

Exploring an RNA binding proteins-associated prognostic model for Gastric Cancer

Ming Liu

Southern Medical University Nanfang Hospital <https://orcid.org/0000-0003-3567-530X>

Jiayi Xie

Southern Medical University

Xiaobei Luo

Southern Medical University Nanfang Hospital

Yaxin Luo

Southern Medical University Nanfang Hospital

Side Liu

Southern Medical University Nanfang Hospital <https://orcid.org/0000-0001-5589-2817>

Zelong Han (✉ hzl198886@163.com)

Department of Gastroenterology, Nanfang Hospital, Southern Medical University, Guangzhou, China
<https://orcid.org/0000-0003-4217-8117>

Primary research

Keywords: RNA binding protein; Biomarkers; Gastric Cancer; Survival Analyses; The Cancer Genome Atlas

Posted Date: June 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-33114/v1>

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

[Read Full License](#)

Exploring an RNA binding proteins-associated prognostic model for Gastric Cancer

Ming Liu #1, Jiayi Xie 1, Xiaobei Luo¹, Yaxin Luo¹, Side Liu§1, Zelong Han§1

Running title: Biomarkers in Gastric Cancer

¹ Department of Gastroenterology, Nanfang Hospital, Southern Medical University,

Guangzhou, China;

First Author: Ming Liu

Email: lm21810329@i.smu.edu.cn

Jiayi Xie

Email: 814174180@qq.com

Xiaobei Luo

Email: luoxiaobei63@126.com

Yaxin Luo

Email: 3104761566@qq.com

§ Corresponding Author: Side Liu

Email: liuside2011@163.com

Corresponding Author: Zelong Han

Email: hzl198886@163.com

Abstract

Background: Gastric cancer (GC) is one of the most prevalent malignant cancers around the world. Given that abnormal RNA binding proteins (RBPs) are involved in the tumorigenesis, we aimed to explore the potential value of RBPs-associated genes in gastric cancer.

Methods: RNA-seq and clinical data were retrieved from The Cancer Genome Atlas (TCGA) database and differentially expressed RBPs genes were screened. GO and KEGG pathway enrichment analyses were implemented to elucidate the roles of RBPs in GC. The protein-protein interaction (PPI) networks of RBPs were carried out, and the hub genes were determined by MCODE built in Cytoscape. The TCGA-STAD dataset was randomly divided into training and testing groups. A prognostic signature including five RBPs was developed within the training cohort after Cox regression and Lasso regression analyses. We used Kaplan–Meier (KM) and receiver operating characteristic (ROC) curves to evaluate the capacity of the model in both groups. Then, a nomogram based on hub RBPs expression was established. Gene Set Enrichment Analysis was performed between the high-risk and low-risk group.

Results: A total of 166 up-regulated RBPs and 130 down-regulated RBPs were identified. Via Cox regression and Lasso regression analysis within the training group, five hub RBPs (RNASE1, SETD7, BOLL, PPARGC1B, MSI2) were screened and the prognostic model was constructed. The risk score was calculated and gastric cancer patients were divided into high-risk and low-risk groups. In multivariate analysis, risk score was still an independent prognostic indicator (HR = 1.80, 95% CI = 1.45-2.22, P

< 0.01). Patients with low risk had favorable survival rate in both training and testing group compared to those at high risk ($P < 0.001$). The areas under the ROC curves (AUC) of the prognostic model are 0.718 in the training cohort and 0.651 in the testing cohort. The hub RBPs-based nomogram model exhibited excellent ability to predict the OS of GC. GSEA illustrated that several cancer-related signaling pathways were enriched in patients with a high-risk score.

Conclusions: This study discovered a five RBPs signature which might provide a potential prognostic value to GC patients.

Keywords: RNA binding protein; Biomarkers; Gastric Cancer; Survival Analyses; The Cancer Genome Atlas.

Background

Gastric Cancer (GC) is the fifth most prevalent cancers diagnosed both in men and women and the third major cause of cancer-related death in the world. In 2018, over 1,000,000 people were diagnosed with gastric cancer, which resulted in an estimated 783,000 deaths[1]. Despite advances in the diagnostic and therapeutic approaches, the global burden remains high, especially for the region in Asia, Latin America, and central and eastern Europe. The majority of gastric cancer is often diagnosed at an advanced stage due to the absence of early symptom and useful biomarkers[2]. Studies have shown that if patients are diagnosed at an early stage, they could have a 5-year survival rate of more than 90%[3]. Currently, the main methods to assess the prognosis of GC are through histopathological examination and tumor node metastasis (TNM) classification[4]. However, the existing evaluation methods are unsatisfactory due to the heterogeneous attributes of GC[5]. Therefore, it is imperative to identify highly effective markers that can aid in understanding the molecular mechanisms of GC tumorigenesis and predicting the prognosis of GC patients.

RNA binding proteins (RBPs) are a group of protein family in eukaryotes, which can interact with various types of RNAs. By binding to RNA and forming protein-RNA complexes, RBPs can affect RNA maturation, turnover, modification and localization. RBPs can also regulate alternative splicing and translation by controlling accessibility and activity of basic machineries, such as the splicing enhancer/repressor and translation regulators[6, 7]. Thus, RBPs play vital roles in co-transcriptional and post-

transcriptional gene regulation and contribute to numerous biological processes, such as cell differentiation, proliferation and cell fate transition[8]. Since RBPs perform diverse crucial functions in post-transcriptional events, it comes as no surprise that dysregulation of RBPs is closely linked to the initiation and progression of various diseases. For instance, RBP LIN28 can regulate mRNA translation of HMGA genes, which is associated with insulin resistance and type 2 diabetes[9]. Mutations in RBPs hnRNP A1 and TDP-43 can lead to aberrant protein aggregation, which is a pathological hallmark of many neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)[10]. Besides, RBP ZFP36L1 can regulate HSP70 family members by inhibiting chondrocyte apoptosis to protect against osteoarthritis[11]. Through genome-wide screening, a census of 1542 RBPs has been established in human cells, accounting for 7.5% of all protein-coding genes. However, the roles of many RBPs in the pathogenesis of cancer remain relatively underexplored.

RBPs are found to display abnormal expression in tumors compared to adjacent normal tissues in several studies, and this aberrant expression is associated with patient survival rate[12-14]. A variety of functions of RBPs have been discovered in cancer, such as regulating polyadenylation, alternative splicing, subcellular localization, stability and translation. This functional diversity reflects the critical role of RBPs in cancer. For example, MBNL1 is an RBP that regulate alternative splicing of several critical genes in MLL-rearranged leukemia and inhibitor of MBNL1 can induce selective leukemic cell death[15]. HuR inhibitor KH-3 can inhibit breast cancer invasion and metastasis

via disrupting HuR-FOXQ1 mRNA interaction[16]. PSF can posttranscriptionally regulate target transcript ESR1 and SCFD2, thereby promoting the progression of ER-positive breast cancer[17]. In short, these studies have revealed that the RBPs are involved in the initiation and progression of tumors. Besides, there are also reports using high-throughput bioinformatic profiling in The Cancer Genome Atlas (TCGA) database to reveal a similar pattern of dysregulations in RBPs expression across multiple cancer types[18-20]. However, the majority of RBPs have not been studied comprehensively in gastric cancer via bioinformatic method. Therefore, the STAD (stomach adenocarcinoma) data from The Cancer Genome Atlas (TCGA) was extracted. Through bioinformatic analysis, we identified multiple aberrantly expressed RBPs associated with STAD and explored the potential molecular functions and clinical implication of RBPs in STAD. These STAD-related RBPs may shed light on the pathogenesis of gastric cancer, which might serve as potential biomarkers for diagnosis and prognosis.

Methods

Data acquisition

A total of 1542 RBPs were collected in this study[21]. The RNA-sequencing data of 32 normal gastric tissue samples and 375 STAD samples with corresponding clinical data were downloaded from The Cancer Genome Atlas database (TCGA, <https://portal.gdc.cancer.gov/>). Linear models for microarray data (*limma*) (<http://bioconductor.org/packages/2.4/bioc/html/limma.html>) is an R package applied

to perform this analysis[22]. Based on the negative binomial distribution, the Limma package fits a generalized linear model for each gene and utilizes empirical Bayes shrinkage for stable variance estimation. The *Limma* package was applied to identify the differentially expressed RBPs between STAD tissues (Cancer group) and noncancerous tissues (Normal group). Differentially expressed RBPs were selected for further analyses with setting the $|\log_2 \text{fold change (FC)}| \geq 0.5$ & false discovery rate (FDR) <0.05 . [23]. All significant DEGs had an average count value more than 1.

Functional and Signal Pathway Enrichment Analysis

Gene Ontology (GO) analysis includes what is currently known about a gene in terms of biological process (BP), cellular components (CC) and molecular function (MF) [24, 25]. Kyoto Encyclopedia of Genes and Genomes (KEGG) [26] annotates a gene or a set of genes with their respective KEGG pathways. To illustrate the biological functions of these differently expressed RBPs, GO and KEGG pathway enrichment analyses were carried out using the *clusterProfiler* package [27] in R. Both P and FDR values were less than 0.05 was considered statistically significant.

Protein-protein interaction (PPI) network and gene module screening

Search Tool for the Retrieval of Interacting Genes (STRING) [28] is an open access database designed to evaluate the protein-protein interaction (PPI) messages of DEGs. We uploaded and mapped the list of differently expressed RBPs to STRING website (<http://www.string-db.org/>). Then PPIs of DEGs were chosen with a medium

confidence >0.4 [29]. After that, Cytoscape 3.7.2 software was used to further construct and visualize the PPI network [30]. The plug-in Molecular Complex Detection (MCODE) in Cytoscape was applied to detect the significant gene modules of the PPI networks with both MCODE scores and count of nodes >4. All $P \leq 0.05$ were considered as statistically significant.

Prognostic model construction

We used univariable Cox proportional risk regression to screen RNA binding proteins significantly related to overall survival (OS) in the TCGA dataset. To reduce overfitting, the lasso regression was applied to further screen the OS related genes by using the R package glmnet. Finally, the multivariable Cox proportional risk regression analysis was carried out to establish the prognosis model of gastric cancer. We used the following formula to calculate the risk score of each patient: Risk score = $\sum_{i=1}^n v_i * c_i$ (the v_i is the expression level of gene i , c_i represents the regression coefficient of gene i in the multivariate Cox regression analysis, and n represents the number of independent indicators). The median risk score was determined as the critical value to divide the STAD dataset into high risk and low risk. In order to determine the role of risk score in predicting the clinical prognosis of CRC patients, Kaplan-Meier Plotters and receiver operating characteristic (ROC) curves were drawn to analyze the prognostic value between high-risk group and low-risk group. The prognostic value of the five RBPs in STAD was verified by using the Kaplan Meier plotter

(<https://kmplot.com/analysis/>) online tool. All $P \leq 0.05$ were considered as statistically significant.

Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) v4.0.3 was carried out with enrichment analysis performed on normalized mRNA expression profiles. Random sample permutations ($n = 1000$) were performed to calculate the P value. A false discovery rate < 0.01 and a nominal $P < 0.05$ were considered significant.

Nomogram construction

The hub RBPs genes were combined to construct a nomogram using rms R package to predict the likelihood of OS. The calibration curves were plotted to evaluate the consistency between actual and predicted survival.

Results

Identification of differently expressed RBPs

The TCGA-STAD data of 375 GC tissues (Cancer group) and 32 noncancerous tissues (Normal group) were analyzed by the *limma* package built in R through the linear model and the contrast model to discover the differently expressed RBPs. A total of 1542 RBPs were included in the analysis, and 296 RBPs were selected by the criteria of $FDR < 0.05$ and $|\log_2FC| > 0.5$, including 166 upregulated and 130 downregulated RBPs (Fig. 1).

GO and KEGG pathway Pathway Enrichment Analysis

To investigate functions and signal pathway enrichment of identified RBPs, the differently expressed RBPs were divided into two groups: up-regulated or down-regulated expression. We further analyzed these two groups using the *clusterProfiler* package in R with criteria of $P < 0.05$. Significant 5 BP, CC and MF enrichment terms, were displayed respectively. As shown in Fig. 2, in Biological Process (BP) term, the up-regulated genes were mainly enriched in the ncRNA metabolic process, ncRNA processing and ribonucleoprotein complex biogenesis, while the down-regulated genes were focused on the mRNA processing, regulation of translation and regulation of cellular amide metabolic process. In Cellular Component (CC) term, the up-regulated genes were mainly enriched in the cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule and nucleolar part, while the down-regulated genes were focused on the ribosome, ribosomal subunit and mitochondrial matrix. In Molecular Function (MF) term, the up-regulated genes were mainly enriched in the catalytic

activity acting on RNA, nuclease activity and ribonuclease activity, while the down-regulated genes were focused on the mRNA binding, mRNA 3'-UTR binding and AU-rich element binding. Moreover, the significant pathways enrichment analysis showed that the up-regulated genes were mainly enriched in the Ribosome biogenesis in eukaryotes, RNA transport and Spliceosome, and the down-regulated genes were enriched in the Ribosome, RNA transport and RNA degradation (Fig. 2c-d).

Module screening from the PPI network

To further explore the interplay of differently expressed RNA binding proteins in GC, 296 RBPs were uploaded into the STRING website and analyzed by Cytoscape software. A total of 261 RBPs and 2568 edges with score >0.4 (medium confidence) were picked out to construct the PPI networks (Fig. 3a). Then, three significant modules were clustered via MCODE APP built in Cytoscape, which consist of 76 nodes and 1071 edges (Fig. 3b). The RBPs in the key module 1 were mainly enriched in ribosome biogenesis, rRNA processing, ncRNA processing, ribonucleoprotein complex biogenesis, rRNA metabolic process, and ncRNA metabolic process.

Survival-related RBPs screening

The 261 differently expressed RBPs used for the construction of PPI network were further analyzed via univariate Cox regression and 19 prognostic-associated candidate hub RBPs were obtained (Fig. 4a). Subsequently, these 19 prognostic-associated candidate RBPs were included in the lasso regression analysis by using glmnet built in

R to screen the genes and 10 genes were selected (Fig. 4c-d). Finally, the multivariable Cox proportional risk regression analysis was performed and 5 hub RBPs (RNASE1, SETD7, BOLL, PPARGC1B, MSI2) were found to be related to the prognosis of STAD (Fig. 4b). The coefficients of each gene are displayed in Table 1. The genes with HR > 1 (RNASE1, SETD7, BOLL) are regarded as dangerous genes, while the gene with HR < 1 (PPARGC1B, MSI2) are considered as protective genes (Fig. 4b). Besides, the univariate analysis showed that stage, regional lymph node involvement, metastasis and risk score were significantly correlated with OS of STAD patients (P < 0.01) (Table 3). However, we only found that risk score was an independent prognostic indicator for STAD through multivariate analysis. (P < 0.01) (Table 3).

Prognosis-related genetic risk score model construction and validation

Through multiple stepwise Cox regression analysis, the five hub RBPs were used to construct the predictive model. The formula for the risk score of each patient was calculated as follows: $Risk\ score = (0.2769 * expression\ value\ of\ RNASE1) + (0.5576 * expression\ value\ of\ SETD7) + (1.9505 * expression\ value\ of\ BOLL) + (-0.5407 * expression\ value\ of\ PPARGC1B) + (-0.7737 * expression\ value\ of\ MSI2)$ A survival analysis was conducted to evaluate the predictive ability. We plotted the KM survival curve to evaluate the difference in STAD survival in both the training and testing groups. Both groups were stratified into high-risk and low-risk subgroup according to the median risk score. Within the training group, the KM survival curve indicated that the patient in the high-risk subgroup have a worse survival rate compared to those in the

low-risk group ($p=6.71e-05$) (Fig. 5a). A similar result can be found in the testing group ($p=1.845e-03$) (Fig. 6a). To further evaluate the accuracy of the model, a time-dependent ROC analysis was carried out. We found that the area under the ROC curve (AUC) of this RBPs risk score model was 0.718 in the training group (Fig. 5b) and 0.651 in the testing group (Fig. 6b), which suggested that it has an acceptable diagnostic ability. The distribution of risk scores in GC patients and the scatterplot of the relation between risk scores and survival status are displayed. (Fig. 5c-d, Fig. 6c-d) A significant rise in the overall mortality rate is observed, as the risk score increases. Heatmap was plotted to show genes expression profiles between high-risk and low-risk STAD groups. As shown in the Fig. 5e-6e, patients with high-risk tend to have higher level of dangerous genes. On the contrary, patients in the low-risk group have a propensity to express the protective gene.

Nomogram construction

Nomogram is a powerful quantitative tool for prognosis prediction. Thus, we integrated the five RBPs signature to build a nomogram (Fig. 7). On the basis of multivariate Cox analysis, the number of points is assigned to a single variable by using the point scale in the nomogram. A horizontal line was drawn to decide the point of each variable and calculate the total scores for each patient by the sum of points of all variables. The estimated 1, 3 and 5 year survival rates of STAD patients can be calculated by plotting the vertical line between the total point axis and each prognosis axis, which might help to make clinical decisions for STAD patients. Then, we drew the calibration plots to

assess the reliability of the nomogram. Our results suggested that the nomogram model predicted outcome was virtually matching with the actual outcome for both 3-year and 5-year survival of STAD (Fig. 7b-c).

Validation of the prognostic value

To further explore the prognostic value of five hub RBPs in STAD, the Kaplan Meier-plotter was used to determine the relationship between hub RBPs and OS. A total of the five hub RBPs (RNASE1, SETD7, BOLL, PPARGC1B, MSI2) were identified by Kaplan Meier-plotter server. The results of log-rank test demonstrated that the three RBPs were associated with the OS in STAD patients (Fig. 8).

Five-RBPs model-associated biological pathways

To further understand the biological function of this signature. GSEA was performed between the high- and low-risk groups to discover the high-risk-related pathways. We found that "angiogenesis," "epithelial-mesenchymal transition (EMT)," "KRAS signaling," "JAK-STAT signaling pathway," and some other tumor-related biological pathways were significantly enriched ($P < 0.05$, Fig. 9). The results indicate that the Five-RBPs model may reflect the status of these biological pathways to predict the overall survival of STAD patients.

Discussion

Currently, even though some progress have been achieved in diagnosis and therapeutic measures in gastric cancer, the survival rate of GC patients has not been improved remarkably, especially for advanced stage patients. Hence, it is pivotal to identify critical genes related to the molecular mechanisms of GC tumorigenesis and improve the prognosis of GC patients by investigating latent diagnostic and therapeutic targets[31]. To date, the dysregulation of RBPs has been in the limelight in numerous cancer researches. However, the relationship between most of RBPs and gastric cancer are still not well characterized[19]. In this study, 261 differently expressed RBPs were identified between tumor and normal tissues based on TCGA-STAD dataset. Through functional annotation enrichment analysis, we systematically analyzed relevant biological functions and pathways. Then, we further elucidated the interaction among differently expressed RBPs via using the STRING tool to construct protein-protein interaction (PPI) networks and the most significant module and the top hub genes were selected from the PPI network. Furthermore, we also performed univariate Cox regression analysis, lasso regression analysis, multivariate Cox regression analysis, survival analysis and ROC analysis of hub RBPs to explore their biological functions and clinical values. Subsequently, based on the five prognostic-related hub RBP genes, we established a risk model to predict GC prognosis. These results may help discover novel biomarkers for predicting the prognosis of patients with GC.

According to the functional enrichment analysis, the differentially expressed RBPs

were significantly enriched in mRNA processing, regulation of translation, regulation of cellular amide metabolic process, RNA splicing, RNA catabolic process, translational initiation, mRNA binding, mRNA 3'-UTR binding, AU-rich element binding, ncRNA metabolic process, ncRNA processing, ribonucleoprotein complex biogenesis etc. The KEGG pathways analysis showed that the aberrantly expressed RBPs were related to Ribosome biogenesis in eukaryotes, RNA transport, Spliceosome, RNA degradation. These results are consistent with the functions of RBPs. The studies on post-transcriptional regulation of RNAs in cancer has increased in the past decades. RBPs play a vital role in practically all aspects of this post-transcriptional regulatory network, affecting the function of each transcript and maintaining cellular homeostasis[32]. Through dynamic interactions with RNAs, RBPs can form the functional units called ribonucleoprotein complexes that regulate RNA splicing, metabolism and translation, which are closely linked to a wide range of cancers. Lin28B could directly bind to NRP-1 3'UTR to increase NRP-1 mRNA stability and expression, thus activating Wnt/ β -catenin signaling to promote gastric cancer stemness[33]. Heterogeneous nuclear ribonucleoprotein L (HNRNPL) could regulate the alternative splicing of a set of RNAs, including those encoding androgen receptor in prostate cancer. The HNRNPL and its RNA targets are causally related to prostate tumorigenesis and could be considered for novel therapeutic approaches in prostate cancer[34]. Additionally, AU-rich element-binding protein HuR can bind to PDCD4 3'-UTR and prohibit miR-21-mediated repression of PDCD4 translation, thereby preventing chronic inflammation and tumorigenesis[35].

Furthermore, a protein-protein interaction network of these differently expressed RBPs was created and three key modules were extracted. These hub RBPs have been found to be involved in the pathogenesis of cancers. Guanine nucleotide binding protein-like 3 (GNL3), a GIP-binding nuclear protein, was reported to facilitate the progression of gastric cancer and high expression of GNL3 is associated with shorter survival rate in GC patients[36]. GTPBP4, also known as Nog1, can interact with p53 and promote gastric cancer cells proliferation and inhibit apoptosis[37]. On the other hand, some RBPs have been reported to be implicated in other cancers. For example, loss of BOP1 (Block of proliferation 1), which was found in the regulation of rRNA biogenesis and ribosome biogenesis, confers resistance to BRAFi in BRAF-mutant melanoma by activating MAPK pathway[38]. Besides, BOP1 can serve as one of the direct Wnt/ β -catenin target genes, thus inducing EMT, migration and metastasis in colorectal cancer cells[39]. Acetyltransferase NAT10 catalyzes mRNA acetylation to enhance translation efficiency and depletion of NAT10 abrogate resistance to DNA-damaging agents in breast cancer[40, 41]. The selected hub RBPs play a critical role in numerous cancers. However, the function of most hub RBPs in gastric cancer remains unknown, which still needs to be explored in the future research.

Besides, the hub RBPs were selected based on univariate Cox regression analysis, lasso regression analysis, survival analysis, and multiple Cox regression analysis. A total of five RBPs were identified as prognosis related hub RBPs, including RNASE1, SETD7,

BOLL, PPARGC1B, MSI2. Previous researches have shown that the expression of RNASE1 and SETD7 were associated with carcinogenesis and progression of gastric cancer, which are in agreement with our results[42, 43]. However, the role of BOLL and PPARGC1B remain to be further elucidated. Then, we established a risk model to predict GC prognosis by multivariate Cox regression analysis based on the five hub RBPs-related genes, trained using the TCGA cohort. The ROC curve analysis showed that this prognostic model has a moderately acceptable diagnostic capability to screen out the GC patients with poor prognosis. However, the molecular mechanism of these five RBPs contributes to gastric cancer progression is still poorly understood, and further exploration of potential mechanisms may be necessary. Subsequently, a nomogram was built to help us predict 1, 3, and 5 years OS more intuitively. Kaplan Meier plotter was used to detect the prognostic value of the five RBPs coding gene, the results were basically consistent with the prognostic analysis results of TCGA cohort. These results suggested that the prognostic model of five genes signature has a certain value in adjusting treatment plans and predicting the prognosis of gastric cancer patients. To the best of our knowledge, this is the first study to establish an RBPs-associated prognostic model for gastric cancer. Based on just five RBPs coding genes, this prognostic model displays certain capability to distinguish patients with poor prognosis, which is economically and clinically feasible. Furthermore, the RBPs-associated gene signature exhibited critical biological function, suggesting that they can potentially be used for clinical assistant treatment. Nevertheless, there are several limitations in this study. Firstly, the results were biased to some extent due to the normal samples were

far less than STAD specimen. Secondly, our result was a single omics study, which requires further multi-omics study to deepen our understanding comprehensively. Thirdly, our prognostic model was simply based on the TCGA dataset. Future validation of clinical patient cohort and other databases is needed. Finally, further in vitro and in vivo experiments are warranted to fully elucidate the molecular mechanisms of RBPs in GC.

Conclusion

To summarize, the purposes of this study were to expound the potential value of differentially expressed RBPs via bioinformatics analyses in GC. The prognostic model of five RBPs coding gene was constructed, which might serve as an independent prognostic factor for GC. Our results might help illustrate the pathogenesis of GC and these biomarkers might be considered for new target therapy and prognostic biomarkers.

Abbreviations

GC: Gastric cancer; RBPs : RNA binding proteins; LASSO: The Least Absolute Shrinkage and Selection Operator (LASSO); TCGA: The Atlas Cancer Genome; OS: Overall survival; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: Protein–protein interaction; BP: Biological process; CC: Cell components; MF: Molecular functions; STAD: Stomach adenocarcinoma.

Acknowledgments

Not applicable.

Authors' contributions

ML conceived, designed, analyzed the data, and wrote the manuscript. JX helped to collect and analyze the data. Zelong Han and Side Liu supervised the research process.

All authors read and approved the final manuscript.

Funding

This project was supported and funded by the Medical Scientific Research Foundation of Guangdong Province, China, grant number A2019538; the President Foundation of Nanfang Hospital, grant number 2017C046; the Guangdong Gastrointestinal Disease Research Center, grant number 2017B020209003; and the Special Scientific Research Fund of Public Welfare Profession of National Health and Family Planning Commission, grant number 201502026..

Ethics approval and consent to participate

No ethical approval was required by the local ethics committees.

Availability of data and materials

The data supporting the findings of this study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Competing interests

All participated authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author details

¹ Department of Gastroenterology, Nanfang Hospital, Southern Medical University, Guangzhou, China;

Figure Legends

Figure 1. The differentially expressed RBPs in gastric cancer and normal samples. **a** Hierarchical clustering distribution of differentially expressed RBPs in normal and tumor samples. **b** Volcano plot of differentially expressed RBPs. Red: up-regulated RBPs, Blue: down-regulated RBPs.

Figure 2. The gene ontology and signal pathway enrichments analysis. **a** The top ten functional enrichment analysis of up-regulated DEGs in biological process, cellular component and molecular function group, respectively. **b** The top ten functional enrichment analysis of down-regulated DEGs. **c** The significantly enriched KEGG pathways of up-regulated DEGs. **d** The significantly enriched KEGG pathways of down-regulated DEGs.

Figure 3. The protein-protein interaction (PPI) networks establishment and significant gene modules analysis. **a** The PPI networks of differentially expressed RBPs (Orange and blue represents the most two significant gene modules, respectively). **b** Critical module from PPI network. Green circles indicate down-regulated RBPs; red circles indicate upregulated RBPs.

Figure 4. Screening of the optimal RBPs for construction of the prognostic model. **a** Univariate Cox regression analysis for identification of hub RBPs in the training dataset. **b** Multivariate Cox regression analysis to identify prognosis related hub RBPs. **c** Lasso coefficient spectrum of 10 RBPs in STAD. Generate a coefficient distribution map for a logarithmic (λ) sequence. **d** Selecting the best parameters for STAD in the Lasso model (λ).

Figure 5. Development of a prognostic signature based on the hub RBPs in the training cohort. **a** Kaplan–Meier analysis of TCGA-STAD training group was stratified by median risk. The STAD cases in the high-risk group had significant shorter OS than those in the low-risk group. **b** ROC curves for predicting OS based on risk score. **c** Risk score distribution of patients with different risks (low, blue; high, red) in the training group. **d** Scatterplots of patients with different survival status in training group. **e** Heatmap of RBPs gene expression profiles in patients with different risks (low, pink; high, blue) in the training group.

Figure 6. Validation of the prognostic signature with the testing cohort. **a** Kaplan–Meier analysis of TCGA-STAD testing group was stratified by median risk. The STAD cases in the high-risk group had significant shorter OS than those in the low-risk group. **b** ROC curves for predicting OS based on risk score. **c** Risk score distribution of patients with different risks (low, blue; high, red) in the testing group. **d** Scatterplots of patients with different survival status in testing group. **e** Heatmap of RBPs gene expression profiles in patients with different risks (low, pink; high, blue) in the testing group.

Figure 7. Establishment of a prognostic RBPs-based nomogram model. The predictive performance of the RBPs-based nomogram was evaluated by calibration curves. The calibration plots revealed that the nomogram model displayed excellent capacity for forecasting the 3-year OS (b) and 5-year OS (c).

Figure 8. Validation of the prognostic value of five RBPs in STAD by Kaplan Meier-plotter.

Figure 9. Gene Set Enrichment Analysis for high-risk patients ($P < 0.05$).

Table I. Five prognosis-related hub RBPs determined by multivariate Cox regression analysis.

Table II. The prognostic value of different clinical parameters.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA: a cancer journal for clinicians* 2018, **68**(6):394-424.
2. Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H: **Gastric cancer.** *Lancet (London, England)* 2016, **388**(10060):2654-2664.
3. Yu H, Yang AM, Lu XH, Zhou WX, Yao F, Fei GJ, Guo T, Yao LQ, He LP, Wang BM: **Magnifying narrow-band imaging endoscopy is superior in diagnosis of early gastric cancer.** *World journal of gastroenterology* 2015, **21**(30):9156-9162.
4. Bando E, Makuuchi R, Irino T, Tanizawa Y, Kawamura T, Terashima M: **Validation of the prognostic impact of the new tumor-node-metastasis clinical staging in patients with gastric cancer.** *Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association* 2019, **22**(1):123-129.
5. Lin X, Zhao Y, Song WM, Zhang B: **Molecular classification and prediction in gastric cancer.** *Computational and structural biotechnology journal* 2015, **13**:448-458.
6. Moore MJ, Proudfoot NJ: **Pre-mRNA processing reaches back to transcription and ahead to translation.** *Cell* 2009, **136**(4):688-700.
7. Fu XD, Ares M, Jr.: **Context-dependent control of alternative splicing by RNA-binding proteins.** *Nature reviews Genetics* 2014, **15**(10):689-701.
8. Glisovic T, Bachorik JL, Yong J, Dreyfuss G: **RNA-binding proteins and post-transcriptional gene regulation.** *FEBS letters* 2008, **582**(14):1977-1986.
9. Nutter CA, Kuyumcu-Martinez MN: **Emerging roles of RNA-binding proteins in diabetes and their therapeutic potential in diabetic complications.** *Wiley interdisciplinary reviews RNA* 2018, **9**(2).
10. Duan Y, Du A, Gu J, Duan G, Wang C, Gui X, Ma Z, Qian B, Deng X, Zhang K *et al*: **PARYlation regulates stress granule dynamics, phase separation, and neurotoxicity of disease-related RNA-binding proteins.** *Cell research* 2019, **29**(3):233-247.
11. Son YO, Kim HE, Choi WS, Chun CH, Chun JS: **RNA-binding protein ZFP36L1 regulates osteoarthritis by modulating members of the heat shock protein 70 family.** *Nature communications* 2019, **10**(1):77.
12. Zhao L, Cao J, Hu K, Wang P, Li G, He X, Tong T, Han L: **RNA-binding protein RPS3 contributes to hepatocarcinogenesis by post-transcriptionally up-regulating SIRT1.** *Nucleic acids research* 2019, **47**(4):2011-2028.
13. Dang H, Takai A, Forgues M, Pomyen Y, Mou H, Xue W, Ray D, Ha KCH, Morris QD, Hughes TR *et al*: **Oncogenic Activation of the RNA Binding Protein NELFE and MYC Signaling in Hepatocellular Carcinoma.** *Cancer cell* 2017, **32**(1):101-114.e108.
14. Wurth L, Papasaikas P, Olmeda D, Bley N, Calvo GT, Guerrero S, Cerezo-Wallis D, Martinez-Useros J, Garcia-Fernandez M, Huttelmaier S *et al*: **UNR/CSDE1 Drives a Post-transcriptional Program to Promote Melanoma Invasion and Metastasis.** *Cancer cell* 2016, **30**(5):694-707.
15. Itskovich SS, Gurunathan A, Clark J, Burwinkel M, Wunderlich M, Berger MR, Kulkarni A, Chetal K, Venkatasubramanian M, Salomonis N *et al*: **MBNL1 regulates essential alternative RNA splicing patterns in MLL-rearranged leukemia.** *Nature communications* 2020, **11**(1):2369.
16. Wu X, Gardashova G, Lan L, Han S, Zhong C, Marquez RT, Wei L, Wood S, Roy S, Gowthaman R

- et al*: **Targeting the interaction between RNA-binding protein HuR and FOXQ1 suppresses breast cancer invasion and metastasis.** *Communications biology* 2020, **3**(1):193.
17. Mitobe Y, Iino K, Takayama KI, Ikeda K, Suzuki T, Aogi K, Kawabata H, Suzuki Y, Horie-Inoue K, Inoue S: **PSF promotes ER-positive breast cancer progression via posttranscriptional regulation of ESR1 and SCFD2.** *Cancer research* 2020.
 18. Kechavarzi B, Janga SC: **Dissecting the expression landscape of RNA-binding proteins in human cancers.** *Genome biology* 2014, **15**(1):R14.
 19. Wang K, Li L, Fu L, Yuan Y, Dai H, Zhu T, Zhou Y, Yuan F: **Integrated Bioinformatics Analysis the Function of RNA Binding Proteins (RBPs) and Their Prognostic Value in Breast Cancer.** *Frontiers in pharmacology* 2019, **10**:140.
 20. Li W, Li N, Gao L, You C: **Integrated analysis of the roles and prognostic value of RNA binding proteins in lung adenocarcinoma.** *PeerJ* 2020, **8**:e8509.
 21. Gerstberger S, Hafner M, Tuschl T: **A census of human RNA-binding proteins.** *Nature reviews Genetics* 2014, **15**(12):829-845.
 22. Smyth G: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol* 2004, **3**(3):Article3.
 23. Reiner A, Yekutieli D, Benjamini Y: **Identifying differentially expressed genes using false discovery rate controlling procedures.** *Bioinformatics* 2003, **19**(3):368-375.
 24. Hale M, Thapa I, Ghersi D: **FunSet: an open-source software and web server for performing and displaying Gene Ontology enrichment analysis.** *BMC Bioinformatics* 2019, **20**(1):359-359.
 25. Carbon S, Dietze H, Lewis S, Mungall C, Munoz-Torres M, Basu S, Chisholm R, Dodson R, Fey P, Thomas P *et al*: **Expansion of the Gene Ontology knowledgebase and resources.** *Nucleic Acids Res* 2017, **45**(D1):D331-D338.
 26. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Res* 2000, **28**:27-30.
 27. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters.** *OmicS* 2012, **16**(5):284-287.
 28. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, Mering CV: **STRING v9.1: protein-protein interaction networks, with increased coverage and integration.** *Nucleic Acids Res* 2013, **41**(Database issue):D808-D815.
 29. Zhou Z, Cheng Y, Jiang Y, Liu S, Zhang M, Liu J, Zhao Q: **Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis.** *Int J Biol Sci* 2018, **14**(2):124-136.
 30. Smoot M, Ono K, Ruscheinski J, Wang P, Ideker T: **Cytoscape 2.8.** *Bioinformatics* 2011, **27**(3):431-432.
 31. Jeon J, Cheong JH: **Clinical Implementation of Precision Medicine in Gastric Cancer.** *Journal of gastric cancer* 2019, **19**(3):235-253.
 32. Pereira B, Billaud M, Almeida R: **RNA-Binding Proteins in Cancer: Old Players and New Actors.** *Trends in cancer* 2017, **3**(7):506-528.
 33. Wang X, Hu H, Liu H: **RNA binding protein Lin28B confers gastric cancer cells stemness via directly binding to NRP-1.** *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* 2018, **104**:383-389.
 34. Fei T, Chen Y, Xiao T, Li W, Cato L, Zhang P, Cotter MB, Bowden M, Lis RT, Zhao SG *et al*: **Genome-**

- wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing.** *Proceedings of the National Academy of Sciences of the United States of America* 2017, **114**(26):E5207-e5215.
35. Poria DK, Guha A, Nandi I, Ray PS: **RNA-binding protein HuR sequesters microRNA-21 to prevent translation repression of proinflammatory tumor suppressor gene programmed cell death 4.** *Oncogene* 2016, **35**(13):1703-1715.
36. Chen J, Dong S, Hu J, Duan B, Yao J, Zhang R, Zhou H, Sheng H, Gao H, Li S *et al*: **Guanine nucleotide binding protein-like 3 is a potential prognosis indicator of gastric cancer.** *International journal of clinical and experimental pathology* 2015, **8**(10):13273-13278.
37. Li L, Pang X, Zhu Z, Lu L, Yang J, Cao J, Fei S: **GTPBP4 Promotes Gastric Cancer Progression via Regulating P53 Activity.** *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 2018, **45**(2):667-676.
38. Gupta R, Bugide S, Wang B, Green MR, Johnson DB, Wajapeyee N: **Loss of BOP1 confers resistance to BRAF kinase inhibitors in melanoma by activating MAP kinase pathway.** *Proceedings of the National Academy of Sciences of the United States of America* 2019, **116**(10):4583-4591.
39. Qi J, Yu Y, Akilli Ozturk O, Holland JD, Besser D, Fritzmann J, Wulf-Goldenberg A, Eckert K, Fichtner I, Birchmeier W: **New Wnt/beta-catenin target genes promote experimental metastasis and migration of colorectal cancer cells through different signals.** *Gut* 2016, **65**(10):1690-1701.
40. Arango D, Sturgill D, Alhusaini N, Dillman AA, Sweet TJ, Hanson G, Hosogane M, Sinclair WR, Nanan KK, Mandler MD *et al*: **Acetylation of Cytidine in mRNA Promotes Translation Efficiency.** *Cell* 2018, **175**(7):1872-1886.e1824.
41. Liu HY, Liu YY, Yang F, Zhang L, Zhang FL, Hu X, Shao ZM, Li DQ: **Acetylation of MORC2 by NAT10 regulates cell-cycle checkpoint control and resistance to DNA-damaging chemotherapy and radiotherapy in breast cancer.** *Nucleic acids research* 2020, **48**(7):3638-3656.
42. Wang L, Zhu JS, Song MQ, Chen GQ, Chen JL: **Comparison of gene expression profiles between primary tumor and metastatic lesions in gastric cancer patients using laser microdissection and cDNA microarray.** *World journal of gastroenterology* 2006, **12**(43):6949-6954.
43. Guo T, Wen XZ, Li ZY, Han HB, Zhang CG, Bai YH, Xing XF, Cheng XJ, Du H, Hu Y *et al*: **ISL1 predicts poor outcomes for patients with gastric cancer and drives tumor progression through binding to the ZEB1 promoter together with SETD7.** *Cell death & disease* 2019, **10**(2):33.

Figures

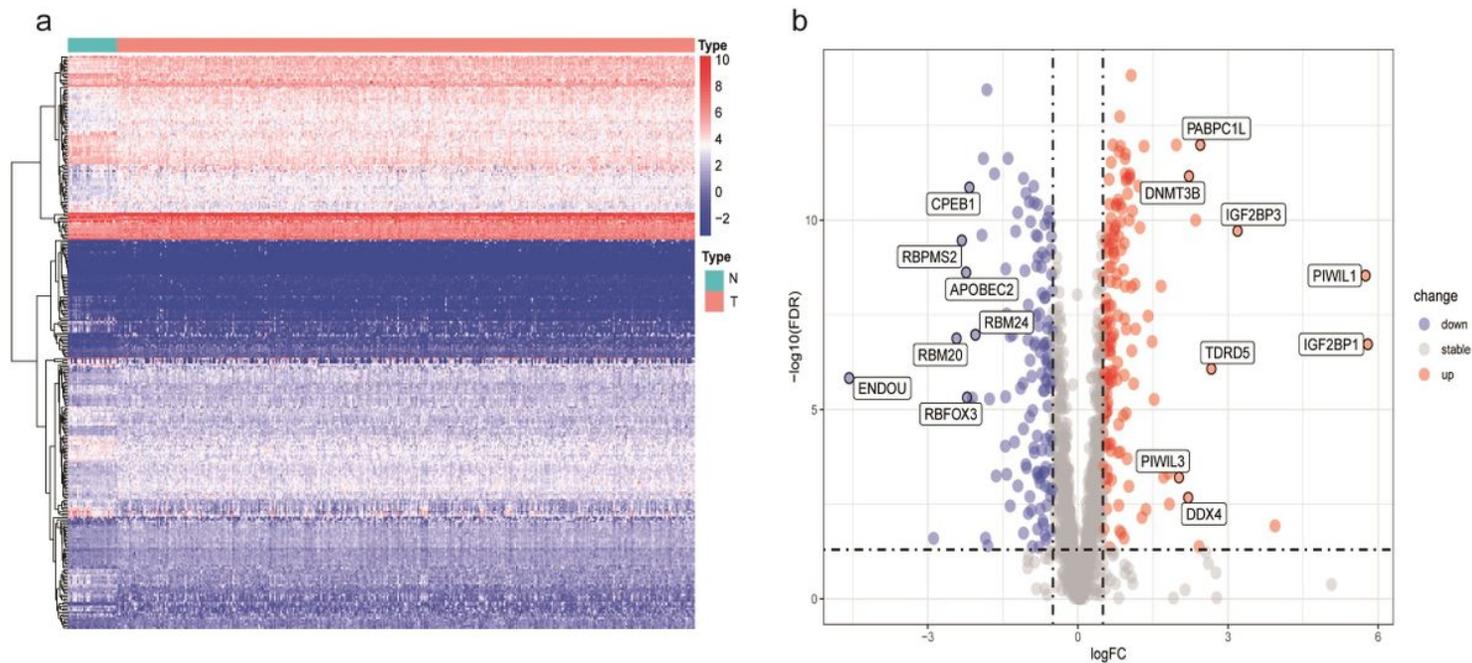


Figure 1

The differentially expressed RBPs in gastric cancer and normal samples. a Hierarchical clustering distribution of differentially expressed RBPs in normal and tumor samples. b Volcano plot of differentially expressed RBPs. Red: up-regulated RBPs, Blue: down-regulated RBPs.

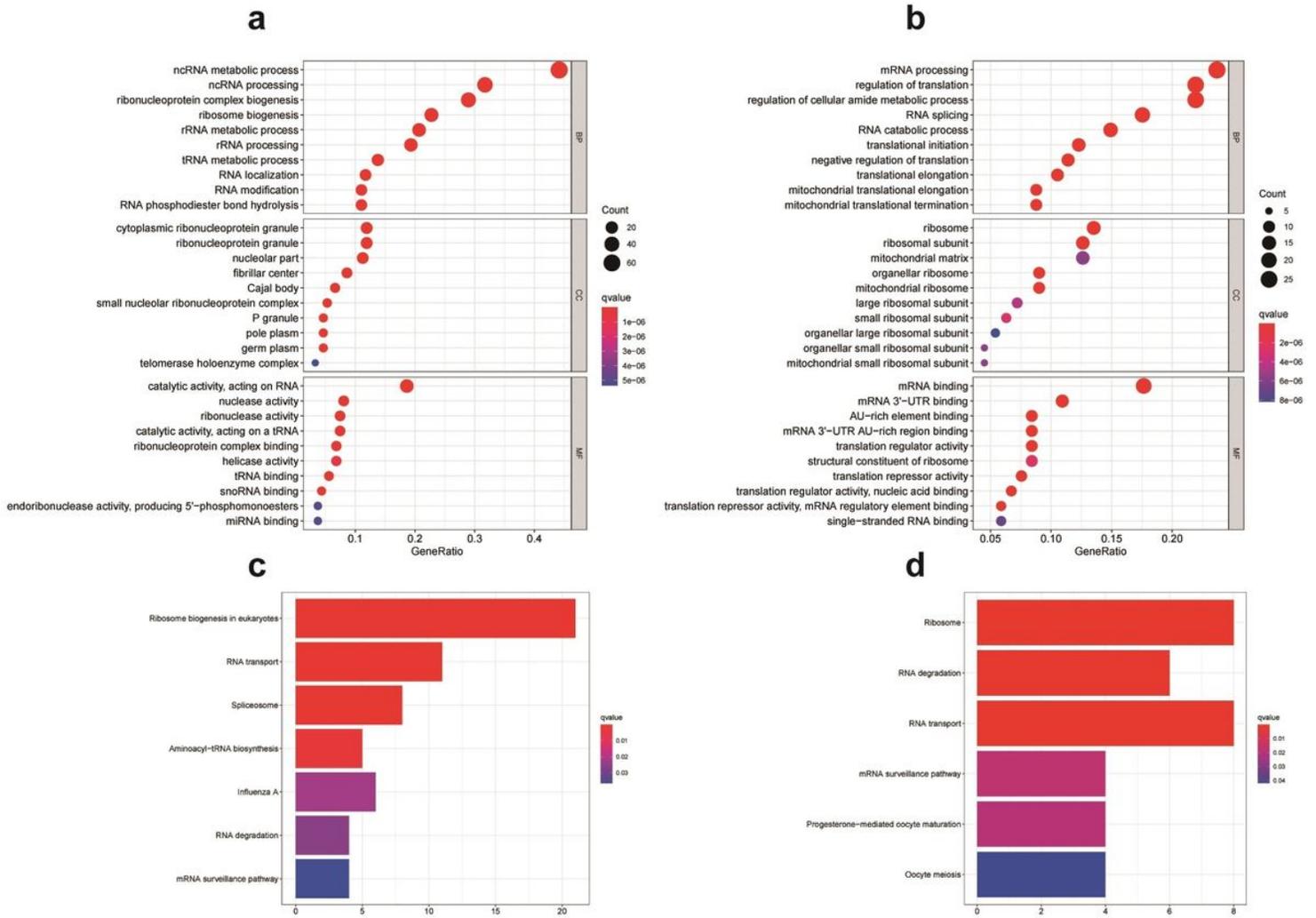


Figure 2

The gene ontology and signal pathway enrichments analysis. a The top ten functional enrichment analysis of up-regulated DEGs in biological process, cellular component and molecular function group, respectively. b The top ten functional enrichment analysis of down-regulated DEGs. c The significantly enriched KEGG pathways of up-regulated DEGs. d The significantly enriched KEGG pathways of down-regulated DEGs.

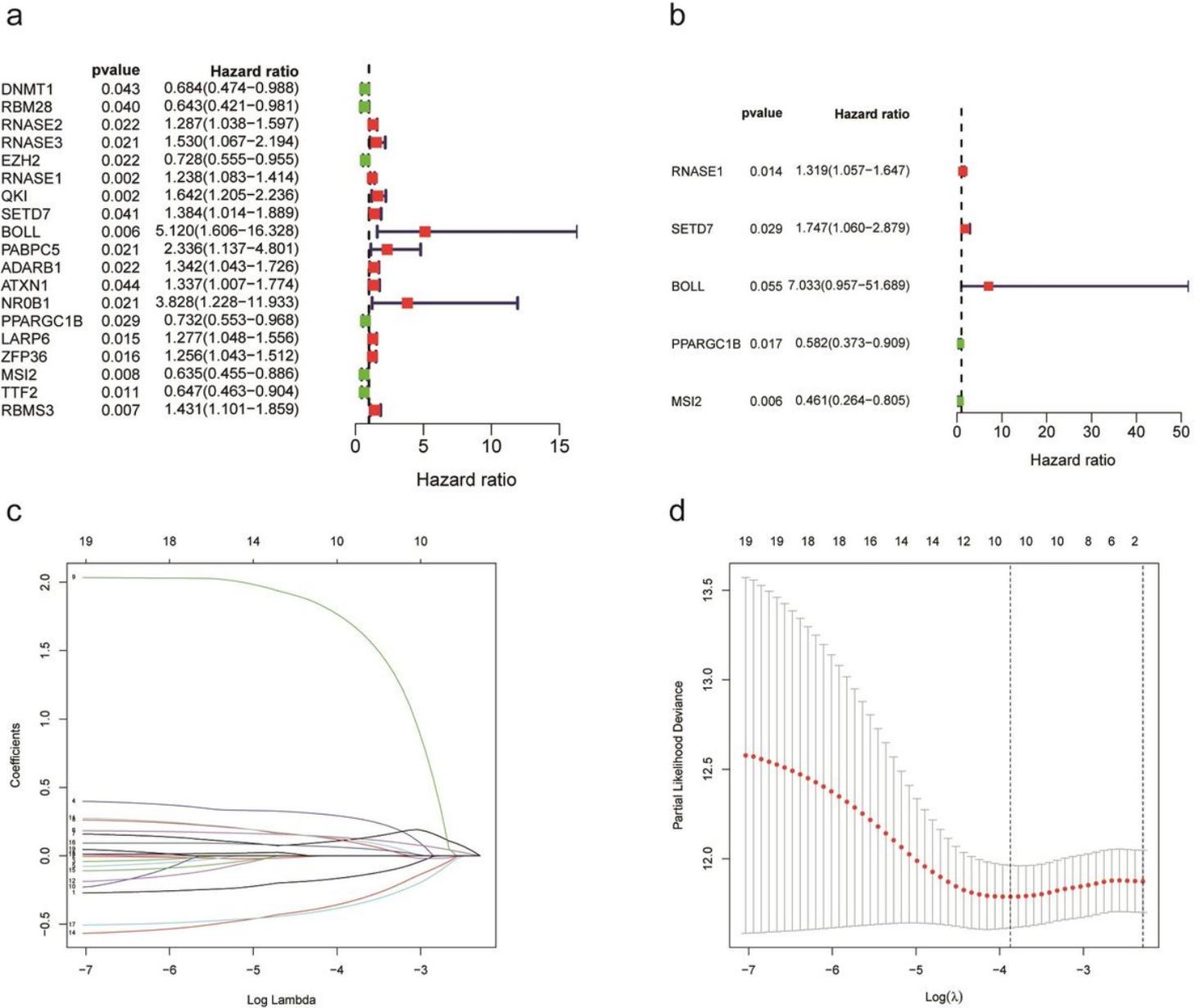


Figure 4

Screening of the optimal RBPs for construction of the prognostic model. a Univariate Cox regression analysis for identification of hub RBPs in the training dataset. b Multivariate Cox regression analysis to identify prognosis related hub RBPs. c Lasso coefficient spectrum of 10 RBPs in STAD. Generate a coefficient distribution map for a logarithmic (λ) sequence. d Selecting the best parameters for STAD in the Lasso model (λ).

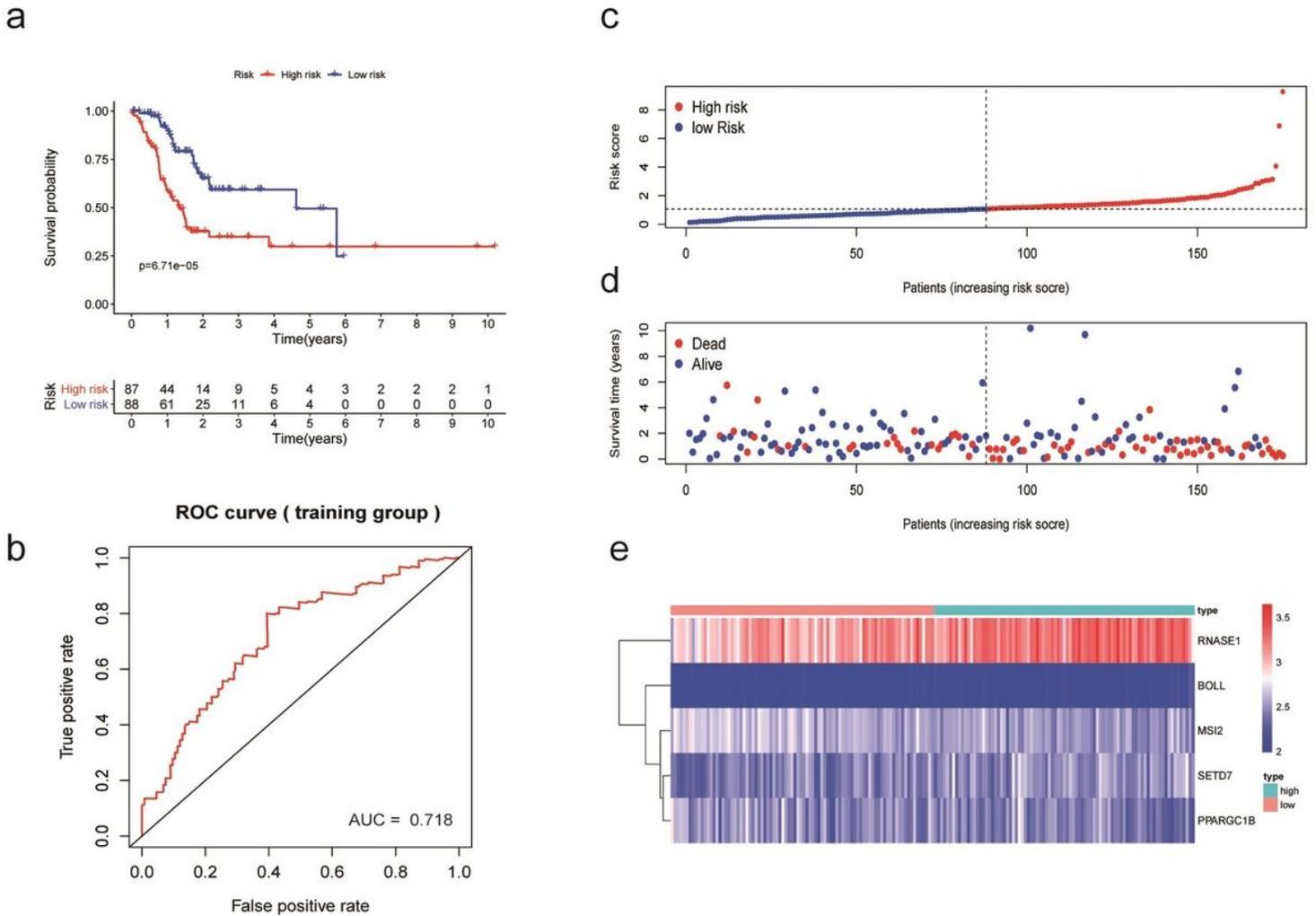


Figure 5

Development of a prognostic signature based on the hub RBPs in the training cohort. a Kaplan–Meier analysis of TCGA-STAD training group was stratified by median risk. The STAD cases in the high-risk group had significant shorter OS than those in the low-risk group. b ROC curves for predicting OS based on risk score. c Risk score distribution of patients with different risks (low, blue; high, red) in the training group. d Scatterplots of patients with different survival status in training group. e Heatmap of RBPs gene expression profiles in patients with different risks (low, pink; high, blue) in the training group.

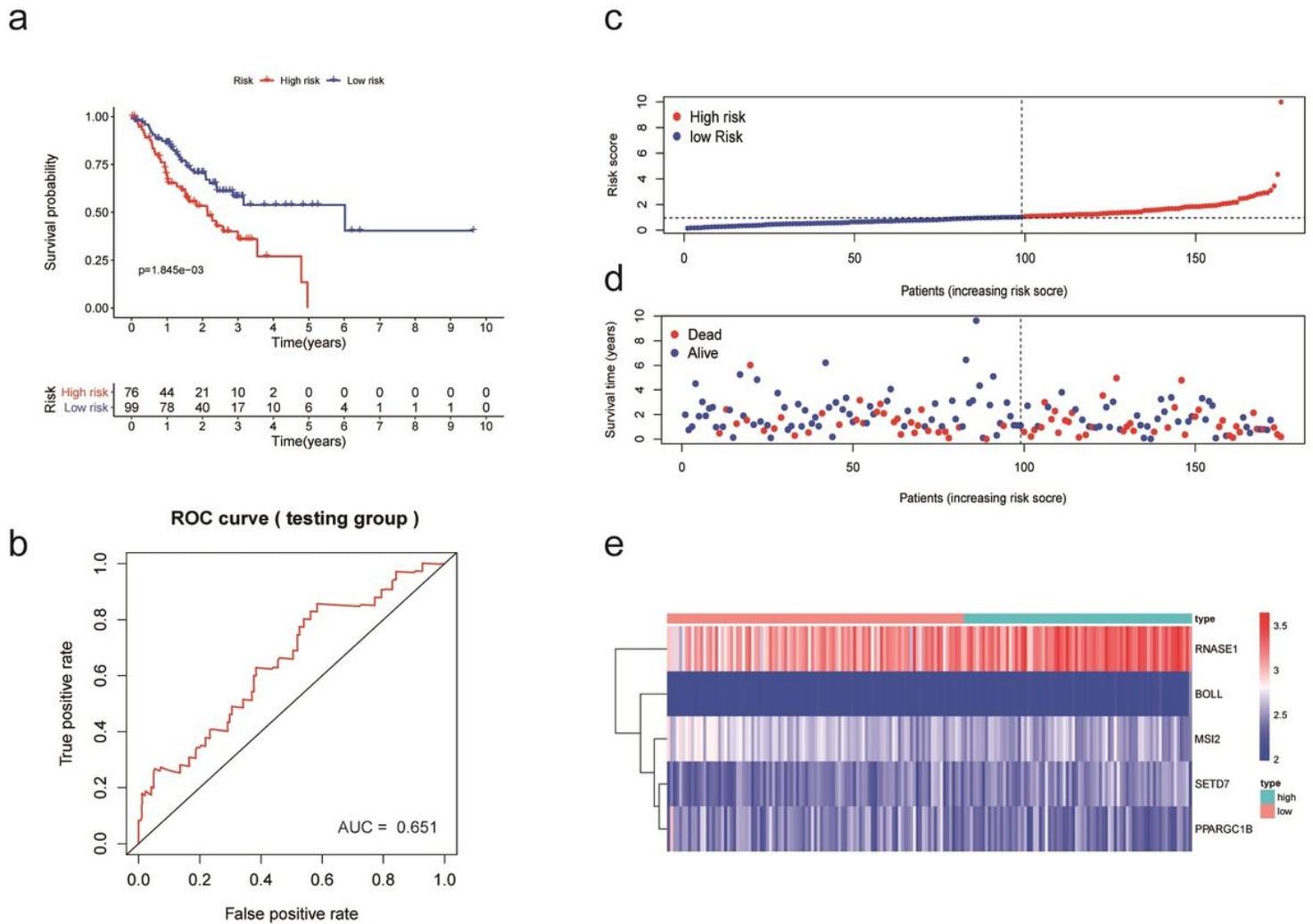


Figure 6

Validation of the prognostic signature with the testing cohort. a Kaplan–Meier analysis of TCGA-STAD testing group was stratified by median risk. The STAD cases in the high-risk group had significant shorter OS than those in the low-risk group. b ROC curves for predicting OS based on risk score. c Risk score distribution of patients with different risks (low, blue; high, red) in the testing group. d Scatterplots of patients with different survival status in testing group. e Heatmap of RBPs gene expression profiles in patients with different risks (low, pink; high, blue) in the testing group.

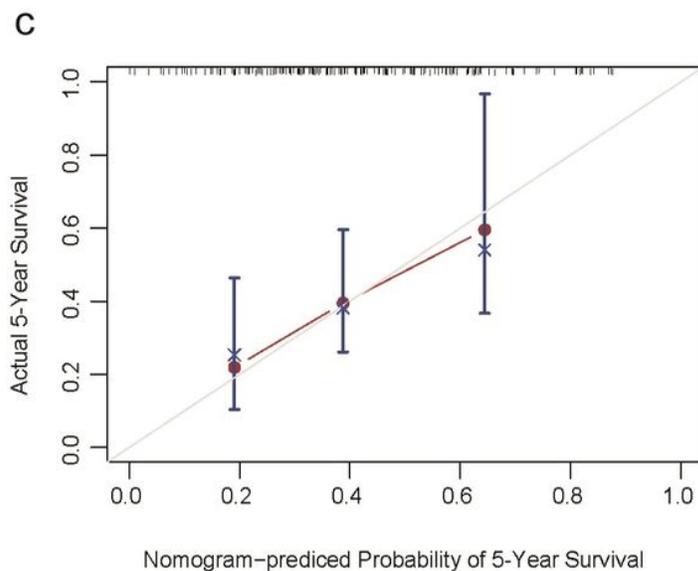
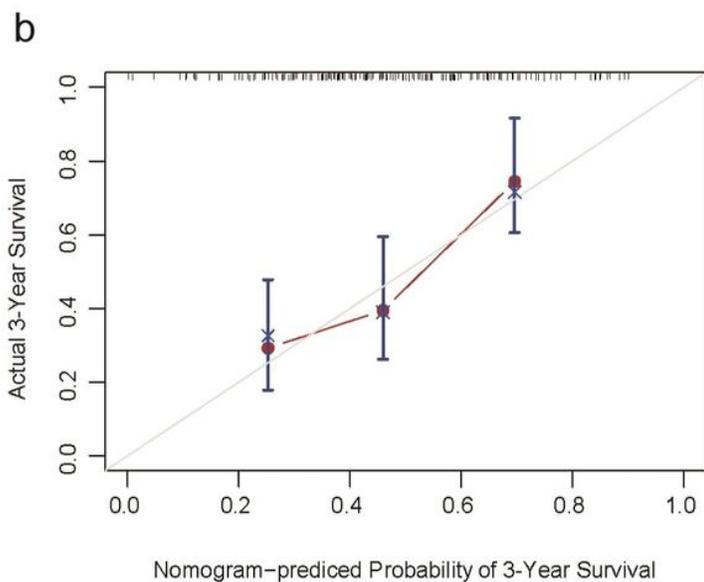
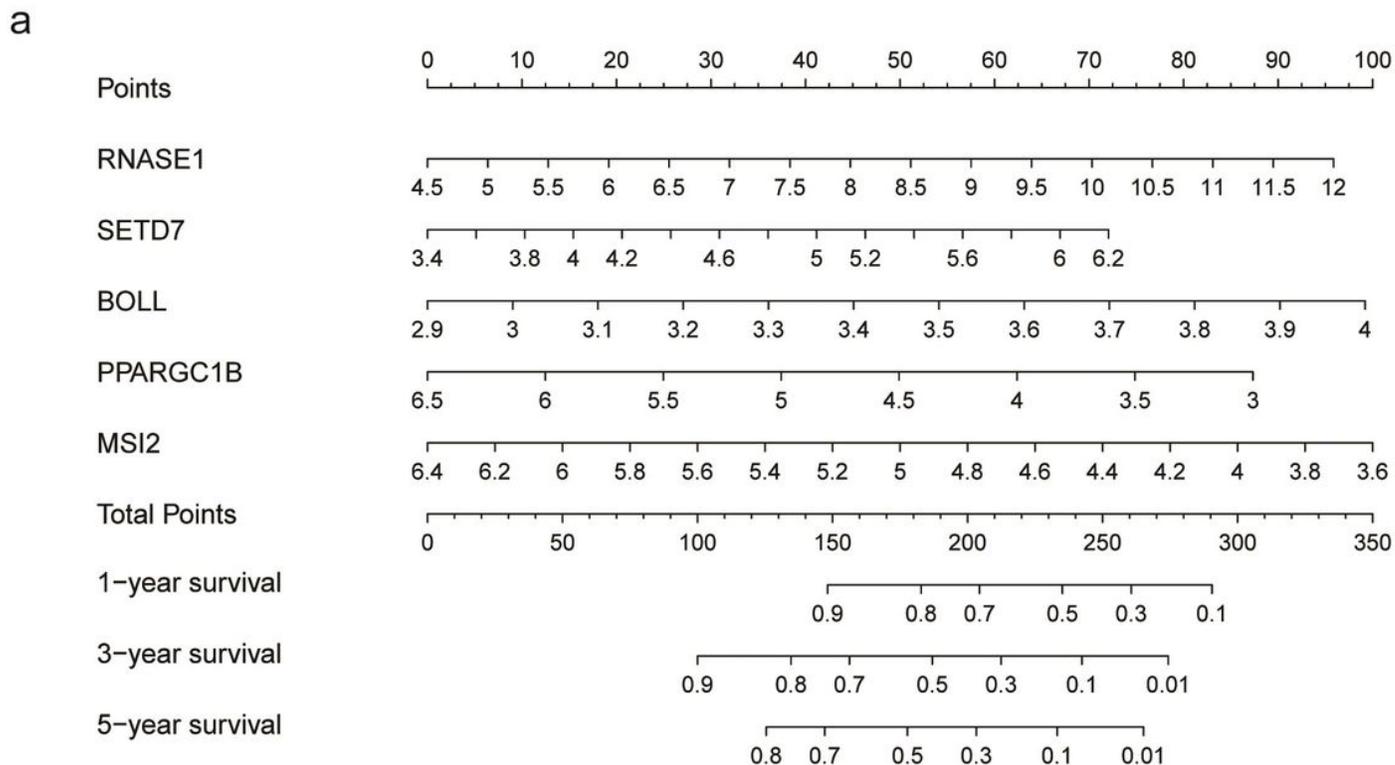


Figure 7

Establishment of a prognostic RBPs-based nomogram model. The predictive performance of the RBPs-based nomogram was evaluated by calibration curves. The calibration plots revealed that the nomogram model displayed excellent capacity for forecasting the 3-year OS (b) and 5-year OS (c).

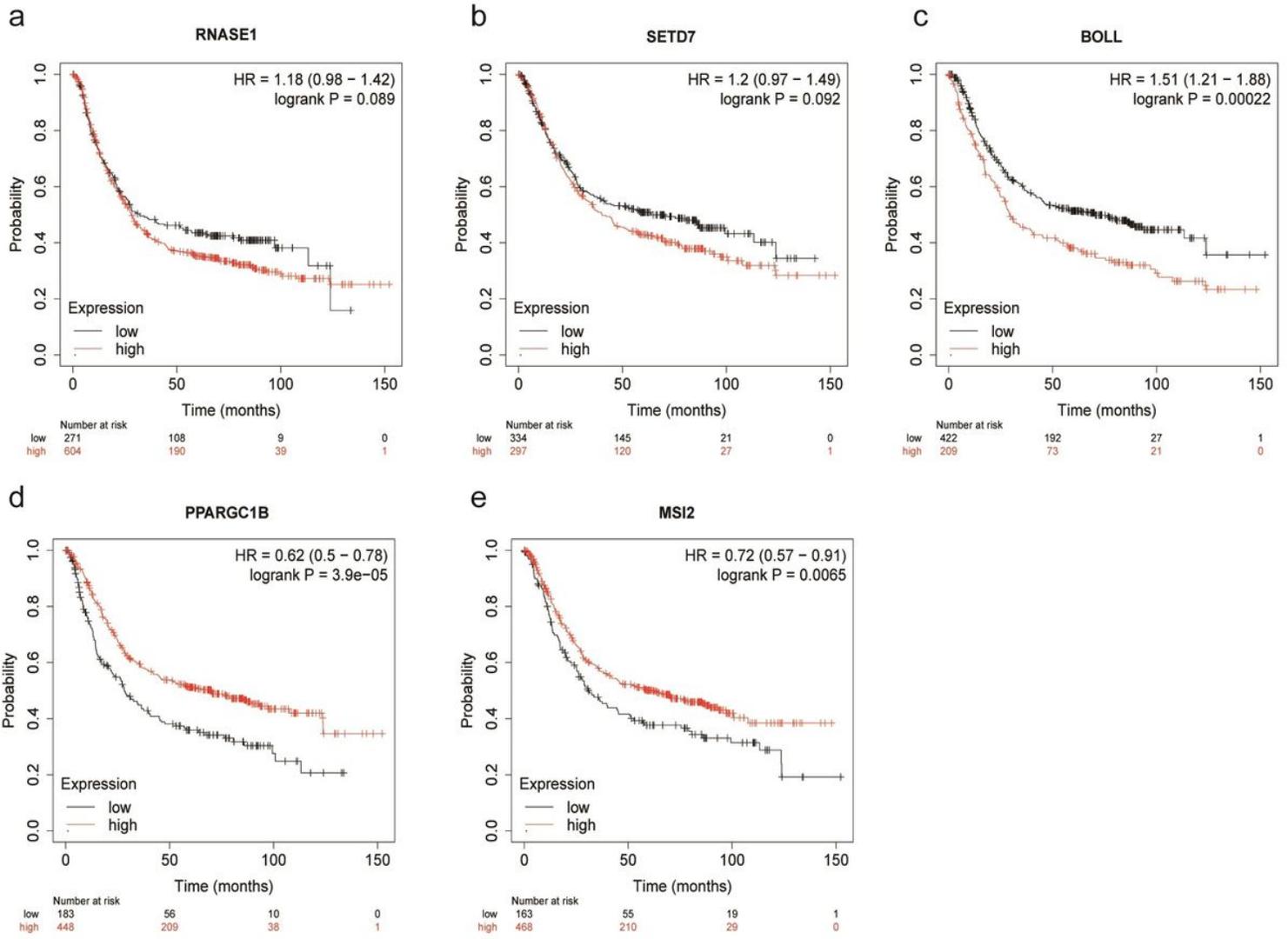


Figure 8

Validation of the prognostic value of five RBPs in STAD by Kaplan Meier-plotter.

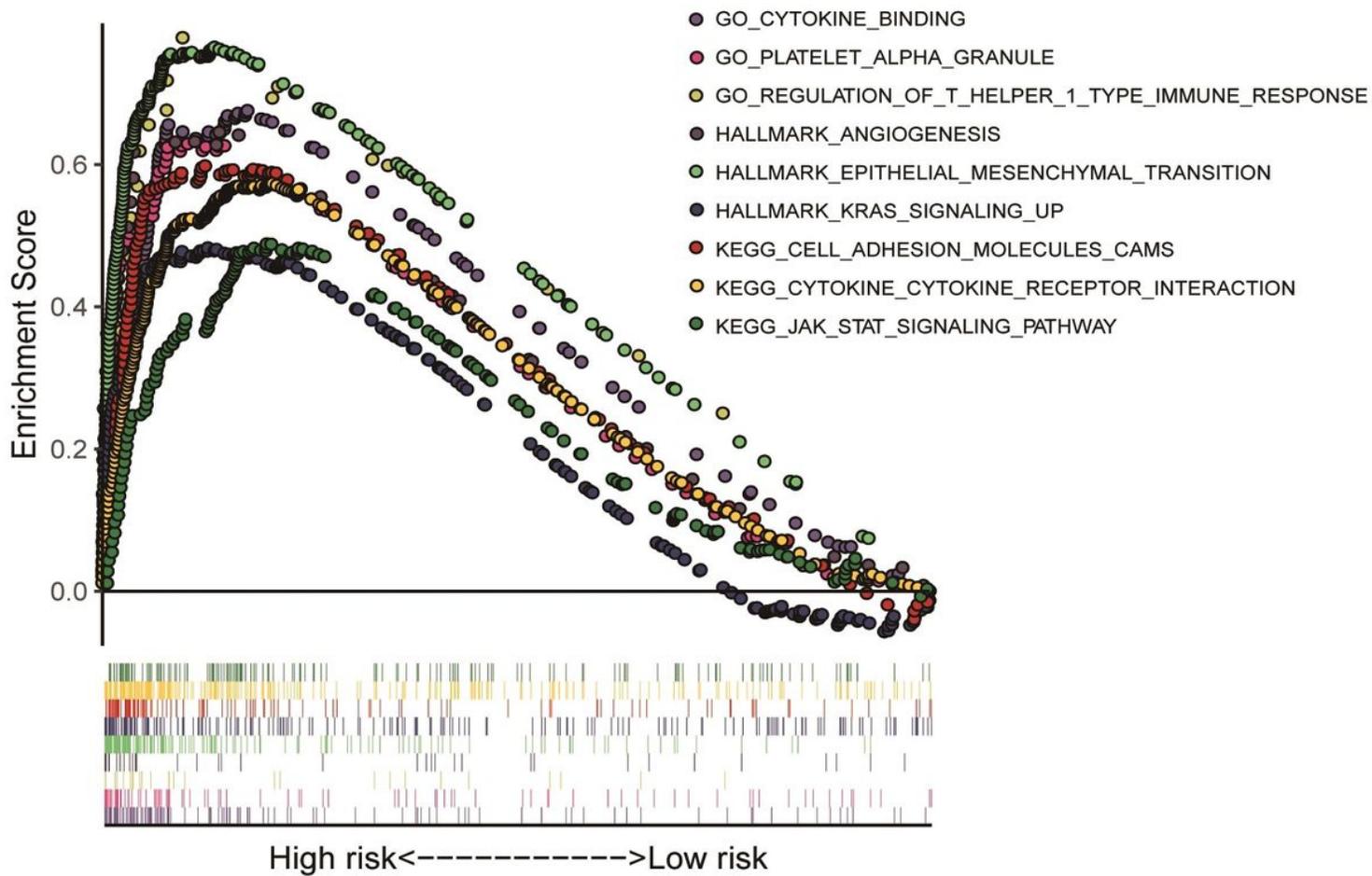


Figure 9

Gene Set Enrichment Analysis for high-risk patients ($P < 0.05$).