Identification of Prognosis-related RNA Binding Proteins to Reveal the Role of RNA Binding Proteins in the Progression and Prognosis of Colon Cancer

Qi Zou (nkzouqi@163.com)
Nanjing Medical University https://orcid.org/0000-0003-1657-6816
Yue Ding
Nanjing Medical University
Yuxiang Dong
Nanjing Medical Univesity
Dejun Wu
Nanjing Medical University
Junyi Wang
Nanjing Medical University
Ping Liu
Nanjing Medical University
Wei Han
Nanjing Medical University
Zhijun Min
Nanjing Medical University
Xiaoping Yang
Nanjing Medical University

Research article

Keywords: RNA binding proteins (RBPs), colon cancer, prognosis, risk model

Posted Date: September 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-54019/v1

Abstract

Background: RNA binding proteins (RBPs) are now under discussion as novel promising bio-markers for patients with colon cancer. The purpose of our study is to identify several RBPs related to the progression and prognosis of colon cancer, and to further investigate the mechanism of their influence on tumor progression.

Methods: The transcriptome data of colon cancer as well as clinical characteristics used in this study were downloaded from the The Cancer Genome Atlas (TCGA) database. Gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis and Gene set enrichment analysis (GSEA) were performed to elucidate the gene functions and relative pathways. Cox and lasso regression analysis were used to analyze the effect of immune genes on the prognosis of breast cancer. Immune risk scoring model was constructed based on the statistical correlation between hub immune genes and survival. Meanwhile, multivariate cox regression analysis was utilized to investigate whether the immune genes risk score model was an independent factor for predicting the prognosis of breast cancer. Nomogram was constructed to comprehensively predict the survival rate of breast cancer. P< 0.05 was considered to be statistically significant.

Results: The results of the difference analysis showed that 473 RBPs exhibited differential expression between normal and colon cancer tissues (p<0. 05). Univariate cox regression analysis revealed 25 RBPs statistically correlated with colon cancer related survival risk (P<0.05). In addition, a 10-RBPs based risk scoring model was constructed through multivariate cox regression analysis. KM curve indicated that patients in high-risk were associated with poor outcomes (p<0.001). ROC curve indicated that the immune risk score model was reliable in predicting survival risk (5-year OS, AUC=0.782). Our model showed satisfying AUC and survival correlation in the validation dataset (5-year OS AUC=0.744). Furthermore, multivariate cox regression analysis confirmed that the immune risk score model was an independent factor for predicting the prognosis of colon cancer. A nomogram was established to comprehensively predict the survival of colon cancer patients with the results of multivariate cox regression analysis. Finally, we found that 10 RBPs and risk scores were significantly associated with clinical factors and prognosis, and were involved in multiple oncogenic pathways.

Conclusion: Collectively, RBPs played an essential role in the progression and prognosis of colon cancer by regulating multiple biological pathways. Furthermore, RBPs risk score was an independent predictive factor of colon cancer, indicating a poor survival.

Introduction

Colon cancer, a major malignancy of the alimentary canal, ranked third among malignant tumors in terms of morbidity worldwide (1, 2). Relevant study revealed that more than 1 million people developed colon cancer each year, and the disease-specific mortality rate in developed countries was approximately 33% (3). Mortality of colon cancer are on the rise due to changes in diet and lifestyle (4, 5). Although colon

cancer treatment options (e.g. surgery, chemoradiotherapy, and immunotherapy) have been greatly improved, the 5-year survival rate remained as high as 50% (6).

RNA binding proteins (RBPs) are a variety of proteins that interact with RNA, which are widely expressed in cells (7–9). Through high-throughput screening, 1542 RBPs were identified, accounting for 7.5% of all protein coding genes (10). These RBPs affect posttranscriptional events and regulate physiological events of cells, thus involved in many biological processes, such as RNA splicing, mRNA stabilization and protein translation (7, 11). As RBPs play a variety of key functions in post transcriptional events, the changes of RBPs are related to the occurrence and development of many human diseases. However, the role of RBPs in the development of colorectal cancer remain unclear.

The Cancer Genome Atlas (TCGA) was regarded as the largest cancer database, containing samples of more than 20,000 primary cancers and normal matched samples of multiple cancer types. Therefore, we can investigate tumor gene data in greater depth with bioinformatics methods. Further, it can be linked to clinical data in order to obtain more valuable and meaningful results. Here, we described the expression and lineage of RBPs in colon cancer, and investigate the mechanisms of RBPs to the development and prognosis of colon cancer.

Methods And Materials

Data acquisition

We identified and downloaded the transcriptome data of patients with colon cancer from the TCGA database through the R package, including 41 cases of paracancerous normal tissue and 473 cases of tumor tissue. Further, relevant clinical information of 473 colon cancer patients were obtained such as age, gender, stage, tumor&Lymph node&metastasis stage, survival status and survival duration (Table 1). Finally, "Limma" package in R software was utilized to correct the transcriptome data we have downloaded.

Table 1
Clinical characteristics of included patients in the study.

Variables	Total (n = 387)	Training cohort (n = 232)	Validation cohort (n = 155)
Age (year)			
< 60	91	53	48
≥ 60	286	179	107
Gender			
FEMALE	184	109	75
MALE	203	123	80
Stage			
I	66	35	31
II	158	100	58
III	102	60	42
IV	61	37	24
T stage			
T1	8	5	3
T2	65	32	33
T3	168	172	96
T4	46	23	23
N stage			
N1	232	139	93
N2	88	51	37
N3	67	42	25
M stage			
M0	326	195	131
M1	61	37	24
Survival			
Dead	73	46	27
Alive	314	186	128

Gene function enrichment analysis

In order to explore the major biological process of selected hub-genes, methods were utilized to conduct the gene functional enrichment analyses including Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene ontology (GO). We utilized Database for Annotation, Visualization, and Integrated Discovery (DAVID, https:// david.ncifcrf.gov/) to identify enriched KEGG and GO themes.

Survival analysis and expression comparison of hub-genes

Clinical information for TCGA colon cancer including survival time, survival state, and TNM staging were also downloaded from the TCGA database (samples with missing information were excluded). Survival R package was applied in survival analyses for hub-genes. For the overall survival rates, the logrank test was used to detect significant differences. The results were visualized using Kaplan-Meier survival curves, and P-value < 0.05 was considered as statistically significant.

Gene set enrichment analysis

Gene enrichment analysis (GSEA) (version 3.0, the broad institute of MIT and Harvard, http://software.broadinstitute.org/gsea/downloads.jsp) was conducted between colon cancer and paracancerous normal tissues to study the biological characteristics of colon cancer. In detail, the “collapse data set to gene symbols” was set to false, the number of marks was set to 1000, the “permutation type” was set to phenotype, the “enrichment statistic” was set to weighted, and the Signal2Noise metric was used for ranking genes. High expression group was used as experimental group and low expression group was used as reference group. “c2.cp.kegg.v7.0.symbols.gmt” gene sets database was used for enrichment analysis. Gene set size > 500 and < 15, FDR < 0.25, and nominal P-value < 0.05 were regarded as the cut-off criteria.

Statistical analysis

All analyses were performed using R 3.6.1. All statistical tests were two-sided, and P value < 0.05 was considered statistically significant. Continuous variables that conformed to the normal distribution were compared with the use of independent t test for comparison between groups, while continuous variables with skewed distribution were compared with the Mann-Whitney U test. The correlation matrix was constructed by R software based on Pearson Correlation Coefficient. The relationship between immune cell infiltration and overall survival was analyzed through the Kaplan-Meier curve which was evaluated by log-rank test. Time-dependent ROC curves were used to analyze the sensitivity and specificity of the recurrence prediction model. The univariate and multivariate regression model was used to analyze the effects of individual variables on survival. The nomogram was constructed with the regression coefficients based on the cox analysis.

Results

Differential expression screening of colon cancer

The transcriptome data of 473 colon cancer cases and 41 adjacent normal tissues cases was obtained from TCGA database for differential expression analysis. A total of 473 RBPs were identified as differentially expressed RBPs (DERBPs) between breast cancer and normal tissues, including 321 up-regulated and 152 down regulated ($p < 0.05$, Fig. 1A, Table 2). The heatmap of the top 10 up-regulated and top 10 down regulated DERBPs was shown in Fig. 1B.

Table 2
Univariate cox regression analysis to screen RNA binding proteins associated with overall survival in colon cancer patients.

Gene	HR	HR.95L	HR.95H	pvalue
CELF4	15.87213555	4.608502177	54.66519864	1.18E-05
LUZP4	344.8138082	15.22018153	7811.770317	0.000242467
KHDC1L	3.309097051	1.701354981	6.436119103	0.000422443
TDRD7	0.416408253	0.233832009	0.741540193	0.002924262
NOP14	0.376647057	0.188141644	0.754022358	0.005830105
TDRD6	0.063910916	0.009002748	0.453706475	0.005954612
EEF1A2	1.369621708	1.090242892	1.720592389	0.006887091
PPARGC1A	0.579714951	0.380038455	0.884303731	0.011384635
NOL3	1.670109583	1.115583263	2.500275965	0.012729851
EXOG	0.292343173	0.109198587	0.782652346	0.014376367
RIOK1	2.361681822	1.18371422	4.711898305	0.014746644
CAPRIN2	1.666977568	1.087328425	2.55563466	0.019078017
EIF2AK3	0.436164581	0.214390243	0.887351677	0.022034038
PNLDC1	1.574919384	1.064983919	2.329022084	0.022884063
PPARGC1B	0.45676183	0.231685293	0.900494662	0.023660181
ZNF385A	1.310439816	1.032491852	1.663211683	0.026225529
AEN	1.694187437	1.059461667	2.709178787	0.027727458
ERI1	0.589956581	0.368421266	0.94470325	0.028037187
SRP14	1.93932163	1.07191063	3.508658541	0.028558768
RRS1	1.620765905	1.039435069	2.527220985	0.033120985
MAK16	0.583949222	0.353353075	0.965031066	0.035830825
AFF3	3.343687337	1.053692809	10.61053554	0.040488238
POP1	0.55488789	0.314556342	0.978840764	0.041970548
ZC3H12C	0.624988088	0.392567184	0.995014676	0.047588085
TRIM25	0.61232798	0.375283794	0.999098712	0.049579544

Functional annotation of these 473 DERBPs

In order to fully understand the biological attributes of these 473 DERBPs, we conducted Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene ontology (GO) analysis. Based on the results of DAVID, the top three enriched GO terms for up-regulated genes and down-regulated genes were: ncRNA processing, ribosome biogenesis and rRNA metabolic process; RNA splicing, regulation of translation and regulation of cellular amide metabolic process, respectively (Fig. 2A). The top biological pathway enriched for up-regulated genes and down-regulated genes were: Ribosome biogenesis in eukaryotes, RNA transport and Spliceosome; Spliceosome, RNA transport and Influenza A pathway, respectively (Fig. 2B).

Establishment of immune prognosis model

For the purpose of revealing the relationship between these 473 DERBPs and overall survival, 25 prognostic DERBPs were identified by utilizing univariate Cox regression analyses (Fig. 3A, Table 2). TCGA colon cancer data were randomly divided into two sets (training set : validation set, 3 : 2). Then, multivariate cox regression analysis was applied to select the independent indicators for the overall survival based on training set and finally we got 10 DERBPs (PPARGC1A, ZNF385A, SRP14, RIOK1, ERI1, NOL3, RRS1, TDRD6, AEN and PNLDC1) for the construction of prognostic index (Fig. 3B, Table 3). After the construction of prognostic index, patients were separated into high risk and low risk (Fig. 3C, Fig. 3D). Heap map was utilized to visualize the difference of gene expression profile in low- and high- risk patients in breast cancer training set (Fig. 3E). The results from K-M analysis indicated that high risk patients had lower overall survival than low risk patients in both training group and validation group ($P < 0.001$) (Fig. 4A, Fig. 4B, Fig. 4C). The ROC curve revealed that the risk model had a good sensitivity and specificity in predicting survival risk (AUC = 0.782, AUC = 0.744 for 5 years overall survival in training and validation group, respectively) (Fig. 4D, Fig. 4E, Fig. 4F). To explore whether the constructed RBPs risk scoring model was independent form age, gender, stage, and other clinical pathological parameters, we performed an univariate and multivariate cox regression analysis for age, gender, stage, TNM and risk score. In univariate Cox model, age, pathological stage, pathological T, N, M stage and high risk score were associated with poor survival (Fig. 5A). In multivariate Cox model, only age, stage and risk score worked as independent predicted factors (Fig. 5B). To better predict the prognosis of breast cancer patients at three and five years post-surgery, we constructed a new nomogram from the variables associated with OS (age, pathological stage, TNM stage and risk score) (Fig. 5C, 5D, 5E).

Table 3
Multivariate cox regression analysis to establish RNA binding proteins risk prediction model.

Gene	Coef	HR	HR.95L	HR.95H	pvalue
PPARGC1A	-0.436526957	0.646277076	0.368761746	1.132639335	0.127288514
ZNF385A	0.347562389	1.415612627	0.979935812	2.044990177	0.064030703
SRP14	0.801814972	2.22958389	0.974868312	5.099195716	0.057477623
RIOK1	1.410546886	4.098196037	1.645449933	10.20706278	0.00244868
ERI1	-0.600993288	0.548266779	0.252246613	1.191676894	0.12920327
NOL3	0.530884133	1.700435056	0.918066471	3.149531618	0.091385734
RRS1	0.934795492	2.546692586	1.27947903	5.068971805	0.007774571
TDRD6	-2.621256381	0.072711452	0.004178531	1.265266376	0.07209323
AEN	0.987247699	2.683837569	1.357006588	5.307994936	0.004549131
PNLDC1	0.686749923	1.987246323	1.280883189	3.083144489	0.002179074

Clinical and prognostic correlation of 10 model genes and RBPs risk score

We further investigated the proportion of each model genes in different pathological stages. We demonstrated that SRP14, PPARGC1A and ER1 were most significantly associated with development of colon cancer (Fig. 6). Regard to the immune genes risk score, a strong correlation with pathological stage, clinical N stage and clinical M stage was identified (Fig. 7).

Gene set enrichment analysis of risk scores

To explore the biological relevance of risk scores involved in progression of colon cancer, we performed a gene set enrichment analysis of risk scores based on the TCGA colon cancer cohort. GSEA analysis indicated high risk scores was associated with MYC_TARGETS_V2, UV_RESPONSE_UP and WNT_BETA_CATENIN_SIGNALING pathway (Fig. 8A, Fig. 8B, Fig. 8C). In addition, low risk scores was associated with PROTEIN_SECRETION pathway (Fig. 8D).

Discussion

Malignant tumors are characterized by uncontrolled cell growth, which is mainly due to the dysregulated expression of cancer driver genes that regulate cell proliferation (12). Posttranscriptional mechanisms can greatly influence gene expression patterns in cancer cells, in which RNA-binding proteins (RBPs) play key roles. They can interact with target mRNAs in a sequence-dependent and structure-dependent manner, and determine cellular behavior by manipulating the processing of these mRNAs (7). It has been reported that RBPs show dysregulated expression in various human cancers (13). However, little is currently

known about the expression patterns and roles of RBPs in colon cancer. Therefore, it was crucial to investigate the RBPs subsets for the evaluation of risk and tumor prognosis in colon cancer.

In our study, we conducted a comprehensive and detailed assessment of RBPs in breast cancer, based on the data from a large set of samples. All gene expression data and patients clinical characteristics information were downloaded from TCGA dataset. We analyzed the 1542 RBPs between colon cancer and normal tissues, eventually, we verified 473 DEIGs. Moreover, we identified and constructed a 10 hub RBPs risk score model for colon cancer via univariate and multivariate cox regression analysis, including PPARGC1A, ZNF385A, SRP14, RIOK1, ERI1, NOL3, RRS1, TDRD6, AEN and PNLDC1. Furthermore, to investigate the prognostic value of the model, we performed the ROC curve and investigate the association between the model and clinical features. As expected, the high-risk group was correlated with worse overall survival and was inclined to have advanced stage and higher histological grade which might manifest poor outcome.

According to the result of the biological functions and pathway enrichment analysis of these differentially expressed RBPs, the upregulated RBPs were significantly enriched in ribosome biogenesis in eukaryotes, RNA transport and Spliceosome while downregulated differentially expressed RBPs were enriched in spliceosome, RNA transport and influenza A pathway. Recently, mechanistic data have emerged, suggesting a broader role for dysregulated ribosome biogenesis in the development and progression of most spontaneous cancers (14). Moreover, it is reported that RNA polymerases are consistently dysregulated in cancer, which is mostly mediated through upstream oncogenetic and tumor suppressive signaling pathways rather than through mutations (15). These results suggest that RBPs can affect the growth of tumor cells by regulating multiple biological pathways.

Several genes in the RBPs model had been investigated in human cancers. RIOK1 activates NF-KB signal transduction, which promotes cell cycle progression and tumor lung colonization in vivo. It is demonstrated that RIOK1 is overexpressed in different subtypes of human lung cancer and breast cancer, suggesting that RIOK1 is a potential therapeutic target, especially in Ras driven cancer (16). Ribosome biogenesis regulatory protein homolog (RRS1) is an important factor in ribosome biogenesis. At the molecular level, RRS1 silencing decreased the expression of M phase inducer phosphatase 3 (CDC25C), cyclin dependent kinase 1 (CDK1) and antigen Ki-67 (Ki67), and increased the protein levels of cyclin dependent kinase inhibitor 1 (CDKN1A) and tumor suppressor p53 (p53). In conclusion, RRS1 may promote the development of colon cancer. Therefore, targeting RRS1 may be a promising treatment strategy for CRC patients (17).

However, there were some limitations in our research. Firstly, the sample size in our study was small and a larger cohort and more abundant sequencing results were needed. Secondly, we only focused on the gene expression and gene mutation level, but ignored other events such as the gene methylation, and copy number amplification, which were also important in tumor progression. Finally, the effect of RBPs on the progression and prognosis of colorectal cancer need to be verified in vivo and in vitro.

In summary, our study sheds light on the utility of RBPs in the prognosis of colon cancer. The constructed RBPs risk scoring model is reliable in predicting the prognosis of colon cancer, and this risk scoring model is an independent influencing factor for the prognosis of colon cancer. With the rapid development of high-throughput technology, we have confidence to believe that our risk scoring model have great potential in clinical practice.

Declarations

Conflicts of Interest:

The authors have no conflicts of interest to declare.

Ethical Statement:

The authors are accountable for all aspects of the work (if applied, including full data access, integrity of the data and the accuracy of the data analysis) in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

Not applicable.

Authors contributions

Qi Zou and Xiaoping Yang designed this work. Yue Ding and Dejun Wu performed the experiment and wrote the manuscript. Yuxiang Dong and Junyi Wang performed the bioinformatics analysis. Ping Liu, Wei Han and Zhijun Min performed the data review. All authors have read and approved the manuscript.

Funding

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work (if applied, including full data access, integrity of the data and the accuracy of the data analysis) in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. Cancer J Clin. 2018;68(6):394–424.
2. Fitzmaurice C, Allen C, Barber RM. A systematic analysis for the global burden of disease study[J]. JAMA Oncol. 2017;3(4):524–48.
3. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012[J]. International journal of cancer. 2015;136(5):E359–86.
4. McGuire S. World cancer report 2014. Geneva, Switzerland: World Health Organization, international agency for research on cancer, WHO Press, 2015[J]. 2016.
5. World Health Organization. Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018[J]. International Agency for Research on Cancer. Geneva: World Health Organization; 2018.
6. Allemani C, Matsuda T, Di Carlo V, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries[J]. The Lancet. 2018;391(10125):1023–75.
7. Masuda K, Kuwano Y. Diverse roles of RNA–binding proteins in cancer traits and their implications in gastrointestinal cancers. Wiley Interdiscip Rev RNA. 2019;10(3):05.
8. New J, Subramaniam D, Ramalingam S, Enders J, Sayed AAA, Ponnurangam S, et al. Pleotropic role of RNA binding protein CELF2 in autophagy induction. Mol. Carcinog. 2019 08;58(8).
9. Otsuka H, Fukao A, Funakami Y, Duncan KE, Fujiwara T. Emerging evidence of translational control by AU-rich element-binding proteins. Front Genet 2019;10.
10. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. Nat. Rev. Genet. 2014 12;15(12).
11. Nahalka J. The role of the protein-RNA recognition code in neurodegeneration. Cell. Mol. Life Sci. 2019 Jun;76(11).
12. Li W, Li X, Gao LN, You CG. Integrated Analysis of the Functions and Prognostic Values of RNA Binding Proteins in Lung Squamous Cell Carcinoma. Front Genet 2020;11.
13. Dong W, Dai ZH, Liu FC, Guo XG, Ge CM, Ding J, Liu H, Yang F. The RNA-binding protein RBM3 promotes cell proliferation in hepatocellular carcinoma by regulating circular RNA SCD-circRNA 2 production. EBioMedicine 2019 Jul;45.
14. Pelletier J, Thomas G, Volarević S. Corrigendum. Ribosome biogenesis in cancer: new players and therapeutic avenues. Nat. Rev. Cancer 2018 01 25;18(2).
15. Bywater MJ, Pearson RB, McArthur GA, Hannan RD. Dysregulation of the basal RNA polymerase transcription apparatus in cancer. Nat. Rev. Cancer 2013 May;13(5).

16. Weinberg F, Reischmann N, Fauth L, et al. The Atypical Kinase RIOK1 Promotes Tumor Growth and Invasive Behavior. EBioMedicine 2017 Jun;20.
17. Wu XL, Yang ZW, He L, et al. RRS1 silencing suppresses colorectal cancer cell proliferation and tumorigenesis by inhibiting G2/M progression and angiogenesis. Oncotarget 2017 Oct 10;8(47).

Figures

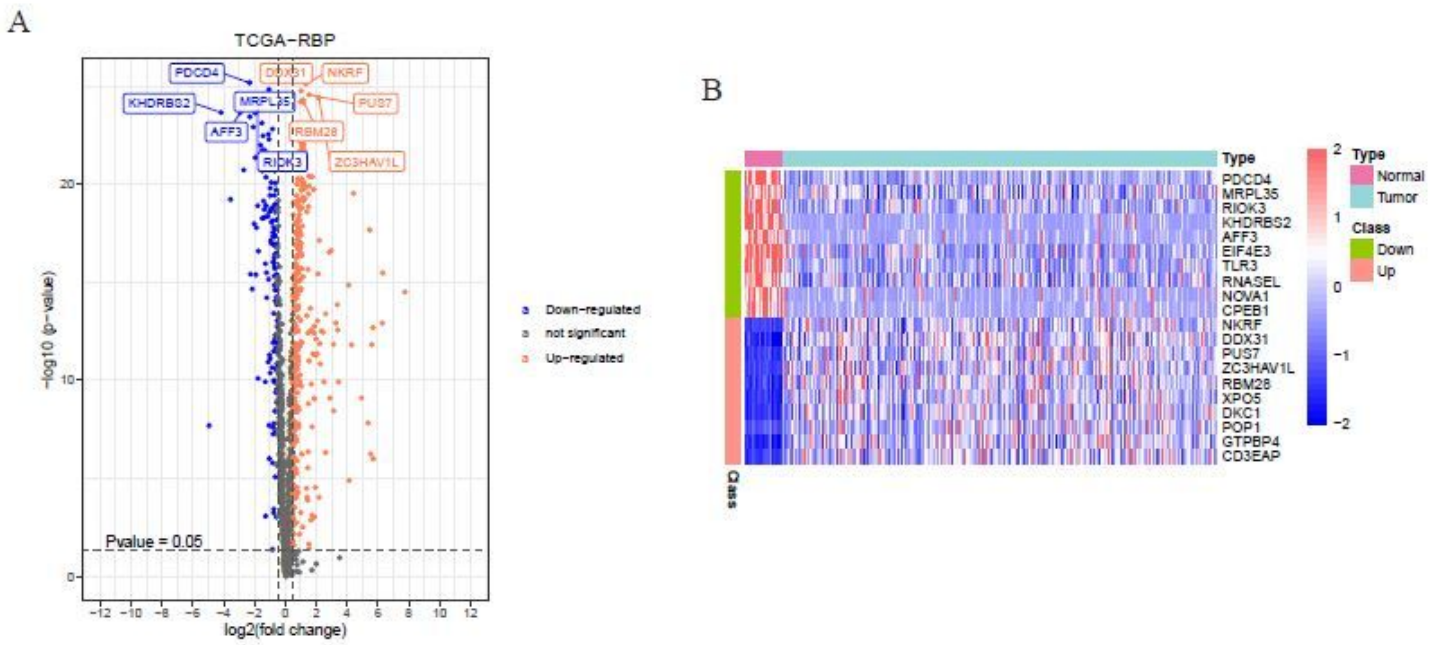

Figure 1

Identification of DERBPs. (A) volcano plots of 473 DERBPs in colon cancer and normal tissues from TCGA database. (B) Heatmap plots of top 10 up-regulated and top 10 down-regulated DERBPs. The colors in the heatmaps from green to red represent expression level from low to high. The red dots in the volcano plots represent up-regulation, the green dots represent down-regulation and black dots represent genes without differential expression

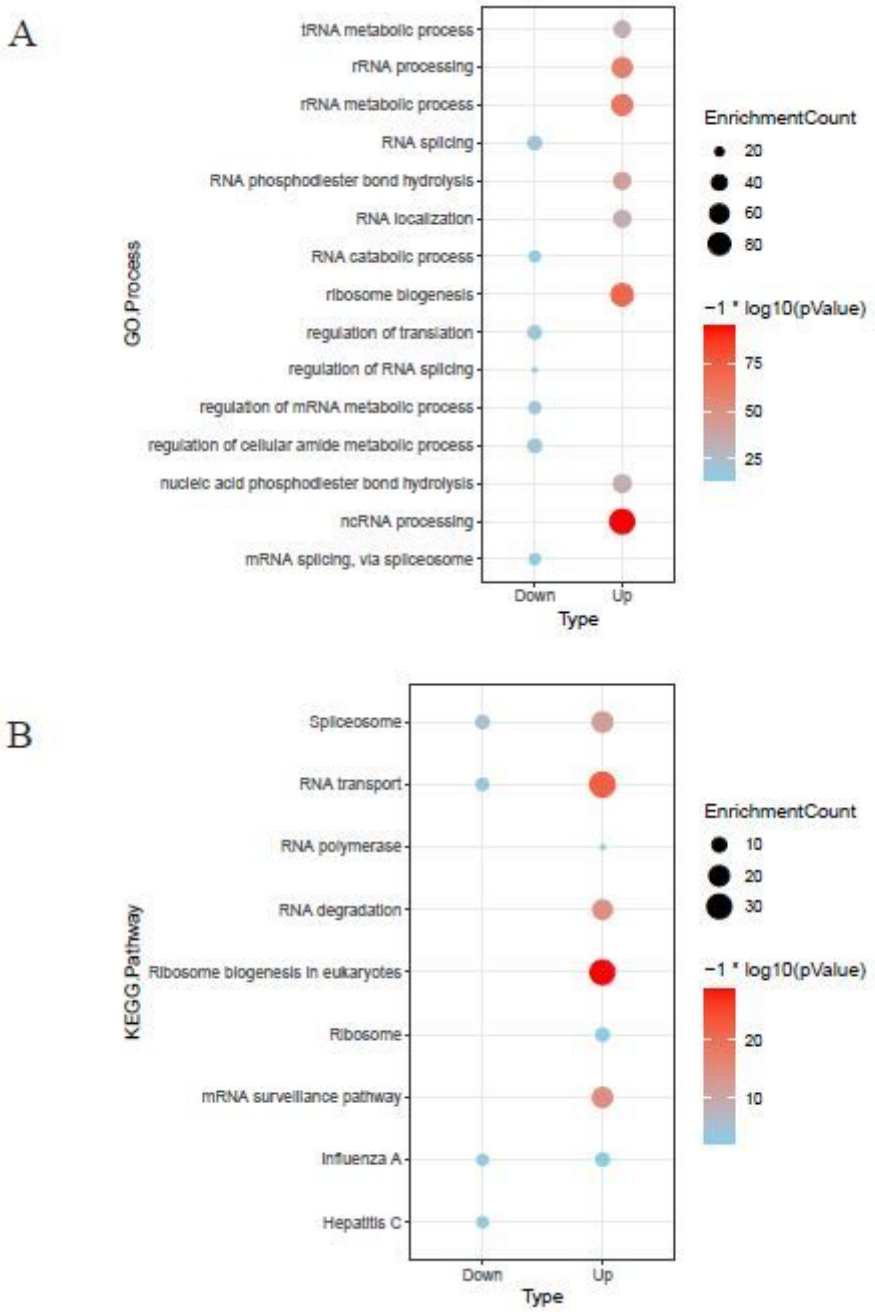

Figure 2

GO (A) and KEGG(B) enrichment analysis of DERBPs.

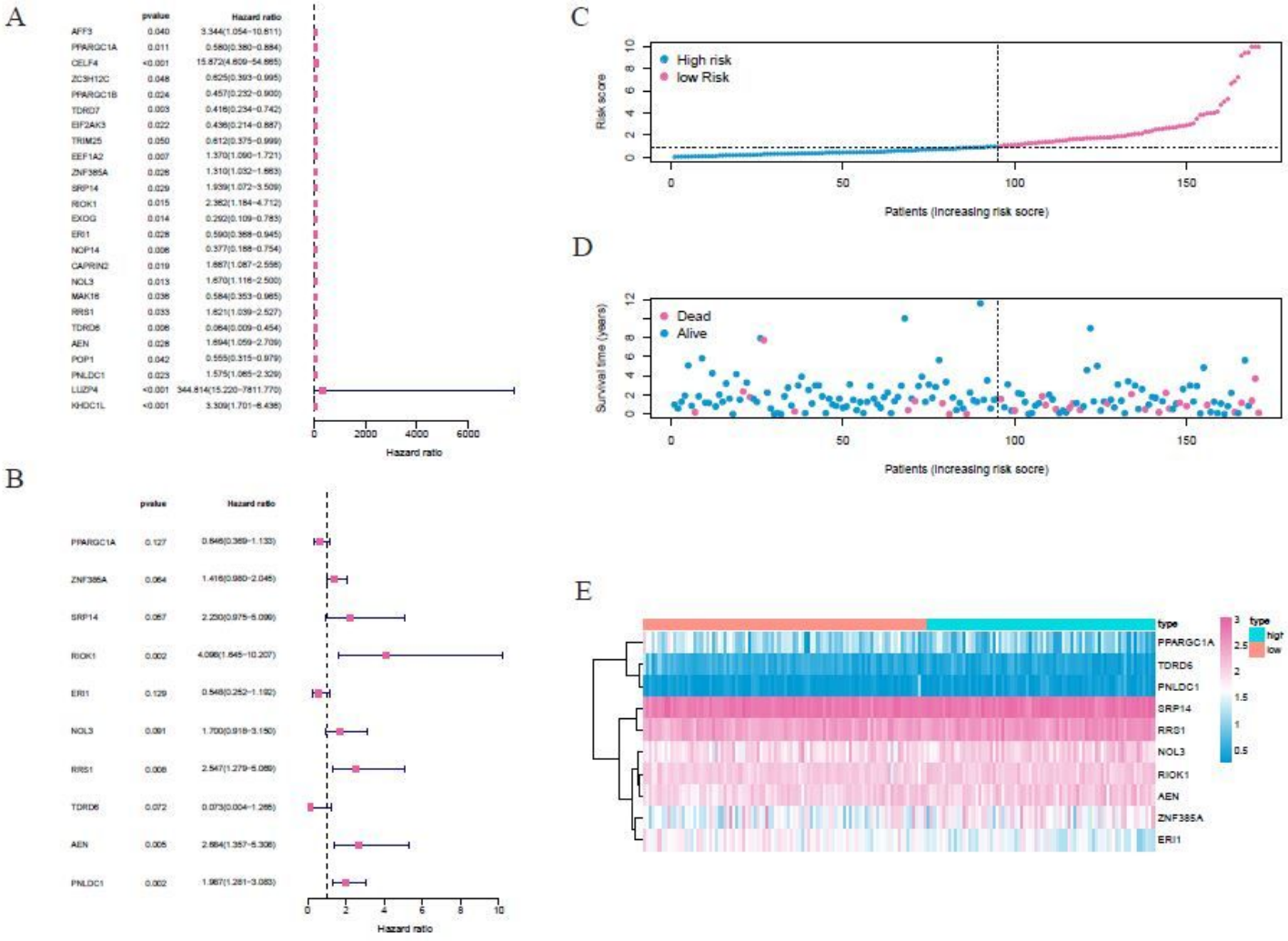

Figure 3

(A) Univariate survival analysis by cox proportional hazards models to select prognostic key RBPs. (B) Multivariate cox regression model for 10 prognostic RBPs used to construct RBPs risk score model. (C) Distribution of immune risk scores in breast cancer patients. (D) Distribution of survival status in colon cancer patients. (E) Distribution of specific risk factors in the high- and low-risk groups (divided by median value). (*P<0.05)

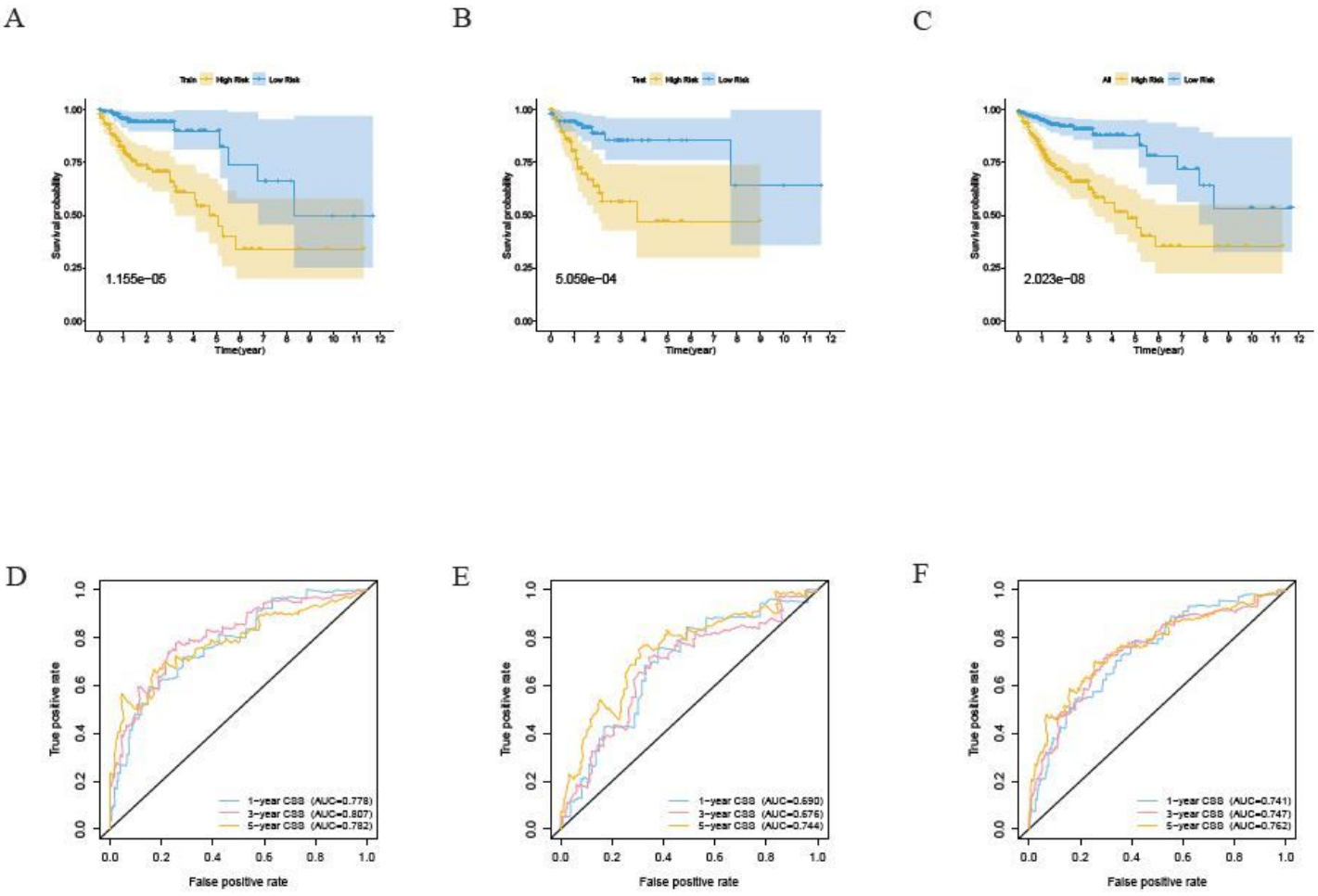

Figure 4

(A) Kaplan-Meier curve analysis of high-risk and low-risk patients in the training cohort. (B) Kaplan-Meier curve analysis of high-risk and low-risk patients in the testing cohort. (C) Kaplan-Meier curve analysis of high-risk and low-risk patients in the entire TCGA cohort. (D) Timedependent ROC curve analysis of the training cohort. (E) Time dependent ROC curve analysis of the testing cohort. (F) Time-dependent ROC curve analysis of the entire TCGA cohort.

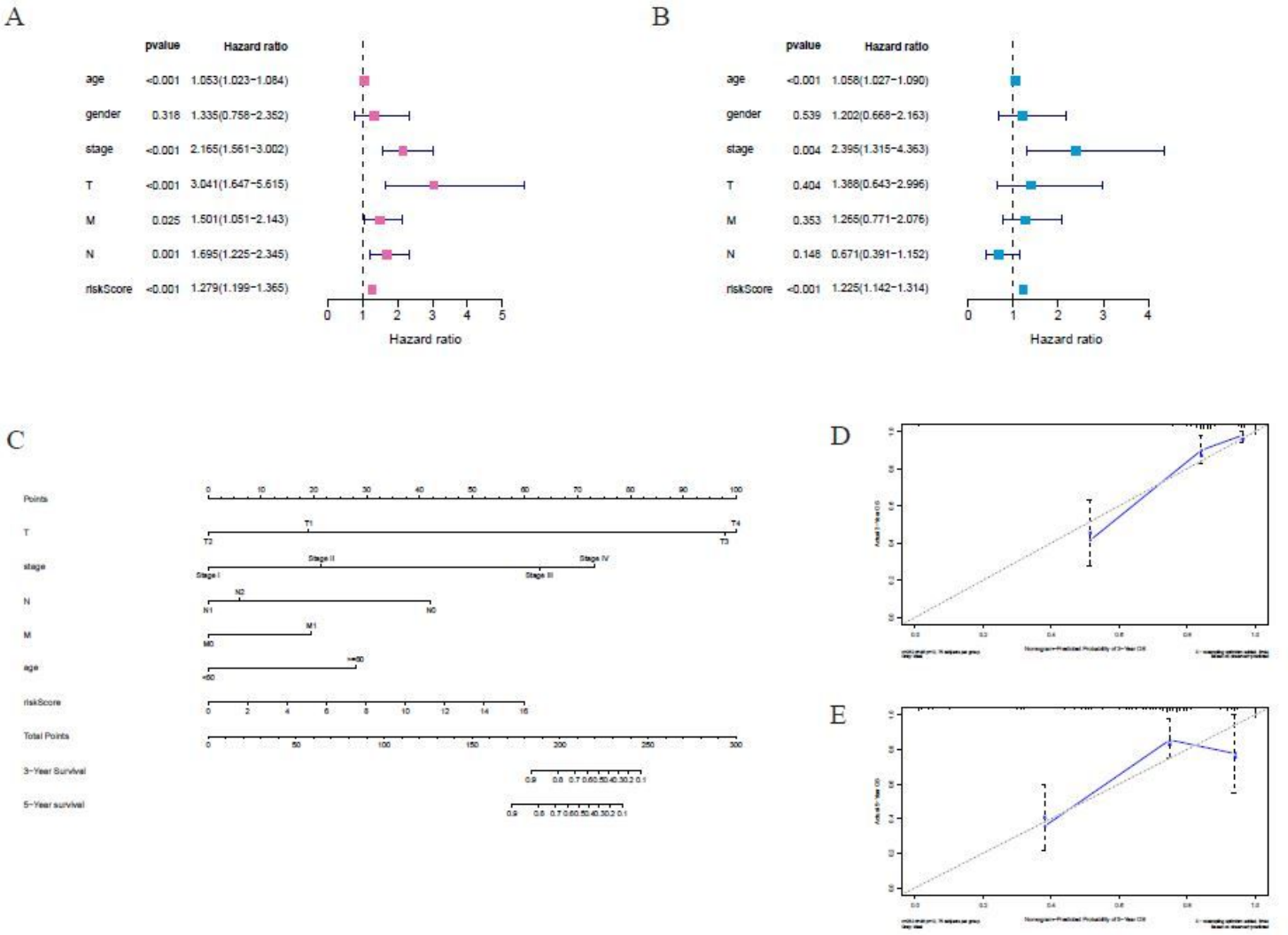

Figure 5

Cox's proportional hazard model of correlative factors in colon cancer patients. (A) Univariate COX regression analysis for seven clinicopathological parameters affecting the overall survival. (B) Multivariate COX regression analysis for seven clinicopathological parameters affecting the overall survival. (C) An established nomogram to predict breast cancer survival based on cox model. (D-E) Plots displaying the calibration of each model comparing predicted and actual 3- and 5-year overall survival.

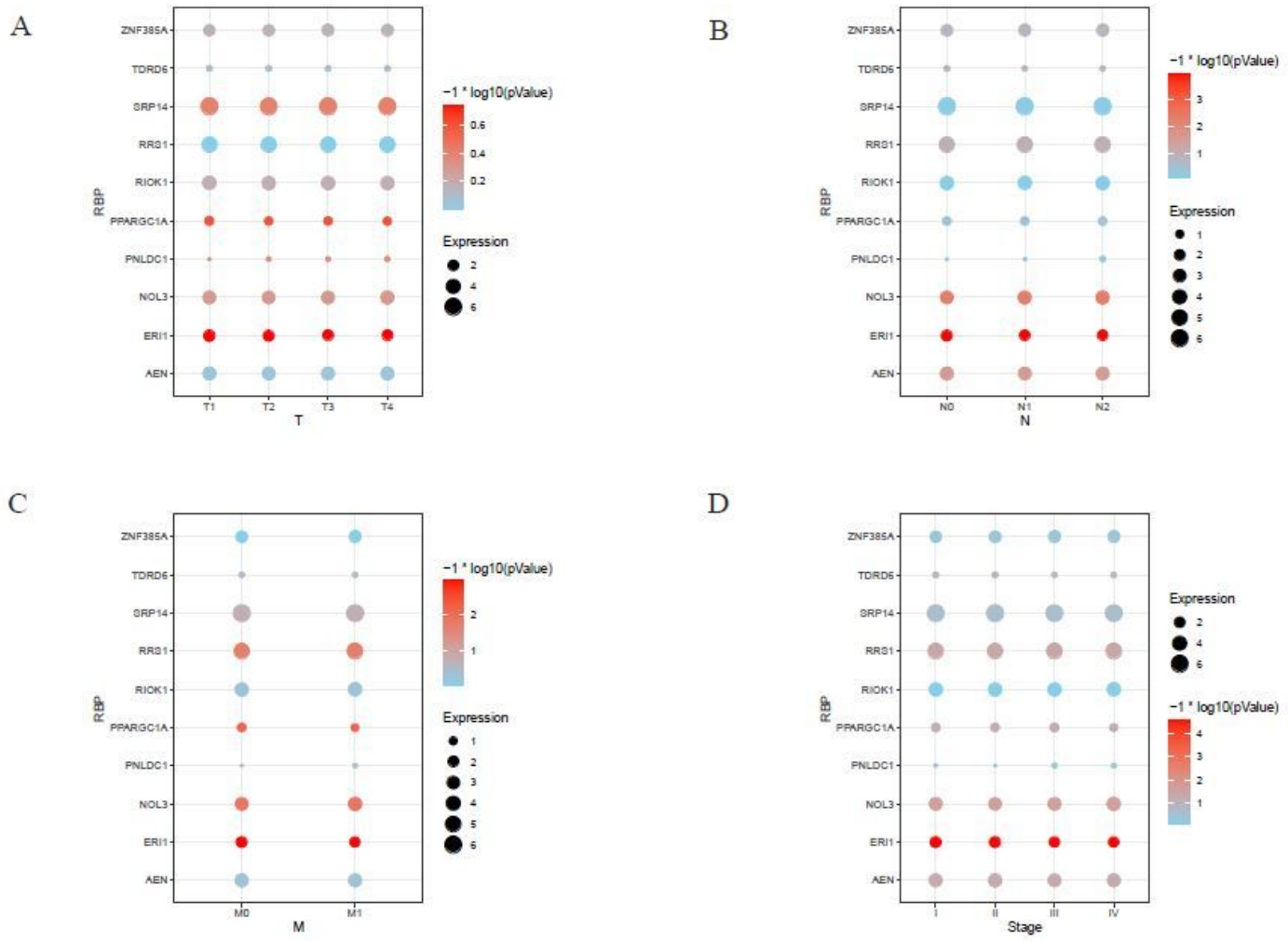

Figure 6

Correlation analysis between TNM&Stage and 10 model genes in colon cancer cases. (A) Correlation analysis between tumor stage and 10 model genes expression in colon cancer cases. (B) Correlation analysis between node stage and 10 model genes expression in colon cancer cases. (C) Correlation analysis between metastasis stage and 10 model genes in colon cancer cases. (D) Correlation analysis between pathologic stage and 10 model genes expression in colon cancer cases.

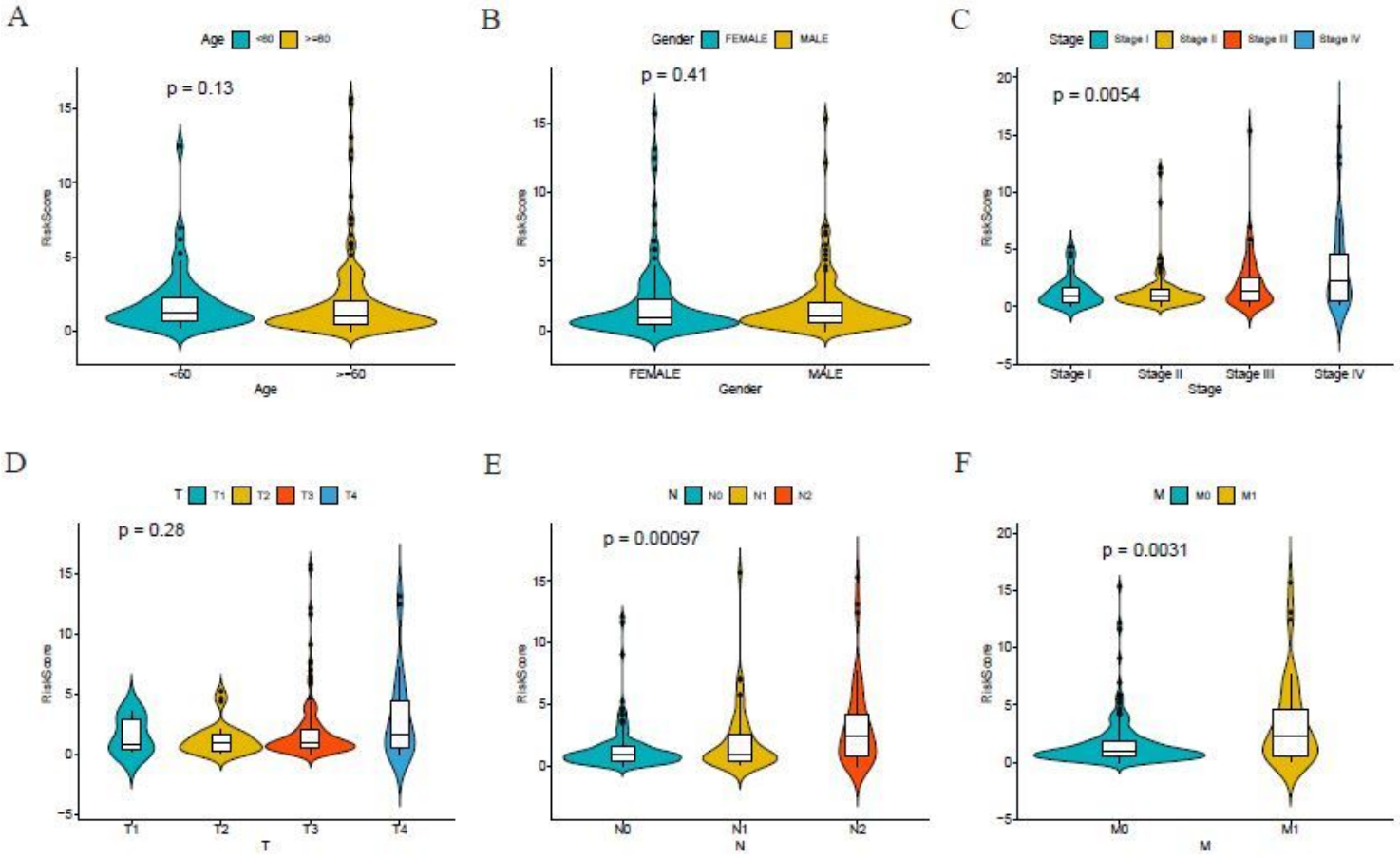

Figure 7

Correlation between RBPs risk scores and various clinical factors. (A) Age. (B) Gender. (C) Stage. (D) T stage. (E) N stage. (F) M stage.

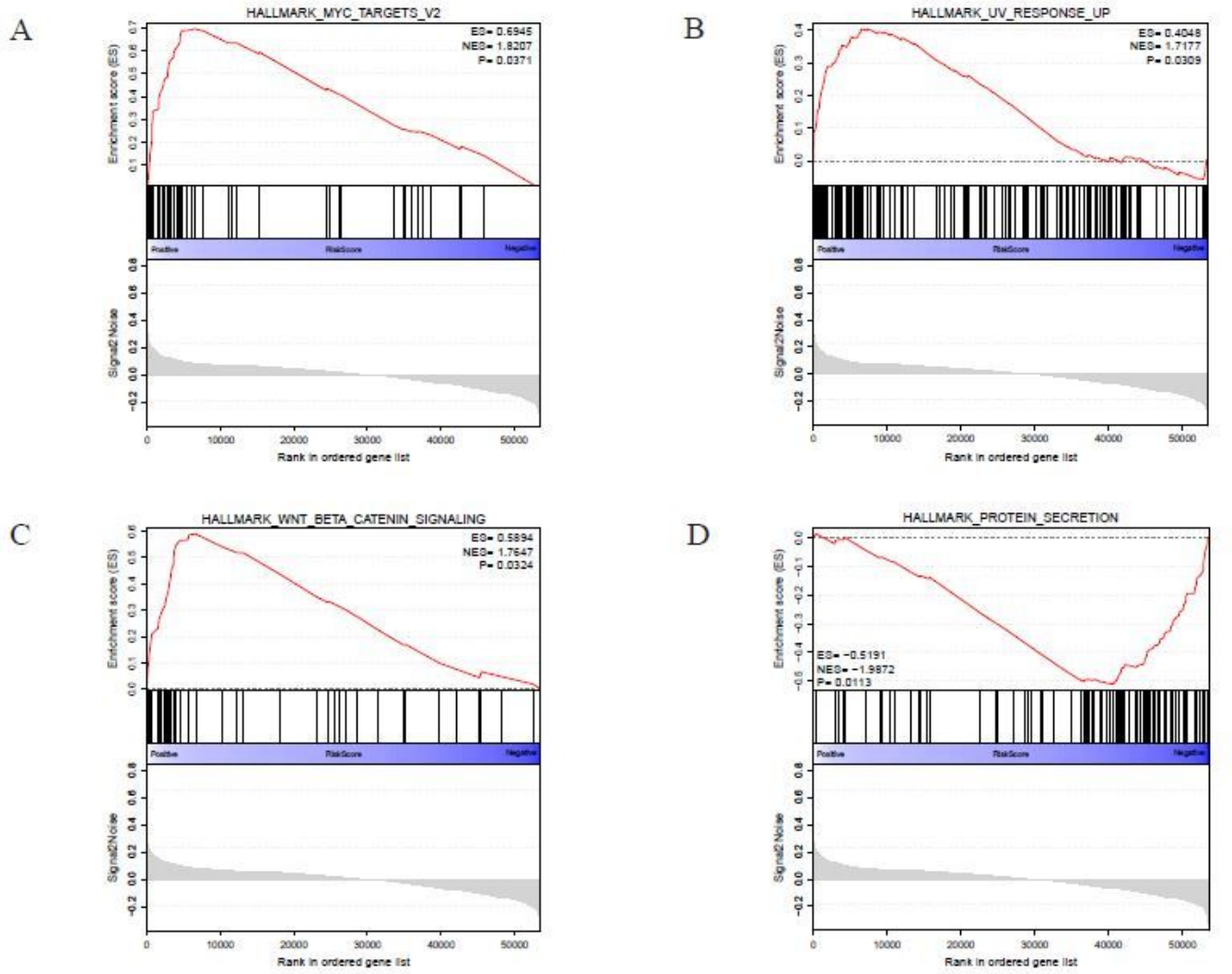

Figure 8

Gene set enrichment analysis of RBPs risk scores. (A-C) high risk scores. (D) low risk scores.