A Prognostic Model Constructed Based on Rna Binding Protein to Predict the Prognosis of Patients With Rectal Cancer

Liqiang Zhou
 Nanchang University Second Affiliated Hospital

You Wu
 Nanchang University Second Affiliated Hospital

Shihao Li
 Nanchang University Second Affiliated Hospital

Dengzhong Wu
 Nanchang University Second Affiliated Hospital

Jinliang Wang
 Nanchang University Second Affiliated Hospital

Yiwu Yuan
 Nanchang University Second Affiliated Hospital

Lin Xin (✉ 351442519007@email.ncu.edu.cn)
 Nanchang University Second Affiliated Hospital   https://orcid.org/0000-0001-5958-6155

Abstract

Background: The incidence of rectal cancer in young people is increasing, and there has been a problem of poor prognosis in recent years. Many studies have shown that RNA binding protein (RBP) is related to the progression of various malignant tumors. However, the role of RBPs in rectal cancer is poorly understood. New prognostic models are urgently needed.

Materials and methods: In the study, we used the RBPTD database, The Cancer Genome Atlas (TCGA) database and the transcription data information and corresponding clinical information of rectal cancer patients in the Gene Expression Omnibus (GEO) database to screen out RBPs that are differentially expressed in tumor tissues and normal tissues. Subsequently, we analyzed the prognostic value of these RBPs using bioinformatics methods. In order to screen the key RBP in the occurrence of rectal tumors and establish a prognostic risk score model. The use of survival analysis shows that assessing the relationship between key RBPs and the patient's overall survival rate. In the TCGA cohort, the prognostic model was further tested. At the same time, the nomogram of the 6 RBP mRNAs in the TCGA cohort was constructed, and the ROC curve was used for verification. Finally, q-PCR was performed on clinical samples to verify the expression of hub genes.

Results: The new 6RBP (EXO1, TOP2A, RUVBL1, NXT1, PACSIN2, WDR4) prognostic model was established to predict the prognosis of rectal cancer. The ROC curve showed good results in the training cohort and validation cohort. The new 6RBP (EXO1, TOP2A, RUVBL1, NXT1, PACSIN2, WDR4) prognostic model was established to predict the prognosis of rectal cancer. The ROC curve showed good survival prediction in both the training cohort and the validation cohort. The constructed nomogram has certain guiding significance for clinical decision-making. In addition, GSEA analysis revealed potential biological functions. The q-PCR verification results showed the consistency with the construction of the prognostic model.

Conclusions: We constructed a six RBPs prognostic model and a nomogram to predict the prognosis of patients with rectal cancer, and performed q-PCR expression testing through clinical samples, which may help clinical decision-making.

1 Introduction

In China, colorectal cancer is the third most common cancer and the third leading cause of cancer-related deaths. More than 50,000 people die from colorectal cancer each year [1]. Research predicts that by 2030, the global burden of rectal cancer will increase by 60%, reaching 1.1 million deaths and 2.2 million new cases [2]. However, recent studies have shown that the incidence of rectal cancer among young people is on the rise, and the long-term prognosis is reduced [3]. At present-day, the diagnosis of rectal cancer relies on colorectal endoscopy, pathological evaluation and tumor markers, which are challenging to meet clinical requirements[4]. In adjuvant therapy, the purpose of treating tumors is achieved by improving the drug resistance of chemotherapeutics and discovering specific tumor-specific targets. At the same time,

exploring effective early diagnosis methods is another important way to improve the clinical prognosis of patients with rectal cancer. Studies have shown that the protocadherin 17 (PCDH17) gene can be used as a potential prognostic indicator for the sensitivity of colorectal cancer patients to resistance in patients who are sensitive to 5-fluorouracil (5-FU)[5]. At the same time, in terms of how to target the KRAS mutation to cause drug resistance, research has found that targeting CDK4 and FYN can cause more KARS-mutated colorectal cancer cells to die, which can bring new directions for tumor treatment[6].Compared with invasive tests, biomarkers with better specificity and sensitivity have extraordinary value in early diagnosis, prognosis prediction and even treatment.

RBPs are composed of more than 2000 proteins, which participate in a variety of biological processes by combining with different RNA types, including RNA shearing, mRNA stability, cytoplasmic localization and protein translation [7]. Due to the critical role of RBPs in the process of post-transcriptional gene expression, RBPs that regulate differential expression may cause different diseases, including cancer Due to the critical role of RBPs in the process of post-transcriptional gene expression, RBPs that regulate differential expression may cause different diseases, including tumor[8].In recent years, whole-genome analysis has shown that the expression level of certain RBPs is closely related to the malignancy of colorectal cancer. For instance, LIN28 is an emerging carcinogenic driver [9]. The two subtypes LIN28A or LIN28B in mammals can promote colon cancer that invasion and progression, thereby affecting the prognosis of patients[10]. Also, Musashi RNA binding protein (MSI) is also upregulated in colorectal cancer and has been confirmed to be related to the invasion and drug resistance of colorectal cancer[11].On the contrary, HuR exerts anti-tumor effects by regulating the tumor suppressor p21 and WNT family protein WNT5A[12]. RBPs can also play an important role in the treatment of colorectal cancer. Tristetraprolin (TTP), RBPs, can be used as a therapeutic target for the anti-tumor drug resveratrol, inducing tumor cell apoptosis, reducing invasion and migration[13]. Nonetheless, most of the RBPs related to rectal cancer have not been clearly studied.

In this study, We collected data about rectal cancer from TCGA and GEO databases, screened out differentially expressed RBPs, and then used a variety of verification methods to validate the risk scoring model, and finally performed functional enrichment analysis to explore its potential biological functions. Our research has identified some RBPs related to rectal cancer. It provides us with potential new ideas for studying the occurrence and development of rectal tumor and may provide potential predictive biomarkers for the early diagnosis and Improved prognosis of rectal cancer in the future.

2 Materials And Methods

2.1 Data processing

First, download the gene names of 2005 RNA Binding Protein (RBP) from RBPTD (http://www.rbptd.com/), and then download the clinical data and FPKM data of RNA Binding Protein Expression and Disease (READ) from the TCGA database (Including 10normal, 167tumor), use the R software package to convert the data into TPM values and remove data less than 0.5; Download the

GSE87211 data set (including 160normal, 203tumor) from the GEO database, use the R software package to remove the data less than 0.5, combine the RBP gene name to extract the preprocessed data, and use the "limma" package to process the extracted data (with P<0.05, |log2FC| >1.0 is the standard). The differentially expressed RBPs are screened for the next analysis[14].

2.2 protein-protein interaction (PPI)network and module analysis

In order to clarify the relationship between the differentially expressed RBPs, the PPI information of the differentially expressed RBPs was analyzed through String Database (https://string-db.org)[15], and Cytoscape 3.7.1[16] was used for visualization. Then use the Cytoscape 3.7.1 plug-in Molecular Complex Detection (MCODE) to pick out the modules whose score, and the number of nodes is greater than 5, and sort out the hub RBPs. P-value ≤0.05, there is a significant difference.

2.3 KEGG pathway and GO enrichment analysis

Through GO enrichment analysis[17] and KEGG pathway analysis[18], the biological functions of the differentially expressed RBPs were determined (subject to P<0.05, FDR<0.05).Use DAVID (https://david.ncifcrf.gov/) to identify essential signal pathways, and submit the analyzed data to Sangerbox (https:Sangerbox.com//) for mapping.

2.4 Cox proportional hazards regression analysis

We used the "survival analysis" R package to perform univariate Cox regression analysis on the key RBPs in the compulsory modules of PPI sorting to screen out RBPs related to the overall prognosis (P<0.05). First of all, the 166 samples obtained after integrating clinical information from TGCA-READ are randomly divided into a training set (n=82) and test set (n=84); the training set is used for the prediction model, and the test set and complete data set will be Used to verify the prediction model. Next, use the multi-factor Cox analysis to further screen out the six-candidate hub RBPs (P<0.05) and the corresponding β value and HR value for constructing the risk ratio model in the training set.

2.5 Prognostic risk model Construction

According to the selected RBPs related to the prognosis. In the training set, we build a proportional hazards model to predict the prognosis of patients with rectal cancer. The risk score calculation formula for each hub RBP is as follows:

$$Risk\ score = \beta 1 * Exp1 + \beta 2 * Exp2 + \beta i * Expi$$

Where βi represents the regression coefficient, Expi represents the expression of RBP. The risk score is positively correlated with the patient's prognosis, so patients are divided into high and low-risk groups. Then the log-rank test was used to compare the significance of OS difference between the two groups, and the Kaplan-Meier survival curve was drawn. Finally, the "SurivivalROC" R language software package

was used to perform univariate and multivariate Cox regression analysis and ROC curve to predict the performance of the model.

2.6 Validation of the potency of the prognostic risk model

The patient scores of the test set and the complete data set are compared with the calculated values of the training set, and each patient is divided into a high-risk group and a low-risk group. Use the "survival analysis" R package to draw ROC curve, Kaplan–Meier survival curve, and single factor and multivariate Cox regression analysis, combined with clinicopathological characteristics to analyze and finally confirm the performance of the prediction model. Then we analyzed the correlation between the risk and clinical traits of the selected hub RBPs genes and the relationship between hub RBPs gene expression and clinical traits and used the R package for visualization. Subsequently, based on the complete model, the "rms" R package was used to construct a nomogram model containing all independent factors to verify the relationship between hub RBPs and the patient's 5-year OS. Finally, we used two data sets (n>100) in the GEO database and the Human Protein Atls(HPA) database (https://www.proteinatl as.org/) on the expression levels of hub RBPs in rectal cancer and normal tissues.

2.7 Gene Set Enrichment Analyses

According to the constructed risk ratio model to determine the best critical value, the TCGA-RBPs samples were divided into High and low-risk score groups. Use GSEA (version 4.0)[19] based on the molecular signature database (Molecular Signatures Database, MSigDB) to provide gene enrichment analysis for the high-risk group and low-risk group (|NSE|>1, FDR<0.05 is considered as statistical Learning meaning).

2.7 Reverse Transcription-QuantitativePolymerase Chain Reaction (RT-qPCR)

TRIzol reagent was used to extract total RNA from tumor tissues and normal tissues adjacent to gastric cancer patients. Use TB Green® Premix Ex Taq™ II (Tli RNaseH Plus) and 7900HT Fast Real-Time PCR System (ThermoFisher) for cDNA analysis. Calculate the expression levels of HEYL, FOXC1, XBP1, PLEK and HOPX compared with the internal reference gene ACTB. And the amplification results were compared and analyzed, and the relative expression levels of EXO1, TOP2A, RUVBL1, NXT1, PACSIN2 and WDR4 were obtained. The primer sequence is shown in **Table 1.**

2.8 Statistical analysis

The data set of our research is analyzed using the R package (version 3.6.1). Bioconductor provides tools for analyzing and interpreting high-throughput genomic data (http://bioconductor.org/). Analyze differential genes using the limma package in R (version 3.61). For continuous variables, t-test and analysis of variance were used for statistical differences; for categorical variables, Pearson's $\chi 2$ test and Fisher's exact test were used for statistical differences. Univariate and multivariate Cox proportional regression analysis uses Survival software package analysis. The Kaplan-Meier method was used to draw a survival curve to evaluate the overall survival rate of high-risk and low-risk rectal cancer. This study uses R (version 3.6.1) and GraphPad Prism (version 7.0) to perform statistical analysis.

3 Results

3.1 Screening of differentially expressed RBPs in rectal neoplasms

This research was carried out according to the flowchart shown in **Figure 1**. Download the gene names of 2005 RBPs from RBPTD (http://www.rbptd.com/). Next, download the RNA sequencing data of 177 samples from the TCGA database and the corresponding clinical information (including 167 rectal cancer samples and 10 normal samples). Then use the limma R software package to handling the original data and screen the differentially expressed RBPs (102 up-regulated RBPs and 34 down-regulated RBPs) according to FDR<0.05 and |log2fold change (FC)|≥1 and the names of RBPs(**Figure** 2A,2B). Then download the GSE87211 data set (including 203 rectal cancer samples and 160 normal tissue samples) from the GEO database, which is the same as the TCGA differential gene screening. Then use the limma R software package for preprocessing, and use FDR<0.05 ; |log2fold change (FC)|≥1 to select the differentially expressed RBPs (including 32 up-regulated and 35 down-regulated) by combining RBPs gene names. (**Figure** 2C,2D).

3.2 PPI Network Construction and functional enrichment analysis of differentially expressed RBPs genes

The biological function of differentially expressed RBPs in rectal cancer. Take the intersection of the differentially expressed genes screened by TCGA and GEO (including 44 up-regulated RBPs and 32 down-regulated RBPs), as shown in **Figure** 3A. According to the information of co-up-regulated and co-down-regulated RBPs in the STRING database, a PPI network with 73 nodes and 394 edges was constructed using Cytoscape software (**Figure** 3B). Subsequently, the MOCODE plug-in in Cytoscape was used to screen out the two compulsory modules with the highest scores from the PPI network; the cytoHUbba plug-in was used to select hub genes (P<0.05), and all hub genes existed in modules ≥1. Therefore, these two key cluster modules represent the critical biological roles of PPI networks.

Use DAVID (https://david.ncifcrf.gov/) to perform GO and KEGG function enrichment analysis on the 76 RBPs genes selected to discover the biological functions of the differential genes. KEGG analysis identified that differentially expressed RBPs are enriched in the biological pathways related to cell cycle, cysteine and methionine metabolism, and microRNA in cancer. The results of molecular function analysis confirmed that the differentially expressed RBPs were significantly enriched in RNA phosphodiester bonds, RNA binding catalytic activity, RNA synthesis, chromatin binding, etc. In terms of cell composition, different RBPs are mainly enriched in nucleoplasm, cytoplasm and nucleolus. In addition, we also found that differentially expressed RBPs are also enriched in RNA hydrolysis, IFN-α biosynthesis forward regulation, IFN-β synthesis forward regulation, rRNA transcription (**Figure.** 4)

3.3 Selection of RBPs related to prognosis

We aimed at identifying 73 differential expression hub RBPs in the PPI network. In order to study the correlation between these RBPs and prognostic significance, we used univariate Cox regression analysis to screen out 13 RBPs with a significant overall prognosis. (**Figure** 5A). Next, we divided the 166 samples

with clinical information of TCGA-READ into a training set (82 samples) and test set (84 samples). And using multivariate Cox regression analysis to analyze their impact on patient survival time and clinical prognosis(**Table 2**.), it was found that six hub RBPs are independent predictors of rectal cancer patients (**Figure** 5B).

3.4 Construction of a genetic risk scoring model related to prognosis

Construct a risk ratio prediction model for the six hub RBPs identified by multivariate cox regression analysis(**Table** 3). Calculate the risk score of each patient according to the following formula:

$$
\begin{aligned}
Risk\ score = &(1.770377 * EXO1) + (-2.3049 * TOP2A) \\
&+ (3.028853 * RUVBL1) + (-1.53016 * NXT1) \\
&+ (-1.53445 * PACSIN2) + (-2.37522 * WDR4)
\end{aligned}
$$

To evaluate the predictive ability of the model. In the training set (n = 82), we risk score based on the median of 82 samples are divided into high-risk and low-risk groups and group survival analysis. The results showed that high-risk group OS is far lower than the low-risk group (**Figure** 6A). Secondly, we further conducted a time-dependent ROC analysis on the prognostic ability of six hub RBPs. The area under the ROC curve of the constructed RBPs risk scoring model is 0.780, indicating that it has a good predictive ability(**Figure** 6B). **Figure** 6C shows the expression heat map of the risk score, patient survival status, and six RBPs in the high-risk group and the low-risk group. Finally, we performed univariate and multivariate Cox regression analysis. The results show that the forecasting model showed moderate independent predictive power.

3.5 Prognostic model performance verification

To validate the predictive value of six key RPBs prognostic models, we use the internal test set (n = 84), the complete data set (n = 166), to assess the predictive ability of the training set. Survival analysis of the two test sets showed that patients in the high-risk group had a worse prognosis than those in the low-risk group (**Figure** 7A,8A), which was consistent with the results of the training set. Time-dependent ROC curve analysis shows that the AUC of the internal test set and the complete set are 0.911 and 0.798, respectively. (**Figure** 8B, 9B). Figure 7C and 8C showed the survival status of patients with risk scores in the prognostic model and the expression heat map of 6 RBPs. Finally, univariate and multivariate Cox analyses involving clinical factors and risk scores showed good independent predictive power in both test sets. In summary, the reliability of the prognostic model is explained (**Table 4, Table 5**).

3.6 Construction and verification of prediction nomogram

To establish a method for quantitatively predicting the survival probability of patients with rectal cancer, we combined six independent prognostic markers of RBPs to construct a nomogram to predict the five-year OS rate of patients (**Figure** 9A). Besides, the ROC curve analysis showed that t n addition, ROC curve analysis showed that AJCC staging (AUC=0.890), tumor status (AUC=0.798), tumor residual (AUC=0.618)

and patient age (AUC=0.615) were all lower than the risk score of the model (AUC= 0.962)，as shown in Figure 9B.

3.7 Analysis of hub genes related to RBPs, clinical features and biological functions

Based on the proportional hazard regression model analysis, six RBPs(PANCSIN2, EXO1, TOP2A, NXT1, RUVBL1, and WDR4) are associated with prognosis. The expression of these 6 RBPs in different risk groups in the TCGA data set is shown in **Figure** 10. We observed significant differences between the low-risk and high -risk subgroups in clinical characteristics such as risk grade, tumor status, tumor residual, and tumor stage. We analyzed 6 RBPs for different clinical characteristics. The results showed that the expression levels of NXT1, TOP2A and WDR4 were significantly different in different risk groups. In different tumor stages and grades, significant differences were found in the expression levels of TOP2A and PACSIN2. The expression of PACSIN2 in different tumor states and tumor invasion blood vessels is different.

To further analyze the potential biological functions of the six RBPs genes, we use GSEA analysis in groups of risk scores. The results showed that the high-risk group was enriched in pancreatic β cells, blood coagulation, myogenesis, and epithelial-mesenchymal transition pathways(**Figure** 11).

3.8 External verification of the prognostic performance and expression of hub RBPs

To further determine the relationship between the expression of these Hub RBPsin rectal cancer and the prognosis, we obtained the immunohistochemical results from the human protein atlas database. Compared with normal rectal tissues, the expression of TOP2A, RUVBL1, PACSIN2, and WDR4 in rectal cancer increased significantly. At the same time, we combined two external data (GES20482, GSE90627) with sample numbers greater than 100. It was found that PANCSIN2, TOP2A, RUVBL1, WDR4 had significant differences in expression in normal tissues compared with rectal cancer tissues. We further demonstrated the expression of 5 TFs screened in clinical specimen tissues (**Figure** 12). The results confirmed that it is consistent with our analysis. EXO1, TOP2A, RUVBL1, WDR4 have higher expression levels in normal tissues, and lower expression levels in tumor tissues. However, PACSIN2 and NXT1 were not statistically different between normal tissues and tumor tissues.

4 Discussion

Colorectal cancer is the third most common malignant tumor in the world. Due to the unknown pathogenesis, the mortality rate of rectal cancer patients is still high[1]. With the development of precision medicine, people in addition to surgical treatment of colorectal tumors are expected to bring new understanding of diseases by deciphering the molecular basis behind the process of cancer[20].Most of this knowledge is obtained through the study of DNA and protein functions, but the study of post-transcriptional functions is not thorough enough. In the past ten years, the understanding of post-transcriptional cancers has increased, especially with the discovery of the diverse biological functions of non-coding RNAs. RBPs play an important biological role by forming an intricate network of RNA

regulators to regulate gene expression before and after transcription [21]. Accumulated studies have shown that RBPs have obstacles to the function or expression of RBP in the progression of various tumors[22-25]. However, the specific functions of most RBPs during tumorigenesis are still unclear[26]. Compared with a single biomarker using a multi-gene signature to construct a prognostic model, it can better predict the patient's prognosis[27].In this study, we integrated data from TCGA and GEO of rectal cancer and identified RBPs that are for screening differential expression RBPs. Subsequently, we constructed a PPI network of vital RBPs and analyzed their biological pathways. Besides, we selected six hub RBPs from 76 critical RBPs through Cox proportional regression analysis to establish a prognostic risk model. In the internal data set, survival analysis and nomogram show that the model has good independent prediction ability. Concurrently, we compared 12 marker genes related to the prognosis of colorectal cancer constructed by Fan et al. [28], and the results showed that the risk prognosis model we constructed was better than its ROC curve (AUC=0.798) than its ROC curve (AUC=0.553). We speculate that these 6 RBPs have the potential to predict the prognosis of rectal tumor.

GO analysis showed that RBPs that are differentially expressed in molecular functions are mainly involved in RNA synthesis, regulating ribozyme activity and participating in the process of chromatin binding, IFN-α biosynthesis forward regulation process, IFN-β synthesis forward regulation process. This is consistent with the reported research that RBPs mainly exert their biological functions by regulating RNA or participating in the process of chromatin regulation[29-31]. IFN forward synthesis pathway is a pathway related to tumor immune escape. In the process of tumorigenesis, inhibiting IFN forward synthesis can reduce the production of IFN and other inflammatory mediators and enhance tumor immune escape[32]. According to recent researches, it is stated that RBPs interact with the IFN forward synthesis pathway, which can affect tumor growth by regulating the immune-inflammatory response[32, 33]. Furthermore, RBPs related to prognosis are mainly enriched in microRNA in cancer, cysteine and methionine metabolism, and cell cycle in the KEGG pathway. MicroRNA is a regulatory non-coding small RNA, and it has been proven to be a promising target for targeted therapy of cancers[34]. Cysteine and methionine metabolism may be a unique metabolic phenotype dependent on the growth of malignant cancers[35]. Studies have found that this metabolic process is related to the prognosis of many tumors[36]. The prognosis of colorectal cancer is related to the expression of cell cycle-related genes, such as changes in the ubiquitin-proteasome system. The prognosis of colorectal cancer patients will also be affected[37]. The above results indicate that RBPs may participate in the pathophysiological process of tumor and are related to the prognosis of cancer patients.

Subsequently, by constructing a PPI network of differential RBPs, further single-factor and multi-factor Cox regression analysis obtained out 6 Hub RBPs (EXO1, TOP2A, RUVBL1, NXT1, PANCSIN2, WDR4) which are related to prognosis. Confirmed in research that the expression of EXO1[38, 39], RUVBL1[40], TOP2A[41] are related to the progression of colorectal cancers, which is consistent with the prediction results of our constructed model. Next, based on the 6 Hub RBPs coding genes in the TCGA training set, through multi-factor Cox proportional hazard regression analysis, we established a proportional hazard model for predicting the prognosis of rectal cancer. ROC curve analysis shows that these six genes have good diagnostic capabilities and can screen outpatients with rectal cancer with acquainting scarcely.

However, the molecular mechanisms of these six RBPs in rectal cancer are still poorly understood, and it is worthwhile to further explore their potential molecular mechanisms in rectal cancer. Then, we constructed a nomogram to help us more intuitive prediction 1– 5-year overall survival. Finally, we analyzed through GSEA that the high-risk group was mainly enriched in pancreatic β-cells, blood coagulation, myogenesis, and epithelial-mesenchymal transition pathways. Among them, the epithelial-mesenchymal transition pathway is an important marker to distinguish benign and malignant tumors [42]. Studies have shown that RBP can regulate the stability of RNA to inhibit the metastasis of rectal cancer cells [43]. We further verified the expression of the six RBPs encoding genes in the high and low risk groups through the HPA database and two external validation sets, which is consistent with the evaluation of the risk prognosis model we constructed. Finally, our q-PCR analysis of clinical samples was basically consistent with our predicted results. In summary, according to the results of external verification and risk model prediction, we speculate that the prognosis of patients in the high-risk group is poor, and the corresponding treatment plan or personalized evaluation treatment needs to be adjusted.

In general, our research affords new ideas for people to further reveal the role of RBPs in the rectal tumor. Also, our prognostic model shows good survival prediction performance, which may contribute to developing new prognostic indicators for rectal cancer. Unlike previous studies, the genetic signatures associated with these RBPs show critical biological functions, which suggests that they may be used in the treatment of rectal cancer in the future. Nevertheless, this study still has several limitations. First of all, due to the heterogeneity of patients with rectal cancer, our prognostic model should be constructed based on a multi-omics platform. Finally, some clinical information missing from the TCGA data set could be reduced the statistical validity and reliability of multivariate Cox regression analysis.

In summary, we systematically explored the expression and prognostic value of differentially expressed RBPs in rectal cancer through a series of bioinformatics analysis. These RBPs may be involved in the occurrence, development, invasion and metastasis of rectal cancer. Six prognostic models of RBPs encoding genes were constructed, which can be used as independent prognostic factors for rectal cancer. According to our research, this is the first report to establish a predictive model related to RBPs and the prognosis of rectal tumor. Future research should further clarify how these RBPs are involved explicitly in rectal cancer pathogenesis and develop specific therapeutic targets and molecular markers for evaluating patient prognosis.

Conclusion

Through a series of bioinformatics analysis methods, we have established new 6RBPs gene markers and nomograms to predict patients' overall survival with rectal cancer. This is further verified by q-PCR of clinical specimens, which may help clinical individualized treatment decisions.

Declarations

Ethics approval and consent to participate

All gastric cancer clinical samples involved were approved by the patient's informed consent and the ethics committee of the Second Affiliated Hospital of Nanchang University.

Consent for publication

All authors agree to the publication of this article.

Availability of data statement and materials

All data can be obtained from the corresponding author's office and public databases.

Competing Interest

The author claims no conflict of interest

Authors' Contributions

Liqiang Zhou and You Wu: Conception, design, data analysis and manuscript writing, Shihao Li, Yiwu Yuan,Dengzhong Wu and Jinliang Wang: Data collection, Lin Xin: Guide, supervise and review

References

1. Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, Abdel-Rahman O, Abdelalim A, Abdoli A, Abdollahpour I, Abdulle ASM *et al*: **Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study**. *JAMA Oncol* 2019.
2. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F: **Global patterns and trends in colorectal cancer incidence and mortality**. *Gut* 2017, **66**(4):683-691.
3. Murphy CC, Singal AG, Baron JA, Sandler RS: **Decrease in Incidence of Young-Onset Colorectal Cancer Before Recent Increase**. *Gastroenterology* 2018, **155**(6).
4. Ladabaum U, Dominitz JA, Kahi C, Schoen RE: **Strategies for Colorectal Cancer Screening**. *Gastroenterology* 2020, **158**(2):418-432.

5. Liu S, Lin H, Wang D, Li Q, Luo H, Li G, Chen X, Li Y, Chen P, Zhai B *et al*: **PCDH17 increases the sensitivity of colorectal cancer to 5-fluorouracil treatment by inducing apoptosis and autophagic cell death**. *Signal Transduct Target Ther* 2019, **4**:53.

6. Wang Y, Lin R, Ling H, Ke Y, Zeng Y, Xiong Y, Zhou Q, Zhou F, Zhou Y: **Dual inhibition of CDK4 and FYN leads to selective cell death in KRAS-mutant colorectal cancer**. *Signal Transduct Target Ther* 2019, **4**:52.

7. Siang DTC, Lim YC, Kyaw AMM, Win KN, Chia SY, Degirmenci U, Hu X, Tan BC, Walet ACE, Sun L *et al*: **The RNA-binding protein HuR is a negative regulator in adipogenesis**. *Nat Commun* 2020, **11**(1):213.

8. Wang Z, Li B, Luo Y, Lin Q, Liu S, Zhang X, Zhou H, Yang J, Qu L: **Comprehensive Genomic Characterization of RNA-Binding Proteins across Human Cancers**. *Cell Rep* 2018, **22**(1):286-298.

9. Zhang H, Zong Y, Qiu G, Jia R, Xu X, Wang F, Wu D: **Silencing Lin28 promotes apoptosis in colorectal cancer cells by upregulating let-7c targeting of antiapoptotic BCL2L1**. *Molecular medicine reports* 2018, **17**(4):5143-5149.

10. Pereira B, Billaud M, Almeida R: **RNA-Binding Proteins in Cancer: Old Players and New Actors**. *Trends Cancer* 2017, **3**(7):506-528.

11. Chiou G-Y, Yang T-W, Huang C-C, Tang C-Y, Yen J-Y, Tsai M-C, Chen H-Y, Fadhilah N, Lin C-C, Jong Y-J: **Musashi-1 promotes a cancer stem cell lineage and chemoresistance in colorectal cancer cells**. *Sci Rep* 2017, **7**(1):2172.

12. Leandersson K, Riesbeck K, Andersson T: **Wnt-5a mRNA translation is suppressed by the Elav-like protein HuR in human breast epithelial cells**. *Nucleic Acids Res* 2006, **34**(14):3988-3999.

13. Lee S-R, Jin H, Kim W-T, Kim W-J, Kim SZ, Leem S-H, Kim SM: **Tristetraprolin activation by resveratrol inhibits the proliferation and metastasis of colorectal cancer cells**. *Int J Oncol* 2018, **53**(3):1269-1278.

14. 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.

15. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P *et al*: **The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible**. *Nucleic Acids Res* 2017, **45**(D1):D362-D368.

16. Su G, Morris J, Demchak B, Bader G: **Biological network exploration with Cytoscape 3**. *Current protocols in bioinformatics* 2014, **47**:8.13.11-24.

17. **Gene Ontology Consortium: going forward**. *Nucleic Acids Res* 2015, **43**:D1049-1056.

18. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K: **KEGG: new perspectives on genomes, pathways, diseases and drugs**. *Nucleic Acids Res* 2017, **45**:D353-D361.

19. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP: **GSEA-P: a desktop application for Gene Set Enrichment Analysis**. *Bioinformatics* 2007, **23**(23):3251-3253.

20. Oronsky B, Reid T, Larson C, Knox SJ: **Locally advanced rectal cancer: The past, present, and future**. *Semin Oncol* 2020, **47**(1):85-92.

21. García-Cárdenas J, Guerrero S, López-Cortés A, Armendáriz-Castillo I, Guevara-Ramírez P, Pérez-Villa A, Yumiceba V, Zambrano A, Leone P, Paz-Y-Miño C: **Post-transcriptional Regulation of Colorectal Cancer: A Focus on RNA-Binding Proteins**. *Frontiers in molecular biosciences* 2019, **6**:65.

22. Culjkovic-Kraljacic B, Fernando TM, Marullo R, Calvo-Vidal N, Verma A, Yang S, Tabbò F, Gaudiano M, Zahreddine H, Goldstein RL *et al*: **Combinatorial targeting of nuclear export and translation of RNA inhibits aggressive B-cell lymphomas**. *Blood* 2016, **127**(7):858-868.

23. Wong CH, Lou UK, Li Y, Chan SL, Tong JH, To K-F, Chen Y: **CircFOXK2 Promotes Growth and Metastasis of Pancreatic Ductal Adenocarcinoma by Complexing with RNA-Binding Proteins and Sponging MiR-942**. *Cancer Res* 2020, **80**(11):2138-2149.

24. Liu T, Wei Q, Jin J, Luo Q, Liu Y, Yang Y, Cheng C, Li L, Pi J, Si Y *et al*: **The m6A reader YTHDF1 promotes ovarian cancer progression via augmenting EIF3C translation**. *Nucleic Acids Res* 2020, **48**(7):3816-3831.

25. 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):e1520.

26. Chen H, Liu J, Wang H, Cheng Q, Zhou C, Chen X, Ye F: **Inhibition of RNA-Binding Protein Musashi-1 Suppresses Malignant Properties and Reverses Paclitaxel Resistance in Ovarian Carcinoma**. *Journal of Cancer* 2019, **10**(6):1580-1592.

27. Zhang X, Zhang W, Jiang Y, Liu K, Ran L, Song F: **Identification of functional lncRNAs in gastric cancer by integrative analysis of GEO and TCGA data**. *J Cell Biochem* 2019, **120**(10):17898-17911.

28. Fan X, Liu L, Shi Y, Guo F, Wang H, Zhao X, Zhong D, Li G: **Integrated analysis of RNA-binding proteins in human colorectal cancer**. *World J Surg Oncol* 2020, **18**(1):222.

29. Huang H, Weng H, Sun W, Qin X, Shi H, Wu H, Zhao BS, Mesquita A, Liu C, Yuan CL *et al*: **Recognition of RNA N-methyladenosine by IGF2BP proteins enhances mRNA stability and translation**. *Nature cell biology* 2018, **20**(3):285-295.

30. Klim JR, Williams LA, Limone F, Guerra San Juan I, Davis-Dusenbery BN, Mordes DA, Burberry A, Steinbaugh MJ, Gamage KK, Kirchner R *et al*: **ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair**. *Nat Neurosci* 2019, **22**(2):167-179.

31. Elliott DJ, Dalgliesh C, Hysenaj G, Ehrmann I: **RBMX family proteins connect the fields of nuclear RNA processing, disease and sex chromosome biology**. *Int J Biochem Cell Biol* 2019, **108**:1-6.

32. Herbert A: **ADAR and Immune Silencing in Cancer**. *Trends Cancer* 2019, **5**(5):272-282.

33. Watanabe T, Imamura T, Hiasa Y: **Roles of protein kinase R in cancer: Potential as a therapeutic target**. *Cancer Sci* 2018, **109**(4):919-925.

34. Mollaei H, Safaralizadeh R, Rostami Z: **MicroRNA replacement therapy in cancer**. *J Cell Physiol* 2019, **234**(8):12369-12384.

35. Graziosi L, Mencarelli A, Renga B, D'Amore C, Bruno A, Santorelli C, Cavazzoni E, Cantarella F, Rosati E, Donini A *et al*: **Epigenetic modulation by methionine deficiency attenuates the potential for gastric cancer cell dissemination**. *J Gastrointest Surg* 2013, **17**(1).

36. Ward NP, DeNicola GM: **Sulfur metabolism and its contribution to malignancy**. *Int Rev Cell Mol Biol* 2019, **347**.

37. Voutsadakis IA: **Pathogenesis of colorectal carcinoma and therapeutic implications: the roles of the ubiquitin-proteasome system and Cox-2**. *J Cell Mol Med* 2007, **11**(2):252-285.

38. Sun X, Zheng L, Shen B: **Functional alterations of human exonuclease 1 mutants identified in atypical hereditary nonpolyposis colorectal cancer syndrome**. *Cancer Res* 2002, **62**(21):6026-6030.

39. Hansen MF, Johansen J, Bjørnevoll I, Sylvander AE, Steinsbekk KS, Sætrom P, Sandvik AK, Drabløs F, Sjursen W: **A novel POLE mutation associated with cancers of colon, pancreas, ovaries and small intestine**. *Fam Cancer* 2015, **14**(3):437-448.

40. Lauscher JC, Elezkurtaj S, Dullat S, Lipka S, Gröne J, Buhr HJ, Huber O, Kruschewski M: **Increased Pontin expression is a potential predictor for outcome in sporadic colorectal carcinoma**. *Oncol Rep* 2012, **28**(5):1619-1624.

41. Abraham AD, Esquer H, Zhou Q, Tomlinson N, Hamill BD, Abbott JM, Li L, Pike LA, Rinaldetti S, Ramirez DA *et al*: **Drug Design Targeting T-Cell Factor-Driven Epithelial-Mesenchymal Transition as a Therapeutic Strategy for Colorectal Cancer**. *J Med Chem* 2019, **62**(22):10182-10203.

42. Derynck R, Weinberg RA: **EMT and Cancer: More Than Meets the Eye**. *Dev Cell* 2019, **49**(3):313-316.

43. Tang L, Zhao P, Kong D: **Muscleblind-like 1 destabilizes Snail mRNA and suppresses the metastasis of colorectal cancer cells via the Snail/E-cadherin axis**. *Int J Oncol* 2019, **54**(3):955-965.

Tables

Table 1. Primers sequences.

Targets		Sequences
ACTB	Forward	5'-CACCATTGGCAATGAGCGGTTC-3'
	Reverse	5'-AGGTCTTTGCGGATGTCCACGT-3'
PACSIN2	Forward	5'-GGTTCAGAAGCACCTAGACCTG-3'
	Reverse	5'-TGATTGGCTCGGAACCACCTCA-3'
EXO1	Forward	5'-TCGGATCTCCTAGCTTTTGGCTG-3'
	Reverse	5'-AGCTGTCTGCACATTCCTAGCC-3'
TOP2A	Forward	5'-GTGGCAAGGATTCTGCTAGTCC-3'
	Reverse	5'-ACCATTCAGGCTCAACACGCTG-3'
NXT1	Forward	5'-CTTCCAGCGAGTTCCAAATCA-3'
	Reverse	5'-CAGATGACAACAAGGACCGTG-3'
WDR4	Forward	5'-CCACCTCCATAGCAAGCAGTG-3'
	Reverse	5'-ACGCTTACTGTCATCGGTTAAAG-3'
RUVBL1	Forward	5'-AGGTGAAGAGCACTACGAAGA-3'
	Reverse	5'-CTACTATGACGCCACATGCCT-3'

Table 2. Multivariate cox analysis to build hub RBPs of risk proportional model

Gene ID	Coef	HR	HR.95L	HR.95H	Pvalue
EXO1	1.77	5.87	1.28	27.01	0.02
TOP2A	-2.30	0.10	0.20	0.51	0.01
RUVBL1	3.02	20.67	1.34	318.06	0.03
NXT1	-1.53	0.22	0.04	1.03	0.06
PACSIN2	-1.53	0.22	0.03	1.81	0.16
WDR4	-2.38	0.09	0.01	0.68	0.20

Coef: coefficients, HR: Hazard Ratio.

Table 3. The Influence of Different Clinical Factors on the Construction of Predictive Model of Training Set

	Univariate analysis				Multivariate analysis			
	HR	HR.95L	HR.95H	*P*alue	HR	HR.95L	HR.95H	*P*value
Age	1.07	1.00	1.14	0.05	1.10	1.02	1.19	0.02
Gender	1.22	0.30	4.89	0.78	2.96	0.49	17.77	0.24
T stage	1.67	0.53	5.24	0.38	1.23	0.22	6.70	0.81
M stage	4.73	1.25	17.93	0.02	13.62	0.54	344.06	0.11
N stage	2.41	1.12	5.18	0.02	2.51	0.59	10.59	0.21
AJCC stage	2.34	1.08	5.06	0.03	0.38	0.04	3.63	0.40
riskScore	1.11	1.04	1.19	0.00	1.17	1.02	1.35	0.03

Table 4. The Influence of Different Clinical Factors on the Construction of Predictive Model of Test Set

	Univariate analysis				Multivariate analysis			
	HR	HR.95L	HR.95H	*P*value	HR	HR.95L	HR.95H	*P*value
Age	1.11	1.04	1.18	0.00	1.11	1.02	1.22	0.02
Gender	0.72	0.22	2.37	0.59	0.64	0.11	3.83	0.63
T stage	5.06	1.28	20.02	0.02	7.47	1.23	45.37	0.03
M stage	2.89	0.86	9.71	0.09	23.23	0.49	1099.03	0.11
N stage	2.44	1.20	4.93	0.01	3.26	0.68	15.60	0.14
AJCC stage	1.94	1.10	3.45	0.02	0.18	0.01	2.43	0.20
riskScore	1.00	1.00	1.00	0.08	1.00	1.00	1.00	0.05

Table 5. The influence of different clinical factors on the complete prediction

	Univariate analysis				Multivariate analysis			
	HR	HR.95L	HR.95H	*P*value	HR	HR.95L	HR.95H	*P*value
Age	1.08	1.02	1.13	0.01	1.16	1.04	1.29	0.01
Gender	1.17	0.40	3.44	0.78	5.34	0.65	43.98	0.12
History of colon polyps	1.12	0.31	4.05	0.87	0.61	0.08	4.74	0.64
T stage	3.04	1.17	7.85	0.02	1.76	0.40	7.80	0.46
M stage	5.03	1.80	14.03	0.00	0.44	0.01	13.47	0.64
N stage	2.33	1.30	4.17	0.00	1.93	0.51	7.26	0.33
AJCC stage	2.37	1.36	4.13	0.00	1.43	0.26	7.98	0.68
Lymphatic invasion	2.10	0.76	5.80	0.15	0.23	0.02	2.95	0.26
Venous invasion	2.17	0.79	6.00	0.13	5.98	0.42	85.00	0.19
Residual tumor	2.55	1.51	4.30	0.00	1.49	0.34	6.55	0.60
Cancer status	14.05	3.16	62.51	0.00	9.64	1.32	70.22	0.03
MSI	1.31	0.36	4.74	0.68	7.48	0.82	68.54	0.07
riskScore	1.03	1.00	1.07	0.06	1.11	1.03	1.20	0.01
Age	1.08	1.02	1.13	0.01	1.16	1.04	1.29	0.01

Figures
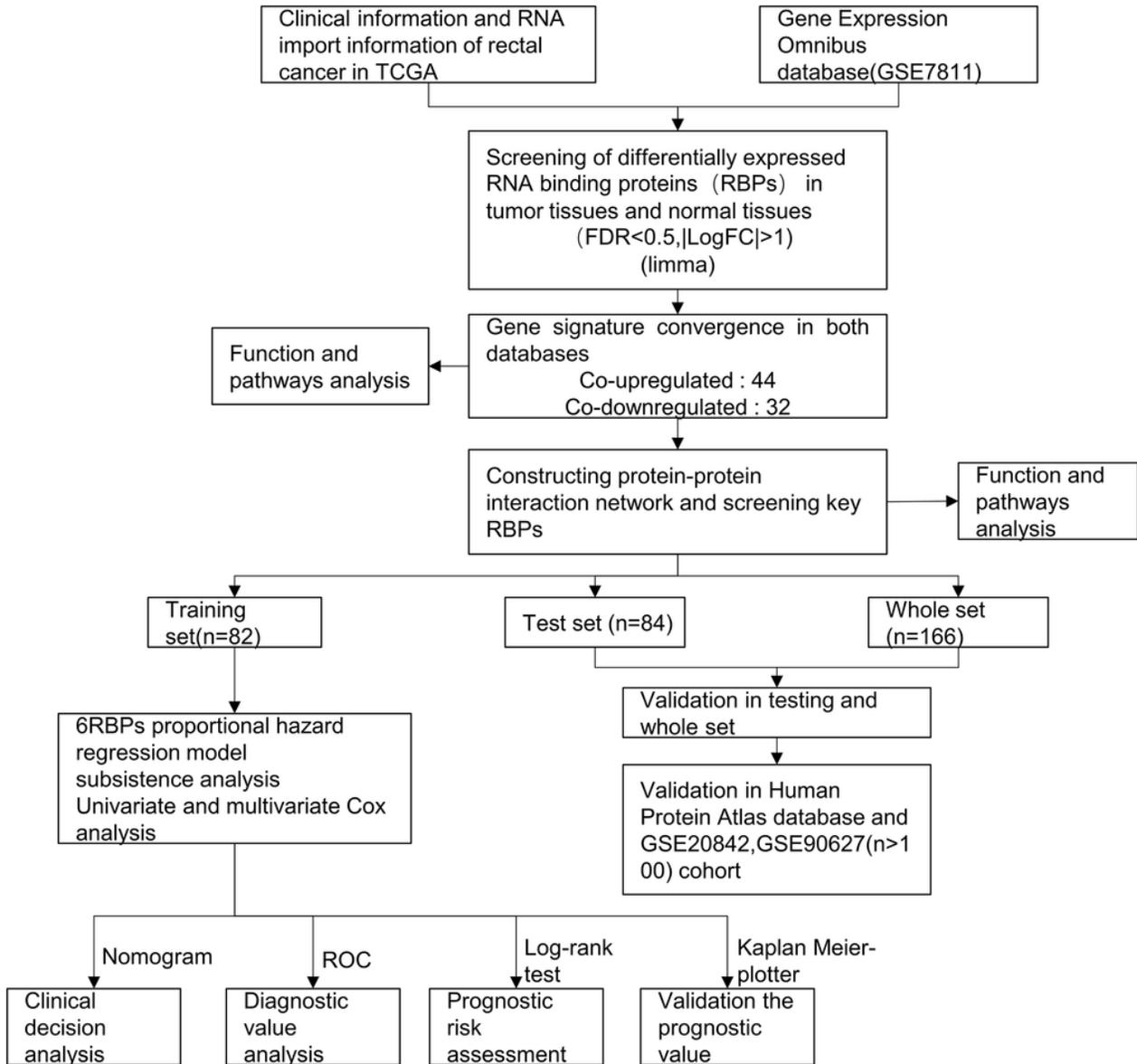
Figure 1

Analyze the flow chart of RBPs in rectal cancer from public data from GEO and TCGA databases
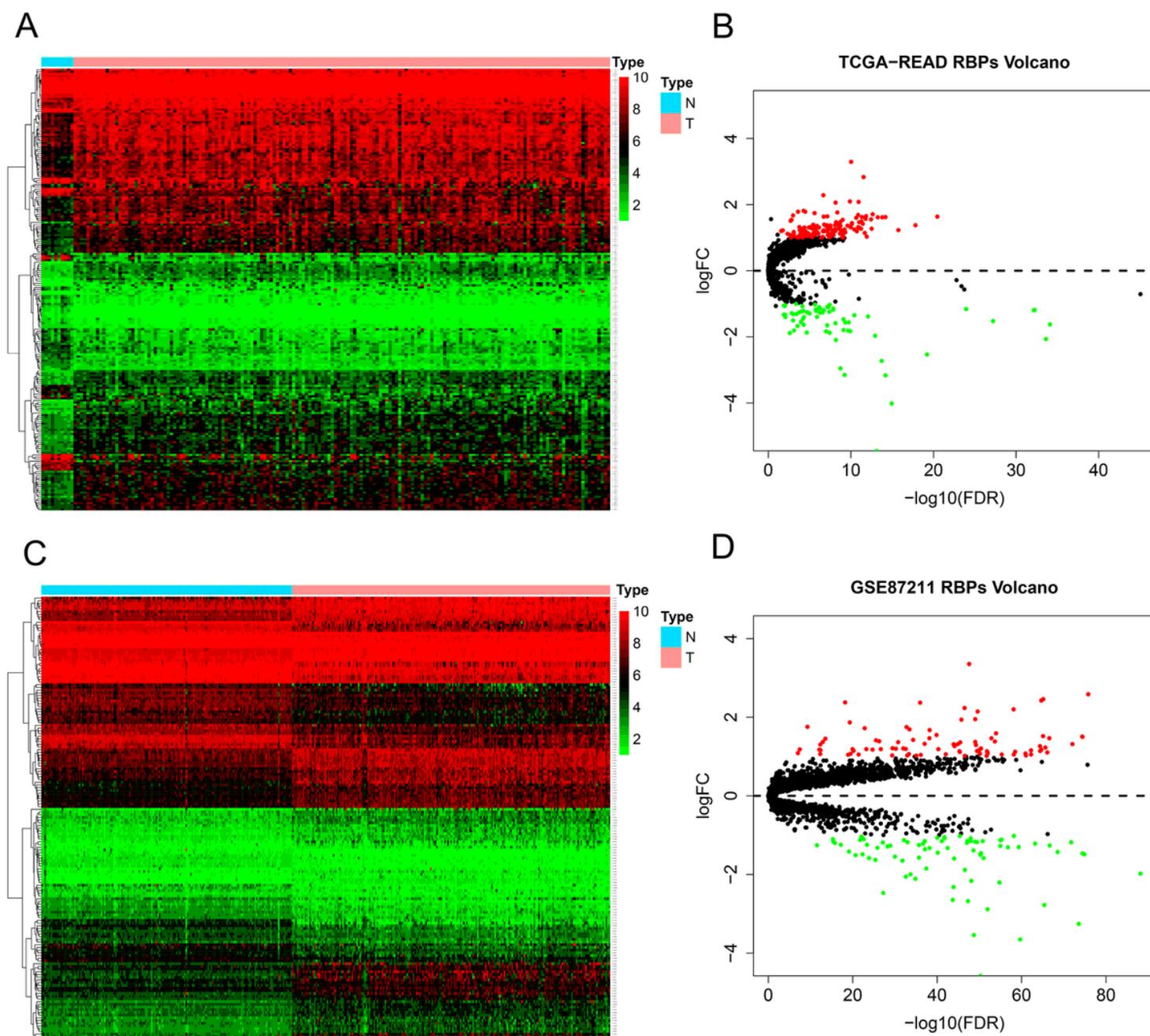
Figure 2

The differentially expressed in prostate cancer. (A) Heat map of RBPs related to differential expression of rectal cancer in TCGA database, (B) Volcano map of RBPs related to differential expression of rectal cancer in TCGA database, (C) Heat map of RBPs related to differential expression of rectal cancer in GSE87211, (C) Volcano map of RBPs related to differential expression in rectal cancer in GSE87211.
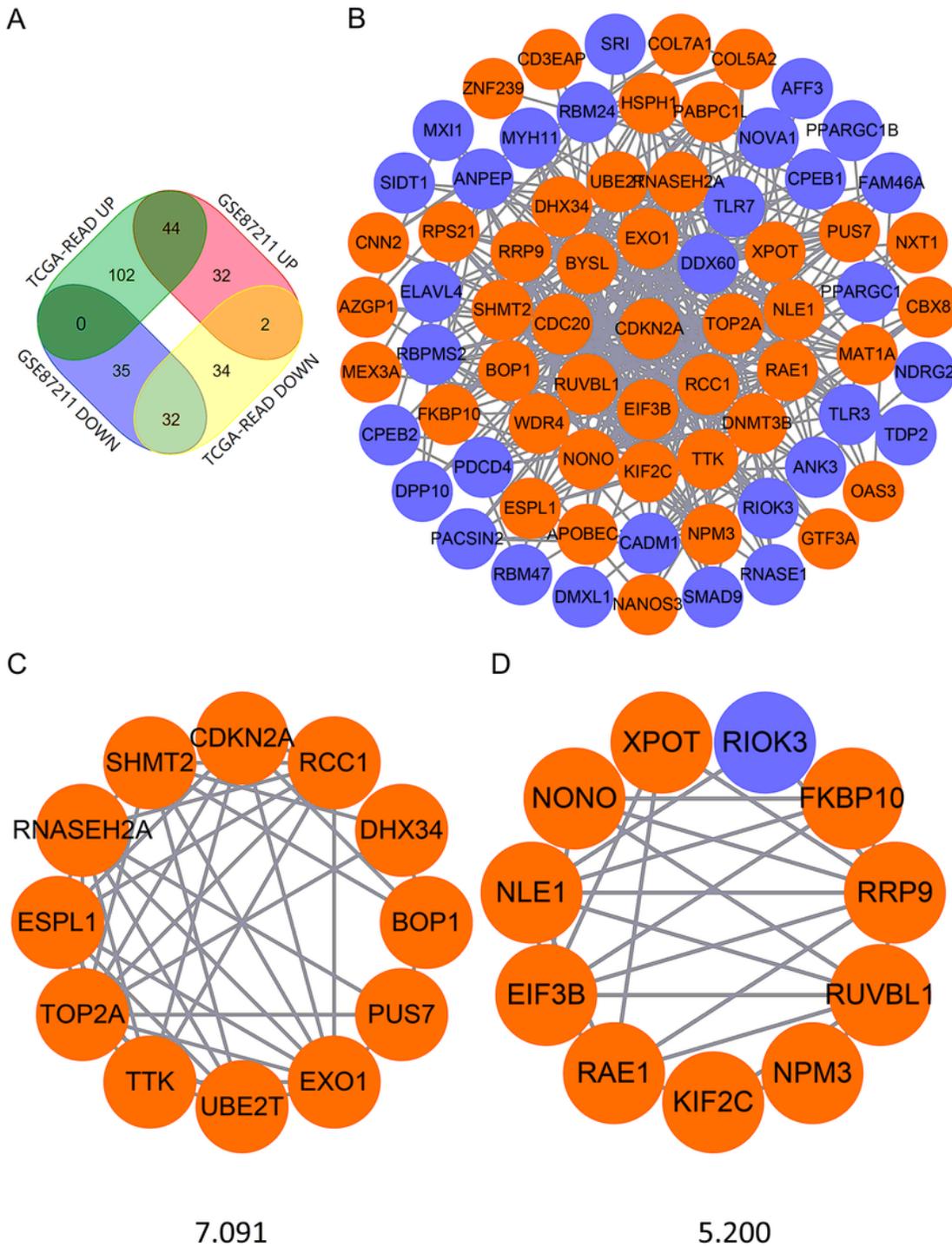
Figure 3

Construction of PPI network and module analysis. (A) The Venn diagram of RBPs that are both upregulated and down-regulated by TCGA and GEO;(B)PPI network construction;(C) critical module 1 in PPI network; (D) critical module 2 in PPI network. The red dot represents the upward adjustment, the blue dot represents down.
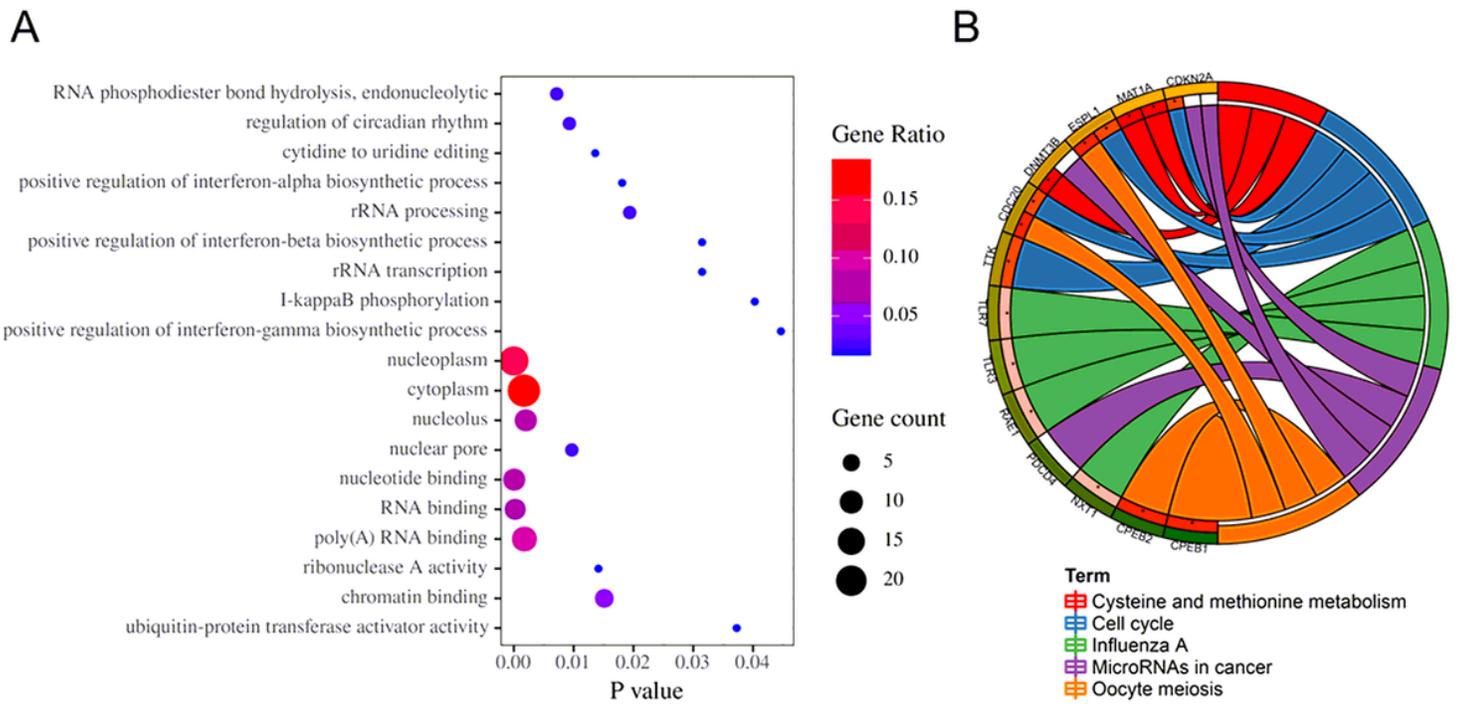
Figure 4

GO enrichment and KEGG pathway analysis of differentially expressed RBPs. (A) GO enrichment analysis, (B) KEGG pathway analysis.

	pvalue	Hazard ratio
KIF2C	0.038	0.673(0.464-0.977)
TTK	0.033	0.636(0.419-0.964)
EXO1	0.033	0.614(0.392-0.962)
TOP2A	0.012	0.633(0.442-0.906)
RUVBL1	0.038	0.515(0.276-0.963)
NXT1	0.012	0.543(0.338-0.873)
NLE1	0.004	0.418(0.229-0.760)
PACSIN2	0.005	0.398(0.210-0.754)
PPARGC1B	0.007	0.426(0.229-0.790)
RCC1	0.005	0.461(0.267-0.796)
WDR4	0.008	0.510(0.311-0.837)
RBPMS2	0.014	1.532(1.091-2.151)
PUS7	0.034	0.534(0.298-0.954)
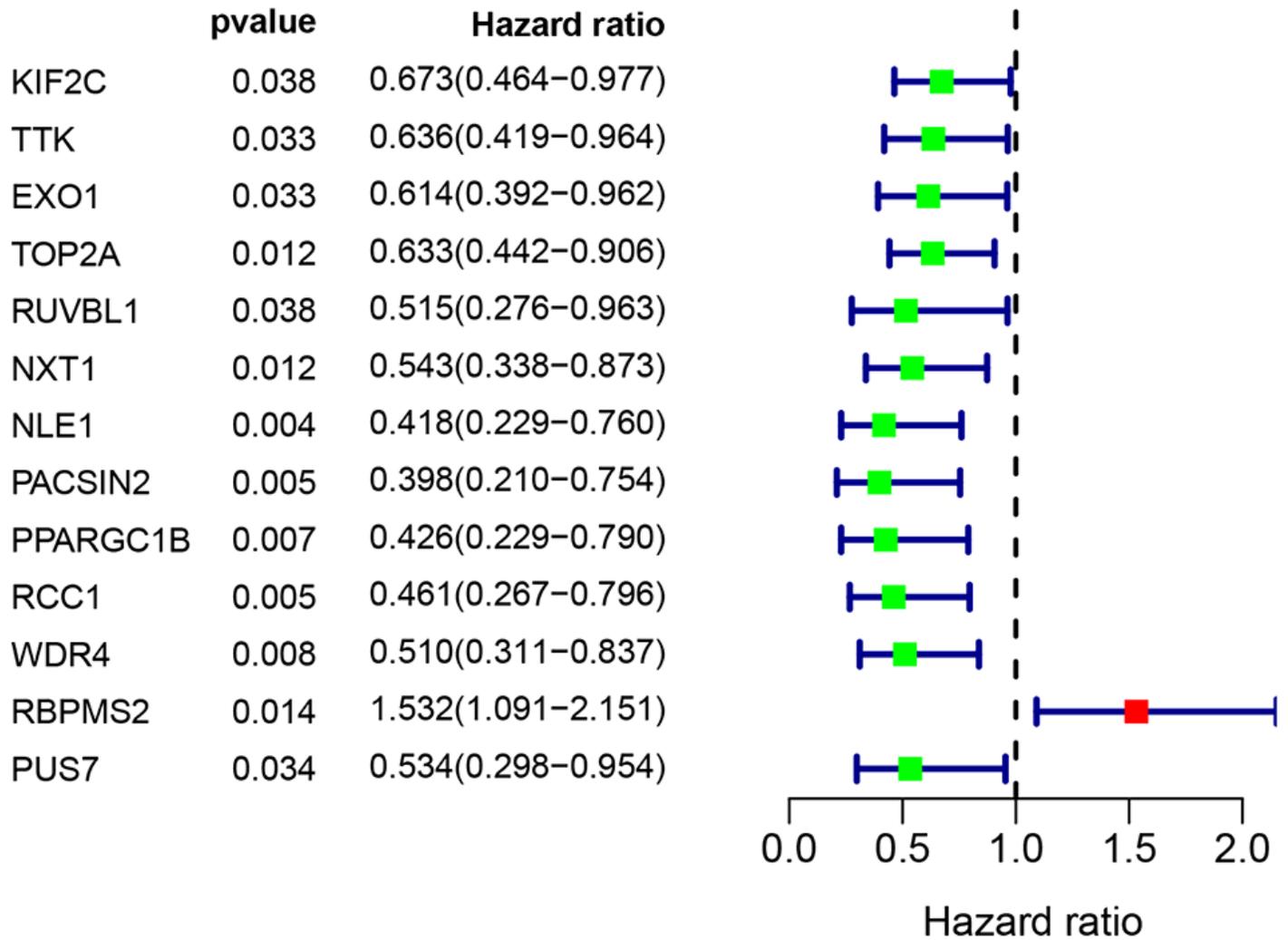
Figure 5

Identify RBPs related to prognosis. Univariate Cox regression analysis to screen the differentially expressed hub RBPs identified by the PPI network.
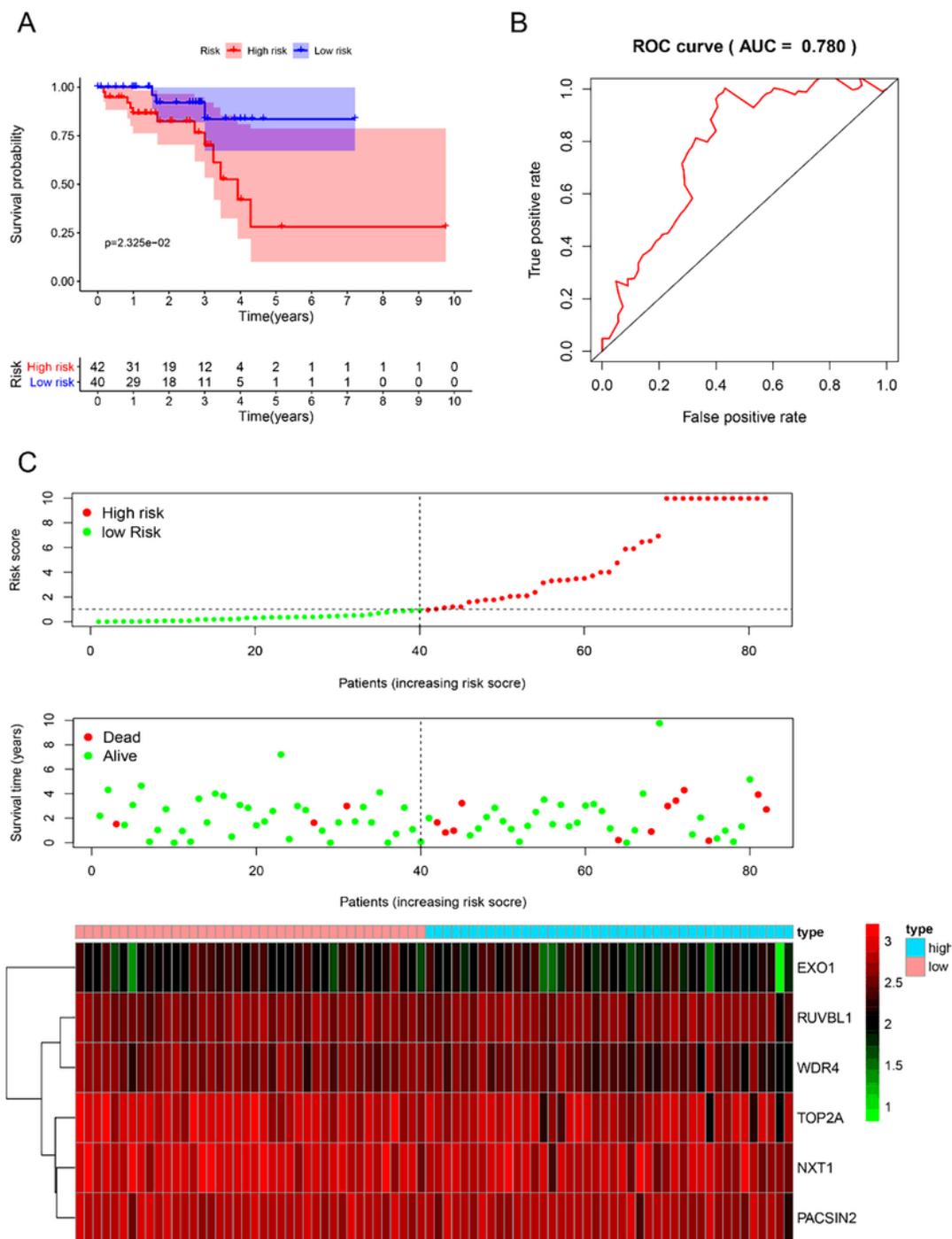
Figure 6

Analysis of Prognostic Risk Proportional Model Based on 6-Gene Training Set. (A) Kaplan–Meier Survival curve analysis of high-risk group and low-risk group;(B) ROC curve for predicting OS from risk score; (C)The relationship between the patient's expression heat map and the patient's overall survival in the training set.
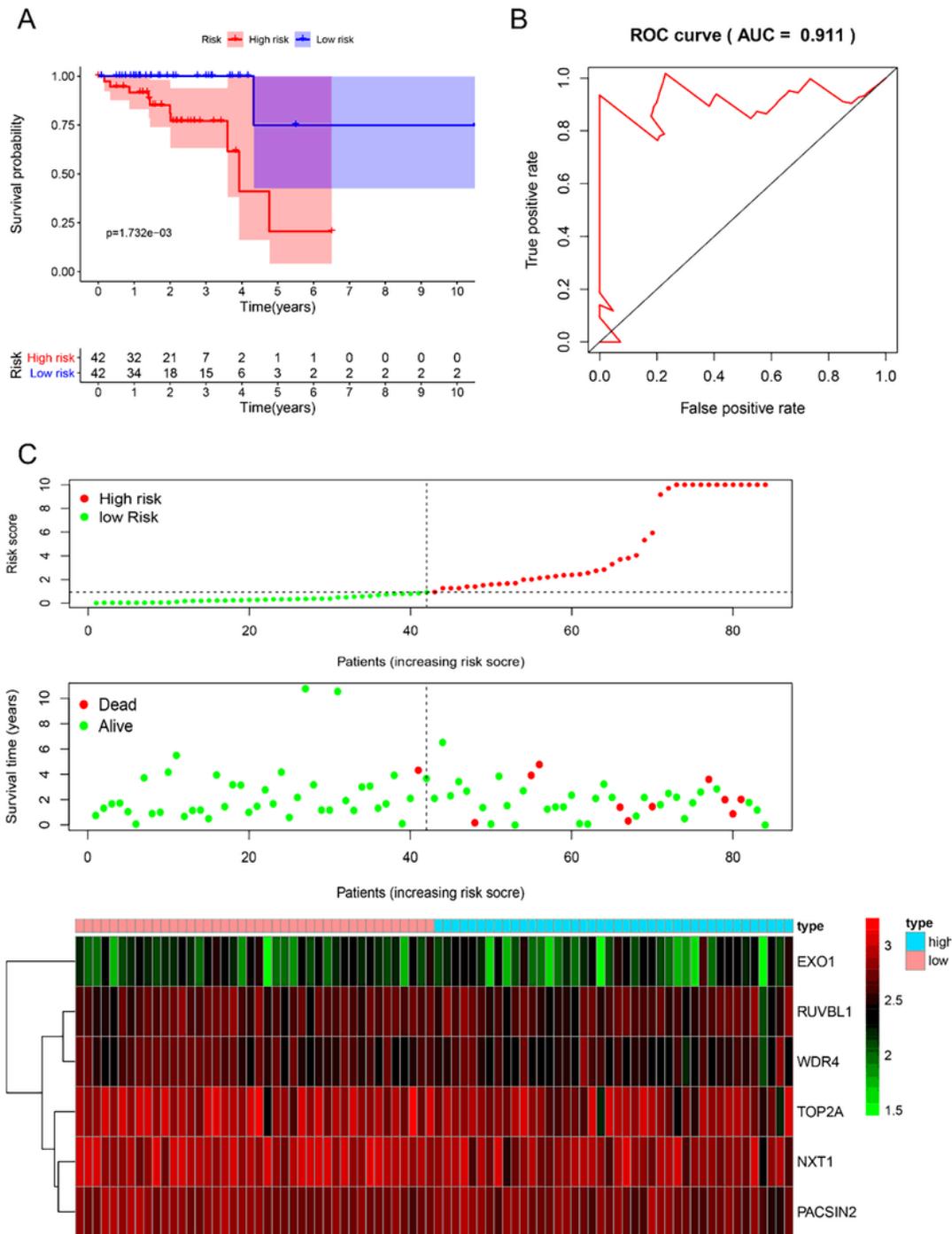
Figure 7

Prognostic risk model validation of 6-RBPs in the test set. (A) Kaplan−Meier Survival curve analysis of high-risk group and low-risk group;(B) ROC curve for predicting OS from risk score; (C)The relationship between the patient's expression heat map and the patient's overall survival in the test set.
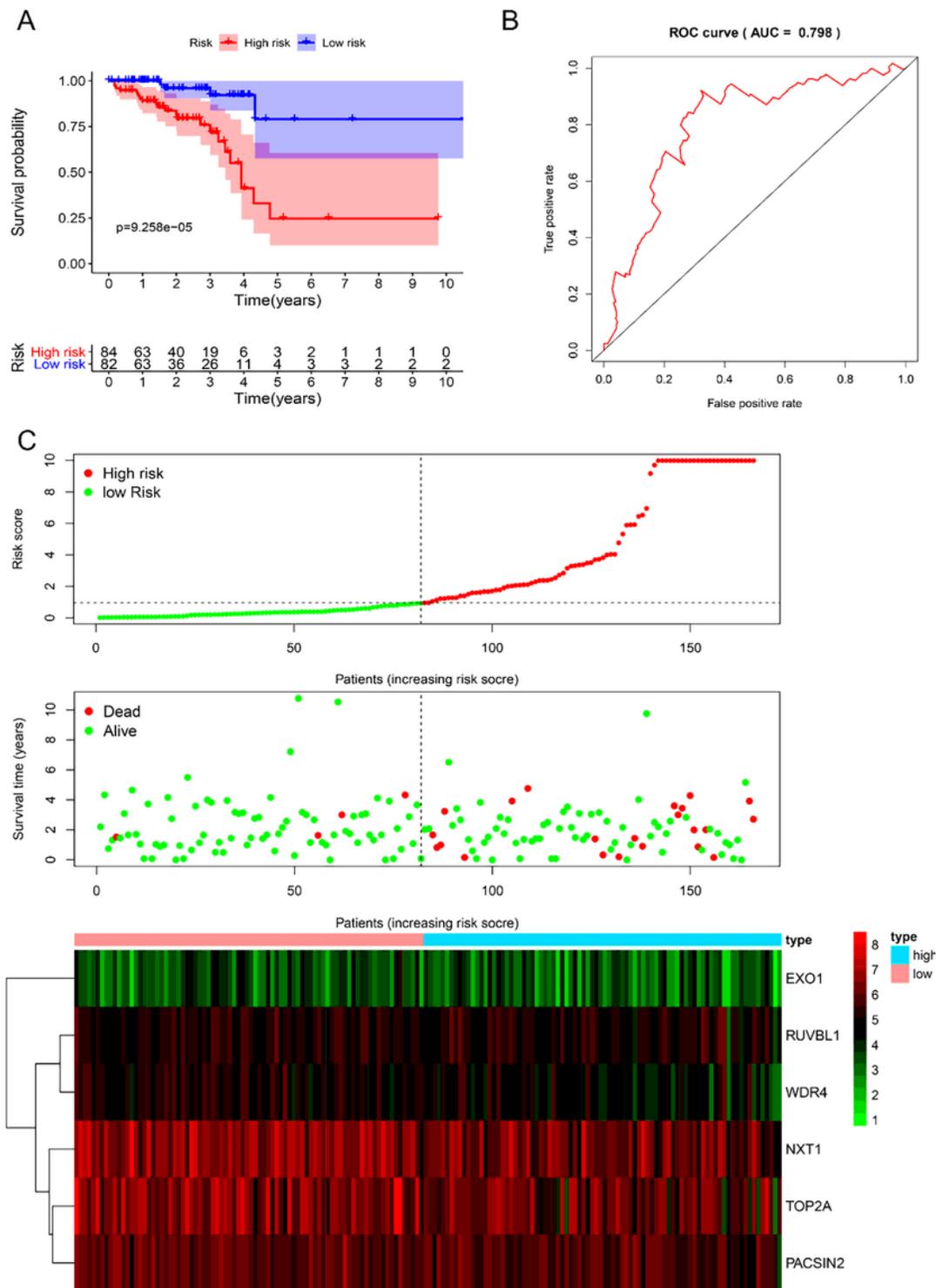
Figure 8

Validation of the prognostic risk model based on 6-RBPs in the whole data set. A) Kaplan–Meier Survival curve analysis of high-risk group and low-risk group;(B) ROC curve for predicting OS from risk score; (C)The relationship between the patient's expression heat map and the patient's overall survival in the whole data set; (D) Univariate and multivariate regression analysis to validate the prognostic model of the whole data set.
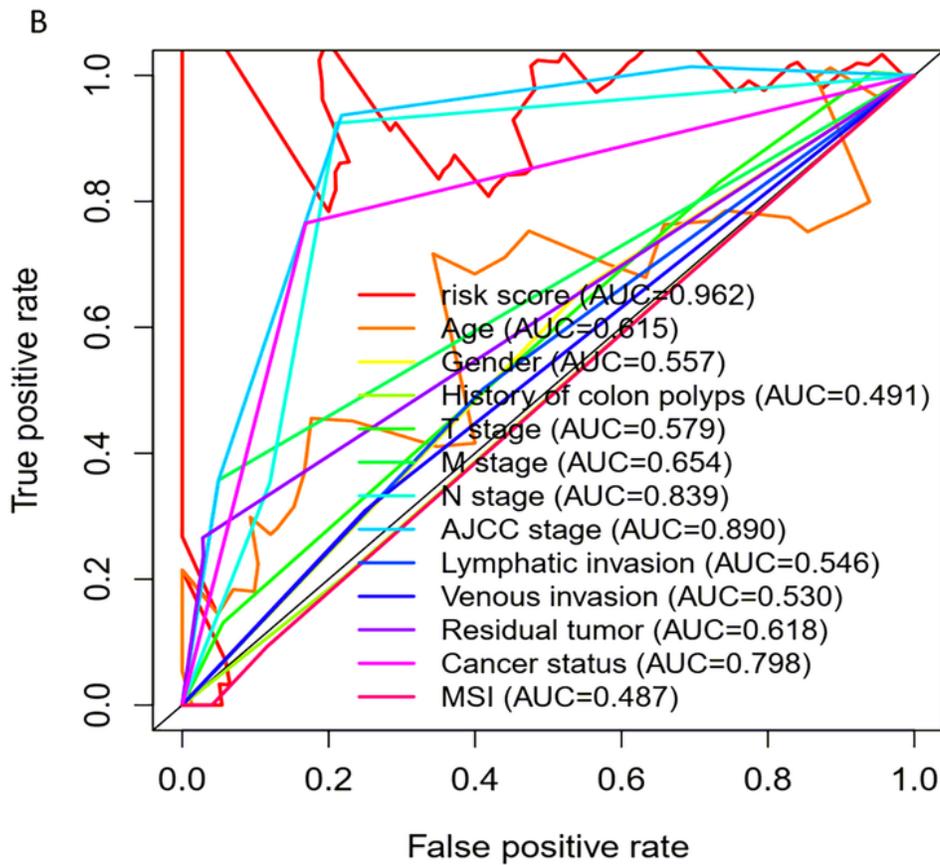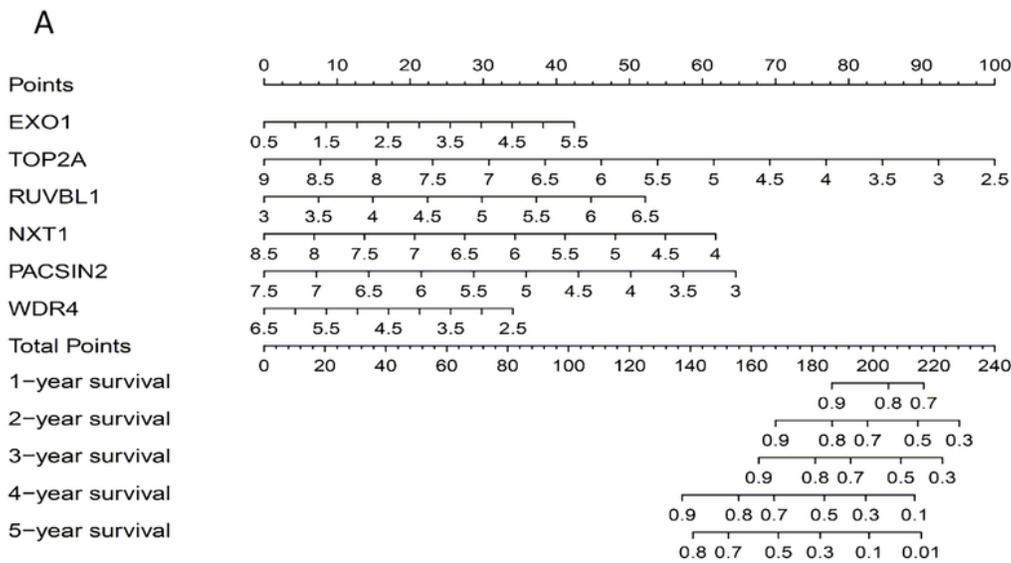
Figure 9

Five-year OS prediction based on nomogram and the prognostic value of 6 RBPs in the whole set. (A) The nomogram shows the five-year OS of patients in the total concentration;(B) The relationship between 6-RBPs and different clinical traits.
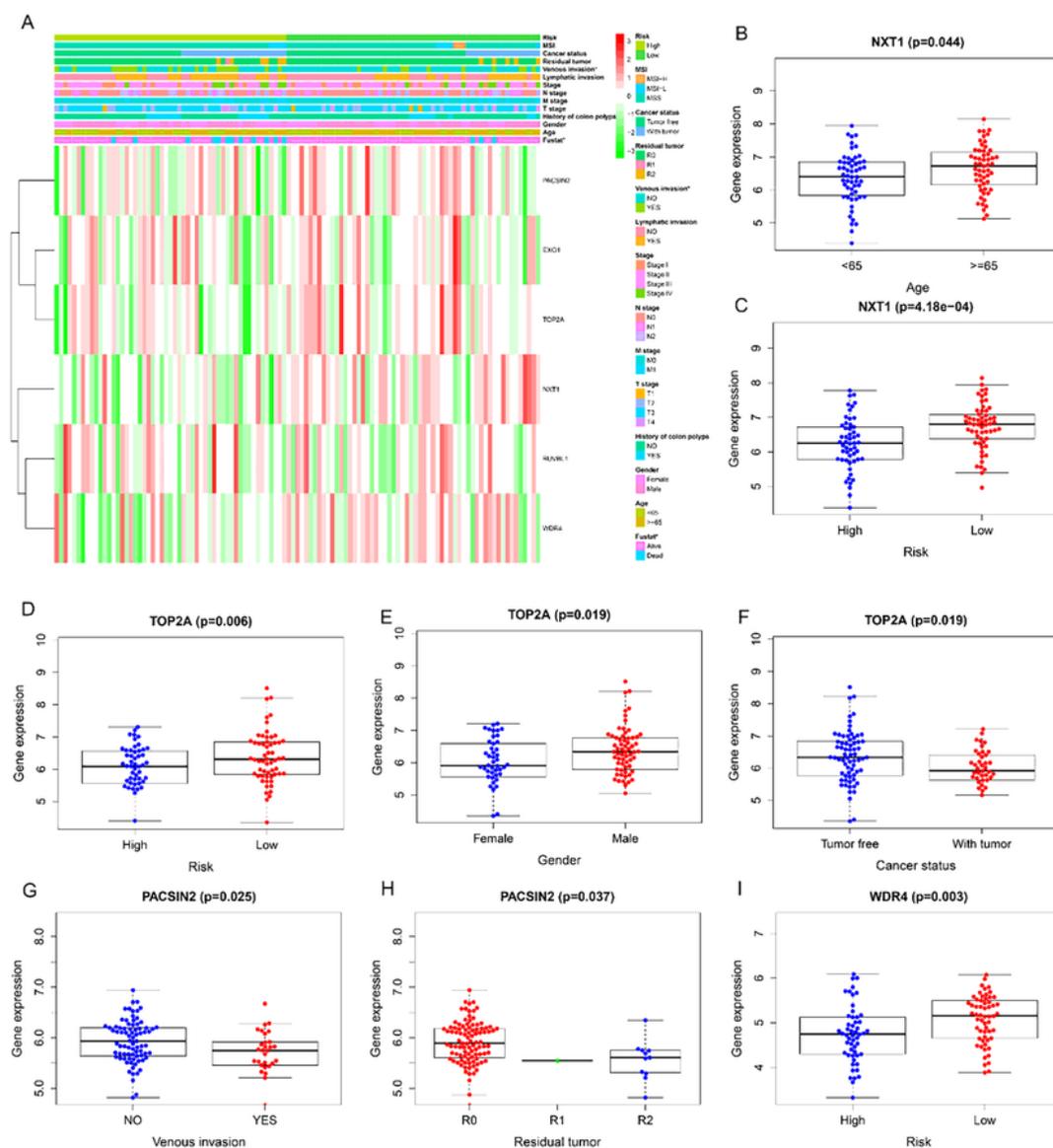
Figure 10

Verification of the correlation between 6 key RBPs and clinical traits.（A）Heatmap showed the expression of the 9 glycolysis-related hub genes in high- and low-risk patients in the TCGA dataset associated with tumor status, grade, histological type and stage.(B-C)NXT1.(D-F)TOP2A .(G-H)PACSIN2. (I)WDR4
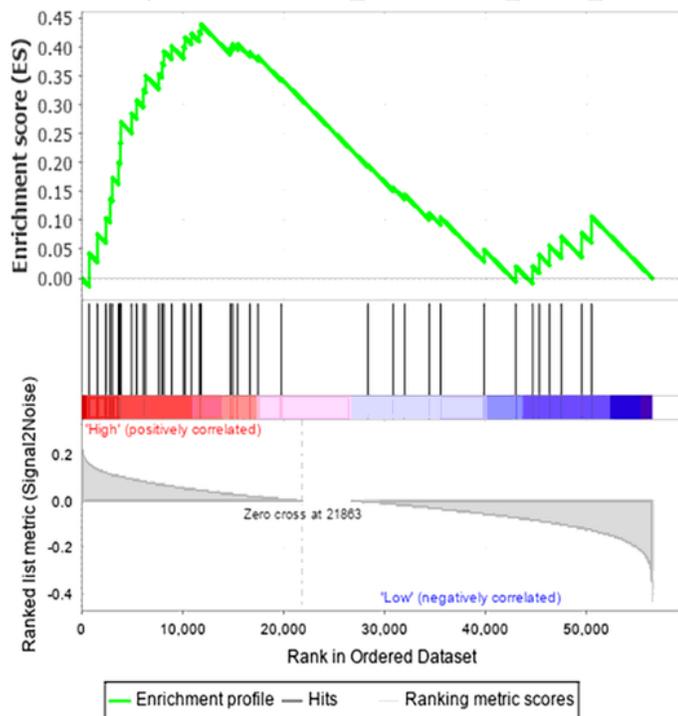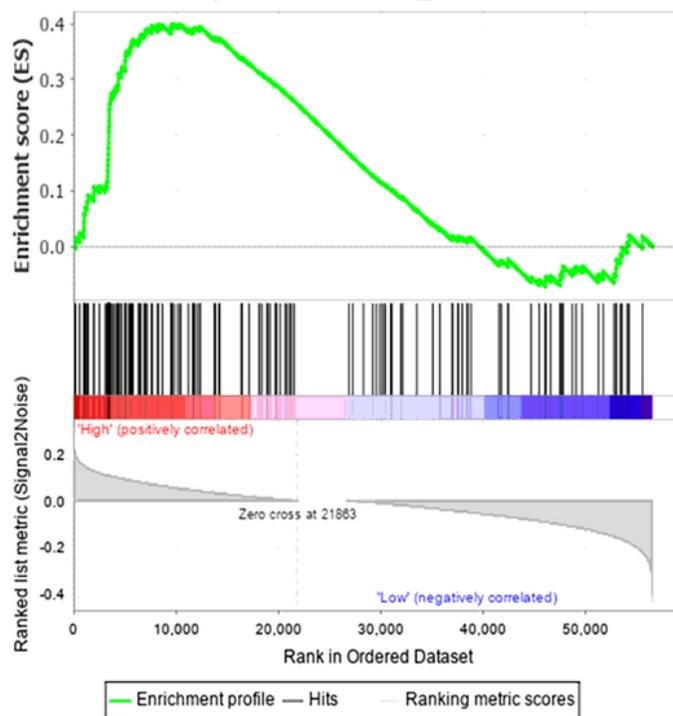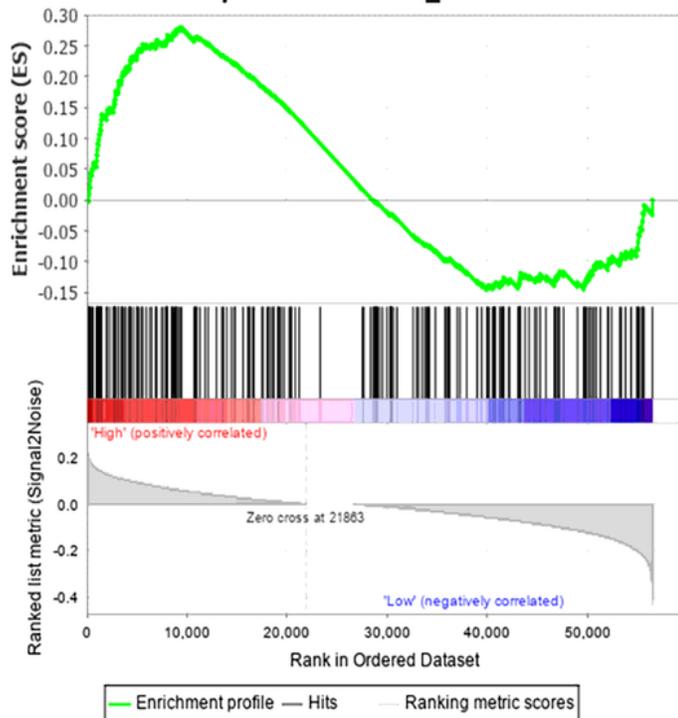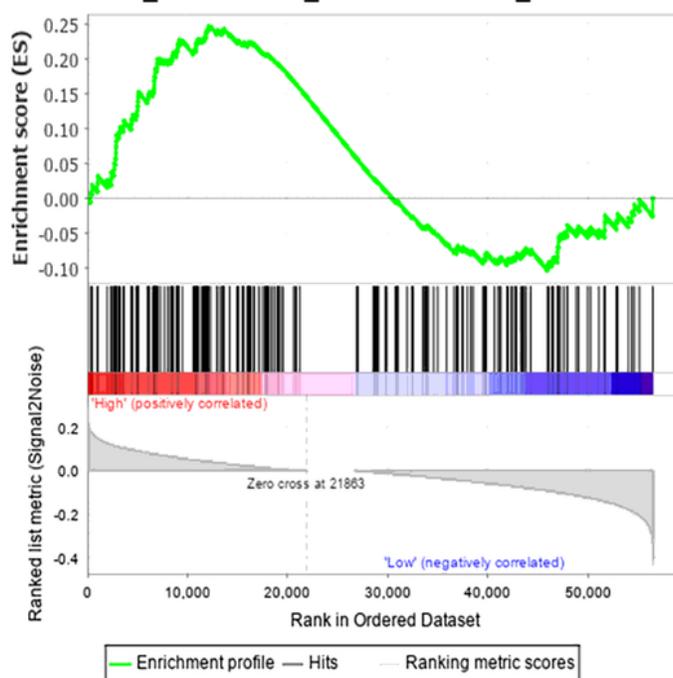
Figure 11

Gene Set Enrichment Analysis for the high-risk group with risk score.
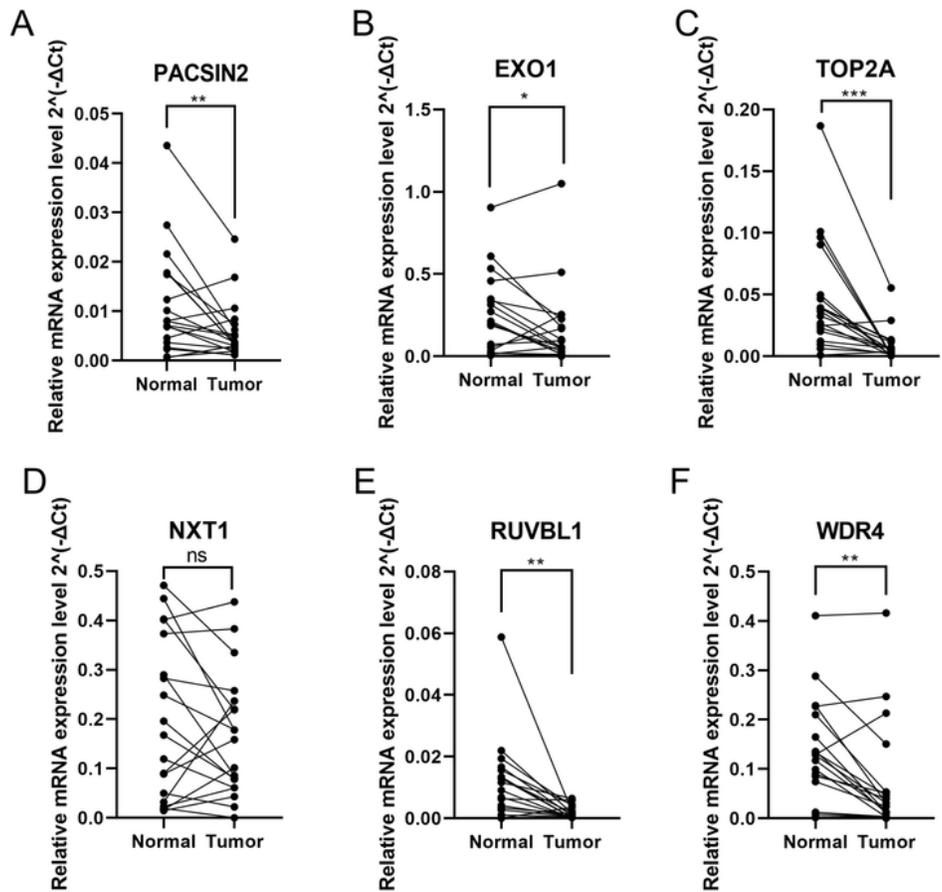
Figure 12

Verify the expression levels of 5 transcription factor genes in tumors and normal tissues. (A)PACSIN2, (B)EXO1,(C)TOP2A,(D)NXT1,(E)RUVBL1.(F)WDR4