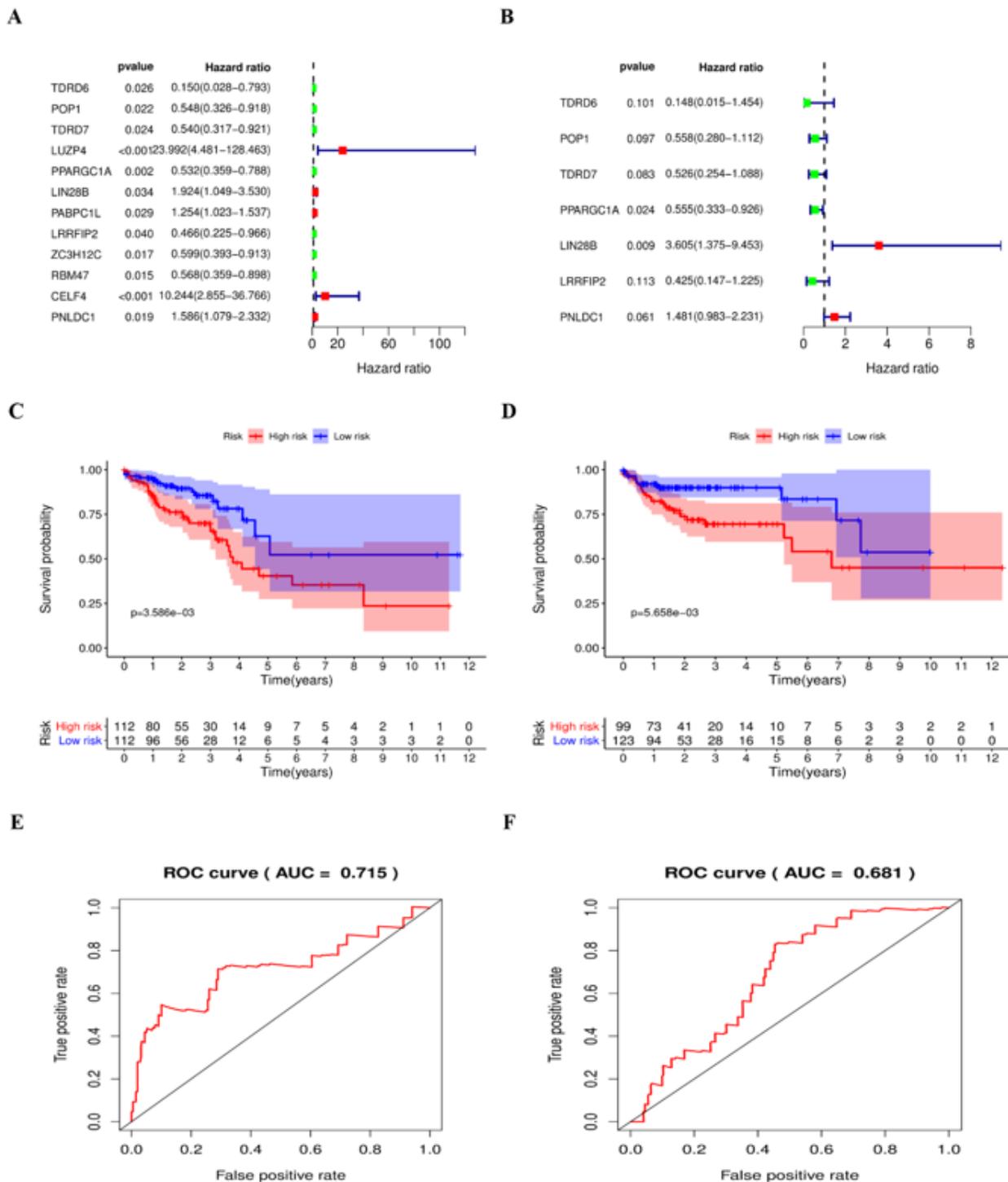
Differentially expressed RBPs. (A) RBPs expression heat map. Red represents high expression, and green represents low expression. (B) RBPs expression volcano graph, red dot represents up-regulation, green dot represents down-regulation, and black dot represents no significant change in expression level. P-value  $<0.05$ ,  $|\log_2 \text{fold change (FC)}| \geq 1$ .



**Figure 2**

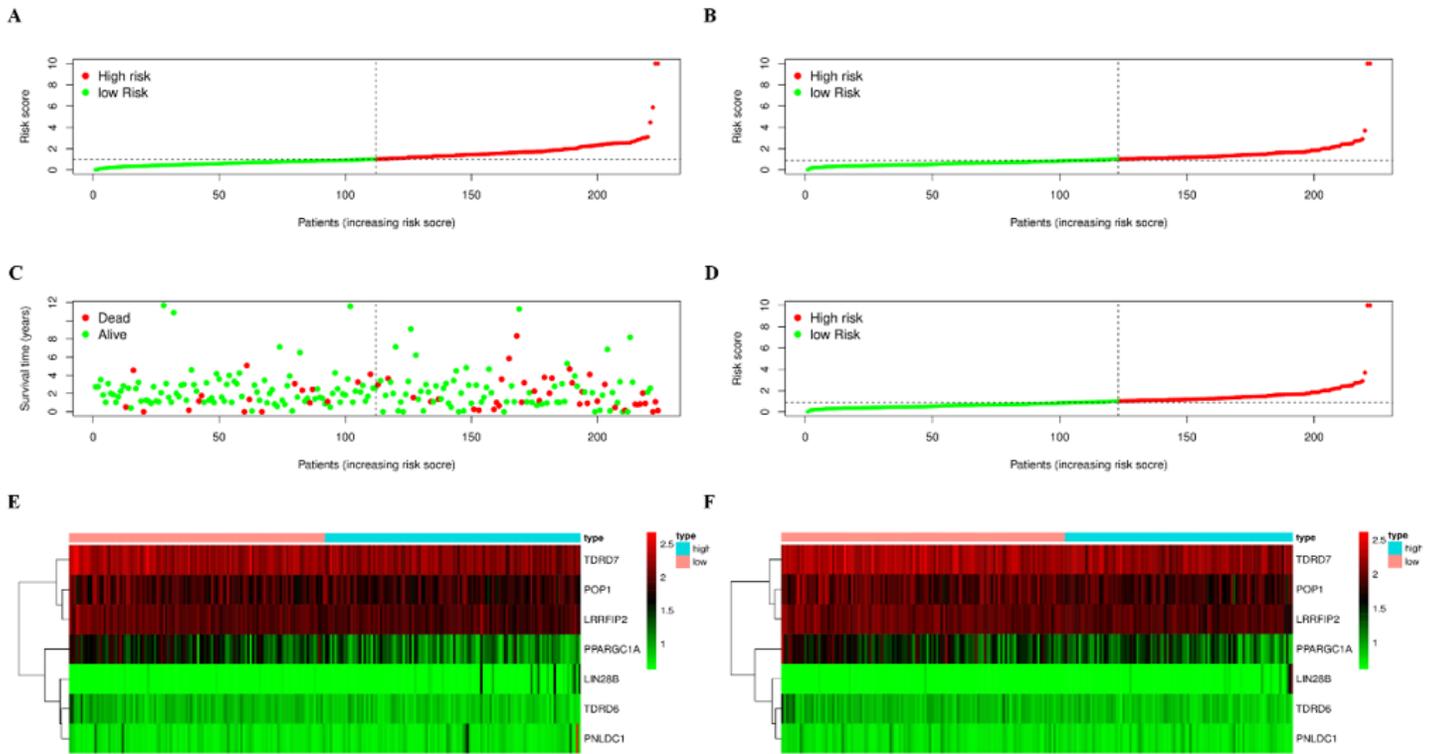
GO and KEGG pathway enrichment analyses. (A, B) BP, CC and MF enrichment analysis of upregulated differentially expressed RBPs. (C, D) BP, CC and MF enrichment analysis of downregulated differentially expressed RBPs. (E, F) KEGG enrichment analysis of upregulated differentially expressed RBPs. (G, H) KEGG enrichment analysis of downregulated differentially expressed RBPs.





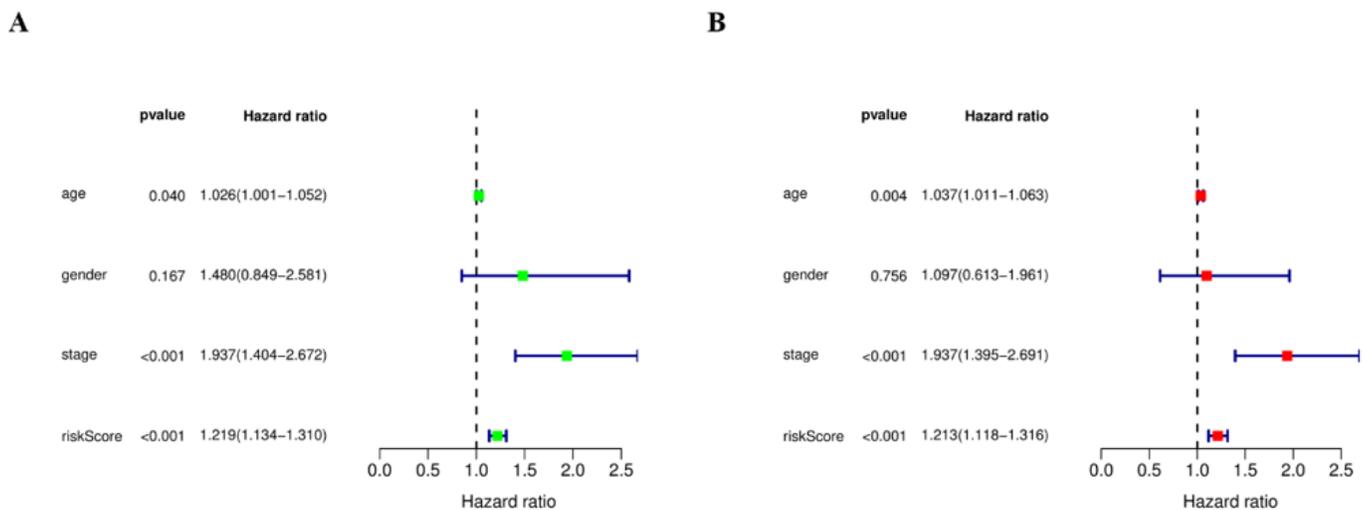
**Figure 4**

Survival analysis and risk model. (A) Univariate Cox regression analysis found RBPs that were significantly associated with survival. (B) Multivariate Cox proportional hazards regression analyses construct risk models. (C) Survival curves of patients in the high-risk group and low-risk group in the train group. (D) Survival curves of patients in the high-risk group and low-risk group in the test group. (E) ROC curves of the train group. (E) ROC curves of the test group.



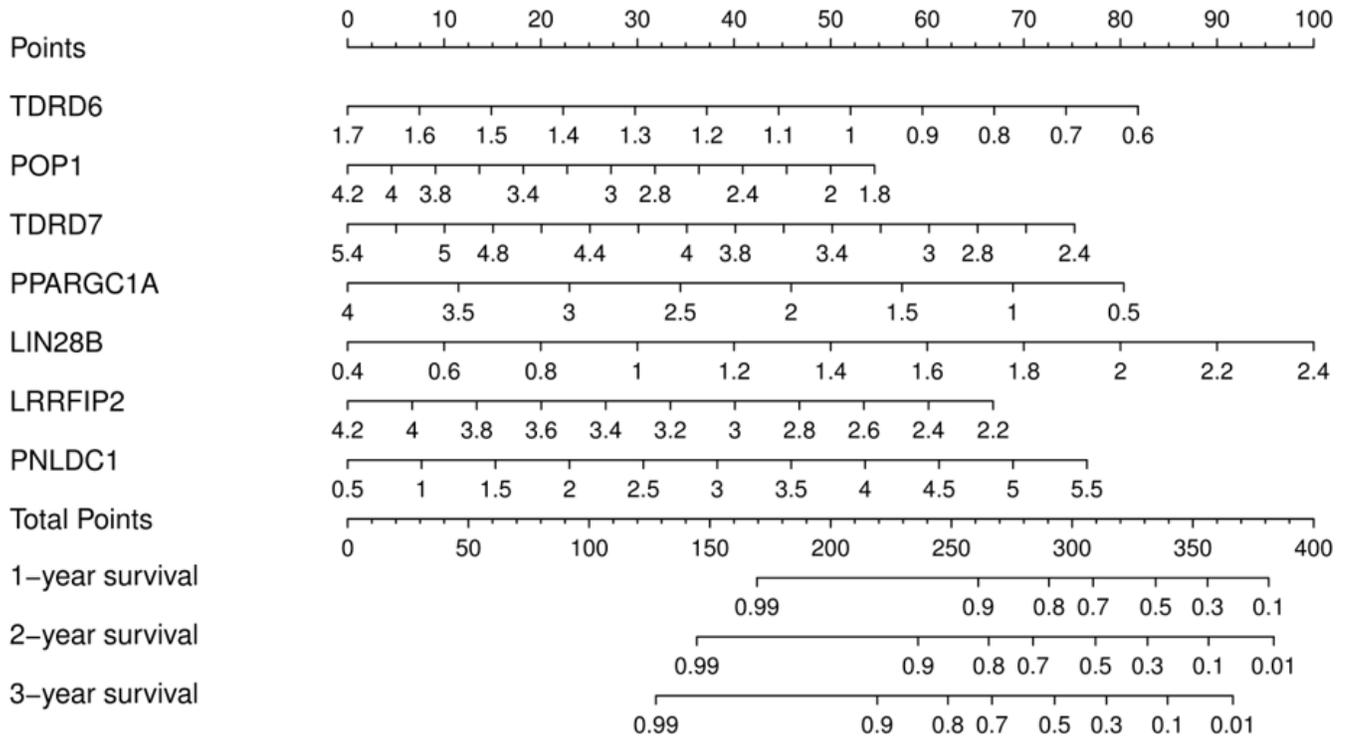
**Figure 5**

Risk score and expression analysis of seven-gene prognostic model. (A) Risk score of patients in the train group. (B) Risk score of patients in the test group. (C) Survival status of patients in the train group. (D) Survival status of patients in the test group. (E) Expression heatmap of the seven RBPs in the train group patients. (F) Expression heatmap of the seven RBPs in the test group patients.



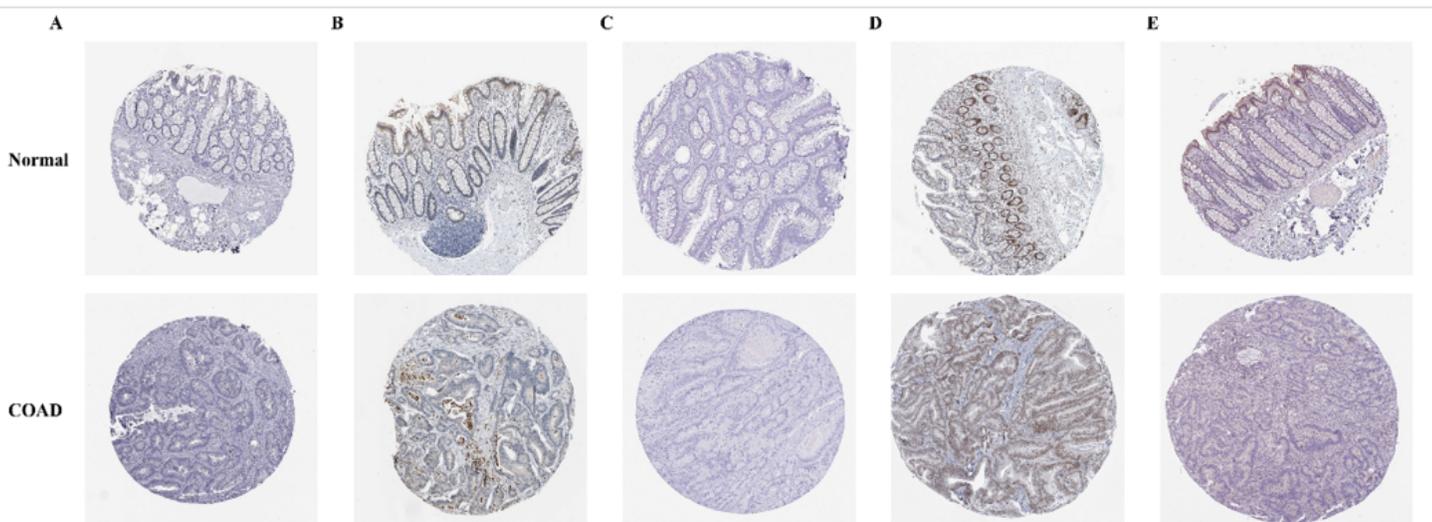
**Figure 6**

Independent prognostic analysis. (A) Single-factor independent prognostic analysis. (B) Multi-factor independent prognostic analysis.



**Figure 7**

Nomogram to predict 1- year, 3-year and 5-year overall survival in the COAD patient.



**Figure 8**

The protein expression of risk model genes in COAD and normal colon tissue. (A) TDRD6. (B) TDRD7. (C) LIN28B. (D) LRRFIP2. (E) PNLDC1.