

Received: 2020.08.25

Accepted: 2020.10.30

Available online: 2020.11.18

Published: 2021.01.20

A Novel RNA-Binding Protein-Based Nomogram for Predicting Survival of Patients with Gastric Cancer

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

AB 1 Maoshu Zhu*
C 2 Jiading Cai*
D 3 Yulong Wu
E 4 Xinhong Wu
AF 5 Lianghua Feng
AG 6 Zhijiang Yin

1 Department of Science and Education, The Fifth Hospital of Xiamen, Xiamen, Fujian, P.R. China
2 Department of Gastroenterology, The Fifth Hospital of Xiamen, Xiamen, Fujian, P.R. China
3 Department of Urology, The Fifth Hospital of Xiamen, Xiamen, Fujian, P.R. China
4 Rehabilitation Department of Traditional Chinese Medicine, The Fifth Hospital of Xiamen, Xiamen, Fujian, P.R. China
5 Department of Rheumatology and Immunology, The Fifth Hospital of Xiamen, Xiamen, Fujian, P.R. China
6 Department of Hand and Foot Microsurgery, The Fifth Hospital of Xiamen, Xiamen, Fujian, P.R. China

* Maoshu Zhu and Jiading Cai equal contributors

Corresponding Authors: Lianghua Feng, e-mail: flh9453@126.com, Zhijiang Yin, e-mail: Yzj701023@163.com

Source of support: This research was supported by the Science and Technology Plan Project of Fujian (2018D0022, 2019D028) and the Xiamen Science and Technology Program of China (3502Z20184041, 3502Z20199150)

Background: We attempted to develop a prognostic model and characterize molecular subtypes for gastric cancer on the basis of ribonucleic acid (RNA)-binding proteins (RBPs).


Material/Methods: RNA sequence data of gastric cancer were obtained from The Cancer Genome Atlas. Univariate Cox regression analysis was used to screen survival-related RBPs, followed by least absolute shrinkage and selection operator Cox modeling. Overall and stratified survival analysis was carried out between high and low risk score groups, followed by receiver operator characteristic curve construction. Univariate and multivariate survival analysis was applied to assess its independent prognostic potential. A nomogram was constructed by combining age and the risk score, which was verified by calibration curves and decision curve analyses for 1-, 3-, and 5-year survival. Molecular subtypes were identified using nonnegative matrix factorization method. Clinical features of the identified subtypes were characterized on prognosis, drug sensitivity, and immune infiltration. An external Gene Expression Omnibus dataset was used to verify the above findings.

Results: On the basis of 44 survival-related RBPs, a robust prognostic 15-RBP signature was constructed. Patients with high risk score had a poorer prognosis than those with low risk score. The risk score had good performance in predicting clinical outcomes for 1-, 3-, and 5-year survival. The signature was effectively independent of other clinical features. The nomogram model combining age and the 15-RBP prognostic model exhibited better practicality and reliability for prognosis. RBP expression data were utilized to define 2 distinct molecular subtypes obviously related to survival outcomes, chemotherapeutic drug sensitivity, and immune infiltration.

Conclusions: Our study provides a nomogram model that consists of age and a 15-RBP signature and identifies 2 molecular subtypes for gastric cancer that possess potential value for preclinical, clinical, and translational research on gastric cancer.

MeSH Keywords: **Nomograms • Prognosis • RNA-Binding Proteins • Stomach Neoplasms**

Full-text PDF: <https://www.medscimonit.com/abstract/index/idArt/928195>

 3815

 2

 9

 31



Background

Gastric cancer is a highly lethal and heterogeneous malignant tumor found globally [1,2]. Despite the progression of innovative treatment technologies such as targeted therapy and immunotherapy, gastric cancer patients' prognosis remains unsatisfactory. For some patients with gastric cancer, even those with the same tumor lymph node metastasis (TNM) stage, the prognosis and treatment response are different [3]. Studies have shown that clinical characteristics such as age, sex, and TNM are not sufficient to accurately predict patients' clinical outcomes [4]. Therefore, a comprehensive analysis of molecular characteristics of gastric cancer is required.

RNA-binding proteins (RBPs) are a group of proteins containing ribonucleic acid (RNA)-binding domains that mediate processes such as RNA maturation, splicing, localization, and translation [5]. RBPs have emerged as major players in cell physiology and various biological processes [6,7]. In addition, it is increasingly recognized that disorder of RBPs is closely related to various human cancers [8]. More important, many studies have confirmed the importance of RBPs in gastric cancer [9–11]. Although the association between RBPs and gastric cancer prognosis has been widely reported, few specific prognostic models have been constructed. Systematic analysis of RBPs may provide novel insights into the underlying mechanisms of gastric cancer.

The advancement of high-throughput sequencing technologies such as RNA-seq and microarrays provides opportunities to fully characterize the molecular characteristics of tumorigenesis [12]. Moreover, the application of these high-throughput technologies has facilitated the identification of promising biomarkers for cancer diagnosis and prognosis assessment. However, these studies usually focus on comparing the differences in gene expression between tumor and normal samples, which are easy to ignore. Those genes are not significantly different but may have crucial biological significance, thereby leading to neglect of biological information and gene regulatory networks. Moreover, considering that individual biomarkers usually have little statistical power, identification of new molecular features may provide more reliable predictions. The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) are 2 public platforms that can collect large amounts of transcriptome data. Bioinformatics analysis based on high-throughput data in public databases can obtain relatively reliable and accurate results through a large clinical sample size. In this study, we downloaded the available data from TCGA database and performed a series of bioinformatics analyses to detect the potential molecular functions and clinical significance of RBPs in gastric cancer. We constructed a nomogram model combining age and a 15-RBP signature and characterized molecular features for gastric cancer on the basis of RBPs,

which were validated through the GSE84437 dataset. Our findings may lay the foundation for the development of new strategies for gastric cancer treatment and prognosis assessment.

Material and Methods

Data acquisition and preprocessing

RNA-seq gene expression data of gastric cancer (version 07-20-2019) generated by the Illumina HiSeq 2000 RNA sequencing platform were downloaded from TCGA by UCSC Xena (<https://xenabrowser.net/>). We obtained a total of 407 samples composed of 375 tumor samples and 32 normal samples. After removing samples with incomplete survival time information, 350 tumor samples were retained for this study. Data were normalized by transcripts per million (TPM) method. We also collected patients' clinical information including sex, grade, stage (7th edition American Joint Committee on Cancer staging system), age, and survival time. The microarray data of GSE84437 dataset that contained 433 gastric cancer samples were retrieved from the GEO (<http://www.ncbi.nlm.nih.gov/geo>) database. TCGA dataset was used as the training set and GSE84437 dataset as the validation set.

Least absolute shrinkage and selection operator (LASSO) Cox model

Univariate Cox regression analysis was used to screen survival-related RBPs by the survival package in R. The log-rank test was performed to assess significant survival-related RBPs for gastric cancer. A LASSO Cox model was constructed for selection of key survival-related RBPs using the Glmnet package in R [13]. Ten-fold cross-validation was performed for each model corresponding to λ . The partial likelihood deviance values corresponding to each λ were calculated. The optimal λ corresponding to the minimum partial likelihood deviance value was selected to construct the model. The risk score of the patients in the model was calculated by the following formula: risk score = $\beta_1\chi_1 + \beta_2\chi_2 + \dots + \beta_p\chi_p$ (where β_i is the nonzero LASSO coefficient corresponding to the optimal λ value and χ_i is the gene expression level corresponding to the nonzero LASSO coefficient). The median risk score value was determined as the cutoff value. In line with the cutoff value, all patients from TCGA and GSE84437 datasets were separated into high and low risk score groups. Overall survival analysis was carried out between the 2 groups using the log-rank test. The receiver operating characteristic (ROC) curve was conducted utilizing the SurvivalROC package in R.

Stratified survival analysis

Correlation analysis between the risk score and different clinical features (sex, grade, and stage) was presented for patients in the training set. Comparisons between 2 groups were tested by the Wilcoxon test, whereas multiple comparisons were tested by the Kruskal-Wallis test. Patients in the training set were divided into different subgroups according to age, sex, grade, and stage. Kaplan-Meier survival analysis followed by log-rank test was carried out between high and low risk scores in different subgroups.

Univariate and multivariate survival analysis

First, univariate survival analysis was utilized to assess the prognostic efficiency of risk score and other clinical characteristics including age, sex, grade, stage, and TNM. Then, multivariate survival analysis was presented to evaluate whether the risk score was an independent risk factor. Risk factors with hazard ratio >1 and $P<0.05$ and protective factors with hazard ratio <1 and $P<0.05$ were defined.

Gene set enrichment analysis (GSEA)

Tumor samples (375) from TCGA database were divided into high and low risk score groups. GSEA of the 2 groups was presented and potential biological functions of RBPs were explored [14]. $P<0.05$, false discovery rate <0.05 , and enrichment score >0.6 were set as the screening criteria.

Nomogram model construction

A nomogram was constructed by combining age and the 15-RBP risk score. The predicted survival probability of 1, 3, and 5 years by the nomogram was compared with the actual survival probability. To verify this nomogram, the total points of all patients were calculated on the basis of the nomogram, and then used as a factor in the Cox regression analysis. Finally, the calibration curves were derived from the regression results. Decision curve analyses were utilized to evaluate the patient's benefits due to age, prognosis model, and nomogram in 1, 3, and 5 years.

Nonnegative matrix factorization (NMF)

Clustering analysis was presented utilizing the NMF package in R [15]. The clustering numbers (k) were set as 2–7. The k value that resulted in the maximum cophenetic correlation coefficient was selected as the optimal clustering number, which was validated using principal component analysis (PCA).

Genomics of drug sensitivity in cancer (GDSC)

GDSC (version 8.3; <https://www.cancerrxgene.org/>) as an open public database provides information concerning molecular biomarkers of drug sensitivity and drug response in different cancer cells and contains drug sensitivity data from almost 75,000 experiments as well as drug response data from nearly 700 cancer cells and 138 anticancer drugs [16]. In this study, the estimated half-maximal inhibitory concentration of 20 kinds of chemotherapy drugs in the 2 molecular subtypes was assessed and was tested by the Wilcoxon test. $P<0.05$ was set as the cutoff value.

Estimation of immune infiltration

CIBERSORT (<http://cibersort.stanford.edu/>) was utilized to assess the abundance of different immune cell types in mixed cell populations [17]. The fraction of 22 immune cells in each sample was calculated and an algorithm was run on 1000 permutations with the LM22 feature matrix. For each sample, the sum of the proportion of all immune cell types was equal to 1.

Results

Construction and evaluation of a 15-RBP signature for gastric cancer

In TCGA-stomach adenocarcinoma dataset ($n=350$), 44 survival-related RBPs were screened out according to univariate Cox regression analysis (**Supplementary Table 1**). LASSO Cox regression analysis was then presented to select robust survival-related RBPs among the 44 candidates. When $\log \lambda=-3.6$, the model had optimal performance but the least number of independent variables (**Figure 1A**). The LASSO coefficients of 44 survival-related RBPs approached zero with the increase of λ (**Figure 1B**). When $\log \lambda=-3.6$, 15 key RBPs that had a distinct impact on prognosis were screened. Finally, a 15-RBP prognostic model was conducted. The risk score of each patient was calculated as follows:

$$\begin{aligned} & [(-0.15669) \times \text{expression value of MS12}] + [(-0.08983) \times \text{expression value of METTL2B}] + [(0.14363) \times \text{expression value of DYNLL1}] \\ & + [(0.25025) \times \text{expression value of SMAD5}] + [(0.03991) \times \text{expression value of PEG10}] + [(0.11085) \times \text{expression value of RNASE1}] \\ & + [(-0.21239) \times \text{expression value of HEXIM2}] + [(-0.21012) \times \text{expression value of TRIM25}] + [(0.09234) \times \text{expression value of ZFP36}] \\ & + [(0.08972) \times \text{expression value of FTO}] + [(-0.07904) \times \text{expression value of FAM98C}] + [(0.39763) \times \text{expression value of RPS4Y2}] \\ & + [(-0.12430) \times \text{expression value of ADAT3}] + [(0.03636) \times \text{expression value of RPS23}] + [(0.03167) \times \text{expression value of IFIT1}] \end{aligned}$$

(where the expression level of each gene was measured by TPM).

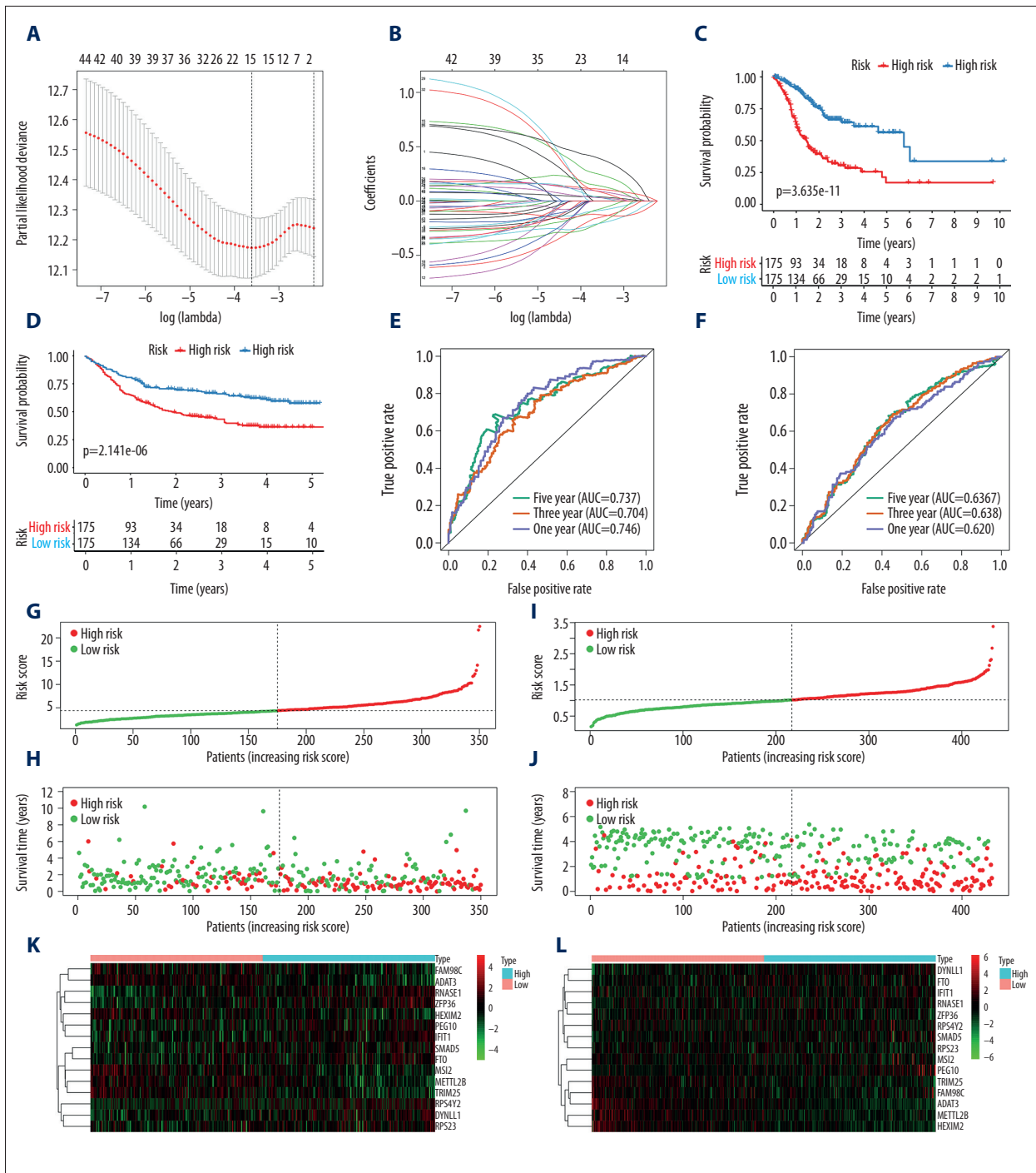


Figure 1. Construction and evaluation of a 15-RBP signature for gastric cancer in the training and validation sets. **(A)** Selection of λ by 10-fold cross-validation. **(B)** Fitting processes of LASSO Cox regression model. Each curve represents the change trajectory of each RBP. The ordinate is the value of the coefficient, the lower abscissa is $\log(\lambda)$, and the upper abscissa is the number of variables in the model corresponding to each λ . **(C, D)** Kaplan-Meier overall survival analysis of patients in the high and low risk score groups for training and validation sets. **(E, F)** Construction and validation of ROC curves for 1, 3, and 5 years. **(G, H)** Distribution of risk scores and survival status for all patients in the training set. **(I, J)** Distribution of risk scores and survival status for all patients in the validation set. **(K, L)** Hierarchical clustering analysis visualizing the difference in expression of 15 RBPs between high and low risk score groups in the training and validation sets. Red represents upregulation and green represents down-regulation.

Table 1. Clinical characteristics of gastric cancer patients in the whole cohort, high-risk, and low-risk groups from TCGA database.

Characteristics		High risk (N=175)	Low risk (N=175)	Total (N=350)	P-value
Age	<65	84	66	150	0.0663
	≥65	91	109	200	
Stage	Stage I	17	32	49	0.429
	Stage II	55	56	111	
	Stage III	84	71	155	
	Stage IV	19	16	35	
T	T1	1	15	16	0.0343
	T2	38	36	76	
	T3	83	78	161	
	T4	49	46	95	
	Tx	4	0	4	
M	M0	156	156	312	0.6092
	M1	13	10	23	
	Mx	6	9	15	
N	N0	40	64	104	0.0244
	N1	53	40	93	
	N2	38	34	72	
	N3	36	35	71	
	Nx	8	2	10	
Gender	Female	55	69	124	0.1463
	Male	120	106	226	
Grade	G1	3	6	9	0.0536
	G2	52	73	125	
	G3	114	93	207	
	Gx	6	3	9	

The median value was used as the cutoff value of the risk score. All patients were divided into high- and low-risk groups on the basis of the median risk score in the training set (cutoff value=4.366934127) and validation set (cutoff value=1.023900687). **Table 1** lists the clinical characteristics of gastric cancer patients in the whole training cohort, high-risk groups, and low-risk groups. There were significant differences in T stage ($P=0.0343$) and N stage ($P=0.0244$) between high- and low-risk groups. Both in the training and validation datasets, patients with high risk score exhibited a poorer prognosis than those with low risk score (**Figure 1C, 1D**). In the training set, ROC curve results showed that the area under the curves (AUCs) of 1, 2, and 5 years were 0.746, 0.704, and 0.737, respectively, suggesting that this 15-RBP model had good performance (**Figure 1E**). We then further validated the efficiency of the model in the validation set. The AUCs of 1, 2, and 5 years were 0.620, 0.638, and 0.636, respectively, implying the high generalization of the model (**Figure 1F**). Furthermore, we

visualized the distribution of risk scores and survival status of all patients in the training and validation sets, respectively (**Figure 1G–1J**). In the training set, the significant difference in expression of these 15 RBPs between high and low risk score groups is shown in **Figure 1K**. Similar results were observed in the validation set (**Figure 1L**).

Clinical features of the 15-RBP model for gastric cancer

The correlation between the risk score and other clinical features including sex, grade, stage, and age was analyzed in the training set. We found that male patients had significantly higher risk scores than female patients (**Figure 2A**). As the grade increased, the patients' risk score also gradually increased (**Figure 2B**). Compared with stage I, there were distinctly higher risk scores for patients with stage II, stage III, and stage IV (**Figure 2C**). These results suggested that the risk score may be related to the severity of the disease. We further conducted

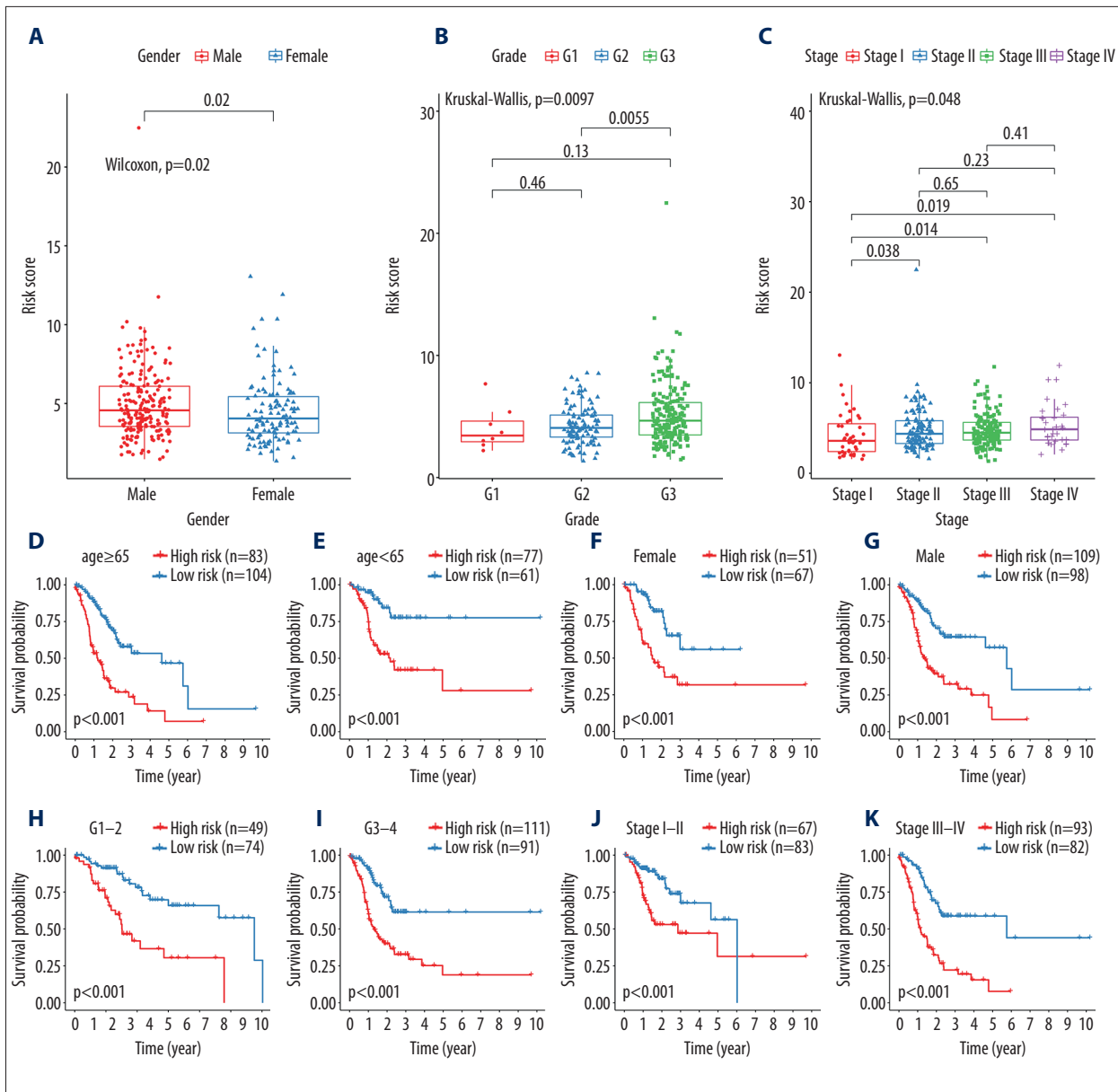


Figure 2. Clinical features of the 15-RBP model for gastric cancer. (A–C) Box plots depicting the correlation between risk scores and different clinical features including sex, grade, and stage. Stratified survival analysis of the risk scores was presented according to age (≥ 65 and < 65 ; D, E), sex (female and male; F, G), grade (grade 1–2 and grade 3–4; H, I), and stage (stage I–II and stage III–IV; J, K).

subgroup analysis to validate the reliability of the model. Both for patients ages ≥ 65 and < 65 , high risk score significantly indicated a poorer prognosis (Figure 2D, 2E). Although there were significant differences in the risk score between male and female patients, high risk scores still strongly predicted poorer prognosis (Figure 2F, 2G). Both for patients with grade 1–2 and grade 3–4, high risk score was distinctly associated with worse survival time (Figure 2H, 2I). Similarly, regardless of stage I–II or stage III–IV patients, those in the high risk score group exhibited poorer prognosis than those in the low risk

score group (Figure 2J, 2K). The above results demonstrated the practicality and reliability of the model.

The 15-RBP signature as an independent prognostic factor for gastric cancer

Univariate Cox regression analysis was utilized to screen prognosis-related clinical factors including age, sex, grade, stage, TNM, and risk score. The results showed that age, stage, T stage, N stage, and risk score were all risk factors for patients with

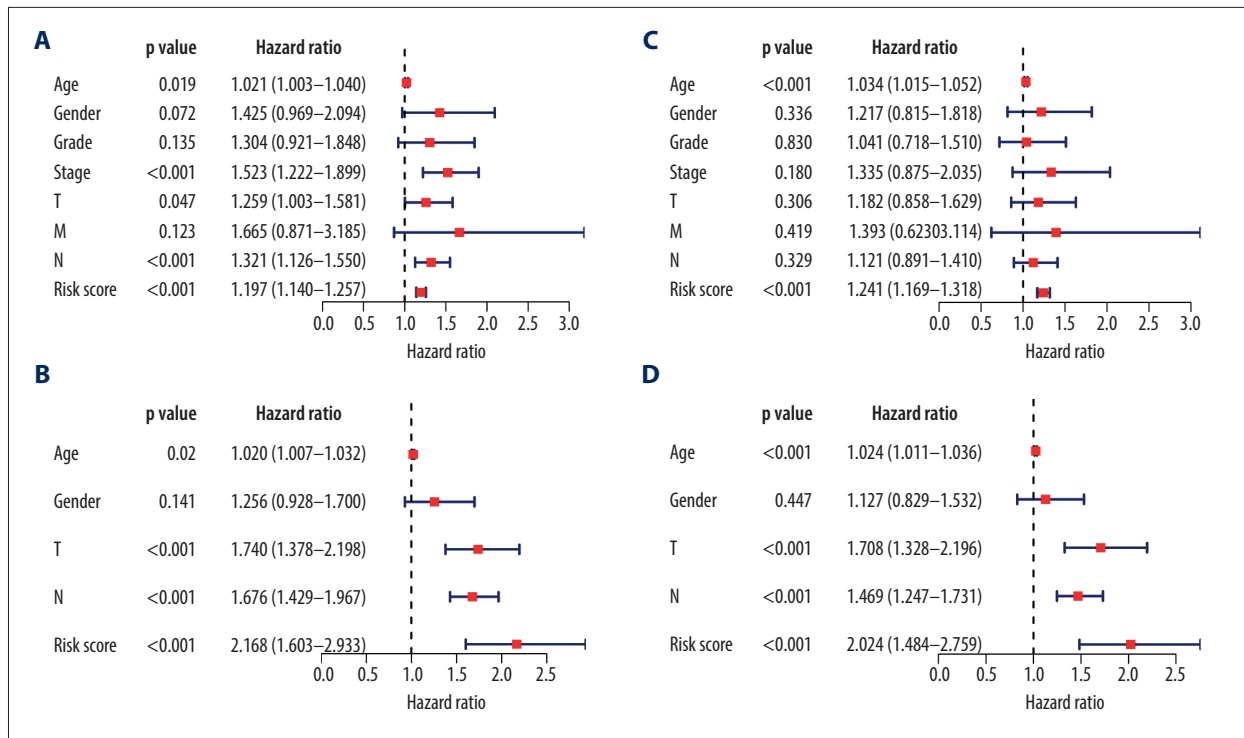


Figure 3. Assessment and validation of the 15-RBP signature as an independent prognostic factor for gastric cancer. (A, B) Univariate Cox regression analysis of clinical features and risk score in the training and validation sets. (C, D) Multivariate Cox regression analysis in the training and validation sets.

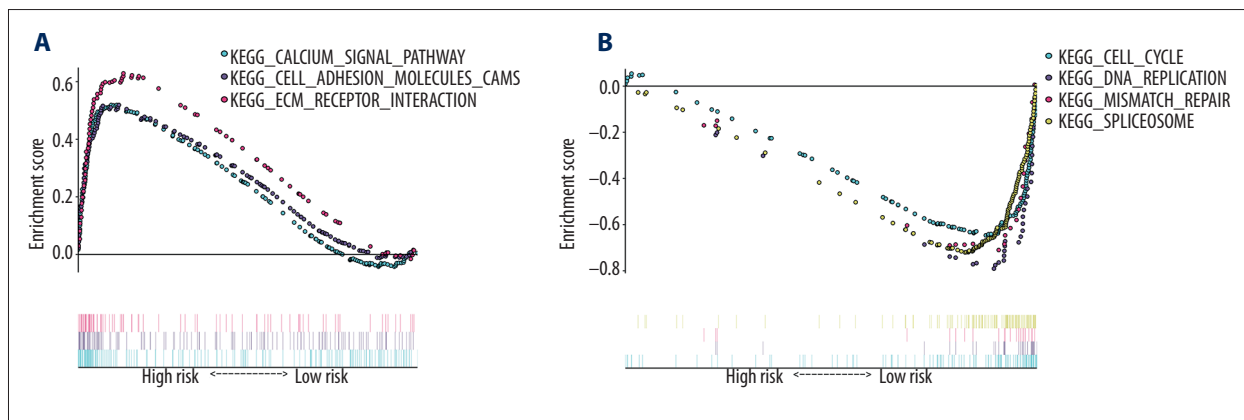


Figure 4. GSEA enrichment analysis of gene sets in the high and low risk score groups. (A) The most distinct pathways correlated with the high-risk score group. (B) The most significant pathways are associated with the low risk score group.

gastric cancer in the training set (Figure 3A), and were confirmed in the validation sets (Figure 3B). After adjusting other clinical factors, the risk score was an independent prognostic factor for gastric cancer in the training set ($P<0.001$, hazard ratio [HR]=1.670, 95% confidence interval [CI]=1.332–2.094; Figure 3C), which was confirmed in the validation set ($P<0.001$, HR=2.144, 95% CI=1.601–2.871; Figure 3D). Furthermore, age, T stage, and N stage were also independent prognostic factors for gastric cancer in the validation set.

Biological functions of these 15 RBPs

GSEA enrichment analysis was presented to probe into the differences in biological functions of RBPs between high and low risk score groups. These results demonstrated that high risk score was significantly related to calcium signaling pathway, cell adhesion molecules, and extracellular matrix (ECM) receptor interaction (Figure 4A). Low risk score was distinctly enriched in cell cycle, deoxyribonucleic acid (DNA) replication, mismatch repair, and spliceosome (Figure 4B).

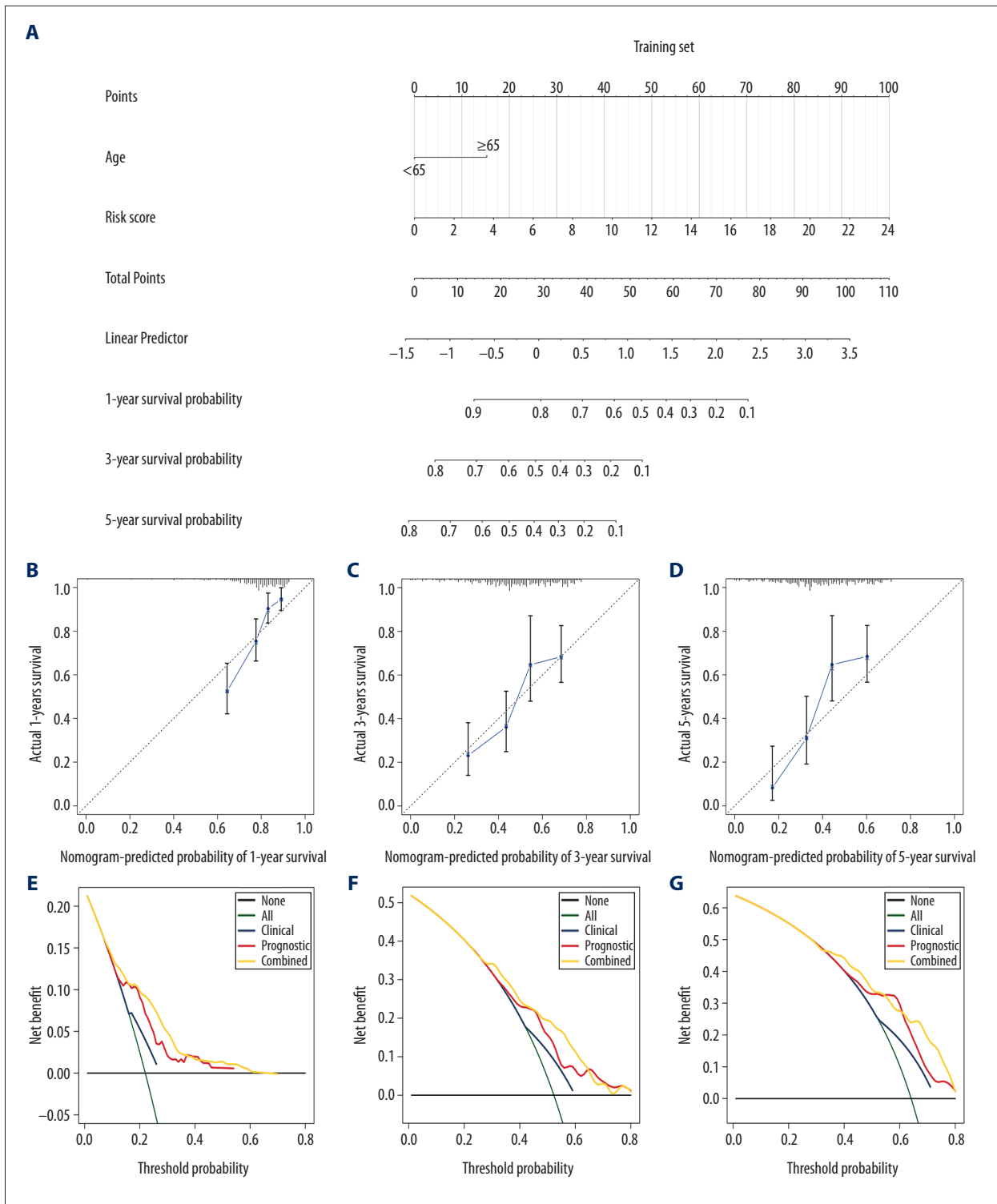


Figure 5. Conduction of evaluation of the nomogram model integrated risk score and age in the training cohort. **(A)** Nomogram for predicting the proportion of patients with overall survival after diagnosis of gastric cancer in the training cohort. **(B–D)** Calibration plots of nomograms in terms of agreement between predicted and actual 1-, 3-, and 5-year outcomes. **(E–G)** Decision curve analysis of the nomogram model, age, and prognostic model for 1-, 3-, and 5-year outcomes.

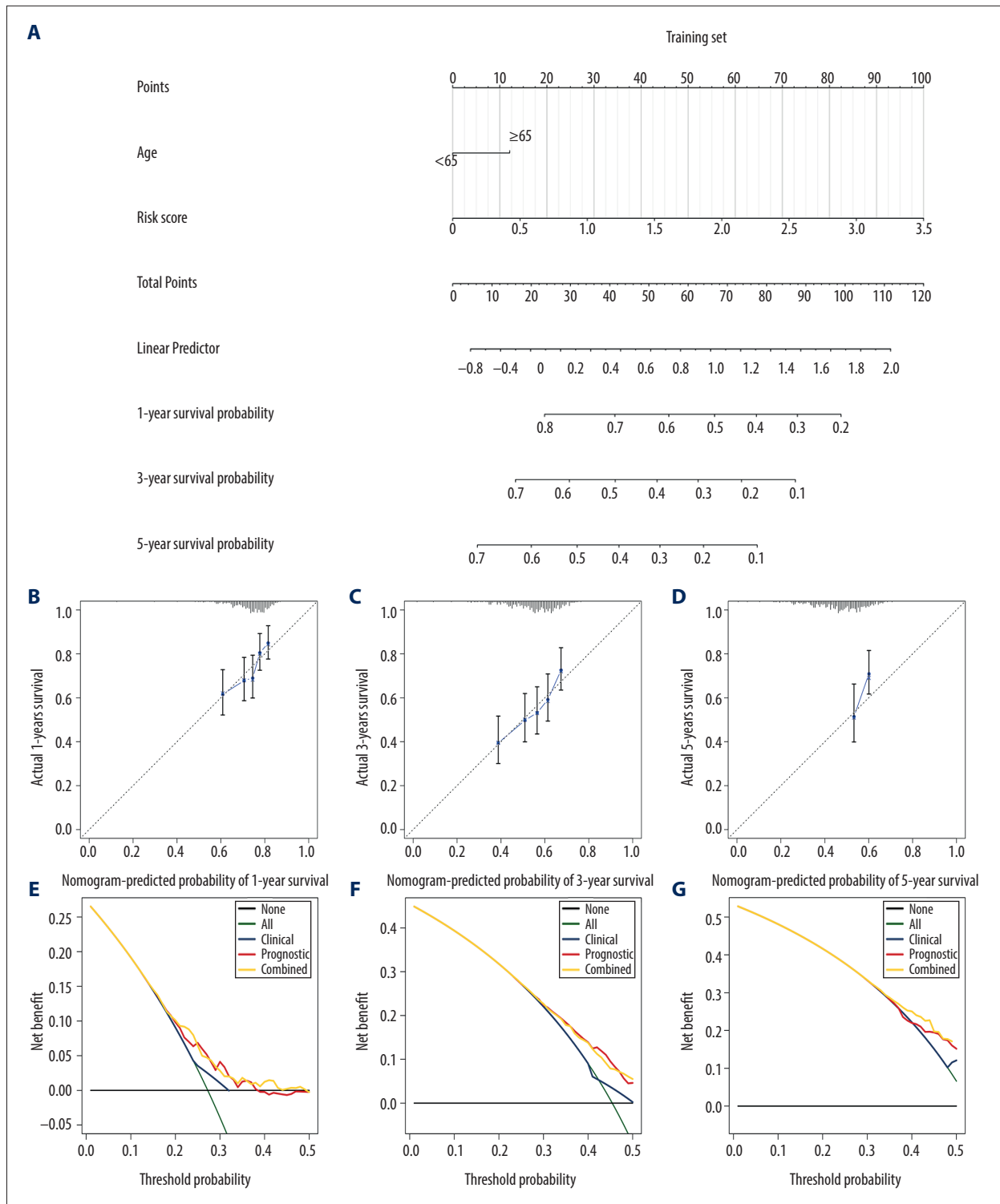


Figure 6. Validation of the nomogram model combining risk score and age in the validation cohort. **(A)** Nomogram for predicting the proportion of patients with overall survival after diagnosis of gastric cancer in the validation cohort. **(B–D)** Calibration plots of nomograms in terms of agreement between predicted and actual 1-, 3-, and 5-year outcomes. **(E–G)** Decision curve analysis of the nomogram model, age, and prognostic model for 1-, 3-, and 5-year survival time.

Construction of a nomogram model for gastric cancer

The nomogram model was constructed to make more accurately personalized predictions for gastric cancer patients. According to multivariate survival analysis results, age and risk score were independent predictors of survival both in the training set and validation set, which were included in a nomogram model for predicting the death risk of gastric cancer patients at the initial diagnosis. Then, a nomogram model was built to predict the 1-, 3-, and 5-year survival rate of patients. As shown in **Figure 5A**, the 15-RBP prognostic model could contribute the most to the prognosis because it had the highest points. The calibration graph showed that the nomogram model possessed high agreement in predicting the 1- (**Figure 5B**), 3- (**Figure 5C**), and 5-year (**Figure 5D**) survival rates of gastric cancer patients because the prediction line almost coincided with the actual line. Furthermore, decision curve analyses were utilized to compare the prediction efficiency of age, the prognostic model, and the nomogram model for 1-, 3-, and 5-year survival time. From the results shown in **Figure 5E–5G**, within a large threshold probability range, whether it was 1, 3, or 5 years, the net benefit rate of the nomogram model was significantly better than the age or the prognosis model, indicating that the nomogram model containing age and prognostic model data exhibited high clinical value.

Validation of the nomogram model

The nomogram model combining age and the 15-RBP prognostic model was externally validated using the GSE84437 dataset (**Figure 6A**). Consistent with the results in the training cohort, the 1- (**Figure 6B**), 3- (**Figure 6C**), and 5-year (**Figure 6D**) survival rates predicted by the nomogram were highly consistent with the actual observations in the calibration plots. Moreover, decision curve analysis results suggested that the nomogram had a higher net benefit rate compared with age or the prognosis model for 1- (**Figure 6E**), 3- (**Figure 6F**), and 5-year (**Figure 6G**) survival time. Thus, the nomogram model exhibited practicality and reliability.

Identification of 2 distinct molecular subtypes for gastric cancer on the basis of prognosis-related RBPs

Forty-four survival-related RBPs from univariate Cox regression analysis were clustered utilizing NMF. The clustering number k was set at 2–7 (**Figure 7A**). The optimal k was determined on the basis of cophenetic, dispersion, and silhouette (**Figure 7B**). For these RBPs, the optimal clustering number was 2 in the training set. There was a distinct difference in clusters 1 and 2, as shown in PCA results (**Figure 7C**). The 2 molecular subtypes were associated with overall survival of gastric cancer patients. Patients in cluster 1 had a poorer prognosis than those in cluster 2 (**Figure 7D**). Heat maps depicted expression patterns for RBP sets (**Figure 7E**). There was

a distinct difference in expression pattern of these RBPs between the 2 molecular subtypes.

Verification of distinct molecular subtypes for gastric cancer

Consistent with NMF results in the training set, 2 distinct molecular subtypes were identified in the validation set (**Figure 8A, 8B**). PCA results depicted a significant difference between the 2 subtypes (**Figure 8C**). However, no obvious difference in prognosis was found between the 2 subtypes (**Figure 8D**). Similarly, there was a prominent difference in expression of RBPs in 2 subtypes (**Figure 8E**).

Chemosensitivity and immune infiltration differ between gastric cancer RBP subtypes

We assessed the sensitivity to 20 kinds of chemotherapy drugs for the 2 clusters. As depicted in **Figure 9A**, cluster 1 had distinctly higher sensitivity to A.770041, ABT.263, AMG.706, AP.24534, AS601245, ATRA, AUY922, Axitinib, AZD.0530, AZD.2281, AZD6482, AZD7762, and AZD8055 compared with cluster 2. Moreover, cluster 2 exhibited higher sensitivity to A.443654, ABT.888, AICAR, AKT inhibitor VIII, AZ628, and AZD6244 in comparison with cluster 1, but was not statistically significant. We compared the differences in fractions of different immune cells between the 2 subtypes. The results showed that the fractions of plasma cells ($P < 0.01$), CD4 memory-activated T cells ($P < 0.001$), follicular helper T cells ($P < 0.05$), resting NK cells ($P < 0.001$), activated NK cells ($P < 0.05$), and monocytes ($P < 0.001$) were significantly higher in cluster 2 than in cluster 1 (**Figure 9B**). We analyzed the differences in immune-related markers between the 2 subtypes. Compared with cluster 2, the expression levels of CCL2 ($P < 0.001$), CD276 ($P < 0.001$), CD4 ($P < 0.01$), CXCR4 ($P < 0.001$), IL6 ($P < 0.001$), and TGFBI ($P < 0.001$) were distinctly higher in cluster 1 (**Figure 9C**). There were higher expression levels of CD274 ($P < 0.05$) and IL1A ($P < 0.05$) in cluster 2 than in cluster 1 (**Figure 9C**).

Discussion

Malignant tumors are characterized by uncontrolled cell growth, mainly due to the dysregulation of RBPs that regulate cell biological processes. Thus, it is valuable to explore important RBPs and understand the potential carcinogenic molecular mechanisms of gastric cancer [18]. In clinical practice, gastric cancer patients with high risk scores need more in-depth monitoring and active treatment. Thus, there is an urgent need for powerful biomarkers to stratify the high-risk population among gastric cancer patients. However, a single biomarker usually has little predictive power. The established clinical prognostic markers had limited accuracy and specificity. In this study, we conducted a nomogram model combined with age and a

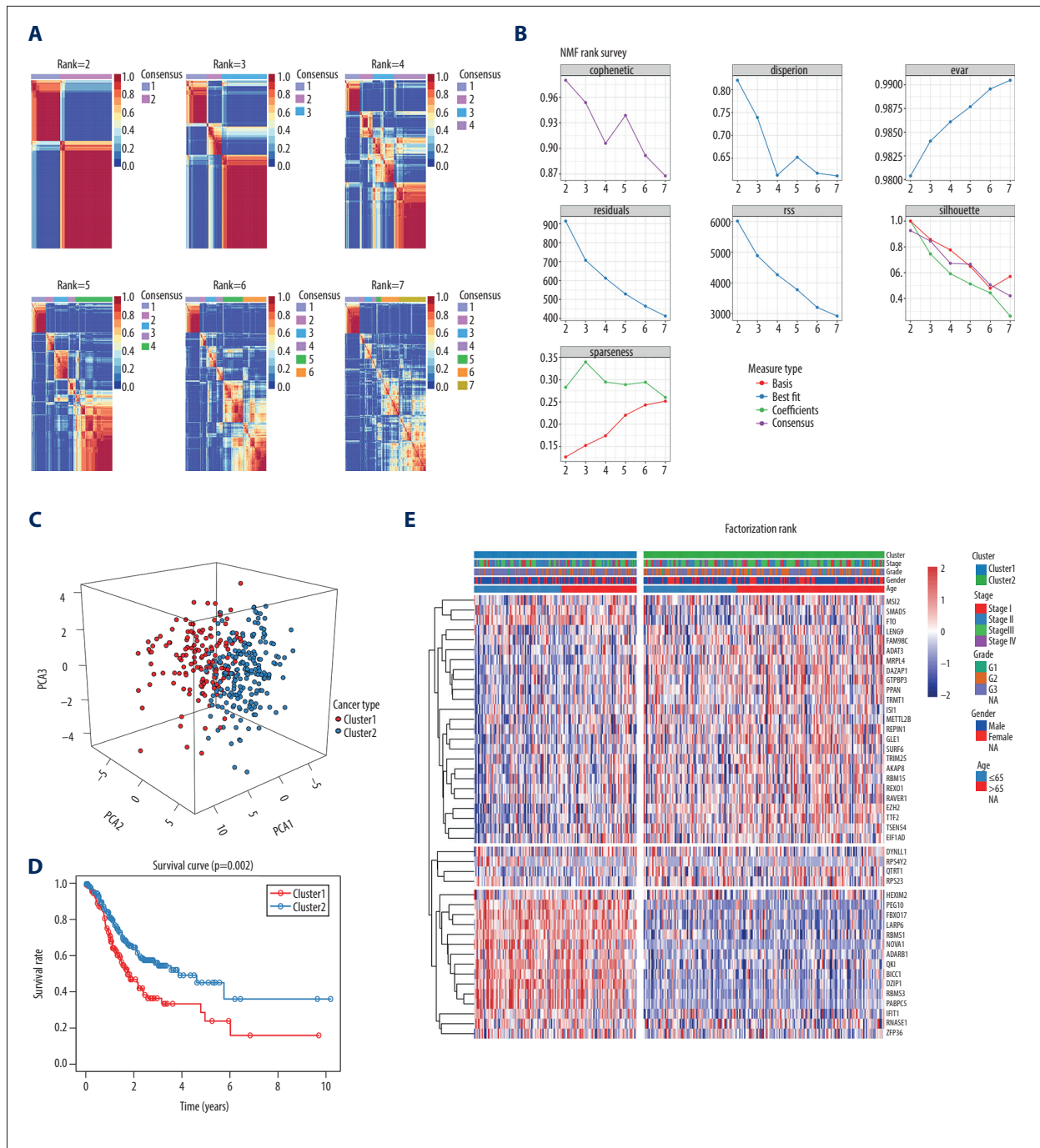


Figure 7. Identification of 2 distinct molecular subtypes for gastric cancer by iCluster in TCGA cohort. **(A, B)** NMF clustering results for prognosis-related RBPs. **(C)** PCA results. **(D)** Overall survival analysis for 2 molecular subtypes. **(E)** Gene expression signatures of 2 distinct molecular subtypes.

15-RBP risk score to robustly predict prognosis of gastric cancer patients, and characterized 2 distinct molecular subtypes on the basis of 44 prognosis-related RBPs.

On the basis of LASSO Cox regression analysis, we developed a risk scoring model based on 15 RBP signatures. In addition,

we investigated the prognostic value of signatures in different subgroups. A signature is a promising index independent of different clinicopathological characteristics. Among 15 RBPs, MSI2 has been found to be overexpressed in gastric cancer tissue samples. MSI2 messenger (m)RNA expression is significantly correlated with TNM staging. Patients with gastric cancer with

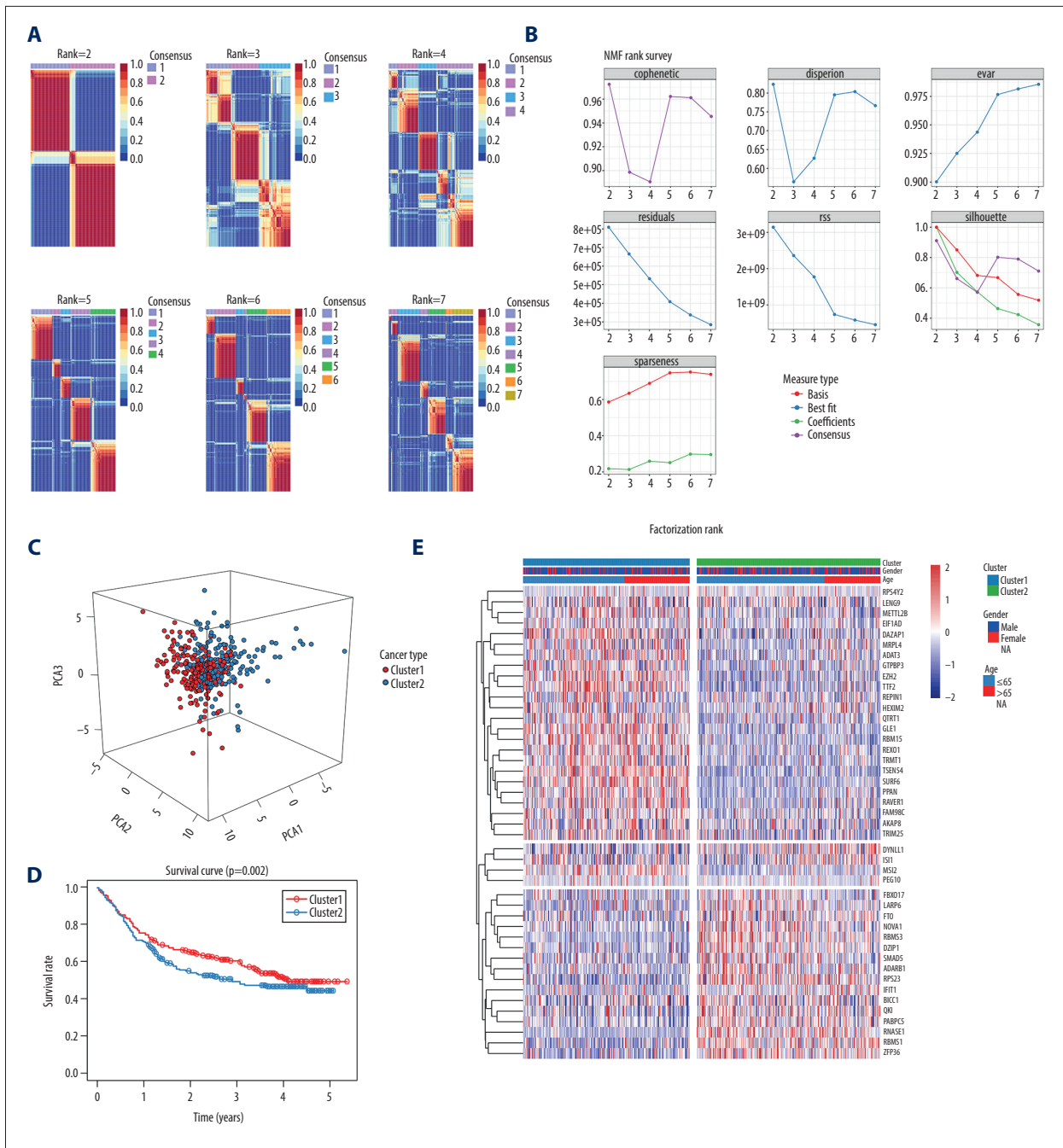


Figure 8. Verification of molecular subtypes for gastric cancer in the validation set. **(A, B)** NMF clustering results for prognosis-related RBPs. **(C)** PCA results. **(D)** Overall survival analysis for 2 molecular subtypes. **(E)** Gene expression signatures of 2 distinct molecular subtypes.

high MSI2 expression exhibit a poor prognosis [19]. SMAD5 is a key element for gastric cancer according to previous bioinformatics analysis [20]. Upregulation of PEG10 has been detected in clinical samples of gastric cancer. PEG10 knockdown effectively inhibits the malignant behaviors of gastric cancer cells [21]. TRIM25 could contribute to gastric cancer metastasis [22]. Several studies have reported that FTO possesses

key clinical value in gastric cancer [23–25]. RPS23, as a hub gene of gastric cancer, is related to the pathology and prognosis of patients [26]. In our research, 15 RBPs were closely related to the prognosis of gastric cancer. However, in gastric cancer-related research, the research on these RBPs was still insufficient. To a certain extent, our research may provide some clues for further research on the biological functions

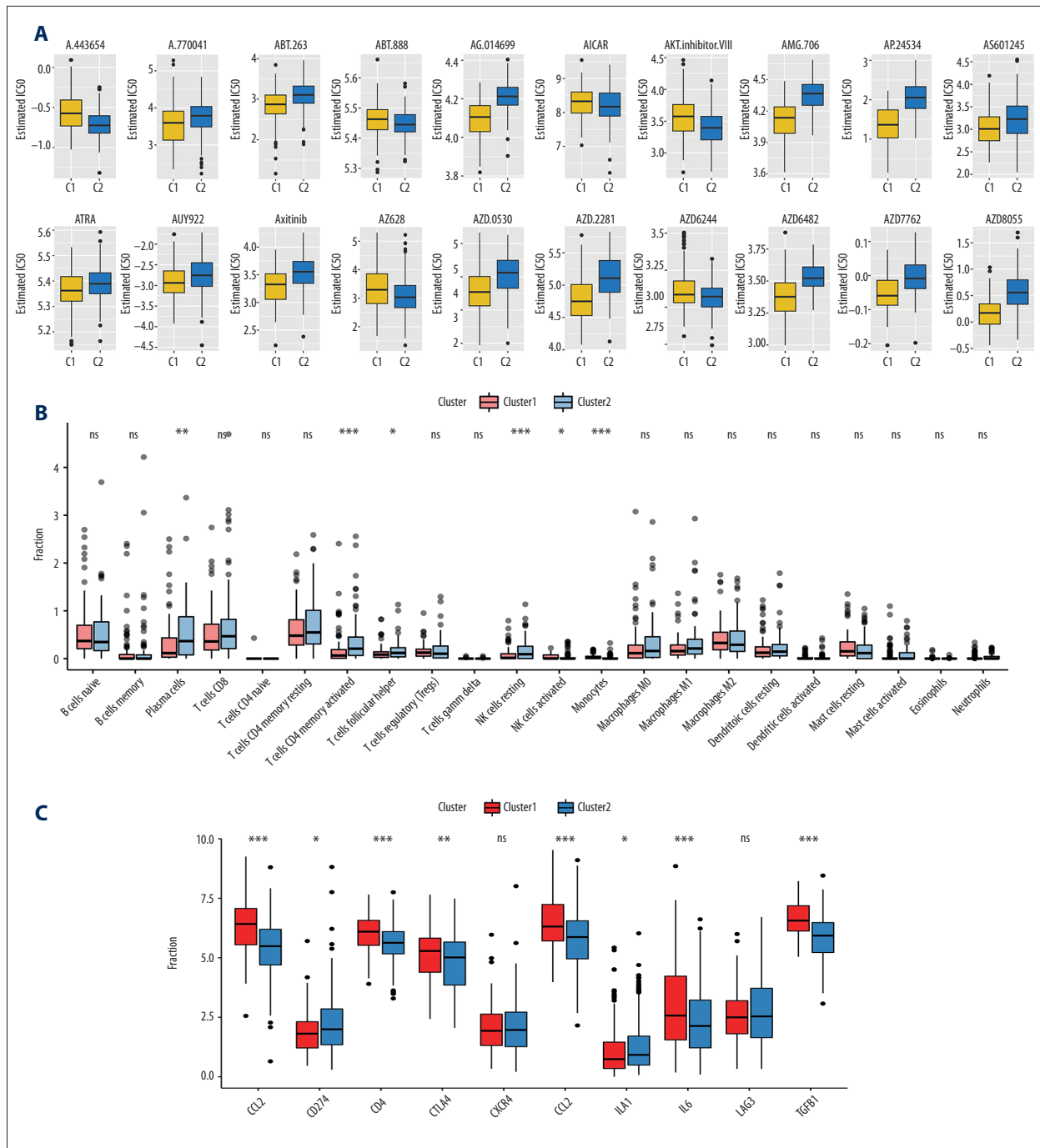


Figure 9. Chemosensitivity and immune infiltration differ between gastric cancer RBP subtypes. **(A)** Box plots depicting the difference in sensitivity to 20 chemotherapy drugs including A.443654 ($P=0.999999985$), A.770041 ($P=0.000181248$), ABT.263 ($P=1.42 \times 10^{-9}$), ABT.888 ($P=0.98841534$), AG.014699 ($P=7.49 \times 10^{-29}$), AICAR ($P=0.959372747$), AKT.inhibitor VIII ($P=0.999999804$), AMG.706 ($P=3.80 \times 10^{-28}$), AP.24534 ($P=6.32 \times 10^{-32}$), AS601245 ($P=1.52 \times 10^{-5}$), ATRA ($P=0.000595412$), AUY922 ($P=0.000114653$), Axitinib ($P=3.53 \times 10^{-11}$), AZ628 ($P=0.999104068$), AZD.0530 ($P=3.53 \times 10^{-11}$), AZD.2281 ($P=2.49 \times 10^{-17}$), AZD6244 ($P=0.995523384$), AZD6482 ($P=9.92 \times 10^{-20}$), AZD7762 ($P=9.29 \times 10^{-17}$), and AZD8055 ($P=4.47 \times 10^{-27}$) between the 2 clusters. IC50, the half-maximal inhibitory concentration. **(B)** Differences in fractions of different immune cells between the 2 subtypes. **(C)** Differences in expression patterns of different immune-related markers between the 2 subtypes. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; ns – no statistical significance.

and clinical significance of these RBPs. In addition to survival prediction, GSEA also showed that there was a distinct difference in cancer features between high- (calcium signaling pathway [27], cell adhesion molecules [28], and ECM receptor interaction [29]) and low- (cell cycle, DNA replication, mismatch repair, and spliceosome) risk score populations, indicating the underlying molecular mechanism of the lethality of these patients. GSEA enrichment analysis confirmed the important role of RBP as a posttranscriptional regulator. By integrating clinicopathological features with 15-RBP signature, we developed a nomogram model to predict the survival probability of patients with gastric cancer. The calibration plots and decision curve analysis results showed that the nomogram model exhibited a higher predictive power than a single factor. Finally, an external GEO cohort was used to verify the prognostic value of the nomogram model, and survival analysis showed the same trend in the validation cohort. Using NMF clustering analysis, we identified 2 distinct molecular subtypes for gastric cancer on the basis of 44 survival-related RBPs, which can help improve the personalized treatment strategies for patients with gastric cancer. Drug sensitivity is the core of personalized cancer chemotherapy [30]. In this study, patients in the 2 subtypes exhibited different sensitivities to chemotherapy drugs. The clinicopathological significance of tumor microenvironment cells has been clarified in predicting clinical outcomes and treatment efficacy of gastric cancer [31]. There were significant differences in immune cell infiltration between the 2 subtypes, which could offer insights into how tumors respond to immunotherapy and may guide the development of new drug combination strategies.

The advantages of this study are as follows. First, we constructed a robust 15-RBP-based signature for prediction of gastric cancer patients' prognosis. In clinical practice, the mRNA expression level of 15 RBPs is detected in gastric cancer tissues by polymerase chain reaction technology, and an individual patient's risk score is calculated. According to the risk score formula, the survival time of the patient can be robustly predicted. Second, a nomogram was built for gastric cancer patients' prognosis, which exhibited high predictive power. The

nomogram model can predict 1-, 3-, and 5-year survival probability by quantifying 2 independent risk factors (age and risk score), which provides an important reference for the formulation of individualized treatment strategies and clinical decisions. Third, our findings were based on the gastric cancer samples from different races and regions, which were confirmed in an independent cohort. Thus, the model we constructed can be extrapolated to any country or region. However, some limitations of this study should be pointed out. First, because of the limited clinical information of patients in the validation set, grade, M stage, and overall stage were not included in our univariate analysis of the validation cohort. Second, the construction and evaluation of the nomogram model was based on public data sets. Third, this model needs to be validated by a large, multicenter prospective clinical cohort in future research. Collectively, in this study we conducted a series of bioinformatics analysis to systematically study the function and clinical features of prognosis-related RBPs on gastric cancer. To our knowledge, this is the first report to develop a nomogram model combining a 15-RBP signature and age for gastric cancer. Moreover, we characterized 2 molecular subtypes related to drug sensitivity and immune infiltration. Our findings may help to understand the pathogenesis of gastric cancer and the development of promising therapeutic targets and prognostic biomarkers.

Conclusions

Taken together, a robust nomogram combining a 15-RBP signature and age was proven to possess high predictive power and could become a predictive tool for clinical outcomes and guide personalized treatment. Two distinct molecular subtypes were characterized for gastric cancer. Therefore, our study could offer insights into developing better treatment options for high-risk gastric cancer patients.

Conflicts of interest

None.

Supplementary Data

Supplementary Table 1. Identification of RBPs for gastric cancer.

id	HR	HR.95L	HR.95H	p Value
DAZAP1	0.686683888	0.48858409	0.96510463	0.030426196
MSI2	0.723377451	0.5639309	0.92790613	0.010805207
RBMS1	1.301253183	1.07614001	1.57345683	0.006584895
RBMS3	1.248001531	1.06541096	1.46188454	0.006050479
METTL2B	0.63476914	0.45908674	0.87768134	0.005974665

Supplementary Table 1 continued. Identification of RBPs for gastric cancer.

id	HR	HR.95L	HR.95H	p Value
AKAP8	0.672242498	0.46986566	0.96178549	0.029766848
REPIN1	0.752226206	0.60481181	0.9355708	0.010515524
DZIP1	1.219898246	1.02570798	1.45085323	0.024646525
DYNLL1	1.570878612	1.0391757	2.37463176	0.032174306
GLE1	0.724504237	0.55534006	0.94519814	0.017530366
FBXO17	1.140083741	1.00764951	1.28992365	0.037441303
MRPL4	0.725316458	0.55855274	0.9418698	0.015985553
QTRT1	0.716486114	0.52461866	0.97852477	0.036039944
GTPBP3	0.74879456	0.56337721	0.99523602	0.046276729
SMAD5	1.321541995	1.00839004	1.73194218	0.043330033
ADARB1	1.234310184	1.01101232	1.50692687	0.038684293
LENG9	0.805249906	0.65750741	0.98619027	0.036221957
REXO1	0.760552426	0.58457613	0.98950328	0.041494431
PEG10	1.101207565	1.01797866	1.19124118	0.01620073
RNASE1	1.235711969	1.08529888	1.40697102	0.00139324
HEXIM2	0.713047442	0.50869589	0.9994904	0.049655189
LARP6	1.231419807	1.05234133	1.44097233	0.009424753
TRIM25	0.683898183	0.53578416	0.87295735	0.002280806
ZFP36	1.259842966	1.06216214	1.49431451	0.00799032
TSEN54	0.776126474	0.60727505	0.99192664	0.042895757
EZH2	0.801477336	0.65912196	0.97457824	0.026550804
QKI	1.302845777	1.07354432	1.58112441	0.007396182
ISY1	0.658027754	0.43714818	0.99051202	0.044893487
PPAN	0.735011184	0.54506752	0.99114591	0.043563537
RAVER1	0.791283656	0.63365456	0.98812486	0.038888882
TRMT1	0.701928304	0.5140833	0.9584115	0.025928361
FTO	1.356312697	1.01745241	1.80802966	0.037715939
FAM98C	0.720244126	0.52150612	0.99471814	0.046358742
TTF2	0.746235779	0.57226427	0.97309559	0.030668901
PABPC5	1.508155718	1.09349215	2.08006401	0.012251672
RBM15	0.748235488	0.56366208	0.99324819	0.044766724
RPS4Y2	1.809163657	1.13375228	2.88693852	0.012902787
ADAT3	0.700272909	0.55952139	0.87643145	0.001857979
BICC1	1.208037358	1.04680029	1.39410953	0.009717571
SURF6	0.730679567	0.55124121	0.96852815	0.029082076
RPS23	1.357583599	1.00360516	1.83641267	0.047332716
IFIT1	1.136935672	1.00501494	1.28617264	0.041404172
NOVA1	1.202726004	1.02291194	1.41414895	0.025473645
EIF1AD	0.680599568	0.46971177	0.98617025	0.041995558

References:

1. Feng RM, Zong YN, Cao SM et al: Current cancer situation in China: Good or bad news from the 2018 Global Cancer Statistics? *Cancer Commun (Lond)*, 2019; 39: 22
2. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2019. *Cancer J Clin*, 2019; 69: 7–34
3. Zhao L, Jiang L, He L et al: Identification of a novel cell cycle-related gene signature predicting survival in patients with gastric cancer. *J Cell Physiol*, 2019; 234: 6350–60
4. Zhou YY, Kang YT, Chen C et al: Combination of TNM staging and pathway based risk score models in patients with gastric cancer. *J Cell Biochem*, 2018; 119: 3608–17
5. Mohibi S, Chen X, Zhang J: Cancer TheRBP'eutics-RNA-binding proteins as therapeutic targets for cancer. *Pharmacol Ther*, 2019; 203: 107390
6. Hentze MW, Castello A, Schwarzl T et al: A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol*, 2018; 19: 327–41
7. Pereira B, Billaud M, Almeida R: RNA-binding proteins in cancer: Old players and new actors. *Trends Cancer*, 2017; 3: 506–28
8. Corley M, Burns MC, Yeo GW. How RNA-binding proteins interact with RNA: Molecules and mechanisms. *Mol Cell*. 2020;78: 9-29.
9. Wang X, Hu H, Liu H: RNA binding protein Lin28B confers gastric cancer cells stemness via directly binding to NRP-1. *Biomed Pharmacother*, 2018; 104: 383–89
10. Yang F, Hu A, Li D et al: Circ-HuR suppresses HuR expression and gastric cancer progression by inhibiting CNBP transactivation. *Mol Cancer*, 2019; 18: 158
11. Zhao Y, Liu Y, Lin L et al: The lncRNA MACC1-AS1 promotes gastric cancer cell metabolic plasticity via AMPK/Lin28 mediated mRNA stability of MACC1. *Mol Cancer*, 2018; 17: 69
12. Zhang W, Yu Y, Hertwig F et al: Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol*, 2015; 16: 133
13. Friedman J, Hastie T, Tibshirani R: Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*, 2010; 33: 1–22
14. Subramanian A, Tamayo P, Mootha VK et al: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*, 2005; 102: 15545–50
15. Gaujoux R, Seoighe C: A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*, 2010; 11: 367
16. Yang W, Soares J, Greninger P et al: Genomics of drug sensitivity in cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res*, 2013; 41: D955–61
17. Newman AM, Liu CL, Green MR et al: Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*, 2015; 12: 453–57
18. Yin Y, Long J, He Q et al: Emerging roles of circRNA in formation and progression of cancer. *J Cancer*, 2019; 10: 5015–21
19. Yang Z, Li J, Shi Y et al: Increased musashi 2 expression indicates a poor prognosis and promotes malignant phenotypes in gastric cancer. *Oncol Lett*, 2019; 17: 2599–606
20. Jing JJ, Wang ZY, Li H et al: Key elements involved in Epstein-Barr virus-associated gastric cancer and their network regulation. *Cancer Cell Int*, 2018; 18: 146
21. Wang J, Chu XQ, Zhang D et al: Knockdown of long non-coding RNA PEG10 inhibits growth, migration and invasion of gastric carcinoma cells via up-regulating miR-3200. *Neoplasma*, 2018; 65: 769–78
22. Chen JJ, Ren YL, Shu CJ et al: JP3, an antiangiogenic peptide, inhibits growth and metastasis of gastric cancer through TRIM25/SP1/MMP2 axis. *J Exp Clin Cancer Res*, 2020; 39: 118
23. Li Y, Zheng D, Wang F et al: Expression of demethylase genes, FTO and ALKBH1, is associated with prognosis of gastric cancer. *Dig Dis Sci*. 2019;64: 1503-513.
24. Su Y, Huang J, Hu J. m(6)A RNA methylation regulators contribute to malignant progression and have clinical prognostic impact in gastric cancer. *Front Oncol*, 2019; 9: 1038
25. Xu D, Shao W, Jiang Y et al: FTO expression is associated with the occurrence of gastric cancer and prognosis. *Oncol Rep*, 2017; 38: 2285–92
26. Dong Z, Pei S, Zhao Y et al: Identification of hub genes in gastric cancer with high heterogeneity based on weighted gene co-expression network. *Crit Rev Eukaryot Gene Expr*, 2020; 30: 101–9
27. Tang B, Wu J, Zhu MX et al: VPAC1 couples with TRPV4 channel to promote calcium-dependent gastric cancer progression via a novel autocrine mechanism. *Oncogene*, 2019; 38: 3946–61
28. Bruner HC, Derksen PWB: Loss of E-cadherin-dependent cell-cell adhesion and the development and progression of cancer. *Cold Spring Harb Perspect Biol*, 2018; 10: a029330
29. Moreira AM, Pereira J, Melo S et al: The extracellular matrix: An accomplice in gastric cancer development and progression. *Cells*, 2020; 9: 394
30. Dong Q, Li F, Xu Y et al: RNAactDrug: A comprehensive database of RNAs associated with drug sensitivity from multi-omics data. *Brief Bioinform*, 2019 [Online ahead of print]
31. Zeng D, Li M, Zhou R et al: Tumor microenvironment characterization in gastric cancer identifies prognostic and immunotherapeutically relevant gene signatures. *Cancer Immunol Res*, 2019; 7: 737–50