

# Identification of a new RNA-binding proteins-based signature for prognostic prediction in gastric cancer

Yuzhi Wang, MM<sup>a</sup> , Weixia Ye, MM<sup>b</sup>, Gang Tian, MD<sup>c</sup>, Yi Zhang, MM<sup>d,\*</sup> 

## Abstract

Gastric cancer (GC) is one of the most common cancers with high incidence and mortality worldwide. Recently, RNA-binding proteins (RBPs) have drawn more and more attention for its role in cancer pathophysiology. However, the function and clinical implication of RBPs in GC have not been fully elucidated. RNA sequencing data along with the corresponding clinical information of GC patients were downloaded from The Cancer Genome Atlas (TCGA) database. Differentially expressed RNA-binding proteins (DERBPs) between tumor and normal tissues were identified by “limma” package. Functional enrichment analysis and the protein-protein interaction (PPI) network were harnessed to explore the function and interaction of DERBPs. Next, univariate and multiple Cox regression were applied to screen prognosis-related hub RBPs and to construct a signature for GC. Meanwhile, a nomogram was built on the basis of the independent factors. A total of 296 DERBPs were found, and most of them mainly related to post-transcriptional regulation of RNA and ribonucleoprotein. A PPI network of DERBPs was constructed, consisting of 262 nodes and 2567 edges. A prognostic signature was built depending on 7 prognosis-related hub RBPs that could divide GC patients into high-risk and low-risk groups. Survival analysis showed that high-risk group had a worse prognosis compared with the low-risk group and the time-dependent receiver operating characteristic (ROC) curves suggested that signature existed moderate predictive capacities of survival for GC patients. Similar results were obtained from another independent set GSE62254, confirming the robustness of signature. Besides, the genetic variation and immune heterogeneity differences were identified between the high-risk and low-risk groups by bioinformatics methods. These findings would provide evidence of the effect of RBPs and offer a novel potential biomarker in prognostic prediction and clinical decision for GC.

**Abbreviations:** AUCs = area under curves, BP = biological process, CC = cellular component, DCA = decision curve analysis, DERBPs = differentially expressed RNA-binding proteins, ESTIMATE = estimation of stromal and immune cells in malignant tumour tissues using expression data, GC = gastric cancer, GEO = Gene Expression Omnibus, GO = Gene Ontology, HPA = Human Protein Atlas, IHC = immunohistochemistry, KEGG = Kyoto Encyclopedia of Genes and Genomes, MCODE = Molecular Complex Detection, MF = molecular function, OS = overall survival, PPI = protein-protein interaction, RBPs = RNA-binding proteins, ROC = receiver operating characteristic, ssGSEA = single-sample gene set enrichment analysis, STRING = Search Tool for the Retrieval of Interacting Genes, TCGA = The Cancer Genome Atlas, TMB = tumor mutation burden.

**Keywords:** gastric cancer, overall survival, prognostic signature, RNA-binding proteins

Editor: Kamalika Mojumdar.

YW and WY contributed equally to this article.

This study was supported by the Luzhou Science and Technology Bureau (CN) (2018LZXNYD-ZK30), Affiliated Hospital of Southwest Medical University (CN) (18057) and Incubated Thesis Capital of the People's Hospital of Deyang City (CN) (FHS202004).

The authors report no conflicts of interest.

The datasets generated during and/or analyzed during the current study are publicly available.

<sup>a</sup> Department of Laboratory Medicine, People's Hospital of Deyang City, Deyang, Sichuan, China, <sup>b</sup> Department of Gastroenterology, Luzhou People's Hospital, Luzhou, Sichuan, China, <sup>c</sup> Department of Laboratory Medicine, Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan, China, <sup>d</sup> Department of Blood Transfusion, People's Hospital of Deyang City, Deyang, Sichuan, China.

\* Correspondence: Yi Zhang, People's Hospital of Deyang City, No. 173, Section 1, Taishan North Road, Deyang City, Sichuan 618000, China (e-mail: 472189926@qq.com).

Copyright © 2022 the Author(s). Published by Wolters Kluwer Health, Inc.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Wang Y, Ye W, Tian G, Zhang Y. Identification of a new RNA-binding proteins-based signature for prognostic prediction in gastric cancer. *Medicine* 2022;101:8(e28901).

Received: 4 October 2020 / Received in final form: 3 January 2022 / Accepted: 1 February 2022

<http://dx.doi.org/10.1097/MD.00000000000028901>

## 1. Introduction

Gastric cancer (GC), one of the most frequently occurring digestive tract tumors, originates in the gastric mucosal epithelium. GC is an important leading cause of cancer-associated death worldwide, with an estimated 1,000,000 new cases and approximately 700,000 deaths every year derived from the cancer statistics of 2018.<sup>[1]</sup> Despite the advances in the diagnosis and treatment of GC, the prognosis for patients with GC remains poor due to most of the diagnosis during its middle to late stages.<sup>[2,3]</sup> Therefore, exploring molecular mechanisms behind the occurrence and progression of GC and discovering new biomarkers are urgently required for early diagnosis and prognosis improvement of GC.

RNA-binding proteins (RBPs) are inherently pleiotropic proteins, interacting with an assortment of types of RNAs include mRNAs, tRNAs, miRNA, and ncRNAs.<sup>[4,5]</sup> RBPs play central roles in RNA structure, localization, stability, and translatability to regulate gene expression post-transcriptionally and other cellular functions.<sup>[6]</sup> It is well established that post-transcriptional deregulation has emerged as a frequent pathological mechanism in numerous diseases, which demonstrate the crucial function of RBPs in human cellular processes.<sup>[7]</sup> In order to explore the construction and function of RBPs, we must comprehensively identify and annotate them first. Given that rapid advances in high-throughput sequencing technologies, over 1500 RBPs have been found and deposited into public databases.<sup>[4]</sup> Numerous studies have indicated that RBPs play essential roles in tumor occurrence and development.<sup>[8,9]</sup> For example, RBP RNPC1 regulates P63 gene stability to inhibit initiation and progression.<sup>[10]</sup> RBP U2AF1 affects pre-mRNA splicing of a good deal of oncogenic drivers to promote tumorigenesis. Overexpression of RBP LIN28A accelerates cell's progress from S to G2/M to enhance colon cancer cell proliferation.<sup>[11]</sup> A systematic study on RBPs may be conducive to understand their contribution to tumors and help discover potential diagnostic or prognostic biomarkers that are what we lack on GC.

In the study, differentially expressed RNA-binding proteins (DERBPs) were investigated between tumor and normal tissues. Subsequently, we performed functional enrichment analysis for DERBPs to explore the biological functions and constructed a co-expression network to reveal the relationship between them. Moreover, we built a model to appraise the predictive value of RBPs for the CC survival, some of which may serve as biomarkers for prognosis and treatment for GC in the future.

## 2. Methods and materials

### 2.1. Data collection and DERBPs analysis

RNA sequence data (fragments per kilobase million format) and corresponding clinical information (Table 1) were downloaded from The Cancer Genome Atlas (TCGA, <https://cancergenome.nih.gov/>) database, which contained 375 GC tissue samples and 32 paracancerous tissue samples. The genes expression data included more than 60,000 genes annotated by Ensemble gene ID (<http://www.ensembl.org/>). RBPs expression levels were extracted and recorded in a matrix. Next, "limma" package in R software (version 4.0.0) was applied to standardize RBPs expression and estimate DERBPs with  $PDR < 0.05$  and  $|\text{fold change}| > 1$  between GC tissues and adjacent non-cancerous tissues. All data in this study were obtained from public databases, eliminating the need for ethics approval.

**Table 1**

**Clinical characteristics of the TCGA and GEO cohorts.**

Characteristics	TCGA cohort (N=371)		GEO cohort (N=300)	
	Number	Percentage	Number	Percentage
Age				
<60	111	30%	106	35%
≥60	257	70%	194	65%
Gender				
Male	238	64%	101	34%
Female	133	36%	199	66%
T				
T1-T2	96	26%	188	63%
T3-T4	267	74%	112	37%
N				
N0	108	64%	38	13%
N1-N3	245	31%	262	87%
M				
M0	328	93%	273	91%
M1	25	7%	27	9%
Stage				
Stage I-II	161	46%	126	42%
Stage III-IV	187	54%	174	58%
Grade				
G1-G2	144	40%	NA	NA
G3	218	60%	NA	NA

### 2.2. Gene ontology and pathway enrichment analysis

In order to research the most associated biological functions and pathways for DERBPs, we performed Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis with "clusterProfiler" package in R version 4.0.0.<sup>[12-14]</sup> The GO term enrichment analysis included 3 categories: molecular function (MF), cellular component (CC), and biological process (BP). Both  $P$  value  $< .05$  and  $FDR < 0.05$  were used as the screening criteria.

### 2.3. Establishment of protein-protein interaction (PPI) network and modules selection

Search Tool for the Retrieval of Interacting Genes (STRING) is a public database that contains interactions between known and predicted proteins (<http://www.string-db.org/>).<sup>[15]</sup> DERBPs were uploaded to STRING database to provide a global perspective for the interactive relationship of them. Protein-protein interaction (PPI) network was constructed and visualized using Cytoscape software (version 3.7.0), and the most important modules that both Molecular Complex Detection (MCODE) score  $\geq 4$  and  $\geq 6$  nodes were selected by the MCODE plug.<sup>[16]</sup> A  $P$  value of less than .05 was considered statistically significant.

### 2.4. Construction and validation of overall survival (OS) risk prognostic model

Identification of candidate prognostic RBPs was performed from DERBPs in the PPI network by the univariate Cox regression analysis. Multiple Cox regression was executed to identify hub prognostic RBPs further based on the results from the univariate Cox regression analysis. Later, a multiple stepwise Cox regression model was built based on the above hub prognostic RBPs. Risk score for every patient was calculated using the following formula: Risk score value =  $EXP1 * \beta1 + EXP2 * \beta2 \dots + EXPx * \betax$ , EXP presents the expression levels of each RBP and

$\beta$  presents the regression coefficient from the multivariate Cox proportional hazards regression model. The patients were assigned to high-risk and low-risk groups according to the median risk score as the cutoff value. The time-dependent receiver operating characteristic (ROC) curve was applied to evaluate the model prognostic power. In addition, the other cohort GSE62254 includes 300 patients from Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62254>), which was used to justify whether the prognostic value of the model was credible. Statistical significance was judged with  $P < .05$ . The clinicopathological characteristics of the patients in the GEO cohort are summarized in Table 1.

### 2.5. Somatic mutation and immune infiltration analysis

This study used the “maftools” package for calculating and visualizing the mutation data in the TCGA set. Tumor mutation burden (TMB) was identified as the total number of somatic non-synonymous mutations in the coding regions. For each sample, mutation detection was done using the preprocessed and annotated MAF data files generated by the varscan platform for the calculation of the TMB. The estimation of stromal and immune cells in malignant tumor tissues using expression data (ESTIMATE) was an algorithm to evaluate the levels of immune cell infiltration (immune score), the stromal content (stromal score), the stromal-immune comprehensive score (ESTIMATE score), and tumor cells (tumor purity) based on expression data for each GC sample in the TCGA set.<sup>[17]</sup> A single-sample gene set enrichment analysis (ssGSEA) algorithm transforms marker gene expression patterns into quantities of immune cell populations in individual tumor samples.<sup>[18]</sup> ssGSEA was used to quantify the infiltration of each immune cell type each GC sample using the R package “GSVA.” The marker gene set for 24 types of immune cells was obtained from a previous study.<sup>[19]</sup>

### 2.6. Establishment of the nomogram

The clinicopathological parameters of GC patients were added in univariate and multivariate Cox regression analyses for the verification of the independence of the risk score based on the

survival-related APA signature. Afterward, a nomogram signature was constructed using all independent prognostic factors to develop a scoring system to evaluate the OS of patients at 1-year, 3-year, and 5-year. To demonstrate the effectiveness of the system, the calibration curves and time-ROC curves were used for the evaluation of the recognition performance of the system. The decision curve analysis (DCA) was used to assess the clinical applicability of the scoring system.

### 2.7. Verifying expression and prognostic value of the hub RBPs

Kaplan–Meier Plotter (<https://kmplot.com/>), which is an online database that contains data from 1440 GC patients and combined with OS data, was used to assess the prognostic value of hub RBPs in GC.<sup>[20]</sup> The Human Protein Atlas (HPA) is the largest and most comprehensive database (<http://www.proteinatlas.org/>) for evaluating protein distribution in human tissues and cells.<sup>[21]</sup> In this study, the immunohistochemical staining results was analyzed for the hub RBPs in both tumor and normal tissues.

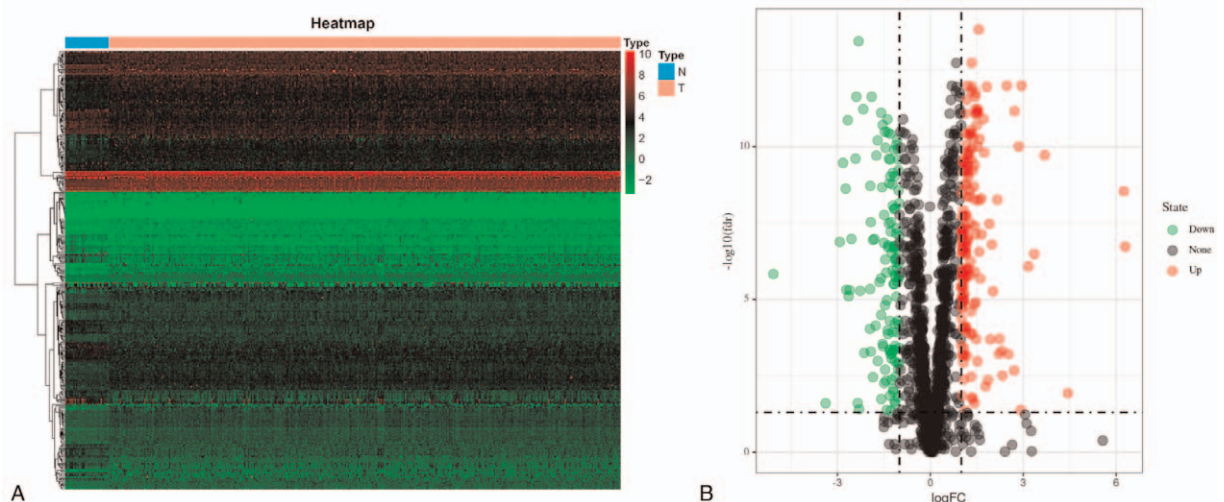
## 3. Results

### 3.1. Identification of DERBPs in GC

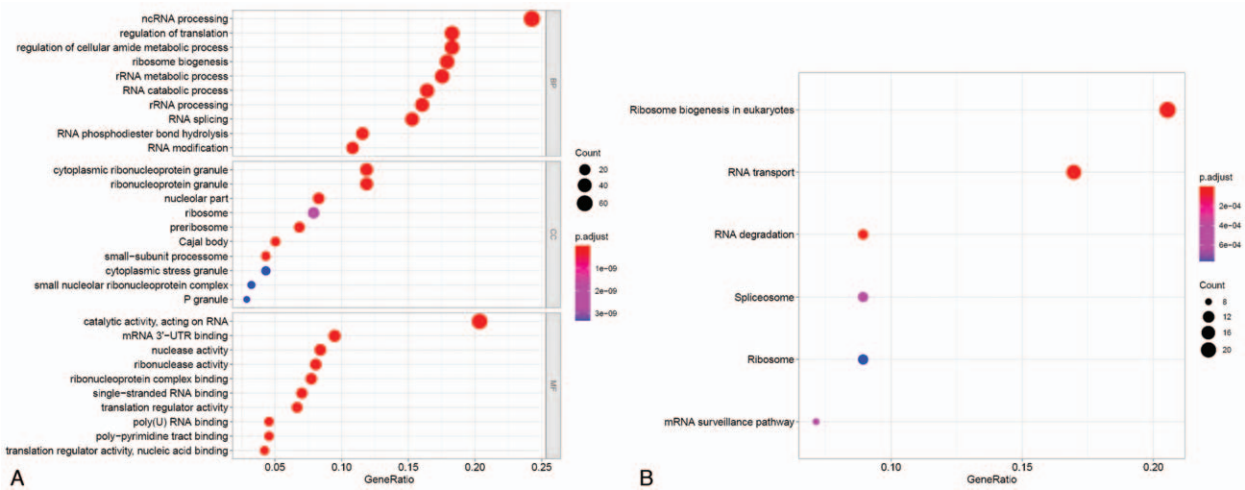
In total, 296 DERBPs were found between tumor and paracancerous tissue using “limma” package at the threshold of  $P$  value  $< .01$  and  $|\text{fold change}| > 1$ , among which 160 RBPs were upregulated and 133 RBPs were downregulated (Fig. 1A). The volcano plot shows the distribution of DERBPs (Fig. 1B).

### 3.2. Functional and pathway enrichment analysis of DERBPs

To research the function and potential mechanism of DERBPs, GO and KEGG analysis was done using “clusterProfiler” package in R software. The results suggest that DERBPs in BP terms primarily enriched in ncRNA processing, rRNA metabolic process, rRNA processing, ribosome biogenesis, and regulation of translation (Fig. 2A). Within the molecular function, DERBPs were notably enriched in catalytic activity on RNA, mRNA 3’-



**Figure 1.** The significantly altered RBPs in GC samples. (A) Heatmap. (B) Volcano plot. GC=gastric cancer, RBP=RNA-binding proteins.



**Figure 2.** Functions and pathways analysis of differentially expressed RBPs. (A) Significant enriched GO terms of differentially expressed RBPs. (B) Significant enriched KEGG pathways of differentially expressed RBPs. GO=Gene Ontology, KEGG=Kyoto Encyclopedia of Genes and Genomes, RBPs=RNA-binding proteins.

UTR binding, translation regulator activity, ribonuclease activity, and single-stranded RNA binding (Fig. 2A). At the cellular component level, DERBPs were mostly enriched in cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, preribosome, nucleolar part, and small-subunit processome (Fig. 2A). Moreover, we found that DERBPs were significantly enriched in ribosome biogenesis in eukaryotes, RNA transport, RNA degradation, mRNA surveillance pathway, and spliceosome (Fig. 2B).

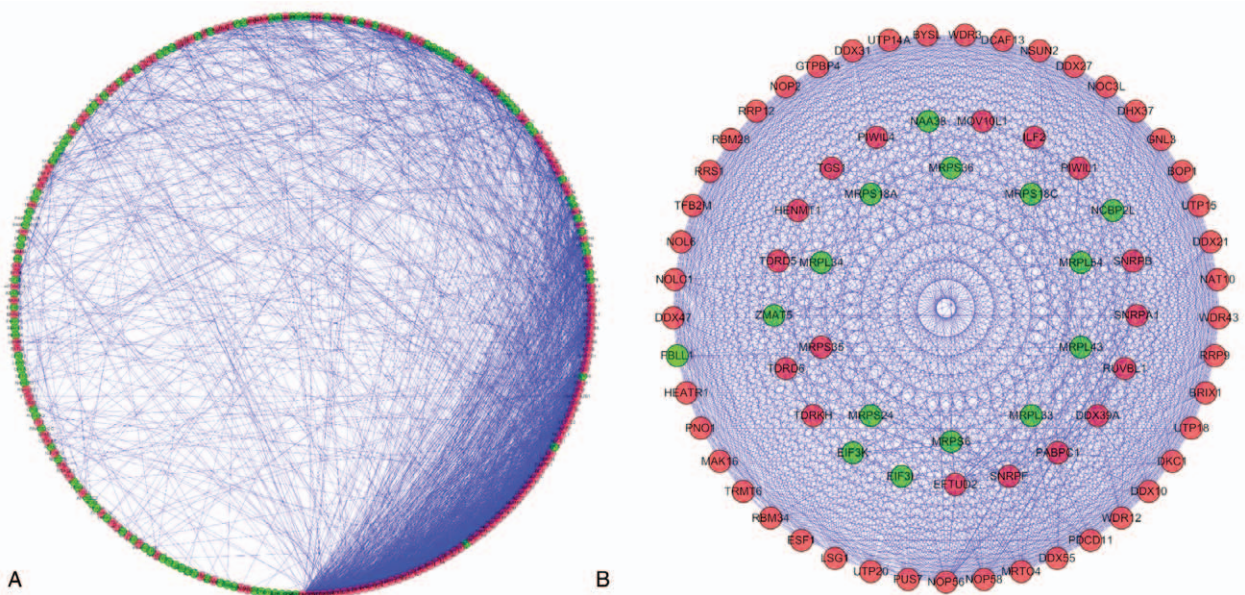
**3.3. PPI network analysis and identification of key module**

PPI network could reflect direct physical interactions and potential molecular functions between genes. We built a PPI network of DERBPs using the STRING database and Cytoscape

software, consisting of 262 nodes and 2567 edges (Fig. 3A). Subsequently, MCODE plug-in was used to screen the vital cluster modules from the PPI network and the key modules were chosen, which incorporated 76 nodes and 1071 edges (Fig. 3B). According to the results of enrichment analysis, the RBPs in the top module were commonly involved in ribosome biogenesis, rRNA processing, preribosome, small-subunit processome, snoRNA binding, RNA helicase activity, and ribosome biogenesis in eukaryotes (Table 2).

**3.4. Construction and validation of the prognostic signature**

We carried out univariate Cox regression analysis for key RBPs in the PPI network to find out prognosis-related RBPs, and attained



**Figure 3.** PPI network construction and significant modules filter. (A) PPI network of the differentially expressed RBPs. (B) The key modules from PPI network. Red nodes represent upregulated RBPs and green nodes represent downregulated RBPs. PPI=protein-protein interaction, RBPs=RNA-binding proteins.

**Table 2**  
GO annotation and KEGG pathway analysis of RBPs in the top module.

ID	Description	Counts	P
GO:0042254	ribosome biogenesis	33	1.24E-53
GO:0006364	rRNA processing	31	1.80E-53
GO:0034470	ncRNA processing	34	3.50E-52
GO:0016072	rRNA metabolic process	31	4.58E-51
GO:0030684	preribosome	18	6.34E-33
GO:0032040	small-subunit processome	12	1.04E-23
GO:0044452	nucleolar part	15	3.23E-20
GO:0030686	90S preribosome	7	4.99E-13
GO:0030515	snoRNA binding	7	2.67E-13
GO:0140098	catalytic activity, acting on RNA	12	1.47E-10
GO:0003724	RNA helicase activity	7	9.74E-10
GO:0004386	helicase activity	7	1.67E-07
hsa03008	Ribosome biogenesis in eukaryotes	15	9.75E-29

19 candidate RBPs, which was associated with GC prognosis (Fig. 4A). Then, the above 19 candidate RBPs were analyzed by multiple Cox regression to identify the independent predictors for GC patients. As a result, 7 hub prognostic RBPs were found, and all of them were used to build a signature for GC by multiple stepwise Cox regression (Fig. 4B). GC patients were ranked according to risk score value and split into high and low-risk groups using the median risk score value as the cutoff point (Fig. 5A). Kaplan–Meier survival plot revealed that the low-risk group had a higher OS than the high-risk group (Fig. 5C). Also, a time-dependent ROC curve demonstrated that area under curves (AUCs) of the 7-RBPs signature of OS for 5 years was 0.692 (Fig. 5E). To determine whether the 7-RBPs signature has similar prognostic value in another GC cohort, we built the model by applying the same formula for the GSE62254 set (Fig. 5B). Kaplan–Meier curve and log-rank test indicated that the low-risk group had longer survival time than the high-risk group in the GSE62254 set (Fig. 5D). Unfortunately, the AUC of time-dependent ROC curve at 5 years was 0.623 (Fig. 5F).

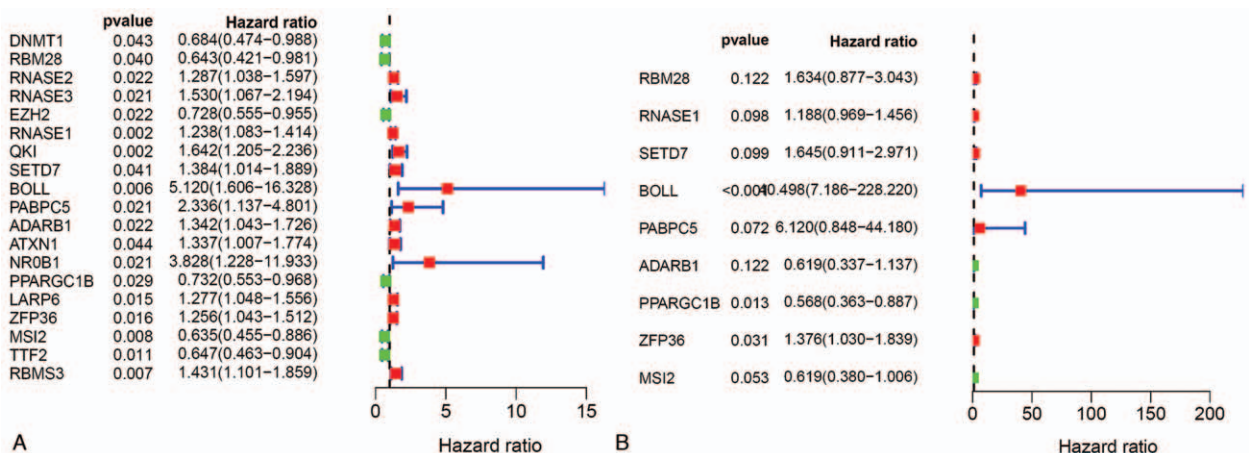
**3.5. TMB and immune infiltration analysis**

We analyzed and visualized somatic mutation data in GC patients by distinguishing between the high-risk and low-risk groups. The

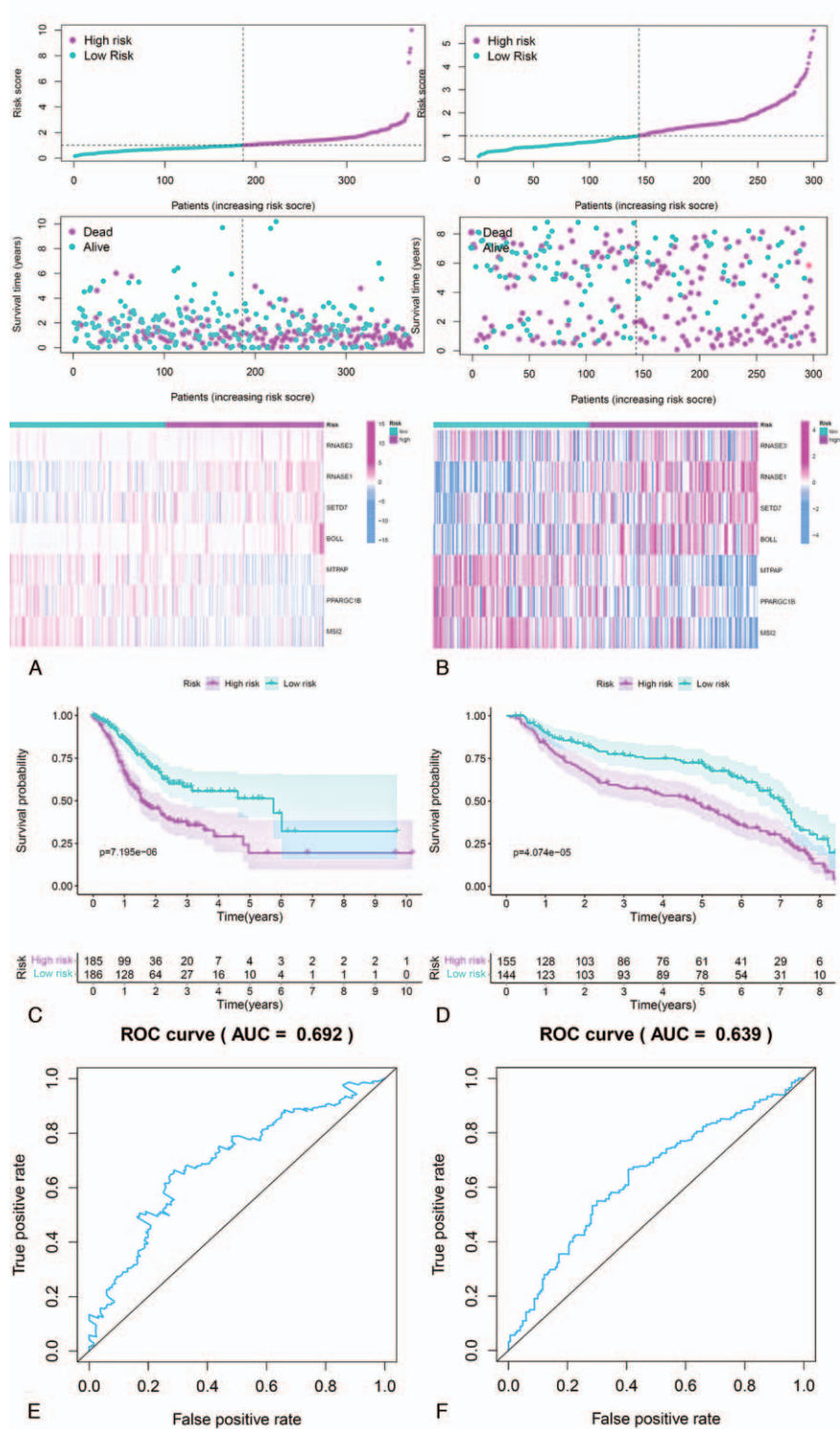
top 10 drive genes with the highest variation frequency in the 2 risk groups are shown in Figure 6A, B. TMB of the 2 risk groups was calculated on the basis of somatic mutation data. The result indicated that the TMB of the low-risk group was higher than that of the high-risk group and the risk score was negatively correlated with TMB (Fig. 6C, D). We then identified whether TMB was a prognostic biomarker for GC patients. GC patients are split into high-TMB and low-TMB groups as per the median TMB value. According to the finding, the TMB can be used alone for predicting patient outcomes (Fig. 6E). Besides, when TMB was combined with the risk score, they could also effectively predict GC patient outcomes (Fig. 6F). The materials and methods mentioned in the algorithm were used to assess the immune status of each GC patient, as shown in the heatmap (Fig. 7A). Moreover, the Wilcoxon test was performed for the comparison of the differences in individual cell markers among the high-risk and low-risk groups. The results indicated that stromal, immune, and ESTIMATE scores in the high-risk group were substantially increased, but tumor purity was decreased when compared with those in the low-risk group (Fig. 7B–E). Concerning the immune cells infiltration, the DC, iDC, macrophages, mast cells, NK cells, Th1 cells, Th2 cells were significantly different ( $P < .001$ ) between low-risk and high-risk groups (Fig. 7F, G).

**3.6. Establishment of nomogram**

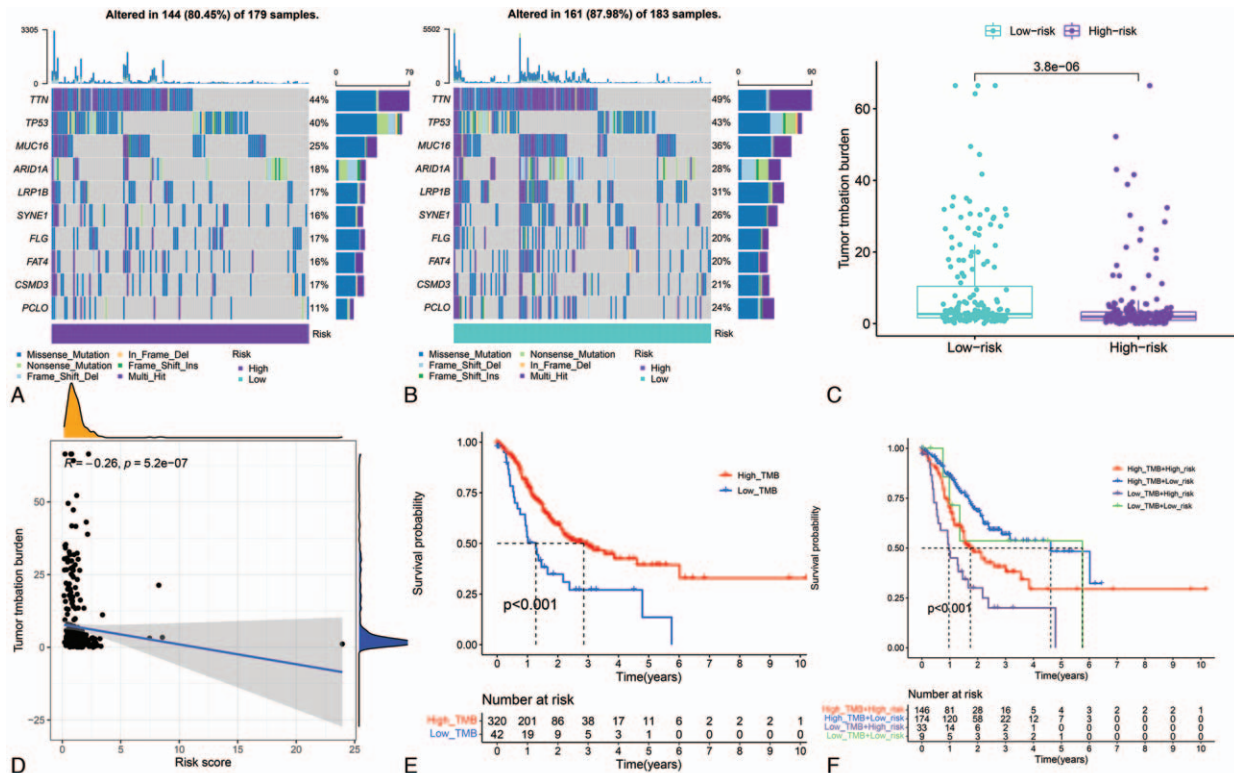
To ascertain whether the 7-RBPs signature was an independent prognostic factor, we explored univariate and multivariate Cox regression analyses for the prognostic value of 7-RBPs signature and clinical factors. Univariate Cox regression analysis found that age, stage, pN, and the risk score were obviously related to OS (Table 3). However, only age and risk score were remarkably related to OS after multivariate Cox regression analysis of GC (Table 3). Subsequently, a nomogram was created on the basis of outcomes of independent factors, including age and risk score. The OS was calculated for all patients and was predicted at 1, 3, and 5-year (Fig. 8A). Importantly, calibration plots showed that the nomogram performed well with the ideal model for predicting GC outcomes (Fig. 8B). As shown in Figure 9C, the nomogram had better predictive power than other independent factors (Fig. 8C). Furthermore, the nomogram could offer the net benefit over the “treat-all” or “treat-none” strategy in clinical (Fig. 8D).



**Figure 4.** COX regression analysis of the RBPs. (A) Univariate Cox regression analysis to screen prognostic candidate RBPs. (B) Multivariate Cox regression analysis to identify hub RBPs for building model. RBPs=RNA-binding proteins.



**Figure 5.** Risk score model analysis of the hub RBPs in TCGA set and GEO set. (A) The distributions of risk score, OS status, and the 7 RBPs expression heatmaps in TCGA set. (B) The distributions of risk score, OS status, and the 7 RBPs expression heatmaps in GEO set. (C) The survival curves of OS difference between the high-group and the low-risk group in TCGA set. (D) The survival curves of OS difference between the high-group and the low-risk group in GEO set. (E) The ROC curves for prediction of 3-year OS in TCGA set. (F) The ROC curves for prediction of 5-year OS in GEO set. GEO = Gene Expression Omnibus, OS = overall survival, RBPs = RNA-binding proteins, ROC = receiver operating characteristic, TCGA = The Cancer Genome Atlas.



**Figure 6.** Comparisons of somatic mutation between high-risk and low-risk groups in the whole set. (A, B) Waterfall plots showing the mutation information of top 10 genes with the highest mutation frequency in 2 groups. (C) TMB variations between the high-risk and low-risk groups. (D) Scatterplots showed relationship between risk score and TMB. (E) Survival curves for the OS of the high-TMB and low-TMB groups. (F) Survival curves for patients stratified by both TMB and signature. OS=overall survival, TMB=tumor mutational burden.

### 3.7. Validation for hub RBPs prognosis and expression

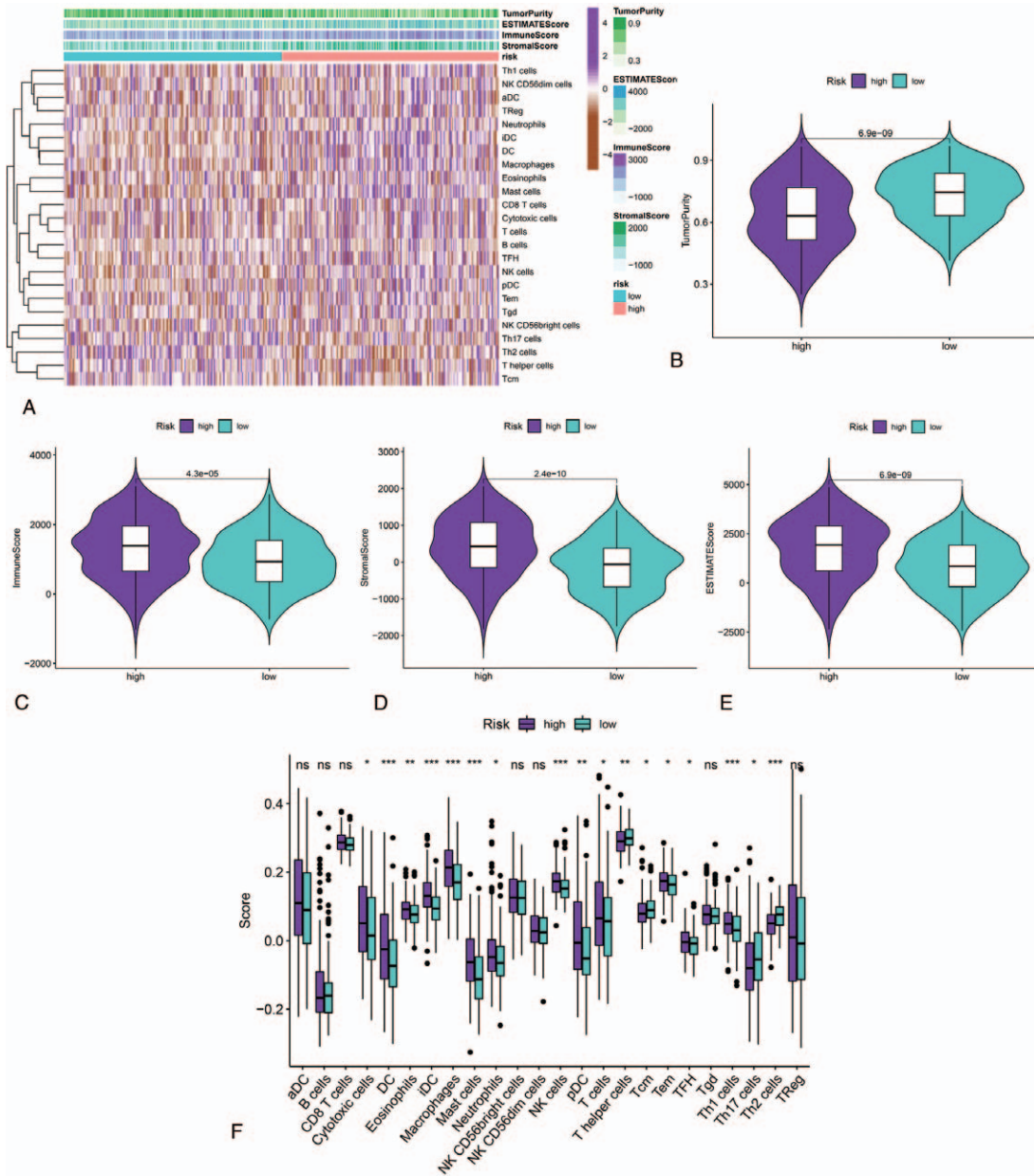
For examining the association of hub RBPs with the survival of GC patients, we performed log-rank test for hub RBPs through an online Kaplan-Meier plotter database (<http://kmplot.com/analysis/>). It was shown that all the hub RBPs were significantly correlated with OS of GC except for SETD7 (Fig. 9). To verify whether the protein of the hub RBPs genes can be detected in GC, immunohistochemistry (IHC) data were retrieved from HPA web portal. The results of IHC revealed that SETD7 protein expression was increased in GC tissue, and MSI2, RNASE1, and RNASE3 protein expression were decreased in GC tissue (Fig. 10). However, BOLL protein expression did not exhibit differential expression in GC and normal tissues (Fig. 10).

## 4. Discussion

GC is one of the most commonly diagnosed cancers with high heterogeneity in the world, seriously endangering human health and life.<sup>[22]</sup> TNM staging is the most commonly used index for evaluating the progression and prognosis of GC patients.<sup>[23]</sup> However, GC patients at the same TNM stage and receiving similar treatment regimens can experience different clinical outcomes,<sup>[24,25]</sup> implying that this indicator provides incomplete prognostic information. Therefore, exploring the molecular mechanism is critical to understand the pathogenesis and help improve the prognosis and treatment of GC. RBPs participate in almost all the steps of the post-transcriptional regulatory layer, regulating the expression and function of each transcript in the cell and ensuring stable maintenance of intracellular environ-

ments. In view of the central role of RBPs in the gene expression, dysregulation of RBPs may lead to several diseases, including cancers.<sup>[26,27]</sup> Several studies have provided some evidence that RBPs dysregulation is common in various cancers.<sup>[28–30]</sup> However, fewer papers have explored the expression and specific functional role for RBPs in GC. In this study, we investigated to identify the DERBPs between GC tissue and adjacent tissues from TCGA database. We implemented gene ontology and pathway enrichment and constructed a PPI network for DERBPs after that. In addition, we utilized univariate and multiple Cox regression analysis to screen relevant prognostic hub RBPs and applied multiple stepwise Cox regression to build a 7-RBPs signature for predicting the survival time of GC patients. Meanwhile, log-rank test analysis and time-dependent ROC analysis were applied to evaluate the discriminating value of the 7-RBPs signature. These findings may provide potential biomarkers and contribute to enlighten the pathological mechanism of GC.

GO and KEGG analysis indicated that DERBPs mainly enriched in ncRNA processing, rRNA metabolic process, rRNA processing, catalytic activity on RNA, mRNA 3'-UTR binding, translation regulator activity, cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, preribosome, ribosome biogenesis in eukaryotes, RNA transport, RNA degradation, mRNA surveillance pathway, and spliceosome. Post-transcriptional regulation, including RNA processing, RNA degradation, and translation, is an essential aspect of the regulation of gene expression. Previous studies demonstrated that the disorder of post-transcriptional regulation is associated with the occurrence and development of various cancers.<sup>[31–33]</sup> Many RBPs bind to sequence-specific motifs



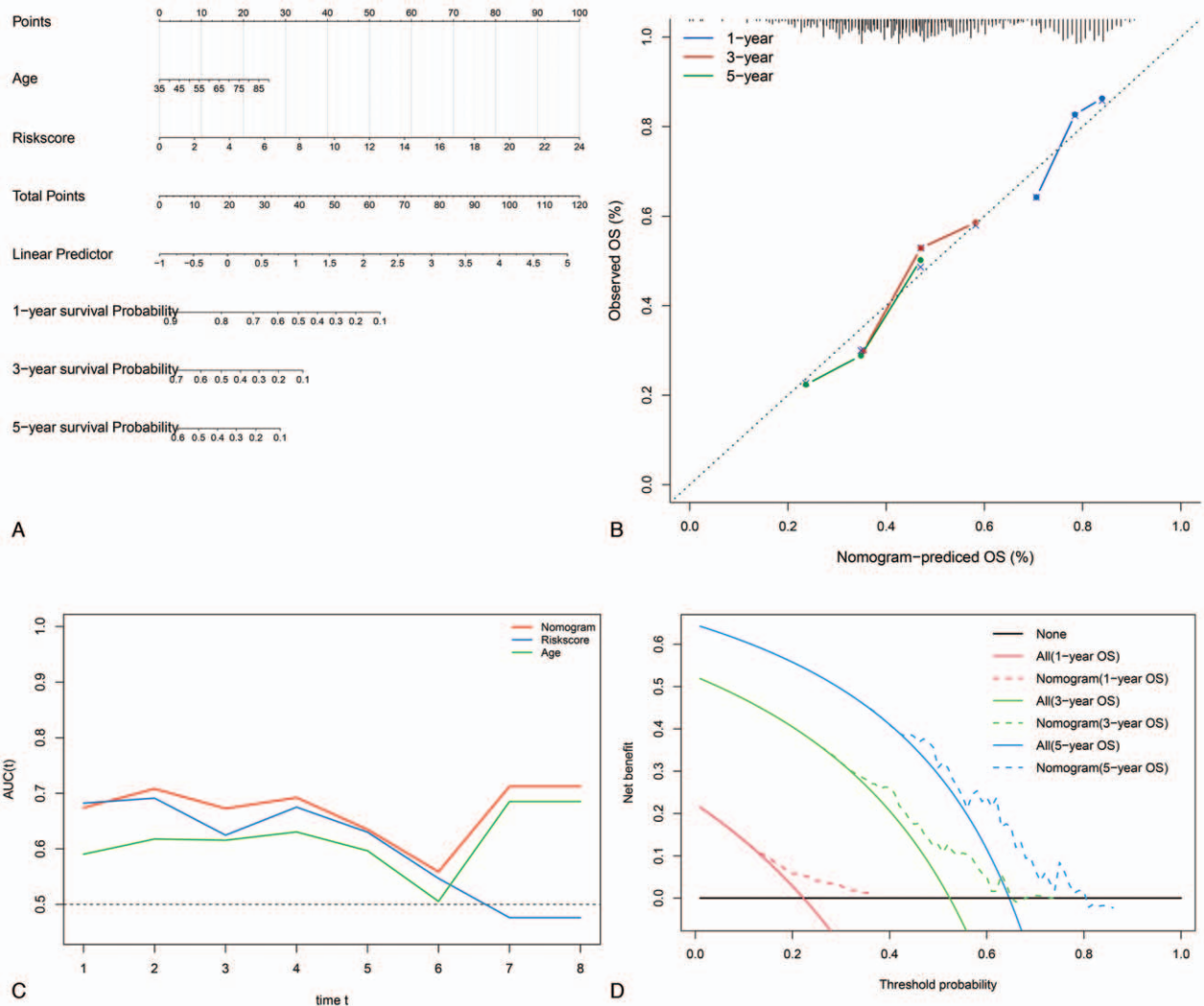
**Figure 7.** Estimation of the immune status and response to immunotherapy based on the signature in the high-risk and low-risk groups for the whole set. (A) Heatmap of the immune scores, stromal scores, tumor purity, ESTIMATE scores, and immune-infiltrating cells in the 2 groups. (B–E) Violin plots for the tumor purity, immune scores, stromal scores, and ESTIMATE scores. (F) The fraction variations of tumor-infiltrating immune cells in 2 groups. (\* $P < .05$ ; † $P < .01$ ; ‡ $P < .001$ ; ns, not significantly different). ESTIMATE=estimation of stromal and immune cells in malignant tumor tissues using expression data.

**Table 3**  
**COX regression analysis of the 7-RBPs signature with OS in the TCGA set.**

Parameters	Univariate COX		Multivariate COX	
	HR (95% CI)	P	HR (95% CI)	P
Age	1.032 (1.008–1.056)	<b>.008</b>	1.043 (1.017–1.070)	<b>.001</b>
Gender	1.225 (0.735–2.040)	.436	1.068 (0.628–1.816)	.809
Grade	1.264 (0.775–2.060)	.348	1.212 (0.698–2.105)	.495
Stage	1.419 (1.076–1.870)	<b>.013</b>	1.295 (0.741–2.261)	.364
T	1.187 (0.870–1.620)	.280	1.021 (0.662–1.574)	.927
M	1.446 (0.658–3.178)	.358	1.433 (0.552–3.720)	.460
N	1.373 (1.104–1.708)	<b>.004</b>	0.327 (0.846–1.605)	.327
Risks core	1.244 (1.146–1.351)	<b>&lt;.001</b>	1.245 (1.138–1.362)	<b>&lt;.001</b>

The significance of bold values is the analysis results of these variables are statistically different ( $P < 0.05$ ).

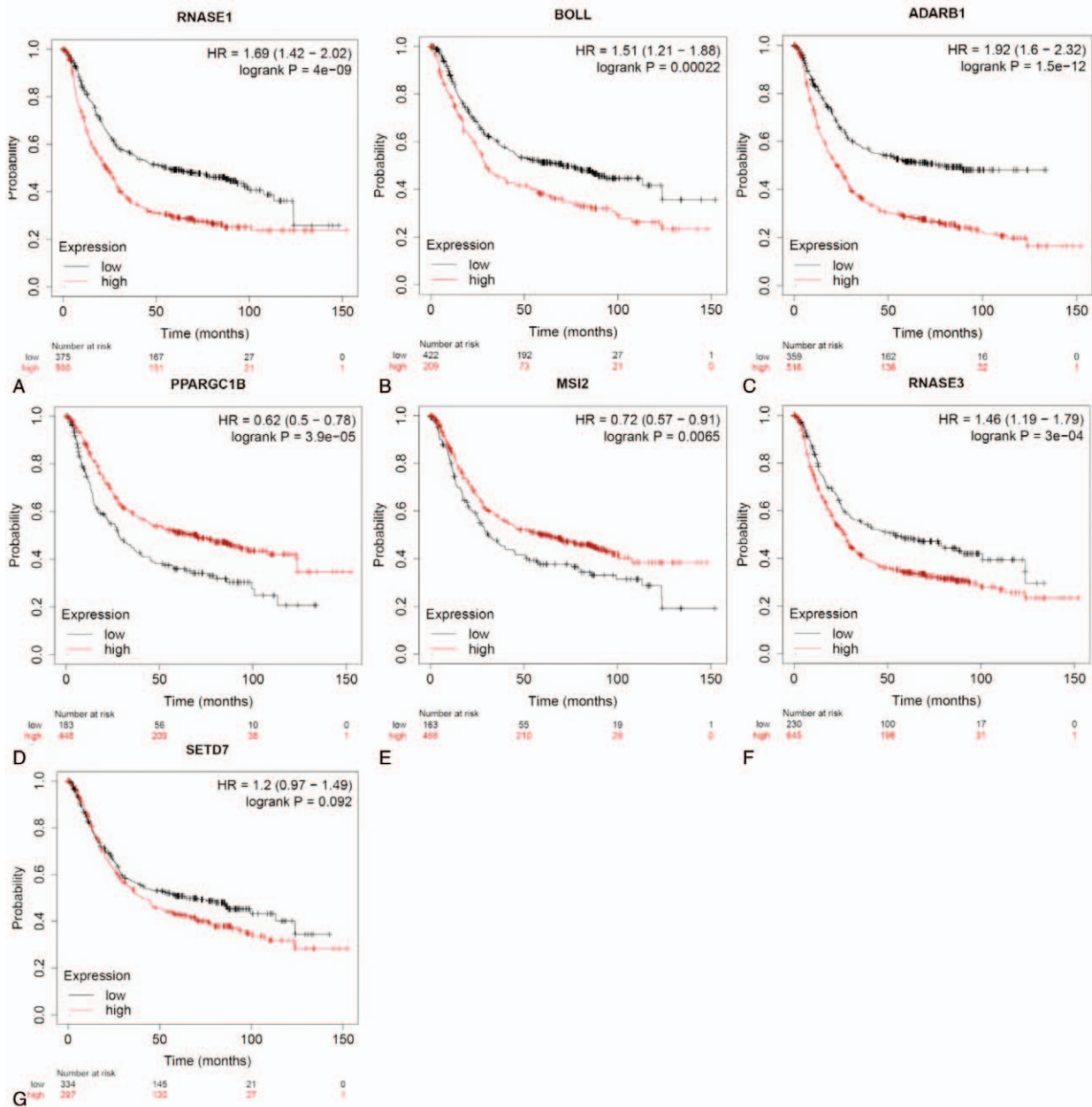




**Figure 8.** Construction of a robust nomogram prediction model based on the independent factors. (A) The nomogram for the prediction of OS. (B) The calibration plot of the nomogram predicted 1, 3, and 5-year OS. (C) AUC values of time-dependent ROC curves of the nomogram changes over time. (D) The DCA curves of the nomogram models at 1, 3, and 5-year. AUC = area under curve, DCA = decision curve analysis, OS = overall survival, RBPs = RNA-binding proteins, ROC = receiver operator characteristics.

or RNA secondary structures through unique modular arrangements of individual RNA-binding domains to play the roles in the homeostatic regulation of gene expression,<sup>[34,35]</sup> which influence the progression of many diseases. For example, RBP MSI2a expression alleviates triple-negative breast cancer invasive abilities through stabilizing TP53INP1 mRNA and inhibiting ERK1/2 activity.<sup>[36]</sup> RBP CASC9 interacts with hnRNPL form a complex to regulate DNA damage signal and PI3K/AKT signaling pathway, influencing tumor cell proliferation and apoptosis in vivo.<sup>[37]</sup> Ribonucleoprotein is the underlying basis for synthesizing all cellular proteins in all living organisms. Some research findings have shown that ribonucleoprotein also is involved in tumor initiation and progression.<sup>[38,39]</sup> The PPI network was constructed on DERBPs by Cytoscape software and employed MCODE tool to select key modules. The functional and pathway enrichment analysis showed that the key module was related to ribosome biogenesis, rRNA processing, preribosome, small-subunit processome, snoRNA binding, and ribosome biogenesis in eukaryotes.

In addition, 7 hub prognostic RBPs were screened by univariate and multiple Cox regression analysis. Compared with single biomarkers, integrating multiple biomarkers into a single model may considerably increase the predictive accuracy.<sup>[40]</sup> Therefore, we built a signature based on 7 hub prognostic RBPs using multivariate stepwise Cox proportional hazards regression analysis for useful and sensitive prognosis of GC patients. The 7-RBPs signature is capable of discriminating the GC patients from high-risk and low-risk groups and may serve as an independent prognostic factor in GC. However, the AUC values of time-ROC curves in the 2 cohorts were not high enough and may be affected by other clinical features. More importantly, the risk stratification capability of 7-RBPs signature was confirmed in another independent set. Besides, the 3 clinico-pathological parameters (age, stage, and pN) as well as the risk score were obviously related to prognosis of GC, and only age and risk score were identified as independent factors. Next, a nomogram based on the independent factors was drawn to quantitatively predict survival time for clinical use. Importantly,

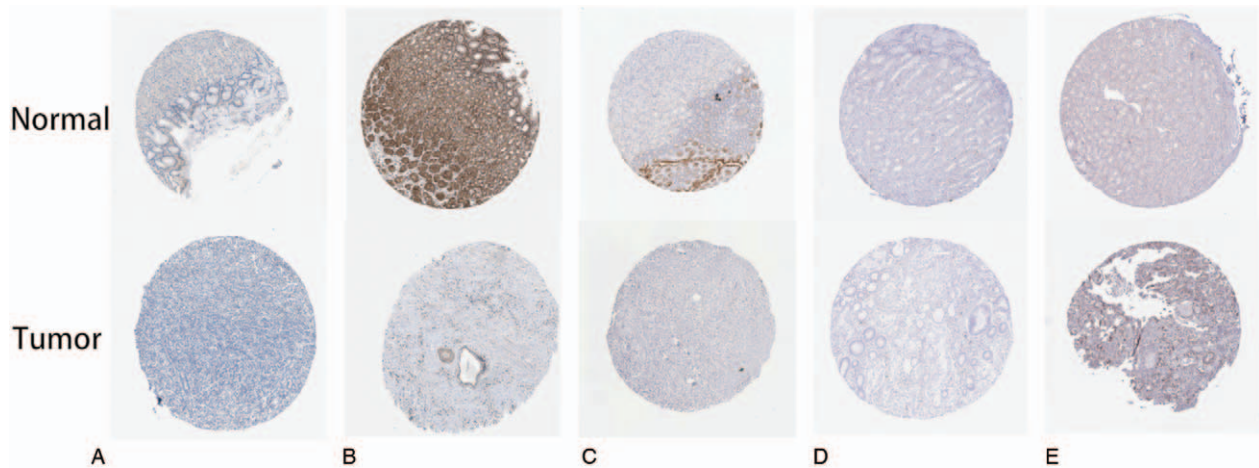


**Figure 9.** Validation of prognosis value of the hub RBPs in GC. (A) RNASE1. (B) BOLL. (C) ADARB1. (D) PPARGC1B. (E) MSI2. (F) RNASE3. (G) SETD7. GC= gastric cancer, RBPs=RNA-binding proteins.

the performance of nomogram is better than the single index (age and 7-RBPs signature) in both the predictive power and clinical benefit possibly because of calibrating other clinical parameters. Kaplan-Meier plotter explored that most of 7 hub RBPs are associated with the survival of GC patients. The protein expression levels of 7 hub RBPs genes in human tissues were obtained from HPA database, and the results indicated that the majority of RBPs were differently expressed between tumor and normal tissues.

In the present study, the 7-RBPs signature has been proven to be significantly connected with OS of GC patients. Among these hub RBPs, many of them have proved to be closely related to tumor occurrence and development. SETD7 is the only lysine methyltransferases 7 family members, which can methylate

transcription factors.<sup>[41]</sup> SETD7 could be used as a prognostic indicator for breast cancer, downregulation of it suppressed expression of antioxidant enzymes and destabilized the redox status.<sup>[42]</sup> SETD7 and ISL1 may combine to form a complex on the ZEB1 promoter to promote tumorigenesis in GC cells.<sup>[43]</sup> BOLL, an ancestral gene of the deleted in azoospermia family, maintains normal functions of sperm.<sup>[44,45]</sup> BOLL functions as an oncogene because of enhancing proliferation and migration activities in the colon cancer cells, and BOLL protein expression was upregulated in colorectal cancer tissue.<sup>[46]</sup> ADARB1 was expressed at high levels in endometrial cancer, and observed a positive correlation between increased expression and invasion degree.<sup>[47]</sup> Another study reported that ADARB1 could inhibit glioblastoma cell growth via regulation of the CDC14B/Skp2/



**Figure 10.** Protein expression of the hub RBPs genes in normal and tumor tissues. (A) BOLL. (B) MSI2. (C) RNASE1. (D) RNASE3. (E) SETD7. RBPs = RNA-binding proteins.

p21/p27 axis.<sup>[48]</sup> MSI2 is a popular molecule in digestive system tumors. MSI2 promoted hepatocellular carcinoma progression via the Wnt/ $\beta$ -catenin signaling pathway and may serve as an indicator to predict outcome of patients with hepatocellular carcinoma.<sup>[49]</sup> MSI2 expression levels were upregulated in GC tissues and associated with poor prognosis of GC. Furthermore, MSI2 induced migration, invasion, and angiogenesis to increase proliferation and invasiveness of GC cells.<sup>[50]</sup>

A number of studies have demonstrated that somatic mutation and the tumor immune microenvironment are significantly related to tumor growth, tumor progression, and drug resistance in individuals with GC.<sup>[51,52]</sup> Therefore, the analysis of somatic mutation and tumor infiltrating cells may help understand the potential mechanisms of GCs and may provide new selection for treatment strategies. TMB refers to the total number of somatic mutations in specific regions of a tumor genome<sup>[53]</sup> and have been regarded as a biomarker of immunotherapeutic responses.<sup>[54]</sup> To our surprise, the somatic mutation frequency and TMB level in the low-risk patients was higher than that of the high-risk patients, and TMB were negatively correlated with risk score. Furthermore, whether applied alone or combined with risk score, TMB could assist with differentiating GC patients with good prognosis from those with poor prognosis. It was widely known that the immune microenvironment can profoundly interfere with the efficiency of immunotherapy,<sup>[55]</sup> so we measured immune status in different risk groups. The infiltration abundance of immune cells was significantly different between the low-risk and high-risk groups in this study. We found that high-risk group had higher immune score, stromal score, ESTIMATE score, but lower tumor purity compared with the low-risk group. Besides, most of the immune cells were highly infiltrated in the high-risk patients, while the abundance of Th cells was higher in the low-risk group. Su et al<sup>[56]</sup> reported that GC tissue or peripheral blood mononuclear cell populations are characterized by an increase in Th1 cell responses, predominantly in patients with metastasized lymph nodes, suggesting that GC development and metastasis may be influenced by Th1 cell infiltration. These findings may provide potential novel targets and promote individualized immunotherapy for GC patients, but the exact mechanisms of how these seven RBPs function together

to influence tumorigenesis and the immune system remains unclear and further studies are demanded.

However, several limitations in this study should not be ignored. On the one hand, this study is a retrospective design; prospective clinical experimental and clinical data are needed to confirm these findings. On the other hand, some potentially significant clinical information, such as treatment plan and perioperative data, were not provided in the TCGA and GEO database. Therefore, we failed to explore the question that the signature predicts response rate in GC patients. Finally, the population in the database mainly came from western countries, and this might present observation bias.

## 5. Conclusion

All in all, we comprehensively analyzed the expression, function, interaction, and prognostic value for RBPs in GC through bioinformatic analysis. In addition, our research not only established an RBPs-based signature but also generated a nomogram to predict prognosis of GC patients. Our findings may provide new insights into the roles of RBPs in GC and develop potential markers for guiding treatment and prognosis.

## Acknowledgments

We thank all the researchers involved in the consolidation and submission of the data from the TCGA database, which may provide convenience and possibility of tumors studies in a large cohort.

## Author contributions

All authors read and approved the final manuscript.

**Conceptualization:** Yuzhi Wang, Yi Zhang.

**Data curation:** Yi Zhang.

**Formal analysis:** Yuzhi Wang, Weixia Ye, Gang Tian.

**Funding acquisition:** Gang Tian, Yi Zhang.

**Investigation:** Yuzhi Wang, Yi Zhang, Gang Tian.

**Methodology:** Yi Zhang, Weixia Ye, Gang Tian.

**Project administration:** Yuzhi Wang, Yi Zhang.

**Resources:** Yuzhi Wang, Weixia Ye.

**Software:** Yi Zhang, Gang Tian.

**Supervision:** Weixia Ye, Gang Tian.

**Validation:** Yuzhi Wang, Weixia Ye.

**Visualization:** Yi Zhang, Gang Tian.

**Writing – original draft:** Yuzhi Wang.

**Writing – review & editing:** Yi Zhang, Weixia Ye.

## References

- [1] Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 2019;144:1941–53.
- [2] Selim JH, Shaheen S, Sheu WC, et al. Targeted and novel therapy in advanced gastric cancer. *Exp Hematol Oncol* 2019;8:25.
- [3] Gamboa AC, Winer JH. Cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for gastric cancer. *Cancers (Basel)* 2019;11:1662.
- [4] Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet* 2014;15:829–45.
- [5] Calabretta S, Richard S. Emerging roles of disordered sequences in RNA-binding proteins. *Trends Biochem Sci* 2015;40:662–72.
- [6] Dreyfuss G, Kim VN, Kataoka N. Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* 2002;3:195–205.
- [7] Wurth L, Gebauer F. RNA-binding proteins, multifaceted translational regulators in cancer. *Biochim Biophys Acta* 2015;1849:881–6.
- [8] Wang J, Liu Q, Shyr Y. Dysregulated transcription across diverse cancer types reveals the importance of RNA-binding protein in carcinogenesis. *BMC Genomics* 2015;16(suppl 7):S5.
- [9] Kechavarzi B, Janga SC. Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biol* 2014;15:R14.
- [10] Zhang J, Jun Cho S, Chen X. RNP1, an RNA-binding protein and a target of the p53 family, regulates p63 expression through mRNA stability. *Proc Natl Acad Sci U S A* 2010;107:9614–9.
- [11] Wang T, He Y, Zhu Y, et al. Comparison of the expression and function of Lin28A and Lin28B in colon cancer. *Oncotarget* 2016;7:79605–16.
- [12] Harris MA, Clark J, Ireland A, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004;32:D258–61.
- [13] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28:27–30.
- [14] Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284–7.
- [15] Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;47:D607–13.
- [16] Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 2003;4:2.
- [17] Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;4:2612.
- [18] Hnzelmann S, Castelo R, Guinney J. Gsva: gene set variation analysis for microarray and rna-seq data. *BMC Bioinform* 2013;14:7.
- [19] Bindea G, Mlecnik B, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 2013;39:782–95.
- [20] Yang M, Chen H, Zhou L, et al. Expression profile and prognostic values of STAT family members in non-small cell lung cancer. *Am J Transl Res* 2019;11:4866–80.
- [21] Thul PJ, Lindskog C. The human protein atlas: a spatial map of the human proteome. *Protein Sci* 2018;27:233–44.
- [22] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019;69:7–34.
- [23] Van Cutsem E, Sagaert X, Topal B, et al. Gastric cancer. *Lancet* 2016;388:2654–64.
- [24] Sasako M, Inoue M, Lin JT, et al. Gastric Cancer Working Group report. *Jpn J Clin Oncol* 2010;40:i28–37.
- [25] Noh SH, Park SR, Yang HK, et al. Adjuvant capecitabine plus oxaliplatin for gastric cancer after D2 gastrectomy (CLASSIC): 5-year follow-up of an open-label, randomised phase 3 trial. *Lancet Oncol* 2014;15:1389–96.
- [26] Castello A, Fischer B, Hentze MW, et al. RNA-binding proteins in Mendelian disease. *Trends Genet* 2013;29:318–27.
- [27] Lukong KE, Chang KW, Khandjian EW, et al. RNA-binding proteins in human genetic disease. *Trends Genet* 2008;24:416–25.
- [28] Yang G, Fu H, Zhang J, et al. RNA-binding protein quaking, a critical regulator of colon epithelial differentiation and a suppressor of colon cancer. *Gastroenterology* 2010;138: 231-40.e1-5.
- [29] Vo DT, Subramaniam D, Remke M, et al. The RNA-binding protein Musashi1 affects medulloblastoma growth via a network of cancer-related genes and is an indicator of poor prognosis. *Am J Pathol* 2012;181:1762–72.
- [30] Hopkins TG, Mura M, Al-Ashtal HA, et al. The RNA-binding protein LARP1 is a post-transcriptional regulator of survival and tumorigenesis in ovarian cancer. *Nucleic Acids Res* 2016;44:1227–46.
- [31] Tripathi V, Shen Z, Chakraborty A, et al. Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet* 2013;9:e1003368.
- [32] Darimipourain M, Wang S, Ittmann M, et al. Transcriptional and post-transcriptional regulation of Sprouty1, a receptor tyrosine kinase inhibitor in prostate cancer. *Prostate Cancer Prostatic Dis* 2011;14:279–85.
- [33] Godbole M, Togar T, Patel K, et al. Up-regulation of the kinase gene SGK1 by progesterone activates the AP-1-NDRG1 axis in both PR-positive and -negative breast cancer cells. *J Biol Chem* 2018;293:19263–76.
- [34] Anantharaman V, Koonin EV, Aravind L. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res* 2002;30:1427–64.
- [35] Matia-González AM, Laing EE, Gerber AP. Conserved mRNA-binding proteomes in eukaryotic organisms. *Nat Struct Mol Biol* 2015;22:1027–33.
- [36] Li M, Li AQ, Zhou SL, et al. RNA-binding protein MSI2 isoforms expression and regulation in progression of triple-negative breast cancer. *J Exp Clin Cancer Res* 2020;39:92.
- [37] Klingenberg M, Groß M, Goyal A, et al. The long noncoding RNA cancer susceptibility 9 and RNA binding protein heterogeneous nuclear ribonucleoprotein L form a complex and coregulate genes linked to AKT signaling. *Hepatology* 2018;68:1817–32.
- [38] Chinnam M, Wang Y, Zhang X, et al. The Thoc1 ribonucleoprotein and prostate cancer progression. *J Natl Cancer Inst* 2014;106:306.
- [39] Jiang P, Li Z, Tian F, et al. Fyn/heterogeneous nuclear ribonucleoprotein E1 signaling regulates pancreatic cancer metastasis by affecting the alternative splicing of integrin  $\beta 1$ . *Int J Oncol* 2017;51:169–83.
- [40] Wang Z, Chen G, Wang Q, et al. Identification and validation of a prognostic 9-genes expression signature for gastric cancer. *Oncotarget* 2017;8:73826–36.
- [41] Benevolenskaya EV. Histone H3K4 demethylases are essential in development and differentiation. *Biochem Cell Biol* 2007;85:435–43.
- [42] Huang R, Li X, Yu Y, et al. SETD7 is a prognosis predicting factor of breast cancer and regulates redox homeostasis. *Oncotarget* 2017;8: 94080–90.
- [43] Guo T, Wen XZ, Li ZY, et al. ISL1 predicts poor outcomes for patients with gastric cancer and drives tumor progression through binding to the ZEB1 promoter together with SETD7. *Cell Death Dis* 2019;10:33.
- [44] Fu XF, Cheng SF, Wang LQ, et al. DAZ family proteins, key players for germ cell development. *Int J Biol Sci* 2015;11:1226–35.
- [45] Zhang C, Xue P, Gao L, et al. Highly conserved epigenetic regulation of BOULE and DAZL is associated with human fertility. *Faseb j* 2016;30:3424–40.
- [46] Kang KJ, Pyo JH, Ryu KJ, et al. Oncogenic role of BOLL in colorectal cancer. *Dig Dis Sci* 2015;60:1663–73.
- [47] Altadill T, Dowdy TM, Gill K, et al. Metabolomic and lipidomic profiling identifies the role of the RNA editing pathway in endometrial carcinogenesis. *Sci Rep* 2017;7:8803.
- [48] Galeano F, Rossetti C, Tomaselli S, et al. ADAR2-editing activity inhibits glioblastoma growth through the modulation of the CDC14B/Skp2/p21/p27 axis. *Oncogene* 2013;32:998–1009.
- [49] Wang MH, Qin SY, Zhang SG, et al. Musashi-2 promotes hepatitis B virus related hepatocellular carcinoma progression via the Wnt/ $\beta$ -catenin pathway. *Am J Cancer Res* 2015;5:1089–100.
- [50] 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.
- [51] Corso G, Carvalho J, Marrelli D, et al. Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer. *J Clin Oncol* 2013;31:868–75.
- [52] Kumagai S, Togashi Y, Sakai C, et al. An oncogenic alteration creates a microenvironment that promotes tumor progression by conferring a metabolic advantage to regulatory T cells. *Immunity* 2020;53:187–203.e8.

- [53] Allgäuer M, Budczies J, Christopoulos P, et al. Implementing tumor mutational burden (TMB) analysis in routine diagnostics: a primer for molecular pathologists and clinicians. *Transl Lung Cancer Res* 2018; 7:703–15.
- [54] Kim H, Hong JY, Lee J, et al. Clinical sequencing to assess tumor mutational burden as a useful biomarker to immunotherapy in various solid tumors. *Ther Adv Med Oncol* 2021;13:1758835921992992.
- [55] Messenheimer DJ, Jensen SM, Afentoulis ME, et al. Timing of PD-1 blockade is critical to effective combination immunotherapy with anti-OX40. *Clin Cancer Res* 2017;23:6165–77.
- [56] Su Z, Sun Y, Zhu H, et al. Th17 cell expansion in gastric cancer may contribute to cancer development and metastasis. *Immunol Res* 2014;58:118–24.