Construction and Validation of a Prognostic Model of Gastric Cancer with 10RBPs Based on 6 Microarrays

Liqiang Zhou  
Second Affiliated Hospital of Nanchang University

Hao Lu  
Second Affiliated Hospital of Nanchang University

Lin Xin  
Second Affiliated Hospital of Nanchang University

Qi Zhou  
Second Affiliated Hospital of Nanchang University

You Wu  
Second Affiliated Hospital of Nanchang University

Shihao Li  
Second Affiliated Hospital of Nanchang University

Yiwu Yuan  
Second Affiliated Hospital of Nanchang University

Research Article

Keywords: RNA binding proteins (RBPs), gastric cancer, Gene Expression Omnibus (GEO), Express database

DOI: https://doi.org/10.21203/rs.3.rs-117996/v1

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

For explore the potential connection of RNA binding proteins (RBPs) to the expression function of gastric cancer (GC). We download the GPL10558 and GPL6947 platform microarray data from Gene Expression Omnibus (GEO) and Express database. Then the system integrates and analyzes the differentially expressed RBPs. And enrich the differentially expressed RBPs to understand the mechanism of its influence on tumors. Univariate Cox, lasso regression and multivariate Cox regression analysis were used to screen independent prognostic parameters to construct prognostic model, and calculate area under time-dependent receiver operating characteristics (AUC) and survival analysis were used to evaluate their prognostic ability. GSE15459, GSE62254 cohorts were used to verify hub signature. Finally, we also verified the prognosis and expression of hub-RBPs. Systematic analysis identified 23 up-regulated and 30 down-regulated RBPs, and enrichment analysis showed that they mainly affect their modification by binding to mRNA, and their stability affects the progression of GC. After multiple statistical analyses, we obtained the prognostic signature constructed by 10 RBPs and determined that it has better predictive performance (AUC = 0.685). Through comprehensive bioinformatics analysis, we have obtained 10 key gastric cancer RBPs as potential prognostic biomarkers, providing new perspectives for the treatment and prognostic of GC.

Introduction

GC is one of the common human diseases and its pathogenesis is related to many factors, such as Helicobacter pylori infection and Epstein-Barr virus infection, high salt intake, low fruit and vegetables. According to the latest data, there are still more than one million new GC worldwide every year. Although the incidence and mortality of GC have declined globally in the past 5 years, however GC is still the third leading cause of cancer-related deaths. In recent years, gastroscopy has been popularized in many countries, the early diagnosis of GC has been improved to a certain extent, but because early GC often has no obvious clinical symptoms, most GC is already in advanced stage at the time of diagnosis and the patient population is gradually younger. According to research, the 5-year survival rate of advanced GC in China is less than 30%. In addition, Traditional tumor markers such as CEA and CA199 lack certain sensitivity and specificity in the early stage of GC. Therefore, it is necessary to find new marks to predict postoperative recurrence of GC. These evidences indicate that RBPs play an indispensable role in tumors.

RBPs are an inherent pleiotropic binding protein, which mainly plays a key role in post-transcriptional gene regulation. They can interact with target mRNA in a sequence and structure-dependent manner, manipulate the processing of these mRNAs to determine cell behavior such as: tumor proliferation, differentiation, invasion and angiogenesis. Therefore, when the function of RBPs is abnormal, it often indicates that it is related to most diseases such as: human fragile X syndrome, myotonic dystrophy, neurological diseases and cancer etc. Humans have identified more than 1,500 cancer-related RBPs in genome screening, which is equivalent to more than 7.5% of the total protein-coding genes in the human genome. About 50% of RBPs have direct or indirect effects on mRNA transcription and post-
transcriptional processes influences. It can be found in most cancers that RBPs show abundant mutations and overexpression such as: GF2BP1, IFIT1B, PABPC1, TLR8, GAPDH, PIWIL4, RNPC3, and ZC3H12C are closely related to the prognosis of lung cancer. The high expression of APOBEC3C and EIF4E3 will have a better prognosis in breast cancer.

Multiple reports suggest that RBPs can affect the occurrence and development of GC. For example, gastric cancer patients with HUR high expression and TTP low expression has a worse prognosis. QKI5 acts as a tumor suppressor gene through H2A-type histone variant. In addition, LIN28 directly binds to neuropilin-1 (NRF-1) 30UTR and enhances its mRNA stability and reduce the sensitivity of platinum drugs. However, there is currently no systematic research on RBPs in GC. In this study, we systematically analyzed the expression of RBPs in GC by integrating the Illumina HumanHT-12 microarray data and analyzed the mechanism of its influence on tumor progression. We also constructed a prognostic signature of 10 RBPs, and verified it on GSE15459 and GSE62254, showing the accuracy and universality of the signature. In addition, we constructed a nomogram based on the signature, and verified the accuracy of its prediction. Finally, we also verified the expression and prognosis of 10 RBPs, and the results are consistent with our results. In short, we provide a new idea for the prognosis and treatment of GC, which will help improve the quality of life of GC patients.

Results

Identification of differently expressed RBPs (DERBPs) in GC patients.

In this study, we downloaded GC-related data from the GEO database and the original data of the chip from the Array Express database. After data processing, a total of 827 tumor samples and 118 normal samples were obtained. According to the analysis of the difference conditions, a total of 1261 probes (1102 genes) were finally obtained. Then we further screened the differentially expressed RBPs and obtained 61 probes (53 RBPs). Further analysis of the data found that 23 of them were up-regulated and 30 were down-regulated (Table S1). According to the differentially expressed RBPs we got, make heat map (Fig. 1).

PPI network construct and DERBPs enrichment analysis.

We construct a PPI network based on these differently expressed RBPs, the red gene in the figure indicates up-regulation, and blue indicates down-regulation (Fig. 2). In total, a network with 46 nodes and 183 edges was obtained (Fig. 2a). Through MCODE we got 2 key modules (Fig. 2b, c). One key Module contains 10 nodes and 20 edges (k = 4.444), and the other key Module contains 11 nodes and 20 edges (k = 4.000). According to the key modules in the PPI network, we conduct GO and KEEG function enrichment analysis on the RBPs of the key modules. According to the threshold, KEGG is not eligible.
About GO enrichment analysis results, mainly involved 39 items (Table S2). We can conclude that RBPs are rich in RNA catabolism and regulation, transfer, degradation, processing, nucleic acid phosphodiester bond hydrolysis, RNA binding, peptide metabolism and negative regulation of cellular amide metabolism content.

**Screening of prognostic hub RBPs.**

According to the clinical information we have collected, a total of 635 samples contain the necessary survival data and will be further studied. For the prognostic significance of these RBPs, the univariate Cox regression analysis was performed first, 18 RBPs related to prognosis were screened. Among them, there are 8 RBPs with HR > 1 and 10 with HR < 1 (Fig. 3a). Then, LASSO regression analysis was performed to further narrow the scope and 14 RBPs were identified (Fig. 3b, c). Next, multivariate Cox regression analysis will screen the hub RBPs that can independently predict the prognosis of GC (Fig. 3d). We finally obtained 10 hub RBPs, five RBPs of which have HR > 1, respectively, ANXA1 ($P = 0.044$, $HR = 1.181$), CDC20 ($P = 0.012$, $HR = 1.267$), EEF1A2 ($P = 0.049$, $HR = 1.086$), ITGB1 ($P = 0.018$, $HR = 1.303$), MYH11 ($P < 0.001$, $HR = 1.181$), and five RBPs HR < 1, respectively, ACO1 ($P = 0.078$, $HR = 0.778$), AUH ($P = 0.091$, $HR = 0.806$), CPEB3 ($P = 0.026$, $HR = 0.713$), DAZAP1 ($P = 0.010$, $HR = 0.804$), KIAA0101 ($P = 0.019$, $HR = 0.811$), MYH11 ($P < 0.001$, $HR = 1.181$).

**Prognosis-related RBPs risk score model construction and analysis.**

The 10 hub RBPs identified from the multivariate Cox regression analysis were used to construct the predictive model. The risk scores of patients were calculated according to the scoring formula:

$$Risk \ score = \ (ACO1*0.250947453) + (ANXA1*0.166026749) + (AUH*-0.215476756) + (CDC20*0.236760628) + (CPEB3*-0.337958059) + (DAZAP1*-0.218248768) + (EEF1A2*0.082820888) + (ITGB1*0.264507736) + (KIAA0101*-0.209547007)+( MYH11*0.16618769).$$

According to the median score, we divide patients into low-risk groups and high-risk groups, the patient's survival time and risk score can be obtained according to the risk score map, it also shows that the low-risk group has a better survival rate (Fig. 4a, b). We can also find that low risk score represents the better prognosis ($P = 1.6631e - 09$) (Fig. 4c). Evaluate the accuracy of the prediction model by the area of AUC of the risk scoring system, we can find that 5 years AUC = 0.685 (Fig. 4d), this result shows certain diagnostic value. Combined with the patient's clinical information, First of all, under the condition that univariate Cox does not exclude other influences, we got that Age($P = 0.002$, $HR 1.020$), N stage($P < 0.001$, $HR 1.676$), T stage($ P < 0.001$, $HR 1.740$), and then under the condition of multivariate Cox, we obtain that Age($P < 0.001$, $HR 1.022$), N stage($P < 0.001$, $HR 1.431$), T stage($ P < 0.001$, $HR 1.521$), and risk scores($P < 0.001$, $HR 1.696$) are all independent prognostic survival factors (Fig. 4e,f). Finally, in addition, to verify the sensitivity and specificity of our constructed model, we use the GSE15459 and GSE62254 databases to
verify that our model obtains the same results through the same formula and method. We found that in the GSE15459 the area under the ROC curve was 0.670, the low-risk group has a significant advantage in survival prognosis than the high-risk group \( P = 1.896 \times 10^{-3} \) and GSE62254 database 5 years AUC = 0.645, low risk scores had better OS than those with high risk scores \( P = 3.769 \times 10^{-4} \), in comparison with other clinicopathological characteristics, the risk score of the model is an independent prognostic factor, and it has a good prognostic ability (Fig. 5, 6). These results indicate that the prognostic model we constructed based on RBPs has good reproducibility and is suitable for most clinical patients.

**Bioinformatics analysis of different risk groups.**

In order to understand the GC patients between high and low risk groups, we first analyze their clinicopathological characteristics. We found that most patients in the high-risk group were in N-stage N2-N3, and most patients in the T-stage were in stage T4 with more deaths. This indicates that the high-risk group of GC patients is more malignant than the low-risk group (Fig. 7a). We further analyzed the relationship between each hub RBPs and clinical case characteristics, and the results suggest that CPEB3 is positively correlated with N staging, CDC20, MYH11 are positively correlated with T staging, and ANXA1, AUH, KIAA9191 are positively correlated with both (Fig. 7b). In addition, we also used GSEA to analyze the differences in the mechanism of GC development between high and low risk groups. We can find that TGF-beta, Angiogenesis, EMT, Wnt/β-catenin, Hypoxia, KRAS, Hedgehog, Myogenesis, Coagulation, Apical surface, Apical junction, NV response Dn and Apoptosis significantly enrichment. The enrichment of these pathways shows that it is closely related to the recurrence and distant metastasis of GC (Fig. 8).

**Construction of a nomogram based on the 10 hub RBPs.**

Based on the results of multi-factor Cox analysis, to visualize the 1–5 years of OS in patients with GC more intuitively, we use nomogram to visualize the regression (Fig. 9a). Then, we use the calibration chart to judge the accuracy of the nomogram. The results show that the slope of the red line is almost 1 (Fig. 9b). It shows that the actual survival rate and the predicted survival rate are almost similar, which suggests that the nomogram we constructed has excellent predictive power. This suggests that the nomogram we constructed can help clinicians make more accurate treatment judgments for patients and improve the quality of life of patients.

**Validation the prognostic value and expression of hub RBPs.**

We use Kaplan Meier-plotter to draw a survival curve for the RBPs constructed to further verify the prognostic value of our RBPs in GC. The survival rate of IREB1, AUH, CPEB3, DAZAP1, KIAA0101 in the high expression group was significantly better than that in the low expression group \( P < 0.05 \), this is
consistent with the results of our analysis. The survival rate of ANXA1, CDC20, EEF1A2, ITGB1, MYH11 in the low expression group was better than that in the high expression group (Fig. 10a). And using GEPIA to verify the mRNA expression levels of 10 RBPs. It can be found that CDC20, DAZAP1, ITGB1, KIAA0101 are significantly increased expression, however, EEF1A2, MYH11 are significantly reduced expression (Fig. 10b). Finally, we used the HPA database to verify the protein levels of the constructed model RBPs in normal tissues and tumor tissues. It can be found that in tumor tissues CDC20, ITGB1 are significantly high expression. ANXA1, AUH low expression; in normal tissues, MYH11 low expression, this is consistent with the results of our analysis (Fig. S1).

**Discussion**

As one of the top five malignant tumors in humans, diagnosis and treatment of GC are still facing challenges. Surgery has been recognized as the gold standard for resectable advanced GC\(^1\)\(^5\). However, there has been a lack of reliable methods for the follow-up of patients after surgery. According to research, RBPs mainly act on the occurrence of cancer by operating the “mRNA life cycle”. Therefore, the abnormal expression of RBPs is usually closely related to the prognosis of cancer patients\(^1\)\(^6\). In our study, firstly, we obtained the differentially expressed RBPs between cancer tissues and normal tissues through data processing and difference analysis. Then, we constructed PPIs of these RBPs and systematically studied the relevant biological pathways. Then, we performed GSEA analysis, K-M survival analysis and ROC analysis to explore the potential biological functions and clinical value of hub RBPs. Finally, we constructed a risk model to predict the prognosis of GC based on the 10 RBPs gene signature.

In our research, by constructing a protein-protein network for differently expressed RBPs and screening key model and key RBPs. These RBPs may lead to GC by regulating RNA and mRNA process. And then we through the GO enrichment pathway, it can be found that involves Biological Process (BP): RNA catabolic process, type I interferon signaling pathway, mRNA catabolic process etc.; Molecular Function (MF): mRNA 3'-UTR binding, translation initiation factor activity, poly(U) RNA binding, single-stranded RNA binding and so on. According to research, the regulation of RNA translation, processing, and catabolism is related to the occurrence of various diseases, and also plays a key role in the development of cancer\(^1\)\(^7\),\(^1\)\(^8\). Type I interferon can activate immune cells and cancer cells feedback inhibition is also closely related to cancer\(^1\)\(^9\). And more and more studies have shown that Type I interferon controls the autocrine or paracrine circuit that forms the basis of cancer immune monitoring. A variety of chemotherapeutic drugs are fully effective in the presence of complete type I IFN signals. Therefore, for type I interferon signaling pathway is worthy of further study\(^2\)\(^0\). In our research, it was also confirmed that RBPs are closely related to GC through the control of RNA and mRNA process, such as mRNA 3'-UTR binding and poly(U) RNA binding etc. Meanwhile, RBPs can also affect the biological process of GC by regulating type I interferon signaling pathway.

In view of the key role that RBPs play in solid tumors such as GC. We used complete statistical methods to screen out 10 RBPs with independent prognosis for GC and constructed a prognostic model. Five RBPs
with HR greater than 1 are considered dangerous genes (ANXA1, CDC20, EEF1A2, ITGB1, MYH11), and the other 5 RBPs with HR < 1 are considered protective genes (ACO1, AUH, CPEB3, DAZAP1, KIAA0101). According to reports, these hub RBPs play an important role in solid tumors. The compulsive ANXA1 expression in GC cells leads to cell growth inhibition, and at the same time acts on the development of GC by regulating the expression of COX-2\textsuperscript{21,22}. CDC20 plays a key role in the occurrence and development of tumors, and high expression of CDC20 is often accompanied by a later stage and a worse prognosis and EEF1A2 is rarely found in normal gastric tissues, but is highly expressed in GC tissues\textsuperscript{23,24}. Zhou et al. have shown that ITGB1 is sensitive to the predictive value of advanced GC\textsuperscript{25,26}, in colorectal cancer, the single nucleotide repeat (C8) of the MYH11 gene has a frameshift mutation which is one of its important mechanisms. In GC, we can also find the same gene mutation\textsuperscript{27}, above genes are also verified in our prognostic model. High expression of above genes indicates a worse prognosis, this is correlate with our research results. Increasing number of research recognizes that the expression of many RBPs in GC tissues has also changed. Moreover, studies have shown that ACO1 can interact with LINC00477 and inhibit the conversion of ACO1 from citrate to isocitrate, leading to the occurrence of GC\textsuperscript{28}. CPEB3 could inhibits epithelial-mesenchymal transition (EMT) by disrupting the crosstalk between colorectal cancer cells and tumor associated macrophages via IL-6R/STAT3 signaling\textsuperscript{29}, however, there is currently no in-depth research on GC and more research is needed in the future. Meanwhile, EMT pathway also enrichment in our study. KIAA0101 was first discovered in 2001. It is well known as a p15PAF proliferating cell nuclear antigen (PCNA) related factor\textsuperscript{30}, The study found that patients with high expression of KIAA0101 had a significantly higher postoperative recurrence rate than other patients. The main mechanism may be and its effect mRNA and protein levels are related\textsuperscript{31}. However, in our study KIAA0101 is a protective gene, which needs further research to prove. Then, we used the GSE15459 and GSE62254 data sets to verify the prognostic signature we constructed. We combined the clinicopathological characteristics for analysis. Whether it is training set or test set, risk score is an independent predictor of the prognosis of GC. The results obtained are consistent with our model, indicating that our prognostic model is accuracy. Compared to other prognostic models such as: ImmunoScore Signature\textsuperscript{32}, Long noncoding RNA (lncRNA) Prognostic of GC\textsuperscript{33}, our research is a completely new point of view.

We further analyzed the high-risk and low-risk groups by GSEA, and it can be found that the following pathways are enriched: TGF-beta signaling, Angiogenesis, Wnt/β-catenin-mediated signaling, epithelial-mesenchymal transition, Hypoxia, KRAS signaling up, Hedgehog signaling Coagulation, Myogenesis, Apical surface, Apical-junction, NV-response-DN, Apoptosis. According to research, TGF-beta signaling is an important regulatory growth factor in our body, which mainly maintains the development of tissues and the homeostasis of the internal environment, so it is related to the onset of many diseases\textsuperscript{34}. Previous reports have shown that RBPs participate in these pathways and affect tumor progression. Among them, RHBDF2 promotes the lysis and high expression of TGF-β by regulating the TGF-β signaling pathway, and accelerates the invasion of GC cells into extracellular matrix and lymphatic vessels, which ultimately increases the high recurrence rate after surgery\textsuperscript{35}. Kyung HoPak et al. research also confirmed
that TGF-β1 can induce VEGF-C in GC to enhance tumor-induced lymph angiogenesis, and ultimately promote the recurrence and metastasis of GC. Therefore, it may also be a potential target for prevention and treatment of GC\textsuperscript{36}. In our research, the EMT pathway is significantly enriched, and multiple enriched pathways are ultimately related to the EMT pathway, so we infer that the EMT pathway plays an important role in the development of GC, which may be related to METTL3-mediated N6-methyladenosine modification and Roles of E-cadherin and Noncoding RNAs\textsuperscript{37,38}. Yoko Katsuno et al. research shows that TGF-β signaling and epithelial-mesenchymal transition in cancer progression. Furthermore, Wen et al.\textsuperscript{39} research indicated that the ectopic activation of Wnt/β-catenin-mediated signaling pathway can lead to a variety of tumors and diseases. This pathway not only plays an important role in proliferation, differentiation, apoptosis, migration, and invasion, but also plays an important role in the invasion and metastasis of cancer in epithelial-mesenchymal transition. Finally, KRAS gene is a kind of RAS family, and its mutation has been proved in the onset of cancer, such as GC, esophageal cancer, and ovarian cancer\textsuperscript{40}. Recent studies have shown that when KRAS is amplified, it often indicates a poor prognosis of cancer\textsuperscript{41–43}, and KRAS and EMT are closely related, and the two can interact at the same time, which reduces the survival rate of patients\textsuperscript{44}. Meanwhile, the stability of the Hedgehog signaling pathway is a prerequisite for the homeostasis of the gastrointestinal microenvironment. Studies have shown that hedgehog is related to a variety of cancers, such as prostate cancer, breast cancer, pancreatic cancer, GC, etc.\textsuperscript{45–48}. In-depth studies of the Hh-signaling pathway have found that high expression often indicates poor prognosis. Coagulation has also studied the recurrence and metastasis of GC. Kentaro et al. retrospectively analyzed the D-dimer level of 448 patients with GC on the 7th day after surgery, and found that the hypercoagulable state has a higher recurrence rate and poor survival rate may be related to the impact of surgical stress on the coagulation system, increasing the chance of micro metastasis\textsuperscript{49}, furthermore coagulation can promote platelet activation, provide vascular endothelial growth factor (VEGF), transforming growth factor-cell growth factor (TGF), platelet-derived growth factor (PDGF) etc. for tumor growth promote recurrence and metastasis\textsuperscript{50}. So, in our research, the above are all highly enrichment. The occurrence and development of tumors are usually the result of the interaction of multiple pathways. For example, in our research, KRAS, TGF-beta signaling, and Wnt/β-catenin-mediated signaling pathway are ultimately closely related to the EMT pathway.

Nomogram for predicting GC recurrence using biomarker gene expression has been confirmed\textsuperscript{51}, then we draw a nomogram to evaluate the survival rate of 1–5 years, and verify the calibration curve. The result shows that the nomogram has good predictive ability. This will help clinicians make precise decisions. Finally, we verified the 10 hub RBPs at the gene level and protein level, and the prognostic relationship obtained was consistent with the results of our analysis. It further illustrates that the prognosis model we constructed is accurate and practical.

However, this study also has some shortcomings: first, we obtained the results through retrospective research and analysis, which lacked a certain degree of sensitivity. Later, prospective studies are needed to verify the results again. Finally, we only download data from GEO database and Array Express database for analysis, which has certain limitations and heterogeneity.
In our current study, we have obtained 10 RBPs after processing the data using bioinformatics technology, which can be used to predict the prognosis of GC, which has a certain promotion significance in the clinic.

**Materials And Methods**

**Dataset processing and differential expressed RBPs (DERBPs) screening.**

We download the original data of GSE26942, GSE29998, GSE38024, GSE8443 microarray datasets from the GEO database (https://www.ncbi.nlm.nih.gov/gds/)\(^{52}\), and download the E-MTAB-1338 and E-MTAB-1440 microarray datasets from the Array Express database (https://www.ebi.ac.uk/arrayexpress/). We use the “lumi” package\(^{53}\) to process the original data to obtain the expression level of each RBPs, use the “sva” package\(^{54}\) to remove the batch effect of each microarray datasets and carry out standardized merging to obtain a combined complete dataset. For the complete data set, analyze the differential RBPs through the “limma” package\(^{55}\). The filter condition is \(|\log_2 \text{Fold Change}| > 0.585, \text{P} < 0.05\) to screen differentially expressed RBPs.

**Protein-protein interaction (PPI) network construction and enrichment analysis.**

Submit DERBPs to the STRING (http://www.string-db.org/) and the minimum correlation coefficient is 0.150 to identify protein-protein interaction information\(^{56}\). Cytoscape 3.7.2 software was used to further construct and visualize the PPI network. By using the molecular complexity detection (MCODE), the key modules and RBPs can be selected in the PPI network. In order to study the role of these key RBPs in GC, we use the "clusterProfiler" package\(^{57}\) to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis for key RBPs. The items that meet \(\text{P} < 0.05\) and \(\text{adj P} < 0.05\) are considered significant.

**Construction of a prognostic model based on RBPs.**

We use univariate Cox regression analysis to screen overall survival related RBPs, Least Absolute Shrinkage and Selection Operator (LASSO) regression to remove linear fitting, and further screen for candidate RBPs that are significantly related to prognosis. Subsequently, based on the above preliminary screened significant candidate RBPs, a proportional hazards prognostic model was constructed through multivariate Cox regression analysis. And calculated a risk score to assess patient prognosis outcomes. The risk score formula for each sample was as follows:
Risk score = \beta_1 \cdot \text{Exp2} + \beta_1 \cdot \text{Exp2} + \beta_1 \cdot \text{Exp2} + \ldots \beta_i \cdot \text{Exp} (\text{Exp is the expression level of each prognostic gene and } \beta \text{ is its regression coefficient}).

Based on the median risk score survival analysis, patients were divided into low-risk and high-risk groups. Firstly, we plot the K-M curve of the high and low risk groups. Additionally, we use R software to draw receiver operating characteristic (ROC) curve and estimate area under ROC curve (AUC), the higher the AUC indicates a better predictability for the model. In addition, we used univariate and multivariate Cox regression to analyze the prognostic ability of risk scores and other clinical characteristics. To confirm the effectiveness of our model, patient samples with reliable prognostic information from the GSE15459, GSE62254 dataset were used as a validation cohort to confirm the predictive capability of this prognostic model.

Gene Set Enrichment Analysis.

Gene Set Enrichment Analysis (GSEA) was constructed by “Subramanian A” in 2005. Compared with single gene, GSEA has obvious advantages in gene set. It mainly consists of three key steps: calculation of an enrichment score; estimate of significance level of enrichment score; adjustment for multiple hypothesis testing \cite{58}. In this study, we use GSEA 4.0 to set normalized enrichment score (NES) > 1, false discovery rate (FDR) < 0.001 as the selection criteria, hallmark7.1 as the comparison gene set, number of permutation = 1000 for enrichment and finally get the difference between high and low risk groups.

Nomogram construction and evaluation.

To develop a quantitative prognostic approach, we constructed nomogram to predict the impact of each gene on 1 to 5-years overall survival. Based on multivariate Cox analysis, point scales in the nomogram were used to assign values to individual variables. We use a horizontal line to determine the points of each variable and calculate the total points for each patient by adding up the points of all variables, normalizing the distribution from 0 to 100. Then, to better evaluate the predicted survival rate and actual survival rate, we compare the predicted and observed results in the calibration curve. When the predicted survival rate is close to the actual survival rate, it indicates that the nomogram has better predictive ability.

Verification of express level and prognostic significance.

The main purpose of The Kaplan Meier plotter (http://kmplot.com) \cite{59} is the discovery and verification of survival biomarkers based on meta-analysis. Therefore, it was used to verify the prognostic value of 10 RBPs in GC. GEPIA provides key interactive and customizable functions including differential expression analysis, profiling plotting, correlation analysis, patient survival analysis, similar gene detection and dimensionality reduction analysis. Using GEPIA (http://gepia.cancer-pku.cn/) \cite{60} to verify the mRNA
expression levels of 10 RBPs, and we use the Human Protein Atlas (HPA) online database (http://www.proteinatlas.org/) to detect the protein expression of 10 hub RBPs at a translational level.

**Declarations**

**Acknowledgements**

We would like to thank everyone who take part in this study.

**Funding**

National Natural Science Foundation of China, Grant/Award Numbers: 81872480, 81760549, 81560492; Science and Technology Research Project of Education Department of Jiangxi Province, Grant/Award Number: GJJ180024.

**Availability of data and materials**

The data and materials can be found from the first author and corresponding author.

**Authors’ contributions**

Liqiang Zhou conceived, designed, analyzed the data, and Hao Lu write the manuscript. Qi Zhou, Shihao Li helped to search for some relevant papers for this research. You Wu and Yiwu Yuan generated the figures and tables. Lin Xin guided the research process and review the manuscript. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

All patient data were derived from online datasets; thus, no ethics approval was required.

**Patient consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.
References

1. Choi IJ, K. M., Kim YI et al. Helicobacter pylori Therapy for the Prevention of Metachronous Gastric Cancer. *N Engl J Med* **22;378(12)**:1085–1095, doi:10.1056/NEJMoa1708423 (2018).

2. Miller, K. D. *et al.* Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin* **69**, 363–385, doi:10.3322/caac.21565 (2019).

3. Gao, K. & Wu, J. National trend of gastric cancer mortality in China (2003–2015): a population-based study. *Cancer Commun* **39**, 24, doi:10.1186/s40880-019-0372-x (2019).

4. Janga, S. C. From specific to global analysis of posttranscriptional regulation in eukaryotes: posttranscriptional regulatory networks. *Brief Funct Genomics* **11**, 505–521, doi:10.1093/bfgp/els046 (2012).

5. Khalil, A. M. & Rinn, J. L. RNA-protein interactions in human health and disease. *Semin Cell Dev Biol* **22**, 359–365, doi:10.1016/j.semcdb.2011.02.016 (2011).

6. Wurth, L. & Gebauer, F. RNA-binding proteins, multifaceted translational regulators in cancer. *Biochim Biophys Acta* **1849**, 881–886, doi:10.1016/j.bbagrm.2014.10.001 (2015).

7. Masuda, K. & Kuwano, Y. Diverse roles of RNA-binding proteins in cancer traits and their implications in gastrointestinal cancers. *Wiley Interdiscip Rev RNA* **10**, e1520, doi:10.1002/wrna.1520 (2019).

8. Li W, Gao LN, Song PP, You CG. Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma. *Aging* **12, No. 4** (2020).

9. Wang, K. *et al.* Integrated Bioinformatics Analysis the Function of RNA Binding Proteins (RBPs) and Their Prognostic Value in Breast Cancer. *Front Pharmacol* **10**, 140, doi:10.3389/fphar.2019.00140 (2019).

10. Kang, M. J. *et al.* NF-kappaB activates transcription of the RNA-binding factor HuR, via PI3K-AKT signaling, to promote gastric tumorigenesis. *Gastroenterology* **135**, 2030–2042, 2042.e2031-2033, doi:10.1053/j.gastro.2008.08.009 (2008).

11. Wang, H. *et al.* Dysregulation of tristetraprolin and human antigen R promotes gastric cancer progressions partly by upregulation of the high-mobility group box 1. *Sci Rep* **8**, 7080, doi:10.1038/s41598-018-25443-3 (2018).

12. Deng, K. *et al.* Tristetraprolin inhibits gastric cancer progression through suppression of IL-33. *Sci Rep* **6**, 24505, doi:10.1038/srep24505 (2016).

13. Li, F. *et al.* QKI5-mediated alternative splicing of the histone variant macroH2A1 regulates gastric carcinogenesis. *Oncotarget* **7**, 32821–32834, doi:10.18632/oncotarget.8739 (2016).

14. Wang, X., Hu, H. & Liu, H. RNA binding protein Lin28B confers gastric cancer cells stemness via directly binding to NRP-1. *Biomed pharmacother* **104**, 383–389, doi:10.1016/j.biopha.2018.05.064 (2018).

15. Smyth, E. C., Nilsson, M., Grabsch, H. I., van Grieken, N. C. T. & Lordick, F. Gastric cancer. *The Lancet* **396**, 635–648, doi:10.1016/s0140-6736(20)31288-5 (2020).
16. Gerstberger, S., Hafner, M., Ascano, M. & Tuschl, T. Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. Adv Exp Med Biol 825, 1–55, doi:10.1007/978-1-4939-1221-6_1 (2014).

17. Siang, D. T. C. et al. The RNA-binding protein HuR is a negative regulator in adipogenesis. Nat Commun 11, 213, doi:10.1038/s41467-019-14001-8 (2020).

18. Jain, A., Brown, S. Z., Thomsett, H. L., Londin, E. & Brody, J. R. Evaluation of Post-transcriptional Gene Regulation in Pancreatic Cancer Cells: Studying RNA Binding Proteins and Their mRNA Targets. Methods Mol Biol 1882, 239–252, doi:10.1007/978-1-4939-8879-2_22 (2019).

19. Laura M. Snell, T. L. M., David G. Brooks. Type I Interferon in Chronic Virus Infection and Cancer. Trends Immunol 38, No. 8, doi:10.1016/j.it.2017.05.005 (2017).

20. Zitvogel, L., Galluzzi, L., Kepp, O., Smyth, M. J. & Kroemer, G. Type I interferons in anticancer immunity. Nat Rev Immunol 15, 405–414, doi:10.1038/nri3845 (2015).

21. Yunshu Gao Ying Chen, D. X., Jiejun Wang and Guanzhen Yu. Differential expression of ANXA1 in benign human gastrointestinal tissues and cancers. BMC Cancer 14:520 (2014).

22. Takaoka, R. T. C. et al. Expression profiles of Annexin A1, formylated peptide receptors and cyclooxygenase-2 in gastroesophageal inflammations and neoplasias. Pathol Res Pract 214, 181–186, doi:10.1016/j.prp.2017.12.003 (2018).

23. Gayyed, M. F., El-Maqsoud, N. M., Tawfiek, E. R., El Gelany, S. A. & Rahman, M. F. A comprehensive analysis of CDC20 overexpression in common malignant tumors from multiple organs: its correlation with tumor grade and stage. Tumour Biol 37, 749–762, doi:10.1007/s13277-015-3808-1 (2016).

24. Yang, S. et al. Overexpression of eukaryotic elongation factor 1 alpha-2 is associated with poorer prognosis in patients with gastric cancer. J Cancer Res Clin Oncol 141, 1265–1275, doi:10.1007/s00432-014-1897-7 (2015).

25. Zhao, Z. S., Li, L., Wang, H. J. & Wang, Y. Y. Expression and prognostic significance of CEACAM6, ITGB1, and CYR61 in peripheral blood of patients with gastric cancer. J Surg Oncol 104, 525–529, doi:10.1002/jso.21984 (2011).

26. Wang, Y. Y., Li, L., Zhao, Z. S. & Wang, H. J. Clinical utility of measuring expression levels of KAP1, TIMP1 and STC2 in peripheral blood of patients with gastric cancer. World J Surg Oncol 11, 81, doi:10.1186/1477-7819-11-81 (2013).

27. Jo YS, K. M., Yoo NJ, Lee SH. Somatic Mutations and Intratumoral Heterogeneity of MYH11 Gene in Gastric and Colorectal Cancers. Appl Immunohistochem Mol Morphol Sep;26(8):562–566 (2018 Sep;26(8):562–566.).

28. Zhao, H. et al. The opposite role of alternatively spliced isoforms of LINC00477 in gastric cancer. Cancer Manag Res 11, 4569–4576, doi:10.2147/CMAR.S202430 (2019).

29. Zhong, Q. et al. CPEB3 inhibits epithelial-mesenchymal transition by disrupting the crosstalk between colorectal cancer cells and tumor-associated macrophages via IL-6R/STAT3 signaling. J Exp Clin Cancer Res 39, 132, doi:10.1186/s13046-020-01637-4 (2020).
30. Yu, P. H. B., Shen, M., et al. p15PAF, a novel PCNA associated factor with increased expression in tumor tissues. Oncogene **20**, 484 ± 489 (2001).

31. Zhu, K. et al. Elevated KIAA0101 expression is a marker of recurrence in human gastric cancer. *Cancer Sci** 104*, 353–359, doi:10.1111/cas.12083 (2013).

32. Jiang, Y. et al. ImmunoScore Signature: A Prognostic and Predictive Tool in Gastric Cancer. *Ann Surg* **267**, 504–513, doi:10.1097/SLA.0000000000002116 (2018).

33. Cai, C. et al. Prediction of Overall Survival in Gastric Cancer Using a Nine-IncRNA. *DNA Cell Bio* **38**, 1005–1012, doi:10.1089/dna.2019.4832 (2019).

34. Zi, Z. Molecular Engineering of the TGF-beta Signaling Pathway. *J Mol Biol* **431**, 2644–2654, doi:10.1016/j.jmb.2019.05.022 (2019).

35. Ishimoto, T. et al. Activation of Transforming Growth Factor Beta 1 Signaling in Gastric Cancer-associated Fibroblasts Increases Their Motility, via Expression of Rhomboid 5 Homolog 2, and Ability to Induce Invasiveness of Gastric Cancer Cells. *Gastroenterology** 153*, 191–204 e116, doi:10.1053/j.gastro.2017.03.046 (2017).

36. Pak, K. H., Park, K. C. & Cheong, J.-H. VEGF-C induced by TGF-β1 signaling in gastric cancer enhances tumor-induced lymphangiogenesis. *BMC Cancer* **19**, doi:10.1186/s12885-019-5972-y (2019).

37. Bure, I. V., Nemtsova, M. V. & Zaletaev, D. V. Roles of E-cadherin and Noncoding RNAs in the Epithelial-mesenchymal Transition and Progression in Gastric Cancer. *Int J Mol Sci* **20**, doi:10.3390/ijms20122870 (2019).

38. Yue, B. et al. METTL3-mediated N6-methyladenosine modification is critical for epithelial-mesenchymal transition and metastasis of gastric cancer. *Mol Cancer* **18**, 142, doi:10.1186/s12943-019-1065-4 (2019).

39. Wen, X., Wu, Y., Awadasseid, A., Tanaka, Y. & Zhang, W. New Advances in Canonical Wnt/β-Catenin Signaling in Cancer. *Cancer Manag Res** Volume 12**, 6987–6998, doi:10.2147/cmar.S258645 (2020).

40. Hewitt, L. C. et al. KRAS status is related to histological phenotype in gastric cancer: results from a large multicentre study. *Gastric Cancer** 22**, 1193–1203, doi:10.1007/s10120-019-00972-6 (2019).

41. Rehkaemper, J. et al. Amplification of KRAS and its heterogeneity in non-Asian gastric adenocarcinomas. *BMC Cancer* **20**, 587, doi:10.1186/s12885-020-06996-x (2020).

42. Wong, G. S. et al. Targeting wild-type KRAS-amplified gastroesophageal cancer through combined MEK and SHP2 inhibition. *Nat Med* **24**, 968–977, doi:10.1038/s41591-018-0022-x (2018).

43. Birkeland, E. et al. KRAS gene amplification and overexpression but not mutation associates with aggressive and metastatic endometrial cancer. *Br J Cancer** 107**, 1997–2004, doi:10.1038/bjc.2012.477 (2012).

44. Shibue, T. & Weinberg, R. A. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat Rev Clin Oncol* **14**, 611–629, doi:10.1038/nrclinonc.2017.44 (2017).
45. Saqui-Salces, M. & Merchant, J. L. Hedgehog signaling and gastrointestinal cancer. *Biochim Biophys Acta* **1803**, 786–795, doi:10.1016/j.bbamcr.2010.03.008 (2010).

46. Walsh, P. C. Hedgehog signalling in prostate regeneration, neoplasia and metastasis. *J Urol* **173**, 1169, doi:10.1097/01.ju.0000156734.69186.57 (2005).

47. Riobo-Del Galdo, N. A., Lara Montero, A. & Wertheimer, E. V. Role of Hedgehog Signaling in Breast Cancer: Pathogenesis and Therapeutics. *Cells* **8**, doi:10.3390/cells8040375 (2019).

48. Thayer, S. P. *et al.* Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis. *Nature* **425**, 851–856, doi:10.1038/nature02009 (2003).

49. Hara, K. *et al.* Postoperative D-dimer elevation affects tumor recurrence and the long-term survival in gastric cancer patients who undergo gastrectomy. *Int J Clin Oncol* **25**, 584–594, doi:10.1007/s10147-019-01603-x (2020).

50. Wojtukiewicz, M. Z., Hempel, D., Sierko, E., Tucker, S. C. & Honn, K. V. Thrombin-unique coagulation system protein with multifaceted impacts on cancer and metastasis. *Cancer Metastasis Rev* **35**, 213–233, doi:10.1007/s10555-016-9626-0 (2016).

51. Jeong, S. H. *et al.* Nomogram for predicting gastric cancer recurrence using biomarker gene expression. *Eur J Surg Oncol* **46**, 195–201, doi:10.1016/j.ejso.2019.09.143 (2020).

52. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* **41**, D991-995, doi:10.1093/nar/gks1193 (2013).

53. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548, doi:10.1093/bioinformatics/btn224 (2008).

54. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883, doi:10.1093/bioinformatics/bts034 (2012).

55. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).

56. Cook, H. V., Doncheva, N. T., Szklarczyk, D., von Mering, C. & Jensen, L. J. Viruses.STRING: A Virus-Host Protein-Protein Interaction Database. *Viruses* **10**, doi:10.3390/v10100519 (2018).

57. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287, doi:10.1089/omi.2011.0118 (2012).

58. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550, doi:10.1073/pnas.0506580102 (2005).

59. Goel, M. K., Khanna, P. & Kishore, J. Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res* **1**, 274–278, doi:10.4103/0974-7788.76794 (2010).

60. Tang, Z. *et al.* GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* **45**, W98-w102, doi:10.1093/nar/gkx247 (2017).
Figure 1

The differentially expressed genes in GC. (a) Heat map of differentially expressed genes; (b) Heat map of differentially expressed genes differential RBPs
Figure 2

Protein-protein interaction network and modules analysis. (a) Protein-protein interaction network of differentially expressed RBPs; (b) (c) key module from PPI network
Figure 3

Regression analysis to select hub-RBPs prognosis of GC. (a) Univariate Cox regression analysis for identification of hub RBPs in the training data set; (b) (c) Selecting the best parameters for gastric cancer in LASSO regression analysis; (d) Multivariate Cox regression analysis to get prognosis related hub 10 RBPs.
Prognostic analysis of 10-RBPs signature in the train cohort. (a) Scatterplots of GC patients with different survival status in training group; (b) Risk score distribution of GC patients with different risks (low, green; high, red) in the testing group; (c) Kaplan–Meier Survival curve of low-risk and high-risk subgroups; (d) 5-year time-dependent ROC for survival prediction models; (e) Univariate Cox regression analysis on the prognosis of clinicopathological characteristics and risk scores in patients with GC; (f) Multivariate Cox
regression analysis on the prognosis of clinicopathological characteristics and risk scores in patients with GC.

Figure 5

Prognostic analysis of 10-RBPs signature in the GSE15459 data set. (a) Scatterplots of GC patients with different survival status in training group; (b) Risk score distribution of GC patients with different risks (low, green; high, red) in the testing group; (c) Kaplan–Meier Survival curve of low-risk and high-risk
subgroups; (d) 5-year time-dependent ROC for survival prediction models; (e) Univariate Cox regression analysis on the prognosis of clinicopathological characteristics and risk scores in patients with GC; (f) Multivariate Cox regression analysis on the prognosis of clinicopathological characteristics and risk scores in patients with GC.

Figure 6
Prognostic analysis of 10-RBPs signature in the GSE62254 data set. (a) Scatterplots of GC patients with different survival status in training group; (b) Risk score distribution of GC patients with different risks (low, green; high, red) in the testing group; (c) Kaplan–Meier Survival curve of low-risk and high-risk subgroups; (d) 5-year time-dependent ROC for survival prediction models; (e) Univariate Cox regression analysis on the prognosis of clinicopathological characteristics and risk scores in patients with GC; (f) Multivariate Cox regression analysis on the prognosis of clinicopathological characteristics and risk scores in patients with GC.

**Figure 7**

The relationship between high and low risk groups, hub RBPs and clinicopathological characteristics. (a) The relationship between clinicopathological characteristics and risk group; (b) The relationship between RBPs expression level and clinicopathological characteristics and risk value.
Figure 8

Enrichment pathways for high-risk groups.

a

| Genes | ACO1 | ANXA1 | APE1 | CDD20 | CPEB3 | DAB2IP | EEF1A2 | ITG8 | KIAA0101 | MEG11 |
|-------|------|-------|------|-------|-------|--------|--------|------|----------|-------|
| Points | 12.6 | 13.9 | 14.5 | 14.0 | 13.5 | 13.0 | 12.5 | 12.0 | 11.5 | 11.0 | 10.5 | 10.0 | 9.5 |

b

Actual Survival vs. Log-Rank Predicted Survival
Figure 9

Nomogram and calibration plots of 10 RBPs. (a) Nomogram for predicting 1-5-years OS of GC; (b) calibration chart for nomogram accuracy.

Figure 10
Kaplan-Meier survival curves and Gepia database validates. (a) Survival curve drawn by Kaplan Meier plotter on RBPs of the 10 RBPs; (b) Gepia database validates the expression results of the 10 RBPs.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1Table1.docx
- Supplementaryfile1.docx
- Additionalfile2figure1.tif