Research Paper

Integrating 13 Microarrays to Construct a 6 RNA-binding proteins Prognostic Signature for Gastric Cancer patients

Liqiang Zhou*, Qi Zhou*, You Wu, Lin Xin

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

*These authors contributed equally to this work.

Corresponding author: Lin Xin, M.D., Ph.D., Department of General Surgery, The Second Affiliated Hospital of Nanchang University, Nanchang 330006, Jiangxi, China. E-mail: xlyxsbs@126.com.

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/). See http://ivyspring.com/terms for full terms and conditions.

Received: 2020.12.16; Accepted: 2021.05.21; Published: 2021.06.11

Abstract

**Background:** It has been confirmed in many tumors that RNA-binding proteins (RBPs) will affect the progress of cancer, but there is still a lack of large-scale research in gastric cancer (GC).

**Methods:** We obtained 13 microarray mRNA expression profiles of the GPL570 platform, and extracted expression from them after integration to analyze the expression differences of RBPs. Enrichment analysis studies the role of these RBPs in GC. Univariate, Lasso and multivariate Cox regression analysis are used to identify independent prognostic hub RBPs, thereby constructing and verifying a prognostic signature. External data and rt-PCR verified the expression of hub RBPs.

**Results:** We have identified 51 dysregulated RBPs in GC. Enrichment analysis shows that it can mainly participate in RNA decomposition, modification, processing, etc. and affect the progress of GC. After multiple statistical analysis, six independent prognostic RBPs of GC were determined and a prognostic signature was developed. According to the median risk value, the training cohort was divided into high-risk and low-risk groups. Considering the clinical characteristics, in training, testing, and complete cohorts, the overall survival rate of the high-risk group was significantly lower than that of the low-risk group, which was confirmed by the time-dependent receiver operating characteristic curve. Univariate and multivariate Cox regression analysis of independent prognostic ability of risk score. In addition, we constructed and verified a nomogram based on the prognostic signature, showing accurate prediction performance. rt-PCR and external data verification are consistent with our conclusions.

**Conclusion:** This study analyzed the overall expression of RBPs in GC and explored its mechanism. A new prognostic signature was developed and verified. A nomogram has also been established and verified, which helps to improve the treatment strategy for GC.

Key words: gastric cancer; RNA-binding protein; prognosis; signature; nomogram; survival

Introduction

Gastric cancer (GC) is a malignant tumor that originates from epithelial cells of the gastric mucosa. The risk factors include: race, environment, Helicobacter pylori infection, high-salt diet, moldy food, obesity, smoking and other factors. According to the latest statistics, there are 1,033,701 new cases of GC and 782,685 deaths each year, accounting for 5.7% and 8.2% of all tumors, respectively [1]. Reports in 2010 showed that the incidence of gastric cancer in China accounted for as high as 42.6% of the world, and it continues to increase [2]. This is because GC only shows upper abdominal discomfort in the early stage, belching and other symptoms are similar to gastritis, gastric ulcer and other chronic gastric diseases [1, 3]. There are no other specific symptoms, which makes it easy to be ignored, and it is mostly in the late stage when discovered. At present, the main ways to treat GC include: surgery, chemotherapy, and molecular targeted therapy. However, these treatments have the characteristics of easy relapse and drug resistance, which make the 5-year survival rate of GC less than 30% [4]. Due to the late detection of
In addition, PTBP3 can mediate CAV1 alternative expression. In cancer cells, RBPs play an important role in the origin and progression of GC. Therefore, we screened for genes with prognostic value, some of which may serve as potential biomarkers for diagnosis and prognosis. In addition, the prognostic signatures constructed based on these prognostic RBPs can accurately predict the development of GC patients and have very good application prospects.

Materials and methods

Identification of differentially expressed RNA binding proteins (DERBPs)

We downloaded the original data containing primary gastric cancer. The 13 microarrays are all on the GPL570 platform, and the gastric cancer samples are all obtained by surgical resection. The data of each microarray is shown in Table 1. The RNA algorithm of the “affy” package is used to extract the expression data of each data set [19], and the “sva” package is used to remove batch effects and merge all data sets [20]. The P value of the differentially expressed genes between the merged data was analyzed by the ebayes test of the “limma” package [21]. The threshold for determining differentially expressed genes (DEGs) is set to |log2 fold change (FC)| ≥ 1.0 and false discovery rate (FDR) <0.05. Based on the RNA-binding protein database RBPTD (http://www.rbptd.com/) [22], we selected differentially expressed RNA-binding proteins (DERBPs) from these DEGs that met the screening criteria.

Table 1. GEO data sets included in this study

<table>
<thead>
<tr>
<th>GEO datasets</th>
<th>Year</th>
<th>Country</th>
<th>Platform</th>
<th>Sample</th>
<th>Tumor (n)</th>
<th>Normal (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE66229</td>
<td>2015</td>
<td>USA</td>
<td>GPL570</td>
<td>GC/Normal</td>
<td>300</td>
<td>100</td>
</tr>
<tr>
<td>GSE54129</td>
<td>2017</td>
<td>China</td>
<td>GPL570</td>
<td>GC/Normal</td>
<td>111</td>
<td>21</td>
</tr>
<tr>
<td>GSE13911</td>
<td>2008</td>
<td>Italy</td>
<td>GPL570</td>
<td>GC/Normal</td>
<td>38</td>
<td>31</td>
</tr>
<tr>
<td>GSE19826</td>
<td>2010</td>
<td>China</td>
<td>GPL570</td>
<td>GC/Normal</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>GSE79973</td>
<td>2016</td>
<td>China</td>
<td>GPL570</td>
<td>GC/Normal</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>GSE51725</td>
<td>2013</td>
<td>Japan</td>
<td>GPL570</td>
<td>GC/Normal</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>GSE35459</td>
<td>2009</td>
<td>Switzerland</td>
<td>GPL570</td>
<td>GC</td>
<td>200</td>
<td>0</td>
</tr>
<tr>
<td>GSE51105</td>
<td>2014</td>
<td>Australia</td>
<td>GPL570</td>
<td>GC</td>
<td>94</td>
<td>0</td>
</tr>
<tr>
<td>GSE35809</td>
<td>2012</td>
<td>Singapore</td>
<td>GPL570</td>
<td>GC</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>GSE57303</td>
<td>2014</td>
<td>China</td>
<td>GPL570</td>
<td>GC</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>GSE34942</td>
<td>2014</td>
<td>Singapore</td>
<td>GPL570</td>
<td>GC</td>
<td>56</td>
<td>0</td>
</tr>
<tr>
<td>GSE22377</td>
<td>2011</td>
<td>Germany</td>
<td>GPL570</td>
<td>GC</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>GSE38749</td>
<td>2012</td>
<td>Brazil</td>
<td>GPL570</td>
<td>GC</td>
<td>15</td>
<td>0</td>
</tr>
</tbody>
</table>

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) gene function enrichment analysis

The GO database has three categories, namely Biological Process (BP), Cellular Component (CC) and Molecular Function (MF). Each describes the molecular functions that gene products may perform. The cellular environment and the biological processes...
involved. The KEGG database helps to research genes and expression information as a whole network. KEGG integrates data on genomes, chemical molecules and biochemical systems, including metabolic pathways (PATHWAY), drugs (DRUG), diseases (DISEASE), gene sequences (GENES), and genomes (GENOME).

**Identify prognostic-related RBPs**

In order to identify RBPs related to prognosis, we used a series of complete statistical analyses. We first used the “survival analysis” package to perform single-factor Cox regression analysis on the probe ID of DERBPs. Subsequently, based on the results of the previous step, Lasso regression was used for further screening to obtain RBPs that are significantly related to the prognosis of GC patients. In addition, we also used multivariate Cox regression analysis to test the results of Lasso regression and found hub RBPs that can independently predict the prognosis of GC.

**Construction and verification of the prognostic model of RBPs**

We randomly divide all GC samples into training set (n=248) and validation set (n=244). We first used multivariate Cox regression analysis in the training set to construct a risk ratio model for predicting the prognosis of GC based on hub RBPs. We first used multivariate Cox regression analysis in the training set to construct a risk ratio model for predicting the prognosis of GC based on hub RBPs. We obtained the HR of hub RBPs, the 95% confidence interval of HR and the regression coefficient (β) of the corresponding gene. According to the expression of each gene and the regression coefficient, we can calculate the risk score of each patient according to: Risk score=β1*Exp1+β2*Exp2+βi*Expi. The GC patients were divided into high-risk and low-risk groups based on the median risk score, and Kaplan-Meier survival curve was drawn. The difference in OS between the two groups was analyzed by log-Rank test. We also drew a 5-year ROC curve with “SurvivalROC” package, and calculate the AUC value to evaluate the predictive ability of the model.

**Bioinformatics analysis**

We used the chi-square test to analyze the relationship between the expression of each hub RBPs and the clinicopathological characteristics of patients with GC, and search for the potential mechanisms that affect tumor progression. In addition, in order to study the differences between different risk groups, we first analyzed the relationship between high and low risk groups and clinicopathological characteristics, and secondly, we also used Gene Set Enrichment Analysis (GSEA) to further analyze the overall differences between different risk groups [23]. We use GSEA4.0 based on the molecular signature database (MSigDB), with hallmark7.1 as the control gene and |NSE| > 1, FDR <0.001 as the screening conditions to identify the development mechanism of GC in the high-risk group.

**RBPs nomogram construction and verification**

We use “rms” to draw a nomogram based on the RBPs prognostic risk model, and obtain the corresponding score by analyzing the expression level of each hub RBPs, and adding the scores of all hub genes to obtain the corresponding total score. By drawing a vertical line on the total score line, we can predict the probability of survival for patients with GC in 1 to 5 years. In addition, in order to test the prediction ability of the nomogram, we draw a 5-year calibration curve by analyzing the survival probabilities of the predicted value and the actual value at the quartile of all GC patients. If the actual value is close to the predicted value, the nomogram has good predictive performance.

**Verification of hub RBPs expression and prognosis**

In order to verify the prognosis of hub RBPs, we used Kaplan-Meier Plotter (http://kmplot.com/) to verify it in the GSE29272 data set [24]. For the verification of hub RBPs expression, we first use the GEPIA network tool for verification, which contains data from the TCGA and GTEx databases [25]. In addition, we also collected surgical samples from 10 pairs of gastric cancer patients. The process was approved by the patient’s informed consent and the ethics committee of the Second Affiliated Hospital of Nanchang University. After homogenizing the clinical samples, the Trizol (Thermo Fisher, USA) method was used to extract total RNA. The obtained RNA was reverse transcribed using reverse transcription kit RR047A (Takara, Japan). ACTB was used as the internal reference gene, and the mRNA expression of hub RBPs was analyzed by fluorescence quantitative PCR using the RR820 kit (Takara, Japan) on the
7900-HT system (Thermo Fisher, USA). The primers used are all synthesized by Shanghai Shenggong Company, and the sequences of all primers are in Supplementary Table 2.

Results

Identify DERBPs

The flowchart of this study is shown in Fig. 1, and the details of the GEO data set included in this study are shown in Table 1. After extracting expression data from each data set, removing batch effects and merging, a single expression data of 179 normal stomach and 1027 GC tissues were obtained. After the difference analysis of the “limma” package, we obtained 1430 probe differences (1102 genes) that met the screening conditions, and used R to draw the heat map and volcano map. After further screening, we got 70 One probe ID (51 genes) representing RBPs, of which 25 RBPs were up-regulated and 26 genes were down-regulated (Table 2).

GO and KEGG gene function enrichment analysis

We performed GO and KEGG functional enrichment analysis on these DERBPs. The results suggest that 51 DERBPs are enriched in posttranscriptional regulation of gene expression, mRNA binding, nucleicase activity, regulation of cellular amide metabolic process, catalytic activity, acting on RNA, nuclear acid phosphodiester bond hydrolysis, RNA catabolic process, AU-rich element binding, RNA modification, debasinase activity, mRNA processing, regulation of mRNA metabolic process, DNA modification, ribonucleoprotein complex biogenesis (Table 3). For the KEGG functional enrichment analysis, we did not find an item that met the filter criteria.

Identification of prognostic-related hub RBPs

In this study, we collected a total of 492 samples with clinical information for further research. First, the univariate Cox regression analysis was used to analyze the relationship between 70 RBP probes and prognosis, and the results suggested that 42 RBP probes were related to the prognosis of GC patients (Fig. 2A). Subsequently, Lasso regression further analyzed the results of the previous step and screened the RBP probes that are significantly related to the prognosis. We obtained 9 key RBP probes that are significantly related to the prognosis of GC patients (Fig. 2B and C). In addition, we also used multivariate Cox regression analysis of these 9 key probes, and the results showed that 6 hub RBP probes (6 RBPs) are independent prognostic factors for patients with GC. Among them, HR>1 includes KIAA0101, WIPF3, COL5A2, RBPMS2, and HR<1 includes DAZ1 and NOVA1.

Construction of a prognostic model of GC based on RBPs

We randomly divided 492 GC samples with clinical information into training set (n=248) and test set (n=244). We first assign the corresponding regression coefficients (β) to the 6 hub RBPs in the training set, and then synthesize their expression levels to construct a prognostic model. Each GC patient gets a risk score according to the following formula: Risk score= -0.058 *ExpDAZ1 + 0.3524 *ExpKIAA0101 + 0.2045 *ExpWIPF3 + 0.2687 *ExpCOL5A2 + 0.4544 * ExpRBPMS2 + -0.2872 * ExpNOV1, divided into high-risk group and low-risk group based on the median Risk group. The KM curve indicated that the survival rate of GC patients in the high-risk group was significantly lower than that in the low-risk group (Fig. 3A), and the 5-year AUC was calculated as 0.713 (Fig. 3B). The distribution of survival status indicated that the survival time and number of patients in the high-risk group was lower in the risk group, low (Fig. 3C).
Figure 2. Identify prognostic-related hub RBPs. A Univariate Cox regression analysis identified prognostic-related RBPs in the entire cohort. B Selecting the best parameters for gastric cancer in LASSO regression analysis. C Multivariate Cox regression analysis identified prognostic-related hub RBPs.

Table 2. 51 differentially expressed RNA binding proteins

<table>
<thead>
<tr>
<th>Upregulate RNA binding protein</th>
<th>Down-regulate RNA binding protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob ID</td>
<td>Gene ID</td>
</tr>
<tr>
<td>203820_s_at</td>
<td>IGF2BP3</td>
</tr>
<tr>
<td>203819_s_at</td>
<td>IGF2BP3</td>
</tr>
<tr>
<td>201291_s_at</td>
<td>TOP2A</td>
</tr>
<tr>
<td>203819_s_at</td>
<td>IGF2BP3</td>
</tr>
<tr>
<td>235004_at</td>
<td>RRM2</td>
</tr>
<tr>
<td>201291_s_at</td>
<td>TOP2A</td>
</tr>
<tr>
<td>201292_at</td>
<td>TOP2A</td>
</tr>
<tr>
<td>204048_at</td>
<td>KIF2C</td>
</tr>
<tr>
<td>211519_s_at</td>
<td>KIF2C</td>
</tr>
<tr>
<td>204822_at</td>
<td>TTK</td>
</tr>
<tr>
<td>202870_s_at</td>
<td>CDC20</td>
</tr>
<tr>
<td>223229_at</td>
<td>UBE2T</td>
</tr>
<tr>
<td>203338_s_at</td>
<td>EZH2</td>
</tr>
<tr>
<td>221258_s_at</td>
<td>KIF18A</td>
</tr>
<tr>
<td>204603_at</td>
<td>EOC1</td>
</tr>
<tr>
<td>221730_at</td>
<td>COL5A2</td>
</tr>
<tr>
<td>221729_at</td>
<td>COL5A2</td>
</tr>
<tr>
<td>225827_at</td>
<td>AGO2</td>
</tr>
<tr>
<td>213310_at</td>
<td>AGO2</td>
</tr>
<tr>
<td>211762_s_at</td>
<td>KPNA2</td>
</tr>
<tr>
<td>201088_at</td>
<td>KPNA2</td>
</tr>
<tr>
<td>204444_at</td>
<td>KIF11</td>
</tr>
<tr>
<td>201926_s_at</td>
<td>CD55</td>
</tr>
<tr>
<td>218984_at</td>
<td>PLL7</td>
</tr>
</tbody>
</table>
Figure 3. Prognostic analysis of 6-RBPs signature in the train cohort. A The survival analysis of 6-RBPs signature in the testing cohort. B 5-year time-dependent ROC for survival prediction models. C The distribution of risk score, OS and OS status of prognostic 6-RBPs signature in the testing cohort. D Univariate Cox regression analysis on the prognosis of clinicopathological characteristics and risk scores in patients with gastric cancer. E Multivariate Cox regression analysis on the prognosis of clinicopathological characteristics and risk scores in patients with gastric cancer.
Using univariate Cox to analyze the relationship between risk value and clinicopathological parameters and overall survival (Fig. 3D), T stage [HR=1.712, 95%CI (1.252-2.342)], N stage [HR=1.910, 95%CI (1.503-2.428)], AJCC stage [HR=2.107, 95%CI (1.622-2.736)] and risk value [HR=1.798, 95%CI (1.460-2.214)] are related to prognosis. Multivariate Cox regression analysis further showed (Fig. 3E) that N stage [HR=1.600, 95%CI (1.078-2.376)] and risk score [HR=1.702, 95%CI (1.342-2.159)] are independent of GC patient prognostic factors. There are 112 samples in the high-risk group in the training set and 132 samples in the low-risk group. The KM curve indicates that the high-risk group has a worse prognosis (Fig. 4A), 5-year AUC=0.681 (Fig. 4B), and the survival status diagram is similar to the training set (Fig. 4C). Single-factor Cox regression analysis shows that the risk value of T stage, N stage, M stage, AJCC stage is related to OS (Fig. 4D). Multivariate Cox regression analysis shows that Age, N stage, M stage, and risk values can independently predict the prognosis of GC patients, and the survival status are related to high and low risk values. Among the 6 hub RBPs, RBPMS2, KIAA0101 and NOVA1 are related to 6 or more clinicopathological features, while WIPF3, COL5A2 and DAZ1 are only related to a few clinicopathological features (Fig. 6B). In addition, in order to study the differences between the high and low risk groups of GC patients, we used GSEA to analyze them. The results indicate (Fig. 7) that the samples of the high-risk group were mostly enriched in epithelial-mesenchymal transition, KRAS signal up-regulation, NOTCH signal, TGF-β signal, Angiogenesis, Hedgehog signal, IL2/STAT5 signal, Hypoxia, Myogenesis, Coagulation, UV response UP, Apical junction, APICAL SURFACE.

### Nomogram drawing and verification

Based on the complete prognosis model we constructed, we established a nomogram to predict the survival rate of GC in 1-5 years (Fig. 8A). By detecting the expression levels of the 6 hub RBPs and obtaining the corresponding scores, the total scores obtained can intuitively predict the short-term survival probability of patients with GC, which will provide clinicians with precise and individualized diagnosis and treatment. In order to verify the accuracy of the nomogram, we calculated the predicted survival rate of 492 samples as the X-axis and the actual survival rate as the Y-axis to draw a 5-year calibration curve. The results indicate (Fig. 8B) that the predicted survival rate of the nomogram is almost equal to the actual survival rate, and the slope of the drawn curve is almost 1, indicating the accurate predictive ability of the nomogram.

### Expression and prognostic verification of hub RBPs

Use K-M plotter to analyze 6 hub gene prognosis prompts (Fig. 9A), high expression of DAZ1, NOVA1 suggests a better prognosis (P<0.05), increased expression of KIAA0101, WIPF3 prompts a poor prognosis (P<0.05), high expression of COL5A2,
RBPMS2 It is negatively correlated with prognosis, but it is not statistically significant (P=0.1). We used GEPIA to verify the expression of 6 hub RBPs, which contains the sequencing data of TCGA and GTEx. GEPIA analysis results showed (Fig. 9B) that the expression of KIAA0101 and COL5A2 in GC tissues was significantly higher than that in normal tissues (P<0.05), and the expression of WIPF3 and RBPMS2 in normal tissues were increased (P<0.05). It is worth noting that DAZ1 and NOVA1 were higher in normal tissues, but they were not statistically significant. rt-PCR verified the expression of hub RBPs in 10 pairs of clinical samples. The results showed that (Fig. 9C) DAZ1, WIPF3, RBPMS2, and NOVA1 were low expressed in gastric cancer (P<0.05), and KIAA0101 and COL5A2 were highly expressed (P<0.05).

Figure 4. Prognostic analysis of 6-RBPs signature in the test cohort. A The survival analysis of 6-RBPs signature in the testing cohort. B 5-year time-dependent ROC for survival prediction models. C The distribution of risk score, OS and OS status of prognostic 6-RBPs signature in the testing cohort. D Univariate Cox regression analysis on the prognosis of clinicopathological characteristics and risk scores in patients with gastric cancer. E Multivariate Cox regression analysis on the prognosis of clinicopathological characteristics and risk scores in patients with gastric cancer.
Figure 5. Prognostic analysis of 6-RBPs signature in the complete cohort. A The survival analysis of 6-RBPs signature in the testing cohort. B 5-year time-dependent ROC for survival prediction models. C The distribution of risk score, OS and OS status of prognostic 6-RBPs signature in the testing cohort. D Univariate Cox regression analysis on the prognosis of clinicopathological characteristics and risk scores in patients with gastric cancer. E Multivariate Cox regression analysis on the prognosis of clinicopathological characteristics and risk scores in patients with gastric cancer.

Discussion

RBPs have diverse structures and functions, which regulate several necessary cellular processes. Some of these RBPs are usually expressed and conservatively evolved to maintain the basic processes of the cell. Unregulated expression of RNAs can cause many diseases, including cancer [26].
abnormality of RBPs has an important influence on tumor phenotype. Such as, mutations in TRBP lead to abnormal expression of miRNA and cancer cell proliferation and differentiation [27]. hnRNPs promote the synthesis of PKM2, thereby enhancing the Warburg effect [28]. RBM38 can inhibit breast cancer metastasis by promoting STARD13-related competitive endogenous RNA network [29]. In GC, only a few literatures have reported that RBPs are dysregulated in GC, there are also reports that RBPs are involved in tumor cell proliferation, metastasis and other phenotypes [16-18]. To this end, we collected 13 GPL570 platform microarray data in the GEO database to systematically identify the differentially expressed RBPs in GC. We have identified 25 up-regulated and 26 down-regulated RBPs. These DERBPs have been reported to be abnormally expressed in GC and play an important role. Among them, EZH2 is up-regulated in GC and is a good prognostic marker, which can inhibit p21 and promote the proliferation of GC [30]. In addition, it can also inhibit FBXO32 epigenetically and contribute to the resistance of GC [31]. However, some studies suggest that it plays a tumor suppressor effect in GC. Research by Zhao et al. suggested that EZH2-mediated epigenetic suppression of EphB3 can down-regulate E-cadherin and vimentin and inhibit GC metastasis [32]. This indicates that the role of RBPs in tumors is heterogeneous.

In order to further understand the possible mechanism of DERBPs affecting the progression of GC, we performed enrichment analysis on 51 RBPs. The results show that these DERBPs are mainly enriched in items such as posttranscriptional regulation of gene expression, mRNA binding, RNA catabolic process, RNA modification, and mRNA processing, some of them have been considered to play an important role in cancer progression. The most important thing in these processes is that RBPs can bind to mRNA, promote its stability, and regulate target gene expression. For example, PTBP3 can regulate the expression of the transcription factor ZEB1 by binding to the 3’UTR of its mRNA, thereby preventing its degradation, inducing the epithelial-mesenchymal transition of breast tumor cells and promoting their invasive growth and metastasis [33]. In addition, RBPs can also promote the degradation of mRNA. IGF2BP3 can accelerate the degradation of elf4e-bp2 mRNA, thereby promoting the proliferation of cervical cancer cells [34]. RBP-mediated selective splicing is a post-transcriptional regulatory mechanism that contributes to protein diversity and mRNA stability. Abnormal or wrong splicing is the main cause of abnormal function of RBPs, which promotes cancer [35]. Research by Liu et al. showed that SRNPB, the core component of alternative splicing, regulates the alternative splicing of intron 7 and its expression in RAB26 mRNA by activating NMD, mediating tumor growth and metastasis [36]. RBPs also participate in the addition of a poly(A) tail to the 3’end of the mRNA. Among them, alternative polyadenylation is a widespread basic regulatory mechanism in eukaryotes. This process can promote mRNA maturation, stability, nuclear transport and efficient translation. In cancer, research suggests that RBPs can regulate polyadenylation of transcripts, extend poly(A) tails and improve translation efficiency of target genes, and promote tumor progression [37].

Figure 6. The relationship between different risk groups and 6hub-RBPs and clinicopathological characteristics. A The relationship between different risk groups and 6hub-RBPs and clinicopathological characteristics. B The relationship between 6-hub RBPs and clinicopathological characteristics.
In addition, RBPs can not only affect tumor occurrence and development, but also serve as prognostic markers. We performed univariate Cox regression, Lasso regression and multivariate Cox regression analysis on all RBPs probe probes, and obtained 4 RBPs (DAZ1, KIAA0101, WIPF3, COL5A2, RBPMS2, NOVA1) that independently predict the prognosis of GC. In addition, we also analyzed their relationship with clinicopathological characteristics, suggesting that they are closely related to T staging and Lauren classification. There are reports that GC patients with high expression of KIAA0101 showed a
high recurrence rate, and accompanied by a poor prognosis, the vitality of GC cells was significantly inhibited after inhibiting its expression [38]. Research by Li et al. suggested that the low expression of NOVA1 in GC is related to lymphatic metastasis and poor prognosis [39]. Inhibiting the expression of NOVA1 in cells promotes the epithelial-mesenchymal transition of GC cells [39, 40]. Although some genes have not been studied in GC, they have been reported in other tumors. Research by Hapkova et al. suggested that RBPMS2 is highly expressed in gastrointestinal stromal tumors, and indicated that it is a new diagnostic marker and a potential cancer treatment target [41]. Among them, some reports are different from our conclusions. COL5A2 is considered to be a favorable prognostic factor for tongue squamous cell carcinoma [42]. However, high expression of this gene in muscle-invasive bladder cancer suggests that the prognosis is poor, and it is closely related to tumor invasion [43, 44]. DAZ1 and WIPF3 have not been reported in tumors, which suggests that they may be new predictive markers and therapeutic targets for GC and require more in-depth research.

Figure 8. The nomogram and the calibration chart are used to predict the OS survival rate of gastric cancer patients. A Nomogram for the prediction of OS at 1-5 year. B Calibration plots for predicting OS at 5 year.

Figure 9. Verify the prognosis and expression of 6 hub-RBPs. A K-M plotter verifies the prognosis of 6 hub-RBPs. B GEPIA verifies the expression level. C rt-PCR verifies the expression of hub RBPs in clinical samples.
Our other result is the construction of a prognostic signature based on RBPs. First we divide the whole cohort into a train set and a test set, build a prognostic signature in the train set based on hub RBPs, and verify it in the test set. The results show that the survival rates of the two concentrated high-risk groups are lower than those of the low-risk group, and the risk scores obtained are independent prognostic factors. The 5-year AUC calculated by the ROC curve is not 0.713 and 0.681, respectively, indicating that the model has broad applicability and accuracy. We also integrated the two sets into a complete prognostic signature, and obtained the same result as the training set after analysis, with a 5-year AUC value of 0.702. We further used the chi-square test to analyze the relationship between clinicopathological characteristics between high and low risk groups. The results show that it is related to AJCC stage, N stage, T stage, gender and survival status. We also used GSEA to analyze the differences in mechanisms affecting GC between high and low risk groups. The results suggest that high-group GC samples are mainly enriched in epithelial-mesenchymal transition (EMT), Kras, Notch, TGF-β, angiogenesis, Hedgehog, IL2-Stat5, hypoxia and other signals. Multiple reports suggest that RBPs mediate the above related signals to affect tumor progression. EMT is a biological process in which differentiated epithelial cells lose their epithelial characteristics and gain mesenchymal cell migration. Tumor cells with this characteristic exhibit high invasiveness and antagonistic ability to radiotherapy and chemotherapy. There are many ways to participate in EMT, including Notch, TGF-β, Hedgehog, Wnt, etc. [45]. It is reported that KIAA0101 can participate in the EMT of liver cancer and induce tumor cell migration and angiogenesis [46]. This may be triggered by the activation of Wnt signal conduction [47-49]. NOVA1 can also participate in EMT transition, which may be triggered by direct binding to β-catenin RNA [50]. Other ways have played an indispensable role in tumor development, so won’t repeat them.

We used the nomogram to optimize the prognostic signature and visualized it. Clinicians can detect the expression of hub RBPs and obtain the corresponding total score, which can predict the survival rate of GC patients in 1-5 years. In addition, we drew a calibration chart. Through the calibration chart, we can find that the predicted 5-year survival rate is almost the same as the actual survival rate, which shows the precise predictive ability of the nomogram. To verify the expression and prognosis of hub-RBPs. We verified the expression of hub RBPs in GEPIA, suggesting that their expression trends are consistent with our analysis. The results of K-M plotter suggest that the predictive ability of hub RBPs for GC is consistent with our analysis.

In a word, we have integrated 13 microarrays on the GPL570 platform, and after systematic bioinformatics analysis, we have identified the differentially expressed RBPs in GC and analyzed their potential mechanisms affecting GC. We also constructed a GC prognostic signature based on RBPs, analyzed and verified its good predictive ability. We also verified the expression of hub RBPs that construct the prognostic signature through external databases and clinical samples. We found that some RBPs are related to the pathogenesis of GC, but there is a lack of relevant literature reports and further studies on GC. This is the limitation of our research.

Supplementary Material


Acknowledgements

National Natural Science Foundation of China, Grant/Award Numbers: 81872480, 81760549, 81560492; Key R&D Program of Jiangxi Province, Grant/Award Numbers: 20203BBG73056; Science and Technology Research Project of Education Department of Jiangxi Province, Grant/Award Number: GJJ180024.

Ethics approval and consent to participate

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

Availability of data statement and materials

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

Competing Interests

The authors have declared that no competing interest exists.

References


http://www.jcancer.org