Am J Transl Res 2020;12(10):6775-6792
www.ajtr.org /ISSN:1943-8141/AJTR0115329

Original Article

Development of RNA binding proteins expression signature for prognosis prediction in gastric cancer patients

Liqiang Zhou*, You Wu*, Lin Xin, Qi Zhou, Shihao Li, Yiwu Yuan, Jinliang Wang, Dengzhong Wu

*Department of General Surgery, The Second Affiliated Hospital of Nanchang University, Nanchang 330006, Jiangxi, China. *Equal contributors.*

Received May 30, 2020; Accepted September 22, 2020; Epub October 15, 2020; Published October 30, 2020

Abstract: It was reported that the expression of RNA binding proteins (RBPs) in malignant tumors is dysregulated and is closely related to tumorigenesis. However, some studies have confirmed the role of RBPs in gastric cancer (GC). We obtained data on gastric cancer in The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx), and identified RBPs that are dysregulated between gastric normal and cancer tissues. Then, we systematically investigated the expression characteristics and clinical prognostic potential of these RBPs through bioinformatics methods. We found 278 dysregulated RBPs in the GC, 91 of which were up-regulated and 181 were down-regulated. We detected 4 hub RBPs (HNRNPL, PABPN1, PCF, SNRPN) are related to overall survival (OS), and 3 hub RBPs (EEF1A2, MRPS5, PCF1) are related to disease-specific survival (DSS), and furthermore, we constructed prognostic signatures. Analysis of the OS and DSS signature showed that the GC patients with high-risk groups have worse OS and DSS than the low-risk groups. The receiver operator characteristic (ROC) curves of the 5-year survival rate of OS and DSS prognosis signature were drawn, and the areas under the two curves were 0.62 and 0.64, respectively. We constructed nomograms to predict OS and DSS, and evaluated by the calibration curve, which showed the GC prediction ability of these two models. Furthermore, the expression of the above six genes was verified by PCR, which is consistent with our results.

Keywords: Gastric cancer, RNA binding proteins, prognostic signature, bioinformatics

Introduction

Gastric cancer (GC) is a heterogeneous gastrointestinal disease. According to global cancer statistics, new cases of GC and death pathology account for approximately 5.7% and 8.2% of all tumors [1]. Among them, the incidence and mortality of gastric cancer in East Asia are prominent higher than other regions. In China, GC is a high-risk tumor disease which incidence is second only to lung cancer, and the mortality rate is lower than that of lung cancer and liver cancer [2]. The majority of GC patients are lack of screening and have no obvious symptom at the early stage [1, 3]. In addition, high postoperative recurrence rate, chemotherapy toxicity, drug resistance, and poor prognosis significantly reduced survival ratio, especially in patients with advanced GC, whose five-year survival rate does not exceed 30% [4]. Therefore, the development of cancer molecular biomarkers for diagnosis, prediction and treatment of GC is essential to improve quality of life and five-year survival rate of patients.

RNA binding proteins (RBPs) accurately recognize RNA binding domains, interact with multiple RNAs and participate in RNA cleavage, transport, sequence editing, intracellular localization and translation control regulate the process [5, 6]. Therefore, RBPs regulate cell functions changes or disturbances, which may lead to disease. Until now, there are over 1,500 RBPs genes have been screened and identified through RNA-seq screening technologies in cancer cells [6]. Due to different affinity or concentration, distinctions in RBPs expression can lead to erroneous interactions with target RNA, thereby forming erroneous RBPs complexes. Such RBPs can affect every post-transcriptional

event in the affected cells and regulate the cell phenotype to a pathological state. Multiple researches indicated that dysregulation of RBP expression has been observed in various human diseases including cancer [7-9]. RBPs made great contribution in the origin and progression of cancer [10, 11]. They significantly impacted the growth and proliferation of tumor cells, avoided immune surveillance, induced angiogenesis and activated metastasis [12]. Although the function of RBPs in other tumors has been thoroughly studied, the mechanism of RBPs affect the progress in gastric cancer has not been eliminated.

In the last few decades, several studies have reported that the RBPs in GC are dysregulated and affect the expression of related proteins, which related to tumor progression [13, 14]. For instance, NF90 protein enhances its stability by directly binding to TMEM98 mRNA and promotes GC progression [15]. PTBP3 mediates the variable shear of CAV1 to affect the invasion and migration of GC [16]. In addition, the DDX6 protein acts as an RNA binding protein of FGFR2 and FGFR2 mRNA, and actively regulates the expression of HER2 and FGFR2 in GC cells in the post-transcription stage [17]. These evidence shows that RBPs play a crucial role in GC, which helps us to initially understand its function in GC. Therefore, we integrated gastric cancer and normal gastric tissue sequencing and clinical pathology data from The Cancer Genome Atlas (TCGA) database and the Genotype-Tissue Expression (GTEx) database, and analyze the abnormal expression of RBPs between tumor and normal samples through RNA-seq bioinformatics. After a systematic analysis, we identified a group of RBPs that are dysregulated in stomach adenocarcinoma (STAD) to investigate their molecular biological functions and potential mechanisms that affect STAD. Some of these RBPs may serve as potential biomarkers for STAD for accurate diagnosis and prognosis.

Materials and methods

Datasets and data processing

In this study, we downloaded the STAD transcriptome fragments per Kilobase million (FPKM) data from TCGA (https://tcga-data.nci.nih.gov/tcga/), which contained 375 STAD tissues and 32 normal gastric tissues [18]. In order to increase the normal sample size and strengthen comparative analysis [19], we also downloaded human normal tissue transcripts per kilobase million (TPM) data from GTEx (https://www.gtexportal.org/) and extracted 359 normal stomach tissues [20]. We searched the University California Santa Cruz Xena (UCSC Xena, https://xena.ucsc.edu/) to identify the clinical data, including OS and DSS. Extract 2005 RNA-binding proteins associated with tumors from RBPTD (http://www.rbptd.com/) [21]. For TCGA-STAD data, in order to eliminate the error caused by the quantitative mRNA abundance of FPKM in multiple samples, we convert FPKM to TPM for standardization [22]. After converting the TPM values of TCGA and GTEx to log (TPM + 1), use the combat function of "sva" R package to combine to normalization and remove batch effect [23]. The *P*-values of DEGs between the STAD sample of TCGA combined with GTEx data and the normal sample was analyzed using wilcox test by the "limma" R package [24]. The cut-off threshold in TCGA combined with GTEx was $|\log_2$ fold change (FC)$| \geq 1.0$ and false discovery rate (FDR) < 0.05 to Identify RBPs that are differentially expressed in GC.

Gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis

GO define and describe the functions of genes and proteins, and update the semantic vocabulary standards as research continues. It covers three aspects of biology: cell components (CC), molecular functions (MF), and biological processes (BP). KEGG helps to study genes and expression information as a whole network, and systematically analyzes the metabolic pathways of gene products in cells and databases of the functions of these gene products. The online website DAVID 6.8 (https://david.ncifcrf.gov/) was used for enrichment analysis [25]. The enrichment results with *P*-value and FDR value below 0.05 are considered meaningful. The visual GO and KEGG enrichment results were performed by "GOplot" R packages [26].

Construction and analysis of protein-protein interaction (PPI) network

STRING is a protein interaction database, which used for the interaction between known proteins and predicted proteins [27]. We intro-

duced DEGs into STRING to construct the gene interaction network for discover the core genes. Then, we use Cytoscape software (version: 3.6.1, http://www.cytoscape.org/), the biograph visualization software to build a comprehensive model of biomolecular interactions [28]. Cytoscape's pluggable unit molecular complex detection (MCODE) is a module for screening PPI networks [29]. Take the cut-off degree = 2, cut-off point = 0.2, k-core = 2, max depth = 100 as the network score and cluster search parameter settings, then analyze the key modules.

Prognostic model construction

We used the Survival R software package to analyze Univariate and multivariate Cox regression, and combined them to trace the OS and DSS correlation of all hub RBPs in key modules. Then, according to the results of multivariate Cox regression analysis, we constructed risk proportional models related to OS and DSS respectively. Subsequently, we calculated the risk score of each GC sample by the formula: *Risk score* = β1 × Exp1 + β2 × Exp2 + βi × Expi (β: coefficient value, EXP: gene expression level). Taking the median risk score as the boundary, the STAD patients were divided into low-risk group and high-risk group. Finally, we use the log-rank test to compare the OS differences between the two groups and draw the time-dependence receiver operating characteristic (ROC) curve by "SurvivalROC" R package to evaluate the prognostic performance of these two models on OS and DSS [30].

Nomogram construction and verification

We used the "rms" and "survival" packages in R software to respectively construct a nomogram composed of the hub RBPs to provide clinicians with a basis for judging the prognosis of gastric cancer patients. Subsequently, we drew a calibration curve to evaluate the accuracy of the nomogram to predict survival.

Validation of expression level and prognostic

We analyzed the relationship between the expression levels of OS-related and DSS-related RBPs in TCGA-STAD and used Kaplan-Meier to analyze the potential of two groups of prognostic model hub RBPs to predict OS and DSS. then, we verified the expression level of the protein of hub RBPs in the OS and DSS models through the Human Protein Atlas (HPA) database (http://www.proteinatlas.org/) [31]. To further validate our research, we implemented quantitative real-time PCR to detect the expression of RBPs that constitute the prognostic signature of OS and DSS. We collected 10 patients (n = 20) who suffered curative resection of gastric cancer in the Gastrointestinal Surgery Department General Surgery Department of the Second Affiliated Hospital of Nanchang University, and were diagnosed as gastric cancer by pathology. The study complied with the Declaration of Helsinki and was approved by the Ethics Committee of the Second Affiliated Hospital of Nanchang University. Each patient signed an informed consent form. Follow the instructions, we used TRIzol reagent (Invitrogen, Carlsbad CA, USA) to extract total RNA from frozen samples to determine the concentration and purity of RNA, and utilized RR047A kit (Takara, Japan) for reverse transcription, RR820A kit (Takara, Japan) for qRT-PCR, with β-actin as an internal control, the primer sequences are in Table S1. We calculated 2^(-ΔCt) represents the expression of each gene and used paired t test to analyze the expression difference between normal and gastric cancer tissues.

Results

Selection of dysregulated RBPs

The role and prognostic value of relational RBPs in STAD was analyzed through various statistical calculation methods. The design diagram was shown in **Figure 1**. Combined with the TCGA and GTEx databases, 375 gastric cancer samples and 391 non-tumor control samples were covered in this study. According to the information of RBPTD, 1936 RBPs were extracted for in-depth analysis. A total of 278 RBPs (Table S2) including 97 up-regulated and 181 down-regulated gene that met the exclusion criteria of differential expression (FDR < 0.05, $|\log_2 FC| \geq 1.0$). A heat map and volcano plot for visualization were shown in **Figure 2**.

GO and KEGG functional enrichment analysis

We divided the different expressed RBPs into two groups according to the GO function enrichment analysis for investigate the potential molecular function of them and then upload them to DAVID 6.8 respectively for enrichment

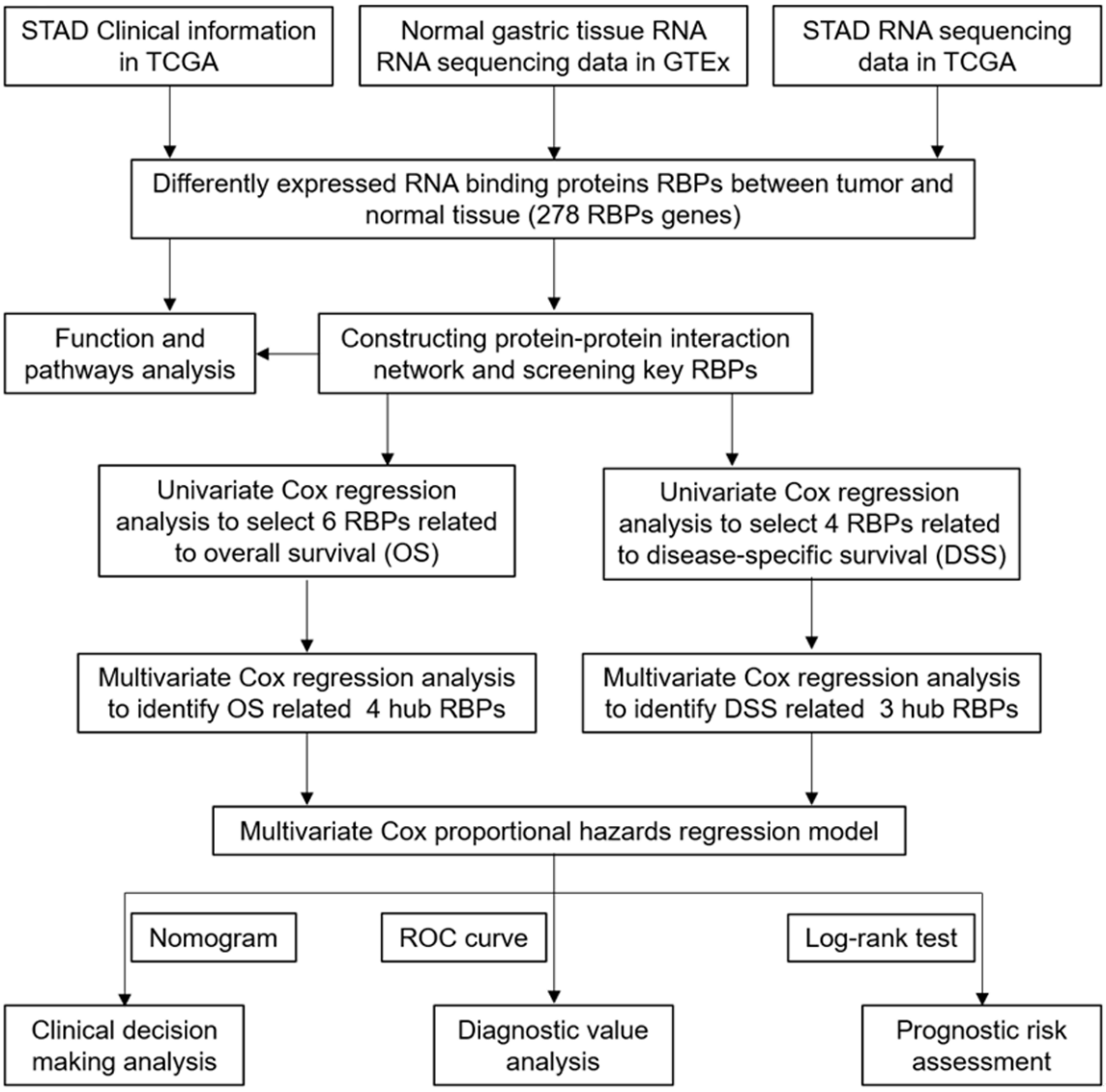


Figure 1. Whole procedures for analyzing RBPs in stomach adenocarcinoma.

analysis. The enrichment results showed that the up-regulated RBPs were Significantly enriched in five biological processes: response to virus, defense response to virus, negative regulation of viral genome replication, RNA splicing and type I interferon signaling pathway (**Table 1**). Down-regulated RBPs were significantly enriched translational initiation, mRNA processing, mRNA splicing via spliceosome, translation, RNA processing and others (**Tables 1** and S3). For molecular function, up-regulated RBPs were enriched in poly(A) RNA binding, RNA binding, 2'-5'-oligoadenylate synthetase activity, double-stranded RNA binding, nucleotide binding (**Table 1**). The down-regulated RBPs were notably enriched in 6 molecular function such as RNA binding, poly(A) RNA binding, nucleotide binding, structural constituent of ribosome, mRNA binding (**Tables 1** and S3). Cellular component analysis showed that Up-regulated RBPs were enriched in cytoplasm, nucleus, nucleoplasm, cytosol, nucleolus (**Table 1**). The down-regulated genes were significantly enriched in 10 cellular components including nuclear speck, nucleoplasm, ribosome, nucleus, and cytoplasm. analysis showed that Up-regulated RBPs were enriched in cytoplasm, nucleus, nucleoplasm, cytosol, nucleolus (**Tables 1** and S3). We conducted pathway enrichment analysis on all identified differen-

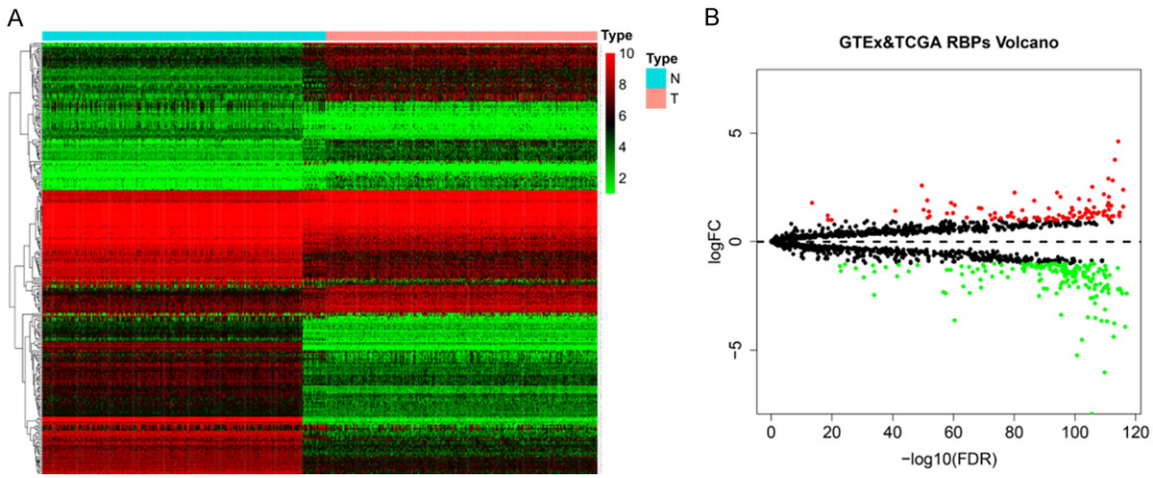


Figure 2. The differentially expressed in stomach adenocarcinoma. A. Heat map. B. Volcano plot.

Table 1. GO and KEGG enrichment analysis result of for differentially expressed RBPs

Expression	Enrichment term	*P* Value	FDR
Up-regulated RBPs GO			
Biological processes	response to virus	5.92E-09	8.80E-06
	defense response to virus	2.02E-07	3.00E-04
	negative regulation of viral genome replication	2.01E-06	2.98E-03
	RNA splicing	2.53E-06	3.76E-03
	type I interferon signaling pathway	2.11E-05	3.13E-02
Molecular function	poly(A) RNA binding	1.88E-23	2.37E-20
	RNA binding	4.66E-09	5.88E-06
	2'-5'-oligoadenylate synthetase activity	6.04E-07	7.62E-04
	double-stranded RNA binding	9.09E-07	1.15E-03
	nucleotide binding	1.78E-05	2.25E-02
Cellular component	cytoplasm	4.17E-09	5.07E-06
	nucleus	4.69E-09	5.70E-06
	nucleoplasm	3.06E-08	3.72E-05
	cytosol	6.98E-08	8.49E-05
	nucleolus	1.37E-07	1.67E-04
Down-regulated RBPs GO			
Biological processes	translational initiation	6.11E-17	1.67E-13
	mRNA processing	6.56E-16	9.99E-13
	mRNA splicing, via spliceosome	4.56E-14	6.82E-11
	translation	5.26E-12	7.88E-09
	RNA processing	5.23E-13	7.83E-10
Molecular function	RNA binding	9.73E-34	1.25E-30
	poly(A) RNA binding	6.65E-32	8.54E-29
	nucleotide binding	5.33E-18	6.85E-15
	structural constituent of ribosome	6.10E-11	7.83E-08
	mRNA binding	3.95E-08	5.07E-05
Cellular component	nuclear speck	2.56E-12	3.17E-09
	nucleoplasm	2.08E-11	2.57E-08
	ribosome	2.47E-10	3.06E-07
	nucleus	1.23E-08	1.52E-05
	cytoplasm	1.86E-07	2.30E-04

Dysregulated RBPS (KEGG)		
Ribosome	1.20E-17	1.51E-15
RNA transport	9.46E-10	5.96E-08
Spliceosome	5.83E-08	2.45E-06
mRNA surveillance pathway	1.23E-07	3.87E-06
RNA degradation	3.62E-05	9.11E-04
Ribosome biogenesis in eukaryotes	1.37E-04	2.88E-03
Aminoacyl-tRNA biosynthesis	1.66E-04	2.99E-03
Influenza A	5.53E-04	8.70E-03

tially expressed RBPs. The results demonstrated that the RBPs were enriched in seven pathways: Ribosome, RNA transport, Spliceosome, mRNA surveillance pathway, RNA degradation, Ribosome biogenesis in eukaryotes, Aminoacyl-tRNA biosynthesis, and Influenza (**Table 1**).

PPI network construction and key modules screening

We further study the role of differential expression RBPs in GC by constructing a PPI network. We uploaded 278 RBPs to the String database, set the minimum required interaction score to 0.40, and obtained a PPI network with 197 nodes and 1484 edges (**Figure 3A**). We used the MODE tool to process PPI networks, identified the key networks and got a network composed of 54 nodes and 721 edges (**Figure 3B**). GO Enrichment analysis showed that the functions of these 54 genes are mainly enriched in poly(A) RNA binding, structural constituent of ribosome, nucleotide binding, RNA binding. In addition, biological processes are enriched in translation, mRNA splicing through spliceosomes, cytoplasmic translation (**Figure 4A**). KEGG pathway enrichment analysis indicated that the RBPs pathway in the core module is enriched in Ribosome, Spliceosome, mRNA surveillance pathway, RNA transport, Legionellosis, Systemic lupus erythematosus (**Figure 4B**).

Identification of RBPs related to prognosis

The core network extracted from the previous step has 54 RBPs. Using univariate Cox regression analysis, we obtained six RBPs related to the overall survival (OS), include: HNRNPL, PABPN1, PCF11, SF3B2, SNRPN, and UPF3A (**Table 2**), Thereinto, HNRNPL, PABPN1, PCF, SNRPN have the greatest potential prognostic ability (**Table 2**). The univariate Cox regression analysis of the RBPs related to DSS revealed that EEF1A2, HNRNPL, MRPS5, PCF11 were correlated with DSS (**Table 3**), and EEF1A2, MRPS5, PCF11 were independent prognostic factors (**Table 3**).

Prognosis model construction and analysis

On account of multivariate Cox regression analysis, four OS-related RBPs were used to construct a prognostic model to predict the OS of GC patients. The risk score of each GC patient is calculated based on the following formula: *Risk score* = (-2.4730 * ExpHNRNPL) + (-1.0385 * ExpPABPN1) + (-1.6102 * ExpPCF11) + (0.3265 * ExpSNRPN).

In view of the risk scores, we divided 370 STAD patients into a high-risk group and a low-risk group and evaluated the predictive ability of the model through the survival analysis. These results verified that the overall survival rate of the high-risk group was lower than that of the low-risk group (**Figure 5A**). We further conducted ROC analysis to evaluate the prognostic performance of the OS-related prognostic models constructed by the four RBPs and analyzed the ROC curve of the five-year survival rate. The results showed that the 5-year area under the ROC curve (RUC) of the OS-related model was 0.62, proved that it had a certain predictive ability (**Figure 5B**). We also plotted the overall survival status, and risk scores of the four hub RBPs expression in the low-risk and high-risk groups, as shown in **Figure 5C**.

Besides, we draw a DSS prediction model. The risk scoring formula for each patient is as follows: *Risk score* = (0.1942 * ExpEEF1A2) + (-2.4877 * ExpMRPS5) + (-2.6142 * ExpPCF11).

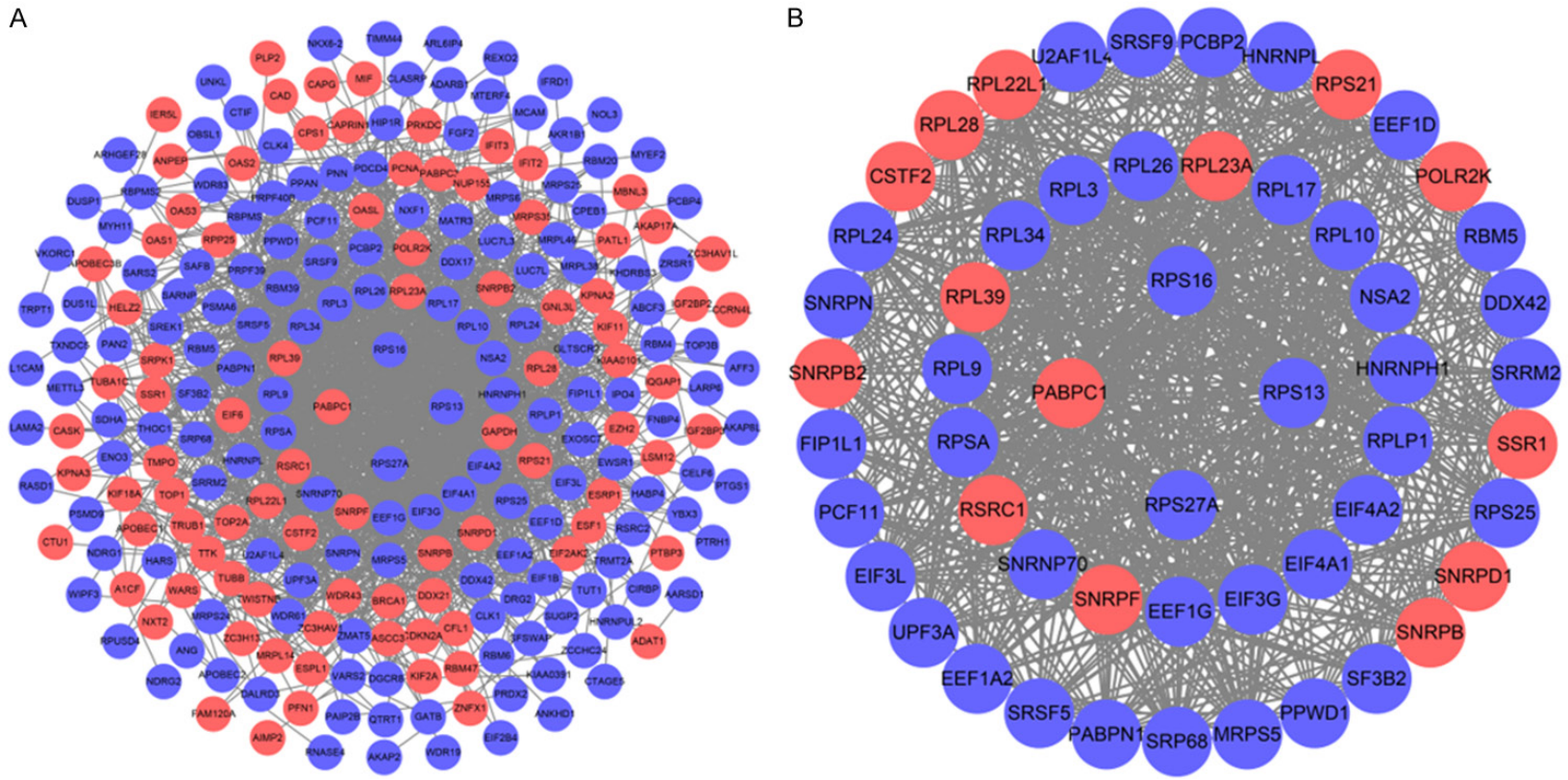
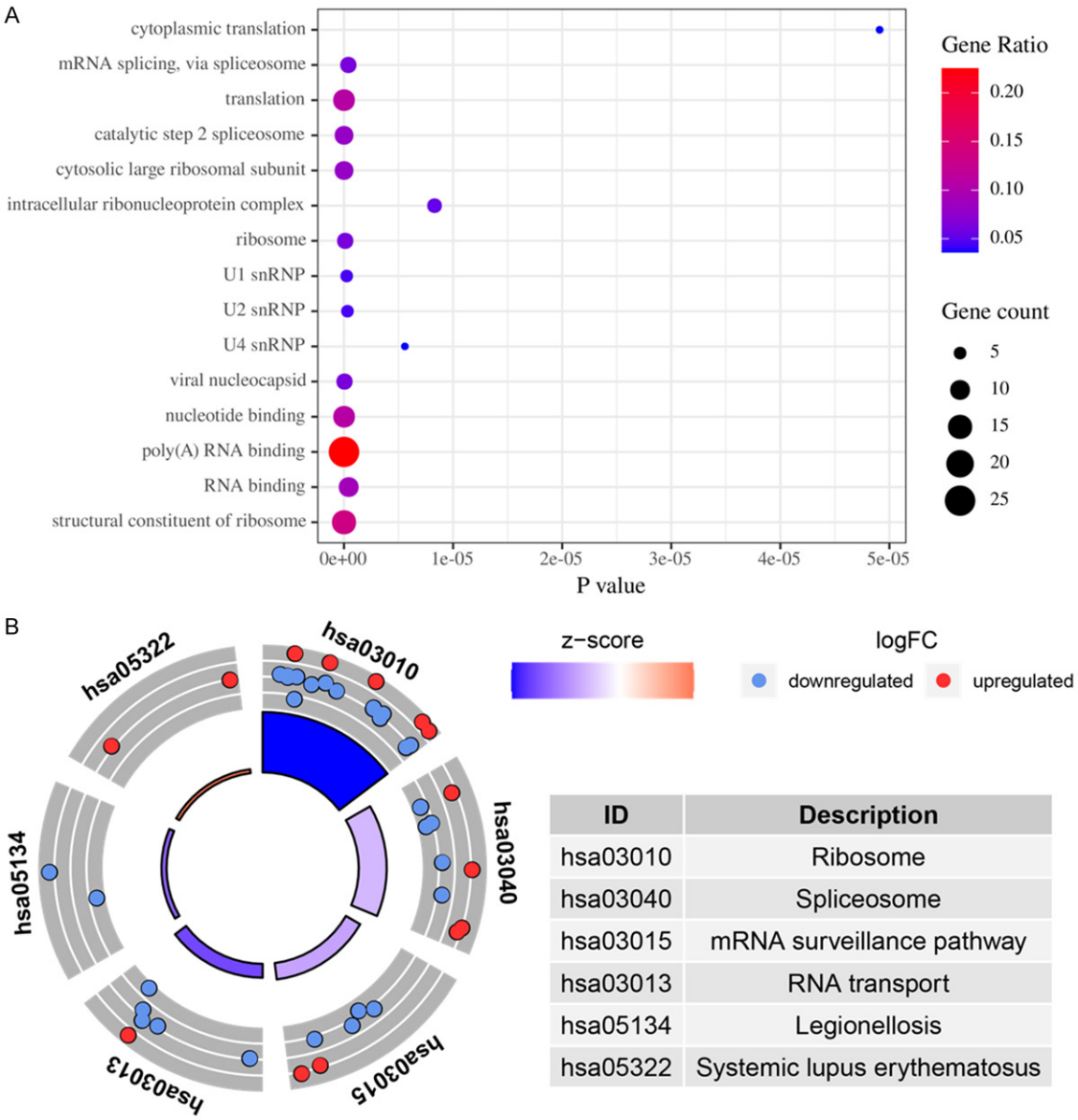


Figure 3. Protein-protein interaction network and key modules analysis. A. Protein-protein interaction network of differentially expressed RBPs. B. Key module from PPI network. Blue circles: down-regulation, red circles: up-regulation.



ID	Description
hsa03010	Ribosome
hsa03040	Spliceosome
hsa03015	mRNA surveillance pathway
hsa03013	RNA transport
hsa05134	Legionellosis
hsa05322	Systemic lupus erythematosus

Figure 4. GO and KEGG pathway enrichment analysis of RBPs in key modules of PPI networks. A. GO. B. KEGG.

We divided the patients into the high-risk and low risk groups through the risk score and conducted a survival analysis. The analysis illustrated that the DSS of the high-risk group is significantly lower than that of the low-risk group (**Figure 6A**). By plotting a 5-year ROC curve, the AUC is 0.64, which can predict the DSS of patient (**Figure 6B**). The disease-specific survival status, and the risk scores of three hub RBPs in the signature were displayed in **Figure 6C**. We stated the significance of multiple clinical features and overall prognosis in TCGA-STAD by univariate Cox regression analysis. The analysis explained that risk score, age, stage N, stage T, and tumor stage were considered as clinicopathological features related to OS in STAD. Further multivariate Cox regression analysis emphasized that risk score and age were independent prognostic factors related to OS (**Figure 5D** and **5E**). For DSS, based on the results of univariate Cox regression analysis, risk score, gender, stage N and tumor stage are related to DSS. Through multivariate Cox regression analysis, only risk score is an inde-

Table 2. Univariate Cox regression analysis and multivariate Cox regression analysis to identify OS-related hub RBPs

Gene ID	Univariate Cox regression analysis			Multivariate Cox regression analysis			
	HR	95% CI	*P*-value	coef	HR	95% CI	*P*-value
HNRNPL	0.0279	0.0019-0.4029	0.0086	-2.4730	0.0843	0.0054-1.3234	0.0783
PABPN1	0.2576	0.0766-0.8665	0.0284	-1.0384	0.354	0.1070-1.1714	0.0890
PCF11	0.1726	0.0476-0.6253	0.0075	-1.6101	0.1999	0.0552-0.7237	0.0142
SNRPN	1.5384	1.0082-2.3475	0.0457	0.3265	1.3862	0.8835-2.1747	0.1553
SF3B2	0.0719	0.0068-0.7620	0.0288	-	-	-	-
UPF3A	0.3779	0.1476-0.9673	0.0424	-	-	-	-

Table 3. Univariate Cox regression analysis and multivariate Cox regression analysis to identify DSS-related hub RBPs

Gene ID	Univariate Cox regression analysis			Multivariate Cox regression analysis			
	HR	95% CI	*P*-value	coef	HR	95% CI	*P*-value
EEF1A2	1.3276	1.0698-1.6474	0.0100	0.1942	1.2143	0.9682-1.5230	0.0928
MRPS5	0.1879	0.0427-0.8275	0.0271	-2.4877	0.0831	0.0157-0.4399	0.0034
PCF11	0.1163	0.0227-0.596	0.0099	-2.6142	0.0732	0.0128-0.4192	0.0033
HNRNPL	0.0270	0.0009-0.8094	0.0374	-	-	-	-

pendent prognostic factor for DSS in patients with GC (**Figure 6D** and **6E**).

Construction of nomogram according to hub RBPs

In order to make the risk scoring model predict the prognosis more accurately, we constructed four RBP signatures related to OS (**Figure 7A**) and three RBP signatures related to DSS (**Figure 7C**) based on multivariate Cox regression analysis Nomogram. Then we allocated the points in the nomogram to the RBPs in proportion to each point, and draw a horizontal line that determines each RBP point and normalized to a distribution of 0 to 100. We calculated the total points of each patient by adding the points of each RBPs in the model and plotted the total point axis, which supported for us to calculate the OS and DSS of STAD patients from one to five years by drawing a vertical line between the total point axis and each prognostic axis. These studies may improve the accuracy of clinicians making clinical decisions for STAD patients. We divided the samples that constitute the OS and DSS-related nomograms into three equal parts, and drawn a 5-year calibration curve to judge the prognosis of the nomogram. The results suggest that whether it is OS nomogram or DSS nomogram (**Figure 7B** and **7D**), the predicted survival rate is almost the same as the actual survival rate, showing its good predictive ability.

Validation the expression of hub RBPs

The joint analysis of GTEx and TCGA 391 paracancerous and 375 cancer tissue data indicated that the expression of EEF1A2, HNRNPL, MRPS5, PABPN1, PCF11, SNRPN in tumor tissue was lower than that in paracancerous tissue (Figure S1). Using Kaplan-Meier to analyze the RBPs related to OS and DSS respectively, HNRNPL, PABPN1, PCF11 are the indicators to predict OS, and EEF1A2, MRPS5, PCF11 are the indicators to predict DSS (Figure S1). In order to further verify the protein expression of the seven hub RBPs related to OS and DSS in STAD, we used HPA immunohistochemistry data. Compared with normal gastric tissues, the expression of EEF1A2, HNRNPL, MRPS5, PABPN1, and SNRPN in gastric cancer was relatively reduced (**Figure 8A**). By qRT-PCR analysis of the expression level of RNA in clinical surgical samples of gastric cancer patients, we found that the expression levels of HNRNPL, PABPN1, PCF11, SNRPN, EEF1A2, MRPS5 in gastric cancer were significantly lower than those in adjacent tissues (**Figure 8B**), the result

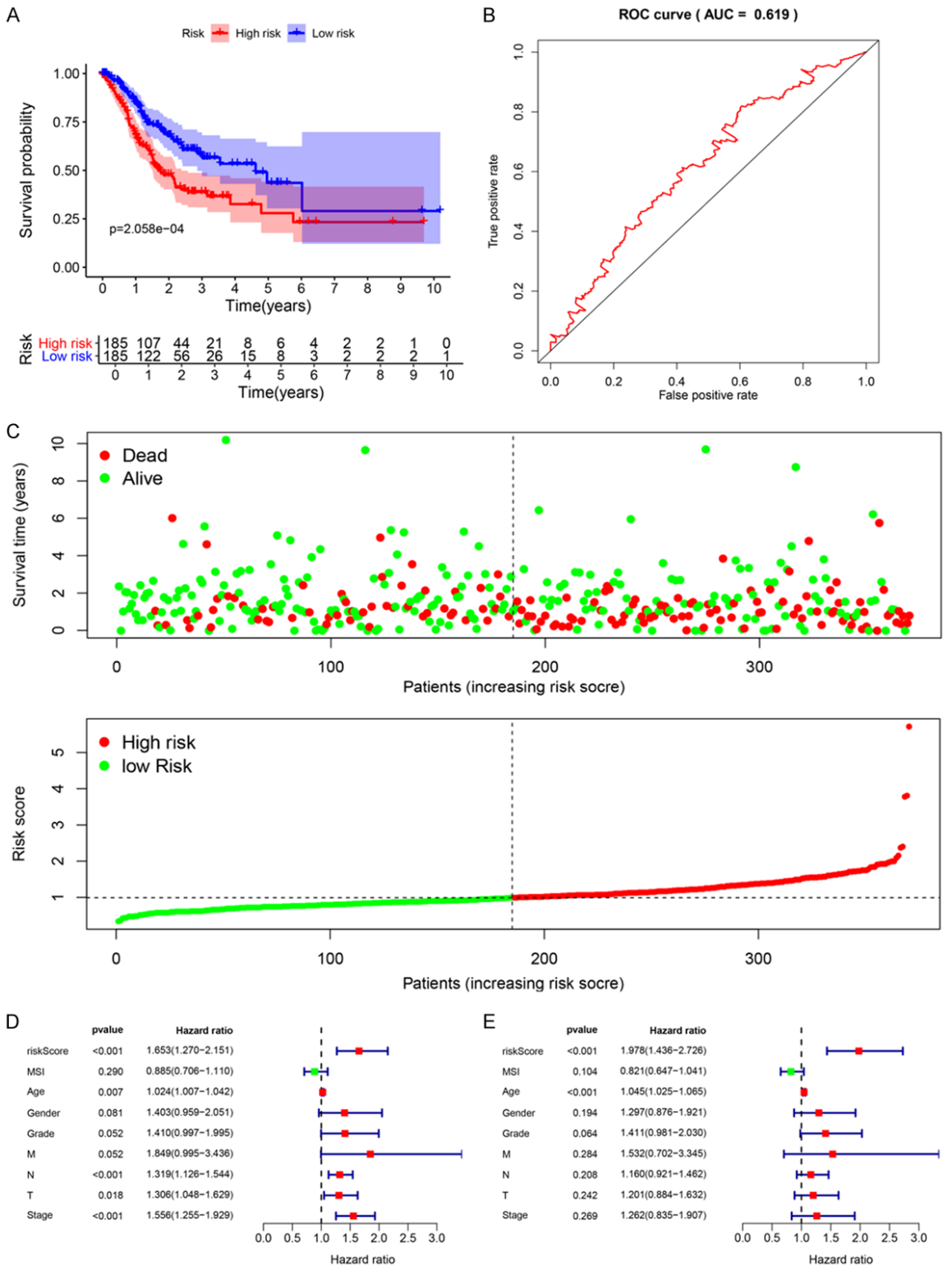


Figure 5. Risk score analysis of overall survival related prognostic models and the clinicopathological prognostic value of OS-related prognostic model in STAD. A. Survival curve for low-risk and high-risk subgroups. B. ROC curve for predicting OS based on risk score. C. Expression heat map, risk score distribution, and survival status. D. Univariate Cox regression analysis. E. Multivariate Cox regression analysis.

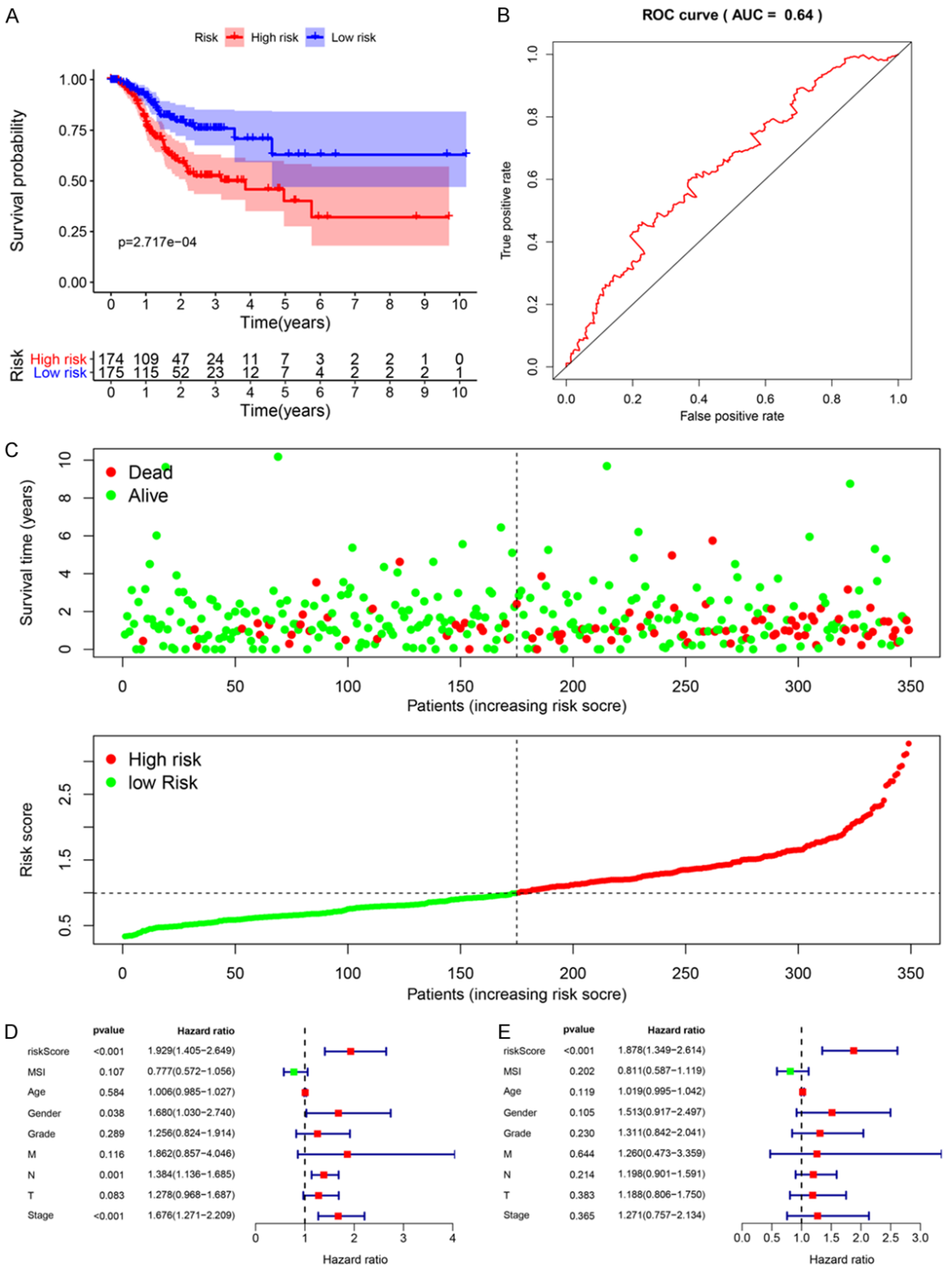


Figure 6. Risk score analysis of disease-specific survival related prognostic models and the clinicopathological prognostic value of DSS-related prognostic model in STAD. A. Survival curve for low-risk and high-risk subgroups. B. ROC curve for predicting DSS based on risk score. C. Expression heat map, risk score distribution, and survival status. D. Univariate Cox regression analysis. E. Multivariate Cox regression analysis.

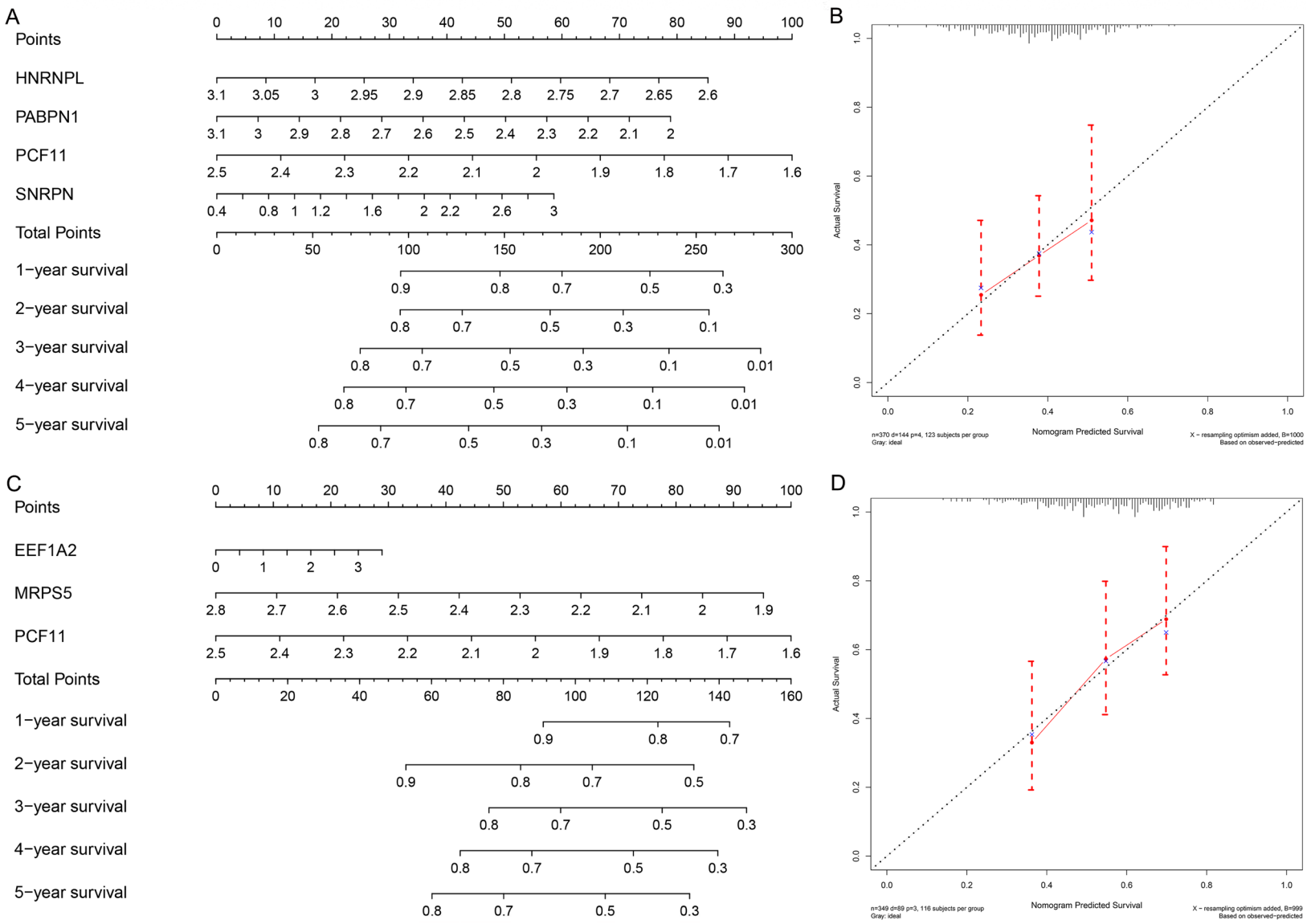


Figure 7. Nomogram for predicting 1-5 year OS and DSS of STAD patients and validation. A. OS Nomogram. B. Calibration curves for the OS nomogram. C. DSS Nomogram. D. Calibration curves for the OS nomogram.

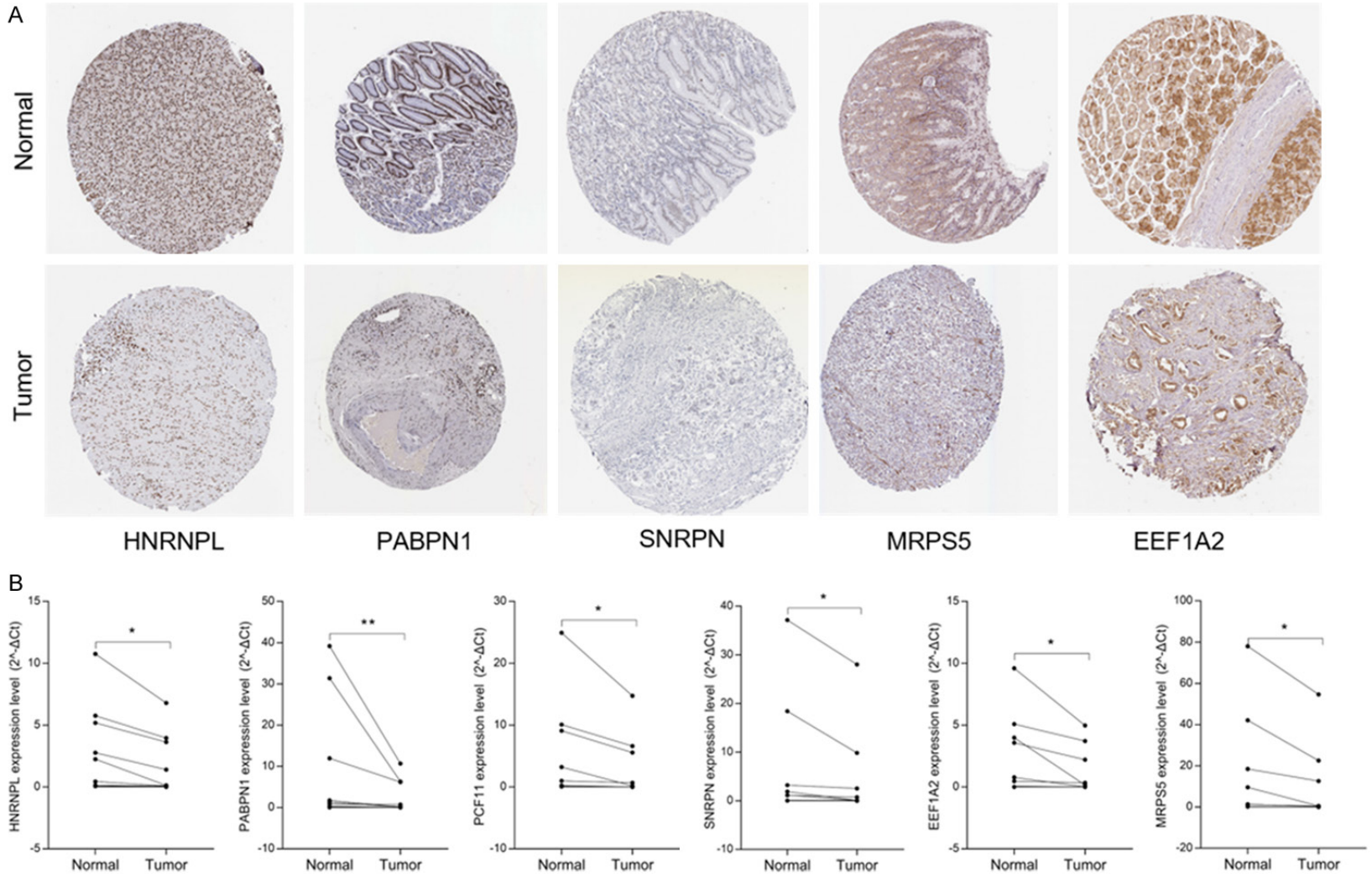


Figure 8. Verification of hub RBPs expression. A. HPA database. B. qRT-PCR.

is in line with our bioinformatics analysis results.

Discussion

Compared with global tumor data, Chinese digestive system tumors still account for a large proportion. The incidence of GC is gradually increasing, and currently ranks second after lung cancer [2]. The main feature of malignant tumors is uncontrolled cell growth, which is due to abnormal expression of oncogenes that regulate cell proliferation and differentiation. Multiple studies in recent years have confirmed that RBPs expression is dysregulated in GC [13, 14]. However, only a small part of the research on the regulatory mechanism of RBPs in GC which makes the expression pattern and role of RBPs in STAD poorly understood. In this study, we integrated RNA-seq data of GTEx and TCGA gastric and GC tissues, and identified 278 dysregulated RBPs. Through systematic analysis of related molecular functions and biological pathways, the PPI network of these RBPs were constructed. Moreover, we used univariate and multivariate Cox regression analysis to determine hub RBPs, and separately constructed OS-related risk prediction signatures composed of four hub RBPs, and based on three hub RBPs to predict the risk model of DSS in STAD patients. These findings can help clinical practitioners develop novel biomarkers to more accurately predict and diagnose STAD patients.

The biological functions and molecular biological mechanism of 278 dysregulated RBPs were obtained by GO and KEGG function enrichment analysis. Firstly, for biological process, differentially expressed RBPs are mainly enriched in translation, RNA processing, splicing, decomposition, and response to viruses. The occurrence and development of GC involve multiple genes and multiple pathways in harmony. The regulation of RNA metabolism, RNA processing, and translation has been shown to participate in the occurrence and progression of various human diseases and play an important role [32-34]. The regulation of RNA stability post-transcription is considered to be an important process in gene expression. RBPs are proteins that enhance the stability of target mRNA and promote gene expression. This ability is mainly achieved by interacting with RNA to form a ribosomal protein complex. This function of RBPs plays an important role in the progression of various diseases. The stability of NF90 protein is enhanced by directly binding to TMEM98 mRNA and promotes GC progression [15]. DDX6 protein acts as an RNA binding protein for FGFR2 mRNA, and actively regulates the expression of HER2 and FGFR2 in GC cells in the post-transcription step [17]. HNRNPR promote the proliferation and metastasis of GC by stabilizing the expression of CCNB1 and CENPF mRNA [35]. The potential link between Epstein-Barr virus (EBV) and GC has passed for thirty years. GC associated with EBV is a specific subtype of GC [36]. Multiple reports indicated that RBPs regulates EBV mRNA transcription [37, 38]. But the relationship between RBPs, EBV and GC is not clear. In terms of molecular function, RBPs can combine with RNA and nucleotides including poly(A) RNA, mRNA, double-stranded RNA to perform its function. In addition, it has structural constituent of ribosome and 2'-5'-oligoadenylate synthetase activity. The ribosome is a key organelle for protein synthesis. Mutations in ribosomal proteins regulate p53 translation and activity, ultimately leading to disease and cancer [39]. The 2'-5'oligoadenylate synthase is closely related to the viral response, which may be related to EBV-related GC. KEGG pathway analysis showed that they are mainly enriched in Ribosome, RNA transport, spliceosome, RNA degradation, mRNA surveillance pathway, ribosome biogenesis in eukaryotes, aminoacyl-tRNA biosynthesis and other pathways to affect the development of GC.

We constructed a PPI network of 278 dysregulated RBPs and obtained a key module containing 54 hub RBPs. Despite the connection between these differentially expressed RBP and GC is uncertain, it is reported that some RBPs are related to other malignant tumors. For example, KIF11 can enhance the characteristics of breast cancer stem cells by activating the Wnt/β-catenin signaling pathway to facilitate the self-renewal of breast cancer cells [44]. KPNA2 accelerates metabolic reprogramming of glioblastoma by regulating c-myc signaling axis [45] and promote the proliferation and tumorigenicity of epithelial ovarian cancer through the c-Myc pathway and FOXO3a [46]. EZH2 silently recruits the non-coding RNA PHACTR2-AS1 of histone methyltransferase SUV39H1, resulting in excessive activation of

ribosomal synthesis and instability of ribosomal DNA, thereby promoting the proliferation and metastasis of cancer cells [47]. The hub RBP in key models also plays an essential role in tumors. RPL22L1 and RPS21 have been identified as candidate biomarkers for diagnosis and prognosis of prostate cancer [40]. Colorectal cancer was confirmed that RPL22L1 is associated with poor prognosis and induces 5-FU resistance [41]. In ovarian cancer, RPL22L1 can trigger epithelial-mesenchymal transition and promote tumor metastasis [42]. Studies in non-small cell carcinoma have shown that the core component of spliceosome SNRPB can formally regulate RAB26 alternative splicing and mRNA expression through the pathway by promoting tumor progression [43]. PPI network key module analysis showed that STAD is related to intracellular ribonucleoprotein complex, poly(A) RNA binding, RNA binding, structural constituent of ribosome, nucleotide binding, translation, MRNA splicing through spliceosomes, cytoplasmic translation.

We determined the center point RBP based on the results of univariate Cox regression analysis combined with the results of multivariate Cox regression analysis. Four OS-related hub RBPs (HNRNPL, PABPN1, PCF11, and SNRPN) and three DSS-related (ESS1A2, MRPS5, and PCF11) were identified. Although these hub RBPs have not been reported in GC, studies have shown that they are associated with other tumors. PABPN1 is an inhibitor of Alternative Polyadenylation, and its down-regulation in lung cancer promotes tumor invasion by releasing cancer cells from microRNA-mediated gene regulation [48]. SNRPN is a key component of the spliceosome and considered to be a crucial factor for tumor growth in pancreatic cancer cells, promoting tumor proliferation [49]. Translation elongation factor EEF1A2 inhibits the activity of RNA-dependent protein kinase PKR and promotes tumor cell survival [50]. These evidences are consistent with the results of our analysis. ROC curve analysis showed that the two risk models have moderate diagnostic ability and can predict the prognosis of STAD patients. Furthermore, we established nomograms to predict OS and DSS separately to help clinicians more intuitively predict operating systems from one to five years the 5-year calibration curve shows a good predictive power of these two nomograms. Through the qRT-PCR and HPA database, we verified the expression of RNA and protein levels of these six genes, which is consistent with the results of our GTEx and TCGA analysis, suggesting that the OS and DSS prognostic models we constructed have potential value in adjusting the treatment plan of GC patients.

Overall, we have constructed a prediction model based on four genes for OS and a prediction model based on three genes for DSS, which has favourable prediction performance, which helps to develop new STAD prognostic indicators. The gene signatures associated with RBPs revealed vital biological functions, which indicated that they can potentially be used for clinical adjuvant therapy. However, our research has some deficiencies. First of all, the GC RNA-seq data of our prediction model only comes from the TCGA database and has not been verified in other databases and clinical cohorts. Secondly, the lack of certain clinical information in the TCGA dataset may lead to the reliability of multivariate Cox regression analysis. Finally, this study is based on retrospective analysis, and prospective studies should be conducted to further verify the results.

Acknowledgements

National Natural Science Foundation of China, Grant/Award Numbers: 81760549, 81872480; Science and Technology Research Project of Education Department of Jiangxi Province, Grant/Award Number: GJJ180024; Nanchang University Graduate Innovation Fund, Grant/Award Number: CX2018196.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Lin Xin, Department of General Surgery, The Second Affiliated Hospital of Nanchang University, Nanchang 330006, Jiangxi, China. Tel: +86-0791-86298907; E-mail: xlyxbs@126.com

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185

countries. CA Cancer J Clin 2018; 68: 394-424.
[2] Chen W, Zhang R, Zhang S, Zhao P, Zeng H and Zou X. Report of cancer incidence and mortality in China, 2010. Ann Transl Med 2014; 2: 61.
[3] Van CE, Sagaert X, Topal B, Haustermans K and Prenen H. Gastric cancer. Lancet 2016; 388: 2654-2664.
[4] Sun P, Xiang JB and Chen ZY. Meta-analysis of adjuvant chemotherapy after radical surgery for advanced gastric cancer. Br J Surg 2009; 96: 26-33.
[5] Wang ZL, Li B, Luo YX, Lin Q, Liu SR, Zhang XQ, Zhou H, Yang JH and Qu LH. Comprehensive genomic characterization of RNA-binding proteins across human cancers. Cell Rep 2018; 22: 286-298.
[6] Gerstberger S, Hafner M and Tuschl T. A census of human RNA-binding proteins. Nat Rev Genet 2014; 15: 829-845.
[7] Lukong KE, Chang KW, Khandjian EW and Richards S. RNA-binding proteins in human genetic disease. Trends Genetics 2008; 24: 416-425.
[8] Neelamraju Y, Hashemikhabir S and Janga SC. The human RBPome: from genes and proteins to human disease. J Proteomics 2015; 127: 61-70.
[9] Darnell RB. RNA regulation in neurologic disease and cancer. Cancer Res Treat 2010; 42: 125-129.
[10] Jang HH, Lee HN, Kim SY, Hong S and Lee WS. Expression of RNA-binding Motif Protein 3 (RBM3) and Cold-inducible RNA-binding protein (CIRP) is associated with improved clinical outcome in patients with colon cancer. Anticancer Res 2017; 37: 1779-1785.
[11] Busà R, Paronetto MP, Farini D, Pierantozzi E, Botti F, Angelini DF, Attisani F, Vespasiani G and Sette C. The RNA-binding protein Sam68 contributes to proliferation and survival of human prostate cancer cells. Oncogene 2007; 26: 4372-4382.
[12] Pareira B, Billaud M and Almeida R. RNA-Binding proteins in cancer: old players and new actors. Trends Cancer 2017; 3: 506-528.
[13] Cheng Y, Jin Z, Agarwal R, Ma K, Yang J, Ibrahim S, Olaru AV, David S, Ashktorab H, Smoot DT, Duncan MD, Hutcheon DF, Abraham JM, Meltzer SJ and Mori Y. LARP7 is a potential tumor suppressor gene in gastric cancer. Lab Invest 2012; 92: 1013-1019.
[14] Bian Y, Wang L, Lu H, Yang G, Zhang Z, Fu H, Lu X, Wei M, Sun J, Zhao Q, Dong G and Lu Z. Downregulation of tumor suppressor QKI in gastric cancer and its implication in cancer prognosis. Biochem Biophys Res Commun 2012; 422: 187-193.
[15] Ao X, Li X, Chen Y, Zang Z, Guo W and Liang J. The TMEM98 mRNA promotes proliferation and invasion of gastric cells by directly interacting with the NF90 protein. Cell Biol Int 2020; 44: 1820-1830.
[16] Liang X, Chen W, Shi H, Gu X, Li Y, Qi Y, Xu K, Zhao A and Liu J. PTBP3 contributes to the metastasis of gastric cancer by mediating CAV1 alternative splicing. Cell Death Dis 2018; 9: 569.
[17] Tajirika T, Tokumaru Y, Taniguchi K, Sugito N, Matsuhashi N, Futamura M, Yanagigara K, Akao Y and Yoshida K. Dead-box protein RNA-helicase DDX6 regulates the expression of HER2 and FGFR2 at the post-transcriptional step in gastric cancer cells. Int J Mol Sci 2018; 19: 2005.
[18] Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C and Stuart JM. The cancer genome atlas pan-cancer analysis project. Nat Genet 2013; 45: 1113-1120.
[19] Wang Q, Armenia J, Zhang C, Penson AV, Reznik E, Zhang L, Minet T, Ochoa A, Gross BE, Iacobuzio-Donahue CA, Betel D, Taylor BS, Gao J and Schultz N. Unifying cancer and normal RNA sequencing data from different sources. Sci Data 2018; 5: 180061.
[20] GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015; 348: 648-660.
[21] Li K, Guo ZW, Zhai XM, Yang XX, Wu YS and Liu TC. RBPTD: a database of cancer-related RNA-binding proteins in humans. Database 2020; 2020: baz156.
[22] Li B, Ruotti V, Stewart RM, Thomson JA and Dewey CN. RNA-seq gene expression estimation with read mapping uncertainty. Bioinformatics 2009; 26: 493-500.
[23] Leek JT, Johnson WE, Parker HS, Jaffe AE and Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012; 28: 882-883.
[24] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43: e47.
[25] Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC and Lempicki RA. The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 2007; 8: R183.
[26] Wslter W, Sánchez-Cabo F and Ricote M. GOplot: an R package for visually combining ex-

pression data with functional analysis. Bioinformatics 2015; 31: 2912-2914.
[27] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ and von Mering C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011; 39: D561-568.
[28] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-2504.
[29] Bader GD and Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003; 4: 2.
[30] Heagerty PJ and Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics 2005; 61: 92-105.
[31] Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charloteaux B, Choi D, Coté AG, Daley M, Deimling S, Desbuleux A, Dricot A, Gebbia M, Hardy MF, Kishore N, Knapp JJ, Kovács IA, Lemmens I, Mee MW, Mellor JC, Pollis C, Pons C, Richardson AD, Schlabach S, Teeking B, Yadav A, Babor M, Balcha D, Basha O, Bowman-Colin C, Chin SF, Choi SG, Colabella C, Coppin G, D'Amata C, De Ridder D, De Rouck S, Duran-Frigola M, Ennajdaoui H, Goebels F, Goehring L, Gopal A, Haddad G, Hatchi E, Helmy M, Jacob Y, Kassa Y, Landini S, Li R, van Lieshout N, MacWilliams A, Markey D, Paulson JN, Rangarajan S, Rasla J, Rayhan A, Rolland T, San-Miguel A, Shen Y, Sheykhkarimli D, Sheynkman GM, Simonovsky E, Taşan M, Tejeda A, Tropepe V, Twizere JC, Wang Y, Weatheritt RJ, Weile J, Xia Y, Yang X, Yeger-Lotem E, Zhong Q, Aloy P, Bader GD, De Las Rivas J, Gaudet S, Hao T, Rak J, Tavernier J, Hill DE, Vidal M, Roth FP and Calderwood MA. A reference map of the human binary protein interactome. Nature 2020; 580: 402-408.
[32] Kim TH, Tsang B, Vernon RM, Sonenberg N, Kay LE and Forman-Kay JD. Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. Science 2019; 365: 825-829.
[33] Siang DTC, Lim YC, Kyaw AMM, Win KN, Chia SY, Degirmenci U, Hu X, Tan BC, Walet ACE, Sun L and Xu D. The RNA-binding protein HuR is a negative regulator in adipogenesis. Nat Commun 2020; 11: 213.
[34] Jain A, Brown SZ, Thomsett HL, Londin E and Brody JR. Evaluation of post-transcriptional gene regulation in pancreatic cancer cells: studying RNA binding proteins and their mRNA targets. Methods Mol Biol 2019; 1882: 239-252.
[35] Chen EB, Qin X, Peng K, Li Q, Tang C, Wei YC, Yu S, Gan L and Liu TS. HnRNPR-CCNB1/CENPF axis contributes to gastric cancer proliferation and metastasis. Aging (Albany NY) 2019; 11: 7473-7491.
[36] Fukayama M, Abe H, Kunita A, Shinozaki-Ushiku A, Matsusaka K, Ushiku T and Kaneda A. Thirty years of Epstein-Barr virus-associated gastric carcinoma. Virchows Arch 2020; 476: 353-365.
[37] Lee N, Yario TA, Gao JS and Steitz JA. EBV noncoding RNA EBER2 interacts with host RNA-binding proteins to regulate viral gene expression. Proc Natl Acad Sci U S A 2016; 113: 3221-3226.
[38] Mure F, Panthu B, Zanella-Cléon I, Delolme F, Manet E, Ohlmann T and Gruffat H. Epstein-Barr Virus protein EB2 stimulates translation initiation of mRNAs through direct interactions with both poly(a)-binding protein and eukaryotic initiation factor 4g. J Virol 2018; 92: e01917-17.
[39] Goudarzi KM and Lindström MS. Role of ribosomal protein mutations in tumor development. Int J Oncol 2016; 48: 1313-1324.
[40] Liang Z, Mou Q, Pan Z, Zhang Q, Gao G, Cao Y, Gao Z, Pan Z and Feng W. Identification of candidate diagnostic and prognostic biomarkers for human prostate cancer: RPL22L1 and RPS21. Med Oncol 2019; 36: 56.
[41] Pei YY, Li GC, Ran J, Wan XH, Wei FX and Wang L. Kinesin family member 11 enhances the self-renewal ability of breast cancer cells by participating in the Wnt/β-catenin pathway. J Breast Cancer 2019; 22: 522-532.
[42] Li J, Liu Q, Liu Z, Xia Q, Zhang Z, Zhang R, Gao T, Gu G, Wang Y, Wang D, Chen X, Yang Y, He D and Xin T. KPNA2 promotes metabolic reprogramming in glioblastomas by regulation of c-myc. J Exp Clin Cancer Res 2018; 37: 194.
[43] Huang L, Wang HY, Li JD, Wang JH, Zhou Y, Luo RZ, Yun JP, Zhang Y, Jia WH and Zheng M. KPNA2 promotes cell proliferation and tumorigenicity in epithelial ovarian carcinoma through upregulation of c-Myc and downregulation of FOXO3a. Cell Death Dis 2013; 4: e745.
[44] Chu W, Zhang X, Qi L, Fu Y, Wang P, Zhao W, Du J, Zhang J, Zhan J, Wang Y, Zhu WG, Yu Y and Zhang H. The EZH2-PHACTR2-AS1-ribosome axis induces genomic instability and promotes growth and metastasis in breast cancer. Cancer Res 2020; 80: 2737-2750.
[45] Rao S, Peri S, Hoffmann J, Cai KQ, Harris B, Rhodes M, Connolly DC, Testa JR and Wiest DL. RPL22L1 induction in colorectal cancer is

associated with poor prognosis and 5-FU resistance. PLoS One 2019; 14: e0222392.
[46] Rao S, Peri S, Hoffmann J, Cai KQ, Harris B, Rhodes M, Connolly DC, Testa JR and Wiest DL. Ribosomal L22-like1 (RPL22L1) promotes ovarian cancer metastasis by inducing epithelial-to-mesenchymal transition. PLoS One 2015; 10: e0143659.
[47] Liu N, Wu Z, Chen A, Wang Y, Cai D, Zheng J, Liu Y and Zhang L. SNRPB promotes the tumorigenic potential of NSCLC in part by regulating RAB26. Cell Death Dis 2019; 10: 667.
[48] Ichinose J, Watanabe K, Sano A, Nagase T, Nakajima J, Fukayama M, Yatomi Y, Ohishi N and Takai D. Alternative polyadenylation is associated with lower expression of PABPN1 and poor prognosis in non-small cell lung cancer. Cancer Sci 2014; 105: 1135-1141.
[49] Ma J, Zhang Z and Wang J. Small nuclear ribonucleoprotein associated polypeptide N accelerates cell proliferation in pancreatic adenocarcinoma. Mol Med Rep 2015; 12: 6060-6064.
[50] Losada A, Muñoz-Alonso MJ, Martínez-Díez M, Gago F, Domínguez JM, Martínez-Leal JF and Galmarini CM. Binding of eEF1A2 to the RNA-dependent protein kinase PKR modulates its activity and promotes tumour cell survival. Br J Cancer 2018; 119: 1410-1420.

Table S1. Primer sequence used in this study

Gene name	Forward primer	Reverse primer
ACTB	5'-CACCATTGGCAATGAGCGGTTC-3'	5'-AGGTCTTTGCGGATGTCCACGT-3'
HNRNPL	5'-TCGATCACCACGGATGTTCTT-3'	5'-AAGCGTGTAGGCTTTGCGT-3'
PABPN1	5'-GGAGCTGGAAGCTATCAAAGC-3'	5'-CCTGGAGGTGGACTCATATTCA-3'
PCF11	5'-GTTGGAAGAGAGTATCTCACTGC-3'	5'-GCTAGACGTATTCACATTGGGG-3'
SNRPN	5'-GCCGAATCTTCATTGGCACCTTT-3'	5'-TCTTCACGCTCTGGTTGCTTCG-3'
EEF1A2	5'-CCATGTGTGTGGAGAGCTTCTC-3'	5'-TCTCCACGTTCTTGATGACGCC-3'
MRPS5	5'-CTGTGGAAAGGCGCTTTAGC-3'	5'-TTCACCAATGATCTGACCCCT-3'

Table S2. Differentially expressed RBPs

gene ID	Ensembl ID	logFC
MIF	ENSG00000240972	4.616325982
TOP2A	ENSG00000131747	3.780138749
PCLAF	ENSG00000166803	2.909026111
KIF11	ENSG00000138160	2.828095106
ANPEP	ENSG00000166825	2.592736520
OAS3	ENSG00000111331	2.532086750
GNL3L	ENSG00000130119	2.394930327
OAS2	ENSG00000111335	2.260364170
OAS1	ENSG00000089127	2.259666074
KPNA2	ENSG00000182481	2.191198016
TTK	ENSG00000112742	2.068519883
PCNA	ENSG00000132646	2.064532744
TUBA1C	ENSG00000167553	1.972380080
KIF18A	ENSG00000121621	1.917793746
CDKN2A	ENSG00000147889	1.902969785
ESPL1	ENSG00000135476	1.887876525
CPS1	ENSG00000021826	1.789897773
ESRP1	ENSG00000104413	1.788758090
PTBP3	ENSG00000119314	1.786968807
APOBEC3B	ENSG00000179750	1.685290903
PABPC3	ENSG00000151846	1.668353769
PABPC1	ENSG00000070756	1.660865269
SNRPB	ENSG00000125835	1.613898815
GAPDH	ENSG00000111640	1.584521653
MEX3D	ENSG00000181588	1.561257398
IFIT3	ENSG00000119917	1.557153351
IGF2BP2	ENSG00000073792	1.535325206
HELZ2	ENSG00000130589	1.532080365
BRCA1	ENSG00000012048	1.510973055
SNRPD1	ENSG00000167088	1.483109456
IGF2BP3	ENSG00000136231	1.477366070
ZC3HAV1L	ENSG00000146858	1.463111431
MOCOS	ENSG00000075643	1.452246364
APOBEC1	ENSG00000111701	1.415989122
CASK	ENSG00000147044	1.415410801
RPL28	ENSG00000108107	1.414092246
RNASE6	ENSG00000169413	1.395210872
TUBB	ENSG00000196230	1.373873340

PFN1	ENSG00000108518	1.358413240
EIF2AK2	ENSG00000055332	1.348743065
AKAP17A	ENSG00000197976	1.348467930
NUP155	ENSG00000113569	1.346807169
SRPK1	ENSG00000096063	1.342565885
PLP2	ENSG00000102007	1.325721311
IER5L	ENSG00000188483	1.322065329
EZH2	ENSG00000106462	1.321904566
NXT2	ENSG00000101888	1.315857633
ZNFX1	ENSG00000124201	1.311757002
RPL22L1	ENSG00000163584	1.310904721
CSTF2	ENSG00000101811	1.289547926
PRKDC	ENSG00000253729	1.274612839
CAPRIN1	ENSG00000135387	1.241542365
RPP25	ENSG00000178718	1.230962338
MAZ	ENSG00000103495	1.229319139
FAM120A	ENSG00000048828	1.226802301
RPS21	ENSG00000171858	1.213074227
RBM47	ENSG00000163694	1.207467660
SSR1	ENSG00000124783	1.195441405
RPL39	ENSG00000198918	1.187617993
TRUB1	ENSG00000165832	1.185852379
IFIT2	ENSG00000119922	1.178894453
PATL1	ENSG00000166889	1.175041875
RSRC1	ENSG00000174891	1.171468008
SNRPB2	ENSG00000125870	1.165931778
MBNL3	ENSG00000076770	1.160002237
AIMP2	ENSG00000106305	1.154779808
TWISTNB	ENSG00000105849	1.152455517
TOP1	ENSG00000198900	1.144230602
DDX21	ENSG00000165732	1.131260099
ESF1	ENSG00000089048	1.122798413
POLR2K	ENSG00000147669	1.120274125
SPATS2L	ENSG00000196141	1.118545281
ASCC3	ENSG00000112249	1.116554595
WARS	ENSG00000140105	1.114064564
ZC3H13	ENSG00000123200	1.112378830
CCRN4L	ENSG00000151014	1.103450358
CAPG	ENSG00000042493	1.101818965
ZC3HAV1	ENSG00000105939	1.095593967
WDR43	ENSG00000163811	1.094079041
CFL1	ENSG00000172757	1.085746608
CTU1	ENSG00000142544	1.085431328
RPL23A	ENSG00000198242	1.081272651
KIF2A	ENSG00000068796	1.081112037
EIF6	ENSG00000242372	1.080196789
CAD	ENSG00000084774	1.068402517
SNRPF	ENSG00000139343	1.067720608
MRPL14	ENSG00000180992	1.049919302

IQGAP1	ENSG00000140575	1.049011091
ADAT1	ENSG00000065457	1.042422536
TMPO	ENSG00000120802	1.027480113
ZFP36L2	ENSG00000152518	1.025828039
A1CF	ENSG00000148584	1.014459224
KPNA3	ENSG00000102753	1.006785997
MRPS35	ENSG00000061794	1.006056703
LSM12	ENSG00000161654	1.002559199
MAP1S	ENSG00000130479	1.001714153
OASL	ENSG00000135114	1.001636952
DALRD3	ENSG00000178149	-1.001342234
SFSWAP	ENSG00000061936	-1.010418679
LRRFIP2	ENSG00000093167	-1.011389492
ABCF3	ENSG00000161204	-1.011816163
UNKL	ENSG00000059145	-1.014432598
SRRM2	ENSG00000167978	-1.015113563
PCF11	ENSG00000165494	-1.016758059
HNRNPL	ENSG00000104824	-1.018949233
RNASEH2C	ENSG00000172922	-1.022081588
MRPS25	ENSG00000131368	-1.025229201
SF3B2	ENSG00000087365	-1.029326564
GKAP1	ENSG00000165113	-1.037233457
MCAM	ENSG00000076706	-1.038100851
FNBP4	ENSG00000109920	-1.047688101
RPL3	ENSG00000100316	-1.057609154
RPS25	ENSG00000118181	-1.057609297
HABP4	ENSG00000130956	-1.057770014
SMAD9	ENSG00000120693	-1.065737823
FAM98C	ENSG00000130244	-1.068728919
SRP68	ENSG00000167881	-1.072442698
DUS1L	ENSG00000169718	-1.075515232
DZIP1	ENSG00000134874	-1.078483833
PRDX2	ENSG00000167815	-1.082253175
ZCCHC24	ENSG00000165424	-1.094368459
KHNYN	ENSG00000100441	-1.098960912
CPEB1	ENSG00000214575	-1.100570707
SREK1	ENSG00000153914	-1.104776762
DGCR8	ENSG00000128191	-1.107347716
ZRSR1	ENSG00000212643	-1.111987204
EXOSC7	ENSG00000075914	-1.114726925
RNF214	ENSG00000167257	-1.120780664
DUSP1	ENSG00000120129	-1.122949224
RPUSD4	ENSG00000165526	-1.124053188
RPL24	ENSG00000114391	-1.124448919
PNN	ENSG00000100941	-1.128915968
YBX3	ENSG00000060138	-1.133205695
ENO3	ENSG00000108515	-1.133943228
TRMT2A	ENSG00000099899	-1.134265998
VARS2	ENSG00000137411	-1.136219092

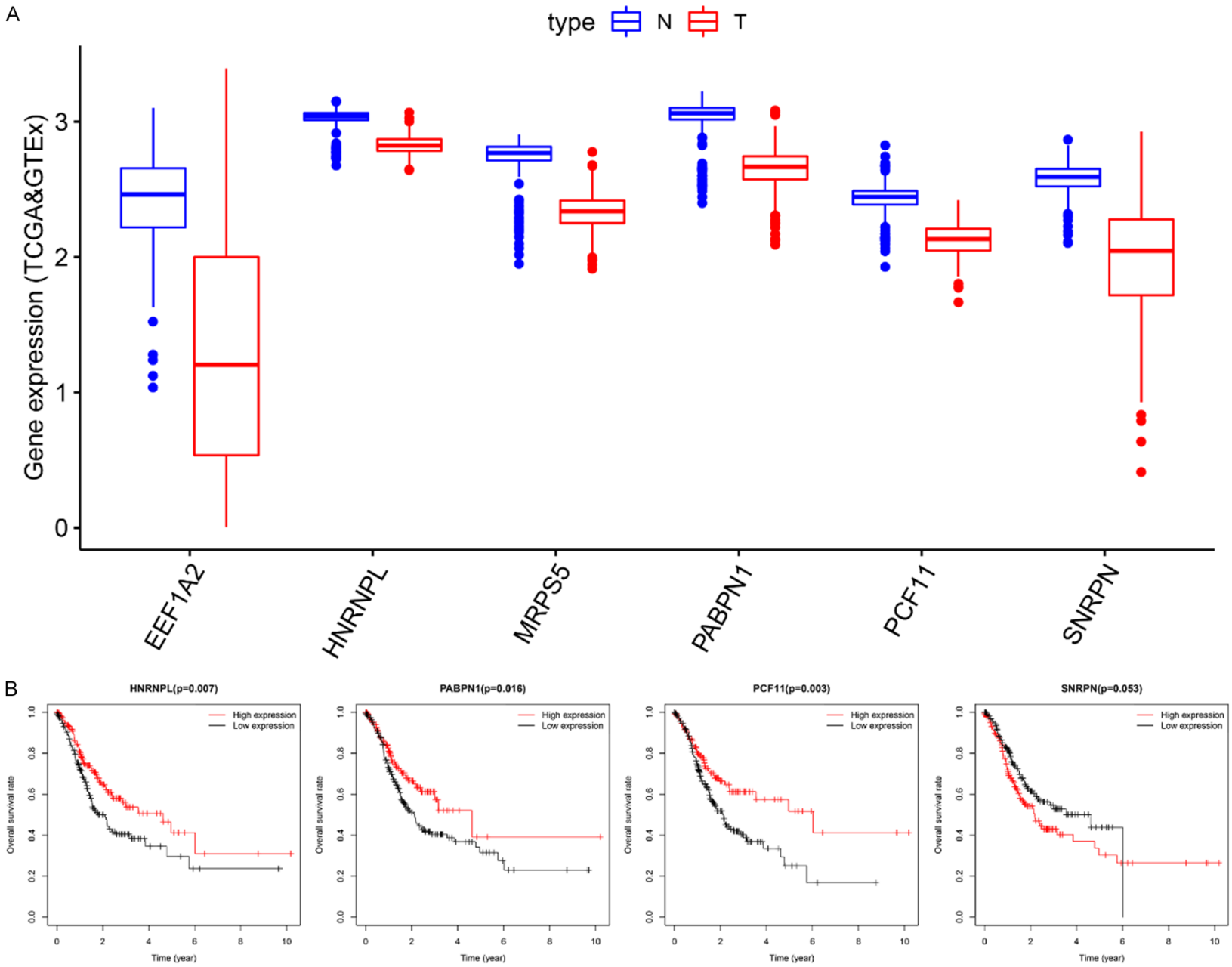
IFRD1	ENSG00000006652	-1.136303582
LUC7L3	ENSG00000108848	-1.139296412
DDX42	ENSG00000198231	-1.156036735
KHDRBS3	ENSG00000131773	-1.164175577
PPWD1	ENSG00000113593	-1.166196155
RBM20	ENSG00000203867	-1.168551751
RBPMS	ENSG00000157110	-1.170810166
UPF3A	ENSG00000169062	-1.182980948
RPSA	ENSG00000168028	-1.183567962
TIMM44	ENSG00000104980	-1.204468629
RPL26	ENSG00000161970	-1.208759051
RPS16	ENSG00000105193	-1.214222057
MTERF4	ENSG00000122085	-1.214515832
FIP1L1	ENSG00000145216	-1.220495176
NSA2	ENSG00000164346	-1.221942663
DDX3Y-	ENSG00000067048	-1.229846910
RSRC2	ENSG00000111011	-1.236127166
NOL3	ENSG00000140939	-1.237009104
MRPS6	ENSG00000243927	-1.237231750
AKR1B1	ENSG00000085662	-1.237277781
C11orf58	ENSG00000110696	-1.245626594
PCBP4	ENSG00000090097	-1.251727031
EEF1D	ENSG00000104529	-1.253346870
RPLP1	ENSG00000137818	-1.256042596
CAPRIN2	ENSG00000110888	-1.260259924
ADARB1	ENSG00000197381	-1.263713353
WDR83	ENSG00000123154	-1.268084039
TRPT1	ENSG00000149743	-1.282488567
THOC1	ENSG00000079134	-1.283713908
ANG	ENSG00000214274	-1.292762298
NKX6-2	ENSG00000148826	-1.294552392
SPG20	ENSG00000133104	-1.296132917
HARS	ENSG00000170445	-1.299155062
EIF3L	ENSG00000100129	-1.315410291
MATR3	ENSG00000015479	-1.321112054
TOP3B	ENSG00000100038	-1.339719523
RPS4Y1	ENSG00000129824	-1.346938672
CSDC2	ENSG00000172346	-1.349408084
PTPRM	ENSG00000173482	-1.361638474
FGF2	ENSG00000138685	-1.369726811
QTRT1	ENSG00000213339	-1.373321530
L1CAM	ENSG00000198910	-1.379817265
SAFB	ENSG00000160633	-1.384041024
EIF1B	ENSG00000114784	-1.384679909
DDX26B	ENSG00000165359	-1.385280599
RPS13	ENSG00000110700	-1.387911154
HPSE2	ENSG00000172987	-1.394976075
RBM39	ENSG00000131051	-1.400040668
PTGS1	ENSG00000095303	-1.400657506

HNRNPUL2	ENSG00000214753	-1.416150565
NXF1	ENSG00000162231	-1.417490958
EWSR1	ENSG00000182944	-1.419740147
PRPF39	ENSG00000185246	-1.425456258
RPL10	ENSG00000147403	-1.447258173
PCBP2	ENSG00000197111	-1.450334238
RPS27A	ENSG00000143947	-1.455386974
DDX17	ENSG00000100201	-1.456848495
CTIF	ENSG00000134030	-1.485382424
RAP1GAP2	ENSG00000132359	-1.501150227
NDRG1	ENSG00000104419	-1.501476804
WDR61	ENSG00000140395	-1.505101600
EIF2B4	ENSG00000115211	-1.514657591
SUGP2	ENSG00000064607	-1.517801795
GATB	ENSG00000059691	-1.537233085
DRG2	ENSG00000108591	-1.545938156
SDHA	ENSG00000073578	-1.575015171
SPARCL1	ENSG00000152583	-1.591361752
METTL3	ENSG00000165819	-1.595485901
HNRNPH1	ENSG00000169045	-1.599533276
WIPF3	ENSG00000122574	-1.600966650
MRPS5	ENSG00000144029	-1.619964609
SRSF9	ENSG00000111786	-1.621080801
WDR19	ENSG00000157796	-1.629942646
EPHB6	ENSG00000106123	-1.634397189
RPL34	ENSG00000109475	-1.638581769
HIP1R	ENSG00000130787	-1.642950635
EIF3G	ENSG00000130811	-1.643036820
REXO2	ENSG00000076043	-1.647302960
KIAA0391	ENSG00000100890	-1.652030440
CLK4	ENSG00000113240	-1.700558827
LARP6	ENSG00000166173	-1.742770428
TUT1	ENSG00000149016	-1.750415458
PDCD4	ENSG00000150593	-1.751064128
MYEF2	ENSG00000104177	-1.795808329
EIF4A2	ENSG00000156976	-1.809176877
AFF3	ENSG00000144218	-1.817523556
CLASRP	ENSG00000104859	-1.842135493
PABPN1	ENSG00000100836	-1.857077442
SNRPN	ENSG00000128739	-1.864782066
PAN2	ENSG00000135473	-1.869587528
PRPF40B	ENSG00000110844	-1.948260521
PPAN	ENSG00000130810	-1.960202878
AKAP8L	ENSG00000011243	-1.982599571
RPL9	ENSG00000163682	-1.996051867
ZGPAT	ENSG00000197114	-1.996446683
VKORC1	ENSG00000167397	-2.005882106
PSMD9	ENSG00000110801	-2.045059573
SARS2	ENSG00000104835	-2.049979488
ARHGEF28	ENSG00000214944	-2.054174771

SIDT2	ENSG00000149577	-2.086031483
LUC7L	ENSG00000007392	-2.094979656
CLK1	ENSG00000013441	-2.095794644
RASD1	ENSG00000108551	-2.119030825
SNRNP70	ENSG00000104852	-2.129182590
RBM6	ENSG00000004534	-2.129798401
PAIP2B	ENSG00000124374	-2.152251356
OBSL1	ENSG00000124006	-2.158659894
LAMA2	ENSG00000196569	-2.170359947
SRSF5	ENSG00000100650	-2.172722112
MRPL46	ENSG00000259494	-2.175444195
RBM5	ENSG00000003756	-2.175649321
GLTSCR2	ENSG00000105373	-2.190032289
U2AF1L4	ENSG00000161265	-2.198950647
ZMAT5	ENSG00000100319	-2.255857690
PTRH1	ENSG00000187024	-2.281115845
HSPB8	ENSG00000152137	-2.319452253
RBPMS2	ENSG00000166831	-2.360506693
RPS19BP1	ENSG00000187051	-2.366509504
EEF1A2	ENSG00000101210	-2.374326718
CIRBP	ENSG00000099622	-2.384140637
IPO4	ENSG00000196497	-2.386449010
NDRG2	ENSG00000165795	-2.389290213
AZGP1	ENSG00000160862	-2.451177416
APOBEC2	ENSG00000124701	-2.492289244
CELF6	ENSG00000140488	-2.502476247
CTAGE5	ENSG00000150527	-2.578420036
MRPL38	ENSG00000204316	-2.595650246
RBM4	ENSG00000173933	-2.833868810
MRPS24	ENSG00000062582	-3.126963746
AARSD1	ENSG00000266967	-3.369139778
ANKHD1	ENSG00000131503	-3.478214046
PSMA6	ENSG00000100902	-3.515921976
MYH11	ENSG00000133392	-3.624940586
RPL17	ENSG00000265681	-3.650876147
TXNDC5	ENSG00000239264	-3.664767164
RNASEK	ENSG00000219200	-3.756526270
AKAP2	ENSG00000241978	-3.918699989
RNASE4	ENSG00000258818	-4.379672757
SARNP	ENSG00000205323	-4.519438989
EIF4A1	ENSG00000161960	-5.240499775
ARL6IP4	ENSG00000182196	-6.028148457
EEF1G	ENSG00000254772	-7.938209368

Table S3. GO and KEGG enrichment analysis of differentially expressed RBPs

Category	Enrichment term	*P* Value	FDR
GO enrichment analysis for up-regulation RBPs			
GOTERM_BP_DIRECT	response to virus	5.92E-09	8.80E-06
GOTERM_BP_DIRECT	defense response to virus	2.02E-07	3.00E-04
GOTERM_BP_DIRECT	negative regulation of viral genome replication	2.01E-06	2.98E-03
GOTERM_BP_DIRECT	RNA splicing	2.53E-06	3.76E-03
GOTERM_BP_DIRECT	type I interferon signaling pathway	2.11E-05	3.13E-02
GOTERM_CC_DIRECT	cytoplasm	4.17E-09	5.07E-06
GOTERM_CC_DIRECT	nucleus	4.69E-09	5.70E-06
GOTERM_CC_DIRECT	nucleoplasm	3.06E-08	3.72E-05
GOTERM_CC_DIRECT	cytosol	6.98E-08	8.49E-05
GOTERM_CC_DIRECT	nucleolus	1.37E-07	1.67E-04
GOTERM_MF_DIRECT	poly(A) RNA binding	1.88E-23	2.37E-20
GOTERM_MF_DIRECT	RNA binding	4.66E-09	5.88E-06
GOTERM_MF_DIRECT	2'-5'-oligoadenylate synthetase activity	6.04E-07	7.62E-04
GOTERM_MF_DIRECT	double-stranded RNA binding	9.09E-07	1.15E-03
GOTERM_MF_DIRECT	nucleotide binding	1.78E-05	2.25E-02
GO enrichment analysis for down-regulation RBPs			
GOTERM_BP_DIRECT	translational initiation	6.11E-17	1.67E-13
GOTERM_BP_DIRECT	mRNA processing	6.56E-16	9.99E-13
GOTERM_BP_DIRECT	mRNA splicing, via spliceosome	4.56E-14	6.82E-11
GOTERM_BP_DIRECT	SRP-dependent cotranslational protein targeting to membrane	3.34E-13	5.01E-10
GOTERM_BP_DIRECT	RNA processing	5.23E-13	7.83E-10
GOTERM_BP_DIRECT	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	5.78E-13	8.66E-10
GOTERM_BP_DIRECT	translation	5.26E-12	7.88E-09
GOTERM_BP_DIRECT	RNA splicing	6.36E-12	9.53E-09
GOTERM_BP_DIRECT	viral transcription	6.21E-11	9.31E-08
GOTERM_BP_DIRECT	rRNA processing	2.65E-09	3.97E-06
GOTERM_BP_DIRECT	mRNA 3'-end processing	8.52E-06	1.28E-02
GOTERM_BP_DIRECT	mRNA splice site selection	1.85E-05	2.76E-02
GOTERM_BP_DIRECT	mitochondrial translation	2.40E-05	3.59E-02
GOTERM_CC_DIRECT	nuclear speck	2.56E-12	3.17E-09
GOTERM_CC_DIRECT	nucleoplasm	2.08E-11	2.57E-08
GOTERM_CC_DIRECT	ribosome	2.47E-10	3.06E-07
GOTERM_CC_DIRECT	nucleus	1.23E-08	1.52E-05
GOTERM_CC_DIRECT	cytoplasm	1.86E-07	2.30E-04
GOTERM_CC_DIRECT	U2-type prespliceosome	2.87E-07	3.56E-04
GOTERM_CC_DIRECT	nucleolus	5.27E-07	6.53E-04
GOTERM_CC_DIRECT	U1 snRNP	5.32E-07	6.59E-04
GOTERM_CC_DIRECT	cytosolic large ribosomal subunit	2.27E-06	2.82E-03
GOTERM_CC_DIRECT	cytosolic small ribosomal subunit	3.91E-06	4.84E-03
GOTERM_MF_DIRECT	RNA binding	9.73E-34	1.25E-30
GOTERM_MF_DIRECT	poly(A) RNA binding	6.65E-32	8.54E-29
GOTERM_MF_DIRECT	nucleotide binding	5.33E-18	6.85E-15
GOTERM_MF_DIRECT	structural constituent of ribosome	6.10E-11	7.83E-08
GOTERM_MF_DIRECT	mRNA binding	3.95E-08	5.07E-05
GOTERM_MF_DIRECT	nucleic acid binding	8.76E-08	1.13E-04
KEGG enrichment analysis for dysregulated RBPs			
KEGG PATHWAY	Ribosome	1.20E-17	1.51E-15
KEGG PATHWAY	RNA transport	9.46E-10	5.96E-08
KEGG PATHWAY	Spliceosome	5.83E-08	2.45E-06
KEGG PATHWAY	mRNA surveillance pathway	1.23E-07	3.87E-06
KEGG PATHWAY	RNA degradation	3.62E-05	9.11E-04
KEGG PATHWAY	Ribosome biogenesis in eukaryotes	1.37E-04	2.88E-03
KEGG PATHWAY	Aminoacyl-tRNA biosynthesis	1.66 E-04	2.99E-03
KEGG PATHWAY	Influenza A	5.53 E-04	8.70E-03


A
type
N
T
Gene expression (TCGA&GTEx)
0
1
2
3
EEF1A2
HNRNPL
MRPS5
PABPN1
PCF11
SNRPN
B
HNRNPL(p=0.007)
PABPN1(p=0.016)
PCF11(p=0.003)
SNRPN(p=0.053)
High expression
Low expression
Overall survival rate
Time (year)

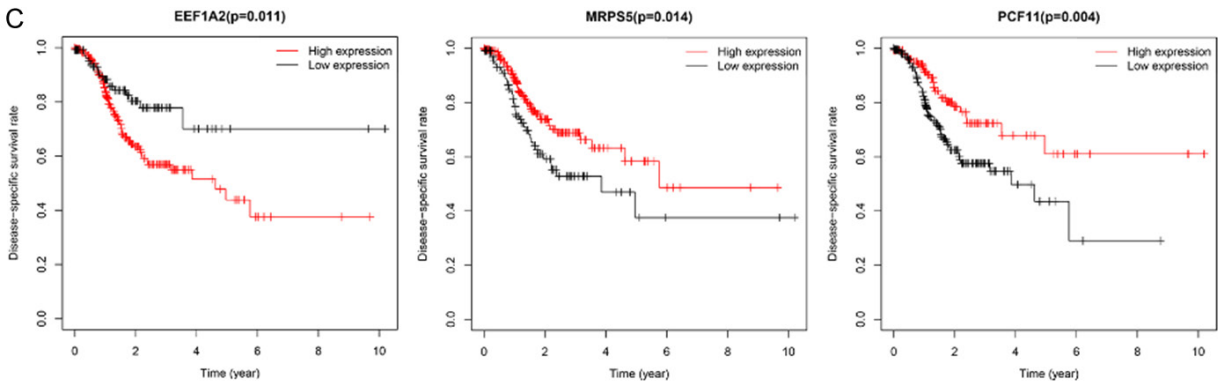


Figure S1. TCGA combined with GTEx database to analyze the expression and prognosis of OS and DSS-related hub RBP. A. The expression of six hub RBPs in gastric cancer tissues was lower than that in normal tissues. B. The relationship between OS-related hub RBPs and the overall survival. C. The relationship between DSS-related hub RBPs and disease-specific survival.