Systematic Construction and Validation of an RNA Binding Proteins-Based Signature for Prognostic Prediction in Gastric Cancer

Yi Zhang (✉ 472189926@qq.com)  
Peoples Hospital of Deyang City  https://orcid.org/0000-0002-0010-1915

Yuzhi Wang  
Peoples Hospital of Deyang City

Chengwen Li  
Southwest Medical University

Tianhua Jiang  
Peoples Hospital of Deyang City

Keywords: Gastric cancer, RNA-binding proteins, Prognostic signature, Overall survival

DOI: https://doi.org/10.21203/rs.3.rs-52319/v1

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

**Background:** 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. In this study, we aim to explore the function and clinical implication of RBPs in GC.

**Methods:** 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 BC. Meanwhile, a nomogram was built based on the same RBPs.

**Results:** 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 depended on seven prognosis-related hub RBPs that could divide GC patients into high- and low-risk groups. Survival analysis showed that the high-risk group had a worse prognosis compared to the low-risk group and the time-dependent receiver operating characteristic (ROC) curves suggested that the signature existed moderate predictive capacities of survival for GC patients. Similar results were obtained from another independent set GSE84437, confirming the robustness of signature. Calibration plots reported good consistency between overall survival (OS) prediction by nomogram and actual observation.

**Conclusion:** The findings of this study would provide evidence of the effect of RBPs on GC as well as offering novel potential biomarkers in prognosis prediction and clinical decision for GC patients.

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 [1]. 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 [2,3]. 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 [4,5]. RBPs have central roles in RNA structure, localization, stability, translatability and regulate gene expression post-transcriptionally and other cellular functions [6]. 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 [7]. In order to explore the construction and function of RBPs, we must comprehensively identify and annotate them firstly. Given that rapid advances in high-throughput sequencing technologies, over 1500 RBPs have been found and deposited into databases [4]. Numerous studies have indicated that RBPs play essential roles in tumor occurrence and development [8, 9]. For example, RBP RNPC1 regulates of P63 gene stability to inhibit initiation and progression [10]. RBP U2AF1 affects pre-mRNA splicing of a good deal of oncogenic drivers to promote tumorigenesis. Overexpression RBP LIN28A accelerates cell's progress from S to G2/M to enhance colon cancer cell proliferation [11]. A systematic study on RBPs may be conductive 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 of 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 survival of GC, some of which may serve as biomarkers for diagnosis and prognosis for GC in the future.

Methods

Data collection and DERBPs analysis

RNA sequence data and corresponding clinical information 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 Genes 82. We applied ‘limma’ package in R version 4.0.0 to standardize RBPs expression and estimate DERBPs with PDR value < 0.01 and |fold change|>0.5 between GC tissues and adjacent non-cancerous tissues.

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. The GO term enrichment analysis included three categories: molecular function (MF), cellular component (CC), and biological process (BP). Both P-value < 0.05 and FDR < 0.05 were used as the screening criteria.

Establishment of protein-protein interaction (PPI) network and modules selection
DERBPs were uploaded to the Search Tool for the Retrieval of Interacting Genes (STRING) database (http://www.string-db.org/) [12] to provide a global perspective for the interactive relationship of them. Protein-protein interaction (PPI) network was constructed and visualized using Cytoscape 3.7.0 software, and the most important modules that both Molecular Complex Detection (MCODE) score ≥ 4 and ≥ 6 nodes were selected by the MCODE plug [13]. A P-value of less than 0.05 was considered statistically significant.

**Construction and validation of overall survival (OS) risk prognostic model**

Identification of relevant prognostic candidate hub RBPs was performed from DERBPs in the PPI network by the univariate Cox regression analysis and multiple stepwise Cox regression. Following, 65a multivariate Cox proportional hazards regression model was built based on the hub RBPs. Risk scores for every patient were calculated according to the formula. Risk score value = EXP1*β1IncRNA1+EXP2*β2…+EXPx*βx, EXP presents the expression levels of each RBP and β presents the regression coefficient from the multivariate Cox proportional hazards regression model. The patients were assigned to high- 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. Additionally, 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) were used to justify whether the prognostic value of the model was credible. Statistical significance was judged with P-value <0.05. Finally, the nomogram and calibration plots were generated using “rms” packages of R software.

**Verifying expression and prognostic value of the hub RBPs**

The prognostic value of the hub RBPs in GC was assessed by an online database Kaplan–Meier plotter (http://kmplot.com/analysis/) [14]. Meanwhile, the Human Protein Atlas (HPA) database (http://www.proteinatlas.org/) was used to detect the expression of them on the translational level in GC tissue.

**Results**

**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 < 0.01 and |fold change|>0.5, among which 160 RBPs were up-regulated and 133 RBPs were down-regulated (Fig. 1a). The volcano plot shows the distribution of DERBPs (Fig. 1b).

**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'-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).

**PPI network analysis and identifying 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 1).

**Construction and validation of the prognostic signature**

We carried out univariate Cox regression analysis for key RBPs in the PPI network to found out prognosis-related RBPs, and attained 19 candidate genes associated with prognosis for further analysis (Figure. 4a). Then, multiple stepwise Cox regression was used to explore the independent prognostic impact of 19 candidate genes for GC. As a result, we obtained seven hub RBPs and established a signature based on them for OS (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 plots revealed that the low-risk group had a higher OS than the high-risk group (Fig. 5c). Also, a time-dependent ROC curve demonstrated the prediction power of the seven-RBPs signature for OS in five years achieved area under curves (AUCs) were 0.692 (Fig. 5e). To determine whether the seven-RBPs signature has similar prognostic value in another GC cohort, we built the model by applying the same formula for the GSE84437 set (Fig. 5b). Kaplan-Meier curves and log-rank test indicated that the low-risk group had longer survival time than the high-risk group in the GSE84437 set (Fig. 5d). Similar, time-dependent ROC curves evaluated that the model had a considerable prediction performance for GC patients (Fig. 5f).

**Establishment of a nomogram based on hub RBPs**
To provide clinicians with a quantitative method for predicting the prognosis in different years for GC patients, we integrated the hub RBPs to construct a nomogram. The nomogram based score was calculated according to the seven RBPs on the point scale. Hence, the 1-year OS, 3-year OS and 5-year OS of each GC patient could be predicted by total points in the nomogram (Fig. 6a). Importantly, calibration plots showed that the nomogram performed well with the ideal model for predicting 3-year OS and 5-year OS of GC (Fig. 6b, c). Furthermore, we explored the prognostic value of seven-RBPs signature and clinical factors for patients with GC from TCGA set by Cox regression analysis. Univariate Cox regression analysis found that age, stage, pN, and the risk scores were obviously related to OS (Table 2). However, only age and risk scores were remarkably related to OS after multivariate Cox regression analysis of GC (Table 2).

**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 expect for SETD7 (Fig. 7). 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. 8). However, BOLL protein expression did not exhibit differential expression in GC and normal tissues (Fig. 8).

**Discussion**

GC is one of the most commonly diagnosed cancers with high heterogeneity in the world, seriously endangering human health and life [15]. Therefore, exploring the molecular mechanism is critical to understand the pathogenesis and help improve the diagnosis 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 ensure stable maintenance of intracellular environments. In view of the central role of RBPs in the gene expression, dysregulation of RBPs may lead to several diseases, including cancers [16, 17]. Several studies have provided some evidence that RBPs dysregulation is common in various cancers [18-20]. However, fewer papers have explored the expression and specific functional role of 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 of DERBPs after that. In addition, we utilized univariate Cox regression analysis and multiple stepwise Cox regression to screen relevant prognostic hub RBPs and build a seven RBPs-based signature for predicting the survival time of GC patients. Meanwhile, log-rank test analysis and time-dependent ROC analysis were applied to evaluate the prognostic value of the 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 [21-23]. Many RBPs bind to sequence-specific motifs or RNA secondary structures through unique modular arrangements of individual RNA-binding domains to play the roles in the homeostatic regulation of gene expression [24,25], 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 [26]. 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 [27]. 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 [28,29]. 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, The seven prognosis related hub RBPs (RNASE3, RNASE, SETD7, BOLL, ADARB1, PPARC1B and MSI2) were screened by univariate Cox regression analysis and multiple Cox regression analysis. Compared with single biomarkers, integrating multiple biomarkers into a single model may considerably increase the predictive accuracy [30]. Therefore, we built a signature based on the seven hub RBPs using multivariate Cox proportional hazards regression analysis for useful and sensitive prognosis of GC patients. The signature is capable of discriminating the GC patients from high- and low-risk groups and may serve as an independent prognostic factor in GC. The time-dependent ROC curve indicated that the signature has a moderate performance in survival prediction for GC patients. More importantly, The risk stratification capability of the signature was confirmed in another independent set. Next, a nomogram was drawn to quantitatively predict survival time for clinical use. Kaplan-Meier plotter explored that most of the seven hub RBPs are associated with the survival of GC patients. The protein expression levels of seven 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 seven-RBPs signature has been proven to be significantly connected with OS of GC. Among these hub RBPs, many of them have proved to be closely related to tumor occurrence and development. SETD7 is the only lysine methyltransferases seven family members which can methylate transcription factors [31]. SETD7 could be used as a prognostic indicator for breast cancer,
downregulation of it suppressed expression of antioxidant enzymes and destabilized the redox status [32]. SETD7 and ISL1 may combine to form a complex on the ZEB1 promoter to promote tumorigenesis in GC cells [33]. BOLL, an ancestral gene of the Deleted in Azoosperma family, maintains normal functions of sperm [34,35]. 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 [36]. ADARB1 was expressed at high levels in endometrial cancer, and observed a positive correlation between increased expression and invasion degree [37]. Another study reported that ADARB1 could inhibit glioblastoma cell growth via regulation of the CDC14B/Skp2/p21/p27 axis [38]. MSI2 is a popular molecule in digestive system tumors. MSI2 promoted hepatocellular carcinoma progression via the Wnt/β-catenin signaling pathway and may serve as an indicator to predict outcome of patients with hepatocellular carcinoma [39]. MSI2 expression levels were upregulated in GC tissues and associated with poor prognosis of GC. Furthermore, MSI2 induced migration, invasion and angiogenesis to increased proliferation and invasiveness of GC cells [40].

However, several limitations in this study should not be ignored. On the one hand, there is a big gap between the number of tumor and adjacent samples, which may affect the accuracy of the results. On the other hand, this study was a retrospective design, prospective clinical experimental and clinical data are needed to confirm these funding further. Moreover, some proverbial potentially significant clinical information, such as treatment plan, vascular invasion and perioperative data, are not provided in the TCGA database. Finally, the population in the database mainly came from western countries, this may present observation bias.

**Conclusions**

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

**Abbreviations**

GC: Gastric cancer; RBP: RNA-binding proteins; DERBPs: Differentially expressed RNA-binding proteins; TCGA: The Cancer Genome Atlas; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; MF: Molecular function; CC: Cellular component; BP: Biological process; STRING: Search Tool for the Retrieval of Interacting Genes; MCODE: Molecular Complex Detection; OS: Overall survival; ROC: Receiver operating characteristic; GEO: Gene Expression Omnibus; HPA: Human Protein Atlas; AUCs: achieved area under curves; IHC: immunohistochemistry

**Declarations**
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. Yi Zhang is especially grateful to Michael Jackson, whose songs have always been a spiritual inspiration to him.

Authors' contributions

T J and Y Z conceived and designed the study; C L and Y W performed the collection and assembly of data; Y Z and Y W analyzed the data; Y Z and Y W wrote the manuscript; All authors read and approved the final manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

No funding was received in this study.

Availability of data and materials

The original data analyzed during the study are all available in the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) and The Cancer Genome Atlas (https://portal.gdc.cancer.gov/) database.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflicts of interest.

Author details

1 Department of Blood Transfusion, People's Hospital of Deyang City, No. 173, Section 1, Taishan North Road, Deyang, Sichuan 618000, China. 2 School of Basic Medical Science, Southwest Medical University, No. 319, Section 3, Zhongshan Road, Luzhou, Sichuan 646000, China.
References

1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer. 2019;144(8):1941-1953.

2. Selim JH, Shaheen S, Sheu WC, Hsueh CT. 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(11).

4. Gerstberger S, Hafner M, Tuschi T. A census of human RNA-binding proteins. Nat Rev Genet. 2014;15(12):829-845.

5. Calabretta S, Richard S. Emerging Roles of Disordered Sequences in RNA-Binding Proteins. Trends Biochem Sci. 2015;40(11):662-672.

6. Dreyfuss G, Kim VN, Kataoka N. Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol. 2002;3(3):195-205.

7. Wurth L, Gebauer F. RNA-binding proteins, multifaceted translational regulators in cancer. Biochim Biophys Acta. 2015;1849(7):881-886.

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(1):R14.

10. Zhang J, Jun Cho S, Chen X. RNPC1, 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(21):9614-9619.

11. Wang T, He Y, Zhu Y, Chen M, Weng M, Yang C, et al. Comparison of the expression and function of Lin28A and Lin28B in colon cancer. Oncotarget. 2016;7(48):79605-79616.

12. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-d613.

13. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003;4(2).

14. Yang M, Chen H, Zhou L, Chen K, Su F. Expression profile and prognostic values of STAT family members in non-small cell lung cancer. Am J Transl Res. 2019;11(8):4866-4880.

15. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7-34.

16. Castello A, Fischer B, Hentze MW, Preiss T. RNA-binding proteins in Mendelian disease. Trends Genet. 2013;29(5):318-327.

17. Lukong KE, Chang KW, Khandjian EW, Richard S. RNA-binding proteins in human genetic disease. Trends Genet. 2008;24(8):416-425.
18. Yang G, Fu H, Zhang J, Lu X, Yu F, Jin L, et al. RNA-binding protein quaking, a critical regulator of colon epithelial differentiation and a suppressor of colon cancer. Gastroenterology. 2010;138(1):231-240.e231-235.

19. Vo DT, Subramaniam D, Remke M, Burton TL, Uren PJ, Gelfond JA, 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(5):1762-1772.

20. Hopkins TG, Mura M, Al-Ashtal HA, Lahr RM, Abd-Latip N, Sweeney K, et al. The RNA-binding protein LARP1 is a post-transcriptional regulator of survival and tumorigenesis in ovarian cancer. Nucleic Acids Res. 2016;44(3):1227-1246.

21. Tripathi V, Shen Z, Chakraborty A, Giri S, Freier SM, Wu X, et al. Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. PLoS Genet. 2013;9(3):e1003368.

22. Darimipourain M, Wang S, Ittmann M, Kwabi-Addo B. Transcriptional and post-transcriptional regulation of Sprouty1, a receptor tyrosine kinase inhibitor in prostate cancer. Prostate Cancer Prostatic Dis. 2011;14(4):279-285.

23. Godbole M, Togar T, Patel K, Dharavath B, Yadav N, Janjuha S, 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(50):19263-19276.

24. Anantharaman V, Koonin EV, Aravind L. Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res. 2002;30(7):1427-1464.

25. Matia-González AM, Laing EE, Gerber AP. Conserved mRNA-binding proteomes in eukaryotic organisms. Nat Struct Mol Biol. 2015;22(12):1027-1033.

26. Li M, Li AQ, Zhou SL, Lv H, Wei P, Yang WT. RNA-binding protein MSI2 isoforms expression and regulation in progression of triple-negative breast cancer. J Exp Clin Cancer Res. 2020;39(1):92.

27. Klingenberg M, Groß M, Goyal A, Polycarpou-Schwarz M, Miersch T, Ernst AS, 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(5):1817-1832.

28. Chinnam M, Wang Y, Zhang X, Gold DL, Khoury T, Nikitin AY, et al. The Thoc1 ribonucleoprotein and prostate cancer progression. J Natl Cancer Inst. 2014;106(11).

29. Jiang P, Li Z, Tian F, Li X, Yang J. Fyn/heterogeneous nuclear ribonucleoprotein E1 signaling regulates pancreatic cancer metastasis by affecting the alternative splicing of integrin β1. Int J Oncol. 2017;51(1):169-183.

30. Wang Z, Chen G, Wang Q, Lu W, Xu M. Identification and validation of a prognostic 9-genes expression signature for gastric cancer. Oncotarget. 2017;8(43):73826-73836.

31. Benevolenskaya EV. Histone H3K4 demethylases are essential in development and differentiation. Biochem Cell Biol. 2007;85(4):435-443.
32. Huang R, Li X, Yu Y, Ma L, Liu S, Zong X, et al. SETD7 is a prognosis predicting factor of breast cancer and regulates redox homeostasis. Oncotarget. 2017;8(55):94080-94090.

33. Guo T, Wen XZ, Li ZY, Han HB, Zhang CG, Bai YH, 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(2):33.

34. Fu XF, Cheng SF, Wang LQ, Yin S, De Felici M, Shen W. DAZ Family Proteins, Key Players for Germ Cell Development. Int J Biol Sci. 2015;11(10):1226-1235.

35. Zhang C, Xue P, Gao L, Chen X, Lin K, Yang X, et al. Highly conserved epigenetic regulation of BOULE and DAZL is associated with human fertility. Faseb j. 2016;30(10):3424-3440.

36. Kang KJ, Pyo JH, Ryu KJ, Kim SJ, Ha JM, Choi K, et al. Oncogenic Role of BOLL in Colorectal Cancer. Dig Dis Sci. 2015;60(6):1663-1673.

37. Altadill T, Dowdy TM, Gill K, Reques A, Menon SS, Moiola CP, et al. Metabolomic and Lipidomic Profiling Identifies The Role of the RNA Editing Pathway in Endometrial Carcinogenesis. Sci Rep. 2017;7(1):8803.

38. Galeano F, Rossetti C, Tomaselli S, Cifaldi L, Lezzerini M, Pezzullo M, et al. ADAR2-editing activity inhibits glioblastoma growth through the modulation of the CDC14B/Skp2/p21/p27 axis. Oncogene. 2013;32(8):998-1009.

39. Wang MH, Qin SY, Zhang SG, Li GX, Yu ZH, Wang K, et al. Musashi-2 promotes hepatitis B virus related hepatocellular carcinoma progression via the Wnt/β-catenin pathway. Am J Cancer Res. 2015;5(3):1089-1100.

40. Yang Z, Li J, Shi Y, Li L, Guo X. Increased musashi 2 expression indicates a poor prognosis and promotes malignant phenotypes in gastric cancer. Oncol Lett. 2019;17(3):2599-2606.

Tables

**Table 1** GO annotation and KEGG pathway analysis of RBPs in the top module
| ID         | Description                              | Counts | P value  |
|------------|------------------------------------------|--------|----------|
| 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 |

Table 2 COX regression analysis of the seven-RBPs signature with OS in TCGA set

| Parameters        | Univariate COX                  | Multivariate COX                  |
|-------------------|---------------------------------|-----------------------------------|
|                   | HR (95% CI)                     | P value                           | HR (95% CI)                     | P value |
| Age               | 1.032(1.008-1.056)              | **0.008**                         | 1.043(1.017-1.070)              | **0.001** |
| Gender            | 1.225(0.735, 2.040)             | 0.436                            | 1.068(0.628, 1.816)             | 0.809   |
| Grade             | 1.264(0.775, 2.060)             | 0.348                            | 1.212(0.698, 2.105)             | 0.495   |
| Stage             | 1.419(1.076, 1.870)             | **0.013**                        | 1.295(0.741, 2.261)             | 0.364   |
| T                 | 1.187(0.870, 1.620)             | 0.280                            | 1.021(0.662, 1.574)             | 0.927   |
| M                 | 1.446(0.658, 3.178)             | 0.358                            | 1.433(0.552, 3.720)             | 0.460   |
| N                 | 1.373(1.104, 1.708)             | **0.004**                        | 0.327(0.846, 1.605)             | 0.327   |
| Risks core        | 1.244(1.146, 1.351)             | **<0.001**                       | 1.245(1.138, 1.362)             | **<0.001** |

Figures
The significantly altered RBPs in GC samples. a Heatmap. b Volcano plot.

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.
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.

|     | pvalue | Hazard ratio |
|-----|--------|--------------|
| DNMT1 | 0.043 | 0.664(0.474–0.966) |
| RBM28 | 0.040 | 0.643(0.421–0.961) |
| RNASE2 | 0.022 | 1.207(1.036–1.550) |
| RNASE3 | 0.021 | 1.530(1.067–2.194) |
| EZH2 | 0.022 | 0.728(0.555–0.955) |
| RNASE1 | 0.020 | 1.236(1.083–1.414) |
| OX1 | 0.009 | 1.642(1.205–2.238) |
| SETD7 | 0.041 | 1.384(1.014–1.889) |
| BOLL | 0.006 | 5.120(1.006–15.328) |
| PABPC5 | 0.021 | 2.336(1.137–4.801) |
| ADARB1 | 0.022 | 1.342(1.043–1.726) |
| ATXN1 | 0.044 | 1.337(1.007–1.774) |
| NR0B1 | 0.021 | 3.828(1.228–11.933) |
| PPARG1B | 0.020 | 0.732(0.555–0.968) |
| LARP6 | 0.016 | 1.277(1.046–1.565) |
| ZFP36 | 0.016 | 1.256(1.043–1.512) |
| MSI2 | 0.006 | 0.635(0.456–0.886) |
| TTP2 | 0.011 | 0.647(0.463–0.904) |
| RBMS3 | 0.007 | 1.431(1.101–1.859) |

Figure 4

COX regression analysis of the RBPs. a Univariate Cox regression analysis to screen prognostic candidate RBPs. b Multivariate Cox regression analysis to identity hub RBPs for building model.
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 seven RBPs expression heatmaps in TCGA set. b The distributions of risk score, OS status and the seven 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.

**Figure 6**

Construction of a robust nomogram prediction model based on the hub RBPs. a The nomogram for the prediction of OS. b The calibration plot of the nomogram predicted 3-year OS. c The calibration plot of the nomogram predicted 5-year OS.
Figure 7

Validation of prognosis value of the hub RBPs in GC. a RNASE1. b BOLL. c ADARB1. d PPARGC1B. e MSI2. f RNASE3. g SETD7.
Figure 8

Protein expression of the hub RBPs genes in normal and tumor tissues. a BOLL. b MSI2. c RNASE1. d RNASE3. e SETD7.