

Development and Validation of a Six-lncRNA Prognostic Signature in Gastric Cancer

Huihui Zeng (✉ zeng_huihui@126.com)

Jinan University First Affiliated Hospital <https://orcid.org/0000-0003-4638-2238>

Ai tao Nai

Jinan University First Affiliated Hospital

Feng Ma

Jinan University First Affiliated Hospital

SHOAIB BASHIR

Jinan University First Affiliated Hospital

Yin Li

Jinan University First Affiliated Hospital

Meng Xu

Jinan University First Affiliated Hospital <https://orcid.org/0000-0001-7387-0734>

Research Article

Keywords: Gastric cancer, Long noncoding RNA, TCGA, Prognostic signature

Posted Date: September 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-842112/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Gastric cancer (GC) has been a leading cause of cancer-related mortality for many years. It is thought that long noncoding RNAs (lncRNAs) can play a significant role in GC. This study aimed to construct a powerful six-lncRNA signature as a prognostic biomarker for GC patients.

Methods: Based on The Cancer Genome Atlas (TCGA), the expression profiles of lncRNAs and the corresponding clinical data of GC patients were obtained. Cox regression and the least absolute shrinkage and selection operator (LASSO) regression model were used to identify the prognostic lncRNA signature. A total of 337 patients were included in the combined dataset ($N = 337$), which was divided into a training dataset ($N = 169$) and a test dataset ($N = 168$). The reliability of the lncRNA prognostic signature was validated in three datasets.

Results: A six-lncRNA prognostic signature was constructed to predict the overall survival (OS) of GC patients. The signature had better discriminability than clinical characteristics. The prognostic risk score was as follows: (expression level of RP11-284F21.7 \times 0.243981) + (expression level of RP11-432J22.2 \times 0.502378) + (expression level of RP4-584D14.5 \times 0.447878) + (expression level of AC093850.2 \times 0.261822) + (expression level of AP000695.6 \times 0.654318) + (expression level of AC098973.2 \times 0.406603). In addition, the signature was confirmed to be a significant predictor for predicting the OS. The nomogram model precisely predicted the OS of GC. Enrichment analysis indicated that the signature was mainly enriched for extracellular matrix-related functions and tumor signaling pathways. The target genes IGFBP7, VCAN, and COL1A1 had prognostic value in GC. AC098973.2 and RP11-284F21.7 was verified for the first time in GC tissues and cell lines.

Conclusions: The six-lncRNA prognostic signature could predict the OS and has high clinical application value in GC.

Background

Gastric cancer (GC) currently ranks fifth in morbidity and fourth in mortality, and accounted for an estimated 1 million new cancer cases and 769,000 cancer-related deaths in 2020 [1]. The mortality rate and disease burden of GC have increased with the aging of the population in China [2]. Despite advancements in the diagnosis and treatment of GC, most patients are diagnosed at the advanced stage, and the 5-year overall survival rate is unsatisfactory. Therefore, it is particularly important to explore new diagnostic markers and therapeutic targets for GC patients.

Fortunately, with the continuous development and improvement of bioinformatics technology, most of the noncoding regions have been found to be widely transcribed, and these transcripts include long and short noncoding RNAs [3]. Long noncoding RNAs (lncRNAs) are longer than 200 nucleotides and cannot be translated into protein [4]. lncRNAs are involved in a variety of physiological processes and are closely related to tumors [5]. In general, lncRNAs have the ability to regulate genes through various mechanisms, such as translation, transcriptional regulation and protein modification [6]. lncRNAs such as HOTAIR [7], GCIInc1 [8], GMAN [9] and MEG3 [10] can be used as oncogenes or tumor suppressor genes, which indirectly influence the occurrence and development of GC. Therefore, lncRNAs might be potential targets for new therapeutic strategies for GC [11]. As another option, in recognition of the heterogeneity and susceptibility of GC, a group of lncRNA biomarkers could be used to evaluate the prognosis more stably and accurately in clinical application.

We studied the corresponding data of 337 patients in TCGA database. We identified differentially expressed lncRNAs for prognosis prediction. Furthermore, we constructed a six-lncRNA prognostic signature and validated the accuracy in the three datasets. We established an individualized nomogram model, and its value in prognosis prediction was

higher than that of clinicopathologic characteristics. Finally, the target genes related to GC were identified by functional enrichment analysis and network construction technology.

Materials And Methods

Data collection and preprocessing

Data for 407 patients with GC (providing 375 cancer tissues and 32 normal tissues) and a total of 14834 lncRNAs were obtained from the TCGA database (<https://portal.gdc.cancer.gov>). The expression profiles of lncRNAs and the corresponding clinical information were also downloaded. The package “edgeR” can be utilized to identify the differentially expressed lncRNAs using log₂ fold change > 0.5 and FDR < 0.05 as a selection threshold. The clinical information included age, gender, histologic grade, tumor (T) stage, lymph node (N) stage, metastasis (M) stage, clinical stage, and survival information. The exclusion criteria were including: (1) pathologic diagnosis was not gastric adenocarcinoma; (2) the follow-up days was less than 30.

Construction and validation of the lncRNA prognostic Signature

Univariate Cox regression analysis was conducted to identify lncRNAs associated with prognosis by the R package “survival” in the training dataset. LASSO regression analysis was used to narrow the gene score by the R package “glmnet”. According to the calculated minimum lambda value, prognosis-related lncRNAs were obtained to represent the best signature. We established a prognostic risk model using multivariate regression analysis. The following formula was applied: risk score = $\sum \beta_i lncRNA_i \times Exp(lncRNA_i)$ (β is the multivariable Cox regression coefficient of each lncRNA, and Exp is the expression level of each lncRNA).

According to the formula, We calculated the risk score of each GC patient in the training dataset. GC patients were further classified into the low-risk group (N = 169) and the high-risk group (N = 168) using the average risk score as the cutoff value. Similarly, the risk score formula was applied in the test dataset and the combined dataset. Kaplan-Meier (K-M) survival analysis was utilized to explore correlations between the signature and overall survival (OS). We used the R package “survival ROC” to verify the accuracy and sensitivity of the prognostic model. Moreover, univariate and multivariate Cox analyses were used to investigate whether the signature was significant predictor for predicting the OS.

Establishment and evaluation of the nomogram

We generated individualized predictions for the OS using the nomogram. The nomogram model for predicting 1-, 3-, and 5-year OS was based on the outcomes of the multivariate analysis. We used the R package “rms” to establish the nomogram and calibration plots. A calibration chart was utilized to evaluate the predictive performance of the nomogram. Receiver operating characteristic (ROC) analysis was applied to assess the accuracy of the predicted nomogram.

Functional enrichment analysis

Functional enrichment analysis was used to elucidate the potential biological mechanism and the pathway of the prognostic six-lncRNAs. The correlations between lncRNAs and those of the co-expressed mRNAs were calculated by Pearson correlation coefficient analysis (Pearson coefficient > 0.4). The differentially co-expressed mRNAs with log₂ fold change ≥ 0.5 and FDR < 0.05 were included. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were utilized to explore the potential related enrichments and pathways.

Protein-protein interaction (PPI) network construction and validation of the target genes

We used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) online database to assess PPIs. The PPI network was visualized by Cytoscape software. The Gene Expression Profiling Interactive Analysis (GEPIA) database (<http://gepia.cancer-pku.cn/detail.php>) and Kaplan-Meier (K-M) Plotter database (<https://kmplot.com/analysis/>) were performed to explore the relationship between the target genes and prognosis of patients.

Expression of the six lncRNAs in GC tissues

The expression levels of lncRNAs in 337 GC tissues and 32 nontumor tissues were analyzed from TCGA database. We also analyzed the differential expression of lncRNAs in 27 pairs of cancer tissues and paracancerous tissues. Moreover, K-M analysis was performed to display the OS of patients according to the expression of the six lncRNAs (high-risk versus low-risk group).

Tissue samples and cell culture

The cancer and paracancerous tissues were taken from 9 GC patients who received gastrectomy at the Department of Gastrointestinal Surgery, the First Affiliated Hospital of Bengbu Medical College. This present research was approved by Medical Ethics Committee of The First Affiliated Hospital of Bengbu Medical College, Anhui, China (Approval number: 2021KY029). No patients received chemotherapy or radiotherapy before operation. All tissue samples were immediately frozen in liquid nitrogen and stored until use. All patients signed informed consent for voluntary participation in the study.

We used human GC cell lines (BGC-803, BGC-823, and MKN-45), and the human normal gastric epithelial cells-1 (GES-1) line. Four cell lines were obtained from our laboratory of Jinan University (Guangzhou, Guangdong, China). The cell lines were maintained at 37°C, in a humidified incubator (95% humidity and 5% CO₂). All Cells were culture in RPMI 1640 medium (GIBCO, Shanghai, China) supplemented with 10% fetal bovine serum (GIBCO, Shanghai, China).

RNA extraction and qRT-PCR analysis

We used the qRT-PCR method to validate the expression of two hub lncRNAs in cancer tissues and cell lines. Total RNA from tissues and cultured cells was isolated using TRIzol (TAKARA, Beijing, China), cDNA was generated using Hifair ® III 1st Strand cDNA Synthesis Super Mix for qPCR (YEASEN, Shanghai, China). Quantitative PCR reactions was performed using SYBR Green Master Mix (VAZYME, Nanjing, China) in order to detect the expression levels of the lncRNAs (AC098973.2 F: 5'CATGCC CTGGATTCTGCTA3', R: 5'TGGCTGGTAGCTC TGATTCT3'; RP11-284F21.7 F: 5'CAGGGACAGGGCAGTATTCC3', R: 5'GCCCAACGACTCAGGTATC3'). GAPDH was selected as an endogenous control. Each sample was independently analyzed in triplicate and each experiment was repeated at least three times. We used the relative quantification $2^{-\Delta\Delta CT}$ method to calculate fold changes.

Statistical analysis

We used R software (version 4.0.3), SPSS 22.0 (SPSS Inc., Chicago, IL, USA) or GraphPad Prism 9 (GraphPad Software Inc., La Jolla, CA, USA) for statistical analyses. K-M analysis and log-rank tests were utilized to compare differences in survival for the different groups. Univariate and multivariate Cox analyses were performed to testify the significant predictors. ROC curve analysis was used to assess the sensitivity and specificity. The prognostic accuracy was evaluated by AUC analysis. Unpaired Student's t-test and Paired Student's t-test were employed to analyze lncRNA expression in cancer tissues and nontumor tissues. Differences with $P < 0.05$ were determined to be statistically significant.

Results

Identification of prognosis-related lncRNAs

The entire flowchart of this study was briefly shown in Fig. 1. A total of 337 samples were included for further analysis. The clinical characteristics of GC patients are described in Table 1. A total of 410 differentially expressed lncRNAs were obtained. Among these lncRNAs, 15 lncRNAs associated with prognosis were identified by univariate Cox analysis (Fig. 2A). Furthermore, the LASSO regression model was conducted with 10-fold cross validation to analyze their expression. Finally, six lncRNAs were ultimately analyzed using LASSO regression analysis (Fig. 2B-C).

Table 1
Clinical information of 337 GC patients
from TCGA database

| Clinical information | N | % |
|----------------------|-----|-------|
| Age(years) | | |
| ≤ 65 | 153 | 45.40 |
| > 65 | 181 | 53.71 |
| Unknown | 3 | 0.89 |
| Gender | | |
| Male | 218 | 64.69 |
| Female | 119 | 35.31 |
| Histologic grade | | |
| G1 | 9 | 2.67 |
| G2 | 120 | 35.61 |
| G3 | 199 | 59.05 |
| Unknown | 9 | 2.67 |
| T classification | | |
| T1 | 15 | 4.45 |
| T2 | 74 | 21.96 |
| T3 | 156 | 46.29 |
| T4 | 88 | 26.11 |
| Unknown | 4 | 1.19 |
| N classification | | |
| N0 | 99 | 29.38 |
| N1 | 91 | 27.00 |
| N2 | 68 | 20.18 |
| N3 | 68 | 20.18 |
| Unknown | 11 | 3.26 |
| M classification | | |
| M0 | 303 | 89.91 |
| M1 | 22 | 6.53 |
| Unknown | 12 | 3.56 |

Construction and validation of the six-lncRNA signature

The six-lncRNA signature was constructed by multivariate Cox regression analysis. The descriptions of the six lncRNAs are listed in Table 2. The risk scores for patients with GC were calculated based on the prognostic model formula in the training dataset, the test dataset and the combined dataset. The prognostic risk score as follows: (expression level of RP11-284F21.7 \times -0.243981) + (expression level of RP11-432J22.2 \times -0.502378) + (expression level of RP4-584D14.5 \times -0.447878) + (expression level of AC093850.2 \times 0.261822) + (expression level of AP000695.6 \times 0.654318) + (expression level of AC098973.2 \times 0.406603). Among the six lncRNAs, RP11-284F21.7, RP11-432J22.2 and RP4-584D14.5 were negative regulatory factors (coefficient < 0). On the other hand, AC093850.2, AP000695.6, and AC098973.2 were positive regulatory factors (coefficient > 0) (Fig. 3).

Table 2
Description of the six lncRNAs

| Ensembl ID | LncRNA | LNCipedia ID | Location | Length |
|-----------------|---------------|--------------|--------------------------|--------|
| ENSG00000229953 | RP11-284F21.7 | lnc-NES-1 | chr1:156646507–156661424 | 758 bp |
| ENSG00000238058 | RP11-432J22.2 | lnc-KCNT1-1 | chr9:135907812–135913513 | 311 bp |
| ENSG00000240449 | RP4-584D14.5 | lncRARRES2-2 | chr7:150363777–150372590 | 577 bp |
| ENSG00000230838 | AC093850.2 | LINC01614 | chr2:215718043–215719424 | 648 bp |
| ENSG00000230479 | AP000695.6 | lnc-CHAF1B-3 | chr21:36430360–36481070 | 893 bp |
| ENSG00000225548 | AC098973.2 | LINC01980 | chr3:27797588–27860325 | 491 bp |

We analyzed the risk score, survival status, and the distribution of lncRNA expression profiles in the three datasets (Fig. 4A-C). The results showed that GC patients in the low-risk group tended to have longer survival times. According to the K-M analysis of the three datasets, the survival rate of the high-risk group was lower than that of the low-risk group ($P < 0.05$) (Fig. 4D). The AUC of model for 5-year survival was 0.748 in the training dataset, 0.723 in the test dataset and 0.745 in the combined dataset (Fig. 4E). Our observations demonstrated that the signature has a certain ability to distinguish risk and to predict the OS of GC patients.

Cox analysis of the six-lncRNA signature

As shown in Fig. 5, Cox analyses were examined to verify the six-lncRNA signature as an independent risk factor for GC in three datasets. First, in the training dataset, univariate Cox analysis showed that clinical stage ($P = 0.018$), N stage ($P = 0.016$) and risk score ($P < 0.001$) were significant for predicting the OS; multivariate Cox analysis showed that the risk score (hazard ratio [HR] = 1.626, 95% confidence interval [CI] = 1.343–1.968, $P < 0.001$) was a significant risk factor (Fig. 5A-B). Second, in the test dataset, univariate Cox analysis showed that clinical stage ($P = 0.020$), N stage ($P = 0.048$) and risk score ($P = 0.029$) were significant for predicting the OS; multivariate Cox analysis showed that age (HR = 1.043, 95% CI = 1.014–1.071, $P = 0.003$) and risk score (HR = 1.161, 95% CI = 1.043–1.292, $P = 0.006$) were significant risk factors for predicting the OS (Fig. 5C-D). Third, in the combined group, age ($P = 0.026$), clinical stage ($P < 0.001$), N stage ($P = 0.001$) and risk score ($P < 0.001$) were significant for predicting the OS. Multivariate Cox analysis showed that age (HR = 1.030, 95% CI = 1.011–1.050, $P = 0.002$) and risk score (HR = 1.234, 95% CI = 1.142–1.333, $P < 0.001$) were significant risk factors for predicting the OS (Fig. 5E-F). In summary, these results demonstrated that the six-lncRNA signature was a significant predictor in GC.

Development and evaluation of the prognostic nomogram

We constructed a nomogram model to predict the survival rates of GC patients at 1, 3 and 5 years. The covariates included age, gender, histologic grade, clinical stage and risk score. The score for each variable was incorporated into the total score by means of the nomogram model. A higher score indicated a poorer prognosis (Fig. 6A). The calibration curve was used to represent the actual probability. (Fig. 6B). The AUC of the predicted nomogram was 0.751, which was higher than that of age (AUC = 0.617), gender (AUC = 0.571), histologic grade (AUC = 0.515), TNM stage (AUC = 0.585), T stage (AUC = 0.546), N stage (AUC = 0.523), and M stage (AUC = 0.498). Overall, we established an individualized predictive nomogram model that can accurately predict the OS (Fig. 6C).

Functional enrichment analysis of prognostic-related lncRNAs in the signature

We identified 184 differentially co-expressed genes associated with lncRNAs. GO analysis enrichment included biological process, cellular component and molecular function. The related biological processes mainly consisted of extracellular matrix organization and extracellular structure organization (Fig. 7A). Cellular components included collagen-containing extracellular matrix (ECM) and endoplasmic reticulum lumen (Fig. 7B). The enriched molecular functions consisted of extracellular matrix structural constituents and glycosaminoglycan binding (Fig. 7C). GO enrichment analysis demonstrated that the biological function may be related to the tumor interaction with the ECM. KEGG pathway indicated that these genes were rich in tumor-related signaling pathways, including the Wnt signaling pathway, focal adhesion and ECM-receptor interaction (Fig. 7D). These results indicated that function of these genes was highly associated with the occurrence of tumor.

Establishment of the PPI network and validation of target genes.

The PPI network was constructed by using the STRING online database to reflect the interaction of 24 target genes (Fig. 8A). We selected three genes (IGFBP7, VCAN, COL1A1) for further research. We verified the expressions and survival correlations of the three genes. In the GEPIA database, the expression levels of IGFBP7, VCAN, and COL1A1 in cancer tissues were higher than those in normal tissues (Fig. 8-B, F, J). In addition, high mRNA expression of these target genes was associated with a poor prognosis (Fig. 8-C, G, K) ($P < 0.05$). According to the K-M Plotter database, the survival rate of patients with high expression of these genes were lower than that of the low expression ($P < 0.01$) (Fig. 8-D, H, L). The expression levels of IGFBP7, VCAN and COL1A1 were related with tumor stages (Fig. 8-E, I, M) ($P < 0.05$). To sum up, we selected three lncRNA target genes that were closely related to prognosis and had an effect on the development of GC.

Expression of the six lncRNAs in GC

We analyzed the six lncRNAs from the TCGA database. The expression levels of the lncRNAs in cancer tissues were significantly higher than those in nontumor tissues ($P < 0.01$) (Fig. 9A). Then, we compared their expression levels in 27 pairs of cancer tissues and paracancerous tissue. These results showed that the six lncRNAs were differentially expressed in GC tissues (Fig. 9B). High expression of RP4-584D14.5, RP11-284F21.7 and RP11-432J22.2 generally indicated better overall survival than low expression. Nevertheless, the survival rates of patients with high AP000695.6 expression were lower than those of patients in the low expression group ($P < 0.05$). Although there was no significant difference in the survival of patients for AC093850.2 and AC098973.2 expression ($P > 0.05$), we observed a trend of poor prognoses in the high expression groups (Fig. 9C). These studies ultimately showed that these lncRNAs can be universally detected in GC tissues.

Validation of two hub lncRNAs in cancer tissues and cell lines

To verify the results of bioinformatics analysis, we selected two hub lncRNAs (AC098973.2 and RP11-284F21.7) for further study. The expressions of the two lncRNAs were determined in tumor tissues and cell lines with qRT-PCR. Compared with adjacent tissues, the expression of the two hub lncRNAs in GC tissues was markedly increased ($P < 0.05$, Fig. 10A-B). Moreover, the expression of the two hub lncRNAs was significantly higher in three GC cell lines than in GES-1 ($P < 0.05$, Fig. 10C, D).

Discussion

GC is a relatively common digestive malignancy, especially in Asian countries. Many patients lose the opportunity for surgery because of the delayed diagnosis. The treatments for advanced GC mainly include chemotherapy and targeted therapy, but the efficacy remains poor. The molecular genetic landscape of GC includes gene mutations, transcriptional changes involving mRNAs and lncRNAs[12]. Many lncRNAs have been discovered in GC tissues, cells, gastric juice and plasma, and lncRNAs may act as potential biomarkers for early diagnosis and the evaluation of prognosis, therapeutic response, and chemotherapy resistance [13]. Furthermore, we can utilize public bioinformatics databases to explore the novel perspective of applying prognostic lncRNAs as biomarkers in GC.

It is well-known that GC is a multistep oncogenic process with characteristics of pathological and molecular heterogeneity. Prognostic-related lncRNA signatures are more reliable than single lncRNAs in prognosis prediction. As shown in Table 3, many lncRNA signatures have been reported in recent years. Cai et al. and Nie et al. identified many differentially expressed lncRNAs, but their models were difficult to implement in clinical practice [14, 15]. Although some studies showed signatures composed of fewer lncRNAs, the accuracy of the prognostic evaluation was not high [16, 17]. Moreover, some scholars have established models that can accurately predict prognosis, but they have not further explored the corresponding target genes [18, 19]. In this study, we established a prognostic signature that had good predictive ability and was an independent risk factor in GC. Compared to the traditional TNM staging system, the nomogram can accurately predict OS and be precisely applied in clinical practice.

Table 3
Comparison of existing prognostic signatures in GC

| Databases | Methods | Signature | Symbols | Survival event | AUC value | References |
|------------------------------|---|-----------|---|----------------|-----------|--------------------|
| TCGA | LASSO Cox Univariate Cox | 3 | CYP4A22-AS1 AP000695.6 RP11-108M12.3 | OS | 0.737 | Cheng et al. 2018. |
| TCGA | Univariate Cox MultivariateCox | 9 | ADAMTS9AS1 LINC01614, CYMP-AS1 OVAAL, LINC01210, LINC02408,FLJ42969LINC01775, LINC01446 | OS | 0.795 | Cai et al. 2019 |
| GSE62254 GSE57303 | Robust likelihood-based survival LASSO Cox | 6 | lncRNA-IPW NCRNA00086 RP11-38P22.2 ERVH48-1 LOC158572 AC004080.17 | OS | 0.77 | Ma et al. 2019 |
| TCGA | LASSO cox Univariable Cox MultivariateCox | 3 | AC007991.4 AC079385.3 AL109615.2 | OS | 0.751 | Zhang et al. 2020 |
| TCGA | Univariable Cox MultivariateCox | 3 | OVAAL, FLJ16779 FAM230D | OS | 0.748 | Wang et al. 2020 |
| TCGA GSE62254 | LASSO cox Univariable Cox MultivariateCox | 14 | ADAMTS9-AS2 FLG-AS1 RNF144A-AS1 LINC00922,C15orf54 ERVMER61-1 POU6F2-AS2 LINC01210,LINC00973 ERICH3-AS1 LINC00326, LINC01208 LINC00645, DSCR10 | OS | 0.737 | Nie et al. 2020 |
| TCGA GSE62254 GSE15459 | Univariate Cox Multivariate Cox | 5 | LINC00205, TRHDE-AS1, OVAAL LINC00106,MIR100HG | OS DFS | 0.734 | Wu et al. 2021 |

Among the six lncRNAs, AC093850.2 and AC098973.2 are known as LINC01614 and LINC01980, respectively. AC093850.2 has been identified as a potential predictor of prognosis in many cancers, such as breast cancer, esophageal squamous cell carcinoma (ESCC), glioma, and osteosarcoma [20–23]. In the latest study, AC093850.2 was found to promote GC cell growth and migration [24], but the mechanism should be explored in GC. AC098973.2 may act as an oncogene in ESCC and hepatocellular carcinoma (HCC) [25, 26]. In addition, the LINC01980/miR-190a-5p/MY05A pathway promotes the development of ESCC [25]. Previous research established a

three-lncRNA prognostic signature including AP000695.6, but the value of AP000695.6 has not been further verified in clinical samples [27]. RP11-284F21.7 was evaluated in lung cancer cells [28]. RP4-584D14.7 was found in renal cell carcinoma cells [29]. In this study, these lncRNAs were identified from GC patients in the TCGA database. The expression levels of the two lncRNAs in tissue samples and cel lines were detected by qRT-PCR. Due to the limited tissue samples, we only verified the expression of two hub lncRNAs in GC tissues. The high expression levels of the two lncRNAs (AC098973.2 and RP11-284F21.7) were carried out in GC tissues and cell lines for the first time. However, a large number of clinical samples from multiple centers need to be assessed, and the specific regulatory mechanisms need to be deeply explored in GC.

Our study evaluated that the biological function of the signature components was related to the ECM. The current study showed that ECM and ECM-related components play an essential role in the occurrence and development of GC [30]. We found that the 24 target genes were mainly composed of ECM-related genes, such as SPARC, BGN, FBN1, SPP1, and FN1. Similarly, the pathways related to the signature included ECM-receptor interactions. The present study suggests that ECM targeting holds great clinical potential as an innovative and effective treatment for GC [31]. KEGG pathway analysis showed that these genes were rich in the Wnt signaling pathway and focal adhesion. Yang et al. revealed that LINC01133 inhibited progression and metastasis via the Wnt/β-catenin pathway in GC [32]. As dysregulation of the Wnt pathway has been observed in approximately 50% of GC tumors [33], the Wnt pathway might offer a new therapeutic target. Likewise, the CBP/β-catenin antagonist PRI-724 has been developed as a Wnt pathway inhibitor [34]. In addition, the integrins and growth factor receptors consisted of the focal adhesions through cytoplasmic signaling networks. Previous research has revealed that targeting focal adhesion proteins would be an effective mechanism of treatment regimens including chemotherapy, radiotherapy and novel molecular therapeutics [35]. Thus, the six-lncRNA signature has clinical potential as a pharmacological target.

IGFBP7, VCAN, and COL1A1 were closely related to prognosis in the study. Recent studies concluded that high expression of IGFBP7 was associated with poor survival in HCC [36], colon cancer [37], and GC [38]. C Rupp et al. provided evidence that IGFBP7 might selectively inhibit colon cancer metastasis through the suppression of epithelial-mesenchymal transition (EMT) [39]. Thus, further research on the pathogenesis and immune mechanism of IGFBP7 needs to be performed in GC. COL1A1 and VCAN are widely related to ECM-receptor interaction. COL1A1 is a member of the type I collagen family and is involved in breast cancer [40], HCC [41], ovarian cancer [42] and GC [43]. COL1A1 is closely correlated with cell invasion and metastasis based on the activation of the TGF-β signaling pathway [44]. VCAN is a chondroitin sulfate proteoglycan and it can provide hydration and a loose matrix during disease progression. Cheng et al. revealed that VCAN was upregulated and had an impact on the progression of GC [45]. Mohamed Salem found that high expression levels of VCAN mediated miR-590-3p and promoted the development of ovarian cancer [46]. Consequently, these target genes have important clinical application value.

Nevertheless, our study revealing a six-lncRNA signature has several limitations. First, we only obtained data from the TCGA, and more databases are required for further validation. Thus, we need large-scale cohorts to verify the prognostic signature in a multicenter prospective clinical study. Third, the six lncRNAs need to be further explored to determine their functions and underlying mechanisms. The six-lncRNA signature might provide new insight for individualized and precise treatment in GC patients.

Conclusions

The results of the current study indicated that the six-lncRNA signature might be a potential biomarker for prognosis in GC. We verified a prognostic six-lncRNA signature by using integrated bioinformatics approaches and experiment. Future investigations will focus on the functional mechanisms of these lncRNAs.

Abbreviations

GC:Gastric cancer; TCGA :The Cancer Genome Atlas; LncRNAs: Long noncoding RNAs; LASSO :The least absolute shrinkage and selection operator; OS: Overall Survival; AUC: Area under the curve ; ROC: Receiver operating characteristic;GEPIA: Gene Expression Profiling Interactive Analysis; K-M: Kaplan-Meier;CI: Confidence intervals; HR:Hazard ratio;ECM: Extracellular matrix; FC:Fold change; FDR:False Discovery Rates; GO:Gene Ontology ; KEGG: Kyoto Encyclopedia of Genes and Genomes ;PPI:Protein–protein interaction;

Declarations

Acknowledgements

Not applicable.

Authors' contributions

Meng Xu conceived and design this study. Huihui Zeng and Aitao Nai were responsible for the collection and analysis of data. Feng Ma contributed to the literature search. SHOAIB BASHIR wrote the draft manuscript. Yin Li revised the manuscript. All authors critically reviewed the manuscript and approved the final manuscript for publication.

Funding

This work was supported by the Science and Technology Foundation of Guangzhou (201803010059), the Natural Science Foundation of Bengbu Medical College (BYKY2019129ZD).

Availability of supporting data

lncRNAs, clinical data of GC patients were downloaded from the TCGA database (<https://portal.gdc.cancer.gov>); the target genes and patient prognosis were downloaded Gene Expression Profiling Interactive Analysis (GEPIA) database (<http://gepia.cancer-pku.cn/detail.php>) and K-M Plotter database (<https://kmplot.com/analysis/>)

Ethics approval and Consent to participate

This study was approved by Medical Ethics Committee of The First Affiliated Hospital of Bengbu Medical College, Anhui, China (Approval number: 2021KY029). All patients signed informed consent voluntarily.

Consent for publication

Not applicable.

Competing interests

The authors declare that they are no competing interests.

Authors details

¹Department of Oncology, The First Affiliated Hospital of Jinan university, Guangdong, Guangzhou, China

² Department of Medical Oncology, The First Affiliated Hospital of Bengbu Medical College, Bengbu, Anhui, China

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray FGlobal. Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021; 71(3):209–249. <https://doi.org/10.3322/caac.21660>.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J.Cancer statistics in China, 2015. CA Cancer J Clin. 2016;66(2):115–32. <https://doi.org/10.3322/caac.21338>.
3. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. Landscape of transcription in human cells. Nature. 2012;489(7414):101–8. <https://doi.org/10.1038/nature11233>.
4. Yoon JH, Kim J, Gorospe M. Long noncoding RNA turnover. Biochimie. 2015;117:15–21. <https://dx.doi.org/10.1016/j.biochi.2015.03.001>.
5. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015;47(3):199–208. <https://doi.org/10.1038/ng.3192>.
6. Peng WX, Koirala P, Mo YY. LncRNA-mediated regulation of cell signaling in cancer. Oncogene. 2017; 36(41):5661–7. <https://doi.org/10.1038/onc.2017.184>.
7. Ono H, Motoi N, Nagano H, Miyauchi E, Ushijima M, Matsuura M, Okumura S, Nishio M, Hirose T, Inase N, et al. Long noncoding RNA HOTAIR is relevant to cellular proliferation, invasiveness, and clinical relapse in small-cell lung cancer. Cancer Med. 2014;3(3):632–42. <https://doi.org/10.1002/cam4.220>.
8. Sun TT, He J, Liang Q, Ren LL, Yan TT, Yu TC, Tang JY, Bao YJ, Hu Y, Lin Y, et al. LncRNA GCIncl Promotes Gastric Carcinogenesis and May Act as a Modular Scaffold of WDR5 and KAT2A Complexes to Specify the Histone Modification Pattern. Cancer Discov. 2016;6(7):784–801. <https://doi.org/10.1158/2159-8290.CD-15-0921>.
9. Zhuo W, Liu Y, Li S, Guo D, Sun Q, Jin J, Rao X, Li M, Sun M, Jiang M, et al. Long Noncoding RNA GMAN, Up-regulated in Gastric Cancer Tissues, Is Associated With Metastasis in Patients and Promotes Translation of Ephrin A1 by Competitively Binding GMAN-AS. Gastroenterology. 2019;156(3):676–91 e611. <https://doi.org/10.1053/j.gastro.2018.10.054>.
10. Dan J, Wang J, Wang Y, Zhu M, Yang X, Peng Z, Jiang H, Chen L. LncRNA-MEG3 inhibits proliferation and metastasis by regulating miRNA-21 in gastric cancer. Biomedicine & Pharmacotherapy. 2018;99:931–8. <https://doi.org/10.1016/j.biopha.2018.01.164>.
11. Yuan L, Xu ZY, Ruan SM, Mo S, Qin JJ, Cheng XD. Long non-coding RNAs towards precision medicine in gastric cancer: early diagnosis, treatment, and drug resistance. Mol Cancer. 2020;19(1):96. <https://doi.org/10.1186/s12943-020-01219-0>.
12. Tan P, Yeoh KG. Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. Gastroenterology. 2015;149(5):1153–1162e1153. <https://doi.org/10.1053/j.gastro.2015.05.059>.
13. Fattahi S, Kosari-Monfared M, Golpour M, Emami Z, Ghasemiyan M, Nouri M, Akhavan-Niaki H. LncRNAs as potential diagnostic and prognostic biomarkers in gastric cancer: A novel approach to personalized medicine. J Cell Physiol. 2020;235(4):3189–206. <https://doi.org/10.1002/jcp.29260>.
14. Cai C, Yang L, Tang Y, Wang H, He Y, Jiang H, Zhou K. Prediction of Overall Survival in Gastric Cancer Using a Nine-LncRNA. DNA Cell Biol. 2019;38(9):1005–12. <https://doi.org/10.1089/dna.2019.4832>.
15. Nie K, Deng Z, Zheng Z, Wen Y, Pan J, Jiang X, Yan Y, Liu P, Liu F, Li P. Identification of a 14-LncRNA Signature and Construction of a Prognostic Nomogram Predicting Overall Survival of Gastric Cancer. DNA Cell Biol. 2020;39(9):1532–44. <https://doi.org/10.1089/dna.2020.5565>.
16. Zhang J, Piao HY, Wang Y, Lou MY, Guo S, Zhao Y. Development and validation of a three-long noncoding RNA signature for predicting prognosis of patients with gastric cancer. World J Gastroenterol. 2020;26(44):6929–44. <https://doi.org/10.3748/wjg.v26.i44.6929>.

17. Wu Y, Deng J, Lai S, You Y, Wu J. A risk score model with five long non-coding RNAs for predicting prognosis in gastric cancer: an integrated analysis combining TCGA and GEO datasets. *PeerJ*. 2021;9:e10556. <https://doi.org/10.7717/peerj.10556>.
18. Ma B, Li Y, Ren Y. Identification of a 6-lncRNA prognostic signature based on microarray re-annotation in gastric cancer. *Cancer Med.* 2020;9(1):335–49. <https://doi.org/10.1002/cam4.2621>.
19. Wang Y, Zhang H, Wang J. Discovery of a novel three-long non-coding RNA signature for predicting the prognosis of patients with gastric cancer. *J Gastrointest Oncol.* 2020;11(4):760–9. <https://doi.org/10.21037/jgo-20-140>.
20. Vishnubalaji R, Shaath H, Elkord E, Alajez NM. Long non-coding RNA (lncRNA) transcriptional landscape in breast cancer identifies LINC01614 as non-favorable prognostic biomarker regulated by TGFbeta and focal adhesion kinase (FAK) signaling. *Cell Death Discov.* 2019;5:109. <https://doi.org/10.1038/s41420-019-0190-6>.
21. Tang L, Chen Y, Peng X, Zhou Y, Jiang H, Wang G, Zhuang W. Identification and Validation of Potential Pathogenic Genes and Prognostic Markers in ESCC by Integrated Bioinformatics Analysis. *Front Genet.* 2020;11:521004. <https://doi.org/10.3389/fgene.2020.521004>.
22. Wang H, Wu J, Guo W. SP1-Mediated Upregulation of lncRNA LINC01614 Functions as a ceRNA for miR-383 to Facilitate Glioma Progression Through Regulation of ADAM12. *Onco Targets Ther.* 2020;13:4305–18. <https://doi.org/10.2147/OTT.S242854>.
23. Cai Q, Zhao X, Wang Y, Li S, Wang J, Xin Z, Li F. LINC01614 promotes osteosarcoma progression via miR-520a-3p/SNX3 axis. *Cell Signal.* 2021;83:1 09985. <https://doi.org/10.1016/j.cellsig.2021.109985>.
24. Chen Y, Cheng WY, Shi H, Huang S, Chen H, Liu D, Xu W, Yu J, Wang J. Classifying gastric cancer using FLORA reveals clinically relevant molecular subtypes and highlights LINC01614 as a biomarker for patient prognosis. *Oncogene.* 2021;40(16):2898–909. <https://doi.org/10.1038/s41388-021-01743-3>.
25. Zhang S, Liang Y, Wu Y, Chen X, Wang K, Li J, Guan X, Xiong G, Yang K, Bai Y. Upregulation of a novel lncRNA LINC01980 promotes tumor growth of esophageal squamous cell carcinoma. *Biochem Biophys Res Commun.* 2019;513(1):73–80. <https://doi.org/10.1016/j.bbrc.2019.03.012>.
26. Zang Y, Yi Q, Pan J, Ding M, Zuo X, Hu X. LINC01980 stimulates the progression of hepatocellular carcinoma via downregulating caspase 9. *J BUON.* 2020;25(3):1395–403.
27. Cheng P. A prognostic 3-long noncoding RNA signature for patients with gastric cancer. *J Cell Biochem.* 2018;119(11):9261–9. <https://doi.org/10.1002/jcb.27195>.
28. Ashouri A, Sayin VI, Van den Eynden J, Singh SX, Papagiannakopoulos T, Larsson E. Pan-cancer transcriptomic analysis associates long non-coding RNAs with key mutational driver events. *Nat Commun.* 2016;7:13197. <https://doi.org/10.1038/ncomms13197>.
29. Flippot R, Mouawad R, Spano JP, Roupert M, Comperat E, Bitker MO, Parra J, Vaessen C, Allanic F, Manach Q, et al. Expression of long non-coding RNA MFI2-AS1 is a strong predictor of recurrence in sporadic localized clear-cell renal cell carcinoma. *Sci Rep.* 2017;7(1):8540. <https://doi.org/10.1038/s41598-017-08363-6>.
30. Yang X, Chen L, Mao Y, Hu Z, He M. Progressive and Prognostic Performance of an Extracellular Matrix-Receptor Interaction Signature in Gastric Cancer. *Dis Markers.* 2020;2020:8816070. <https://doi.org/10.1155/2020/8816070>.
31. Moreira AM, Pereira J, Melo S, Fernandes MS, Carneiro P, Seruca R, Figueiredo J. The Extracellular Matrix: An Accomplice in Gastric Cancer Development and Progression. *Cells.* 2020;9(2). <https://doi.org/10.3390/cells9020394>.
32. Yang XZ, Cheng TT, He QJ, Lei ZY, Chi J, Tang Z, Liao QX, Zhang H, Zeng LS, Cui SZ. LINC01133 as ceRNA inhibits gastric cancer progression by sponging miR-106a-3p to regulate APC expression and the Wnt/beta-catenin pathway. *Mol Cancer.* 2018;17(1):126. <https://doi.org/10.1186/s12943-018-0874-1>.

33. Koushyar S, Powell AG, Vincan E, Phesse TJ. Targeting Wnt Signaling for the Treatment of Gastric Cancer. *Int J Mol Sci.* 2020;21(11).<https://doi.org/10.3390/ijms21113927>.
34. Okazaki H, Sato S, Koyama K, Morizumi S, Abe S, Azuma M, Chen Y, Goto H, Aono Y, et al. The novel inhibitor PRI-724 for Wnt/beta-catenin /C BP signaling ameliorates bleomycin-induced pulmonary fibrosis in mice. *Exp Lung Res.* 2019;45(7):188–99. <https://doi.org/10.1080/01902148.2019.1638466>.
35. Eke I, Cordes N. Focal adhesion signaling and therapy resistance in cancer. *Semin Cancer Biol.* 2015;31:65–75. <https://doi.org/10.1016/j.semcan.2014.07.009>.
36. Akiel M, Guo C, Li X, Rajasekaran D, Mendoza RG, Robertson CL, Jariwala N, Yuan F, et al. IGFBP7 Deletion Promotes Hepatocellular Carcinoma. *Cancer Res.* 2017;77(15):4014–25. <https://doi.org/10.1158/0008-5472.CAN-16-2885>.
37. Rupp C, Scherzer M, Rudisch A, Unger C, Haslinger C, Schweifer N, Artaker M, Nivarthi H, Moriggl R, Hengstschlager M, et al. IGFBP7, a novel tumor stroma marker, with growth-promoting effects in colon cancer through a paracrine tumor-stroma interaction. *Oncogene* 2015;34(7):815–25. <https://doi.org/10.1038/onc.2014.18>.
38. Zhao Q, Zhao R, Song C, Wang H, Rong J, Wang F, Yan L, Song Y, Xie Y. Increased IGFBP7 Expression Correlates with Poor Prognosis and Immune Infiltration in Gastric Cancer. *J Cancer.* 2021;12(5):1343–55. <https://doi.org/10.7150/jca.50370>.
39. Li Y, Xi Y, Zhu G, Jia J, Huang H, Liu Y, Guo Y, Liu L. Downregulated IGFBP7 facilitates liver metastasis by modulating epithelialmesenchymal transition in colon cancer. *Oncol Rep.* 2019;42(5):1935–45. <https://doi.org/10.3892/or.2019.7303>.
40. Meng C, He Y, Wei Z, Lu Y, Du F, Ou G, Wang N, Luo XG, Ma W, Zhang TC, et al. MRTF-A mediates the activation of COL1A1 expression stimulated by multiple signaling pathways in human breast cancer cells. *Biomed Pharmac other.* 2018;104:718–28. <https://doi.org/10.1016/j.biopharm.2018.05.092>.
41. Ma HP, Chang HL, Bamodu OA, Yadav VK, Huang TY, Wu ATH, Yeh CT, Tsai SH, Lee WH. Collagen 1A1 (COL1A1) Is a Reliable Biomarker and Putative Therapeutic Target for Hepatocellular Carcinogenesis and Metastasis. *Cancers (Basel)* 2019;11(6). <https://doi.org/10.3390/cancers11060786>.
42. Li M, Wang J, Wang C, Xia L, Xu J, Xie X, Lu W. Microenvironment remodeled by tumor and stromal cells elevates fibroblast-derived COL1A1 and facilitates ovarian cancer metastasis. *Exp Cell Res.* 2020;394(1):112153. <https://doi.org/10.1016/j.yexcr.2020.112153>.
43. Li J, Ding Y, Li A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. *World J Surg Oncol.* 2016;14(1):297. <https://doi.org/10.1186/s12957-016-1056-5>.
44. Guo Y, Lu G, Mao H, Zhou S, Tong X, Wu J, Sun Q, Xu H, Fang F. miR-133b Suppresses Invasion and Migration of Gastric Cancer Cells via the COL1A1 /TGF-beta Axis. *Onco Targets Ther.* 2020;13:7985–95. <https://doi.org/10.2147/OTT.S249667>.
45. Cheng Y, Sun H, Wu L, Wu F, Tang W, Wang X, Lv C. VUp-Regulation of VCAN Promotes the Proliferation, Invasion and Migration and Serves as a Biomarker in Gastric Cancer. *Onco Targets Ther.* 2020;13:8665–75. <https://doi.org/10.2147/OTT.S262613>.
46. Salem M, O'Brien JA, Bernaudo S, Shawer H, Ye G, Brkic J, Amleh A, Vanderhyden BC, Refky B, Yang BB, et al. miR-590-3p Promotes Ovarian Cancer Growth and Metastasis via a Novel FOXA2-Versican Pathway. *Cancer Res.* 2018;78(15):4175–90. <https://doi.org/10.1158/0008-5472.CAN-17-3014>.

Figures

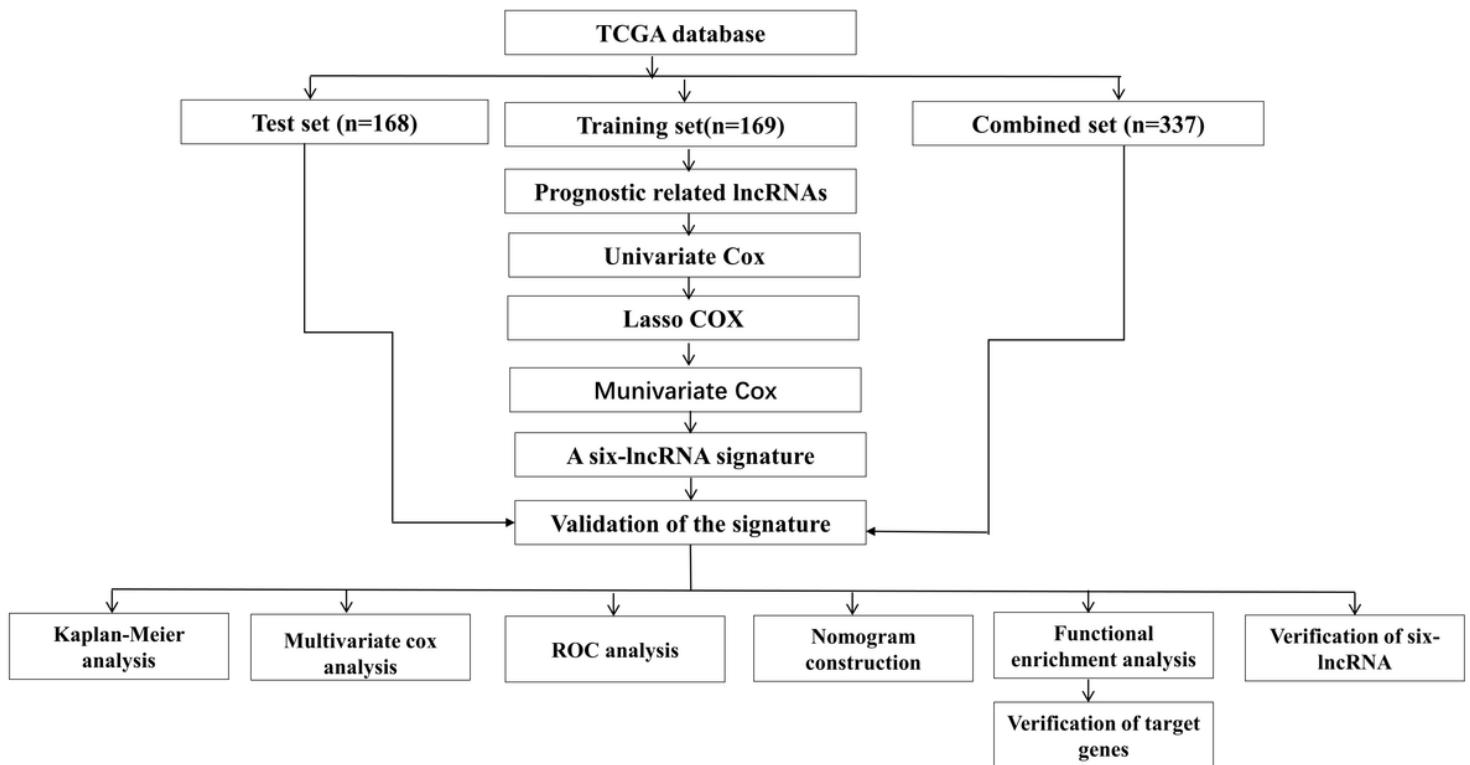
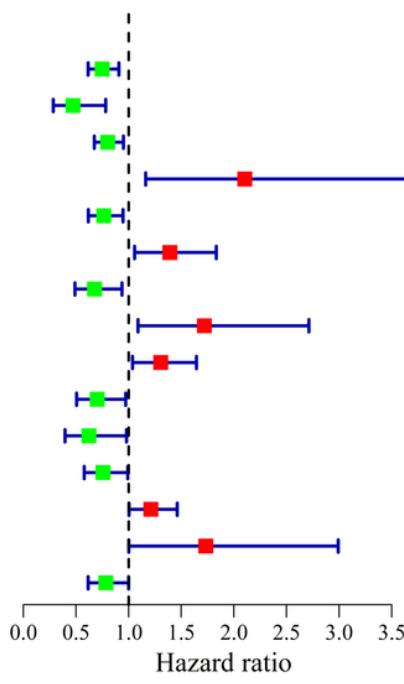


Figure 1

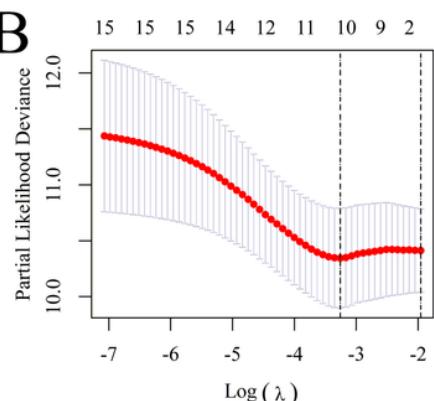
Flow chart of the study

A

| | pvalue | Hazard ratio |
|----------------|--------|--------------------|
| RP11-284F21.7 | 0.003 | 0.748(0.616-0.908) |
| RP4-584D14.5 | 0.004 | 0.473(0.285-0.783) |
| RP11-357H14.17 | 0.011 | 0.801(0.675-0.951) |
| AP000695.6 | 0.014 | 2.104(1.163-3.807) |
| RP11-284F21.10 | 0.014 | 0.765(0.618-0.947) |
| AC093850.2 | 0.019 | 1.392(1.057-1.833) |
| HOXB-AS4 | 0.019 | 0.677(0.490-0.937) |
| RP11-116D2.1 | 0.020 | 1.720(1.090-2.714) |
| AC098973.2 | 0.024 | 1.305(1.036-1.645) |
| HOTTIP | 0.033 | 0.701(0.506-0.972) |
| RP11-432J22.2 | 0.041 | 0.623(0.395-0.981) |
| CTD-2377D24.6 | 0.042 | 0.757(0.579-0.991) |
| RP1-60O19.1 | 0.046 | 1.212(1.003-1.463) |
| RP11-598F7.5 | 0.049 | 1.732(1.002-2.991) |
| RP3-523K23.2 | 0.050 | 0.784(0.615-1.000) |



B



C

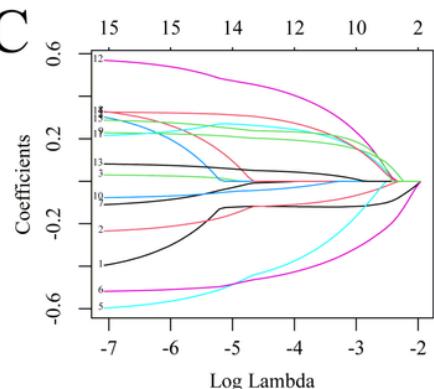


Figure 2

Identification of prognosis-related lncRNAs A Univariate Cox regression of 15 lncRNAs. B The selection of the tuning parameter λ based on the LASSO model. C. The LASSO model coefficient profiles

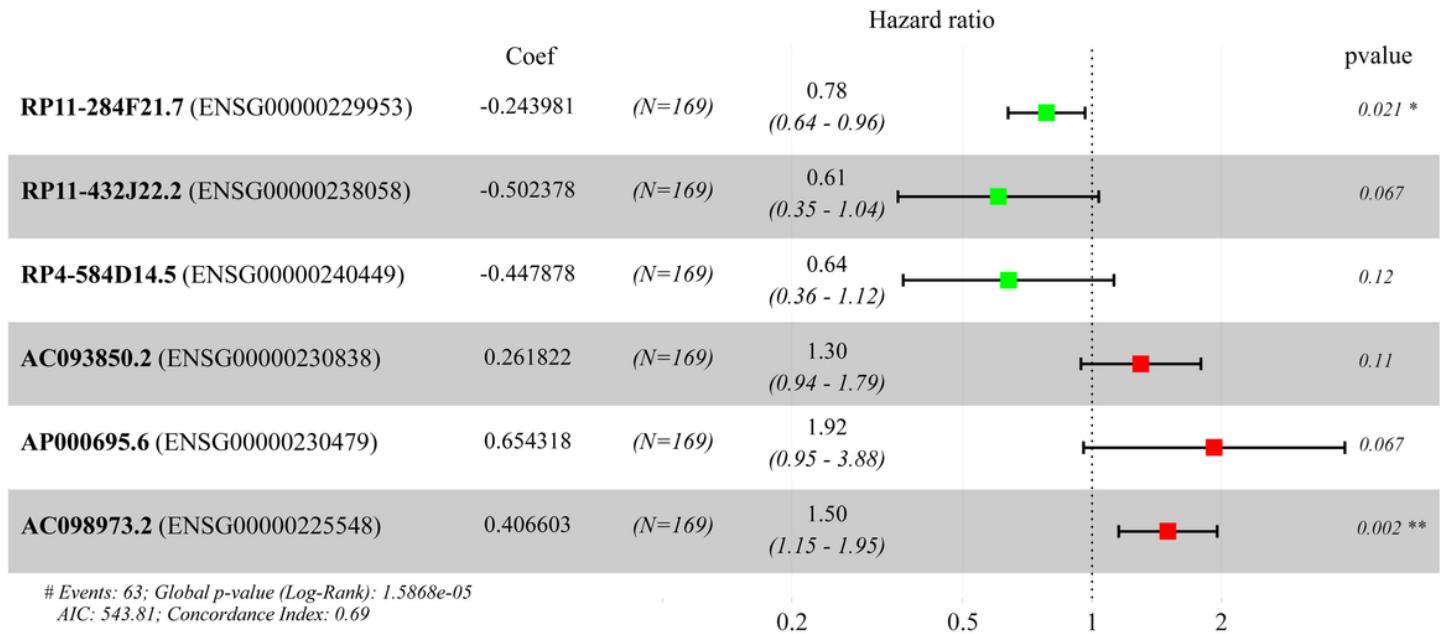


Figure 3

Forest plot of a six-lncRNA prognostic signature

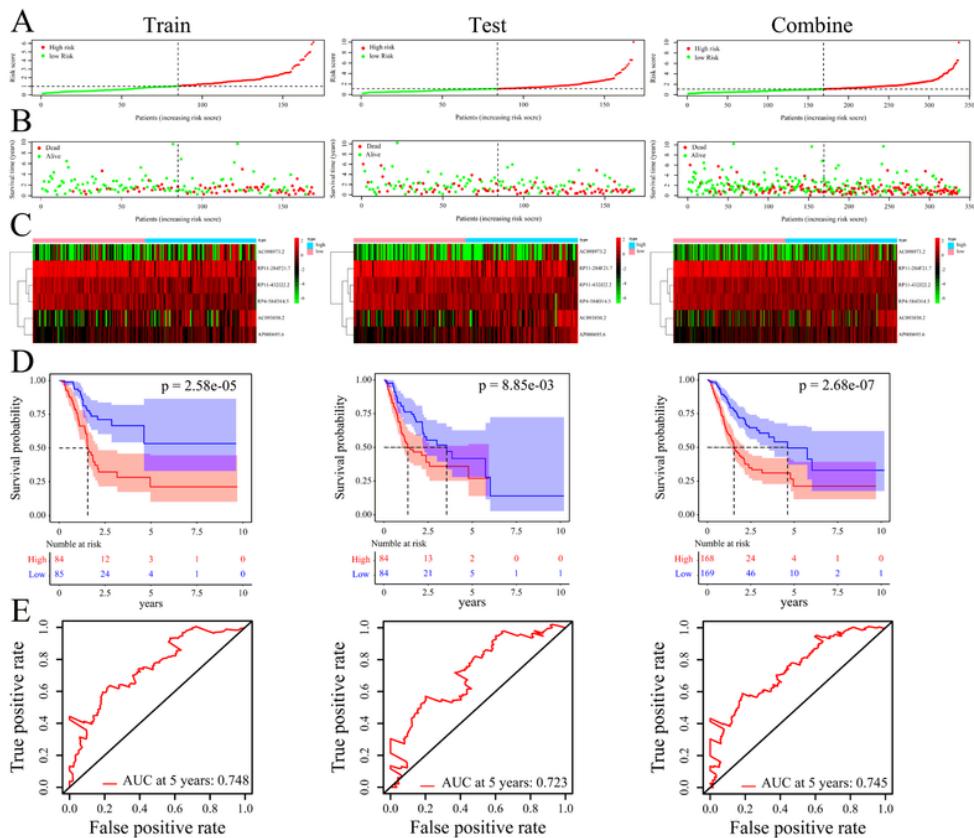
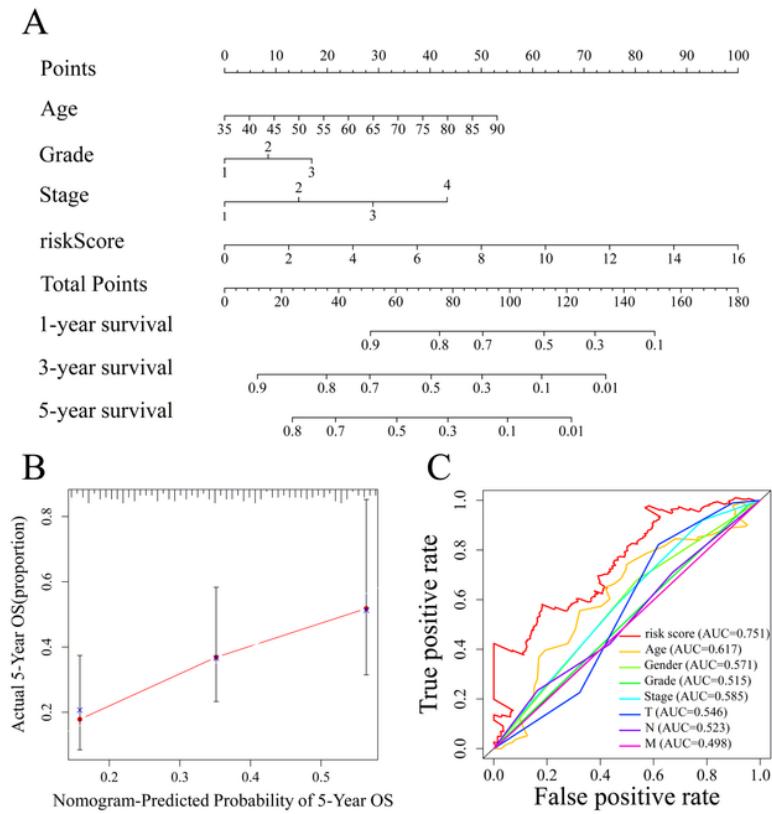


Figure 4

Evaluation of the six-lncRNA signature in the three datasets. A Distribution of six-lncRNA risk score in the three datasets. B Distribution of survival status for patients in the three datasets. C Heat map of the expression profiles of the six lncRNAs in the three datasets; D K-M analysis of the survival in the three datasets. E ROC curve of the six-lncRNA signature in the three datasets

**Figure 5**

Cox analysis of the six-lncRNA signature. A, B Univariate analysis and multivariate analysis in the training dataset. C, D Univariate analysis and multivariate analysis in the test dataset. E, F Univariate analysis and multivariate analysis in the combined dataset

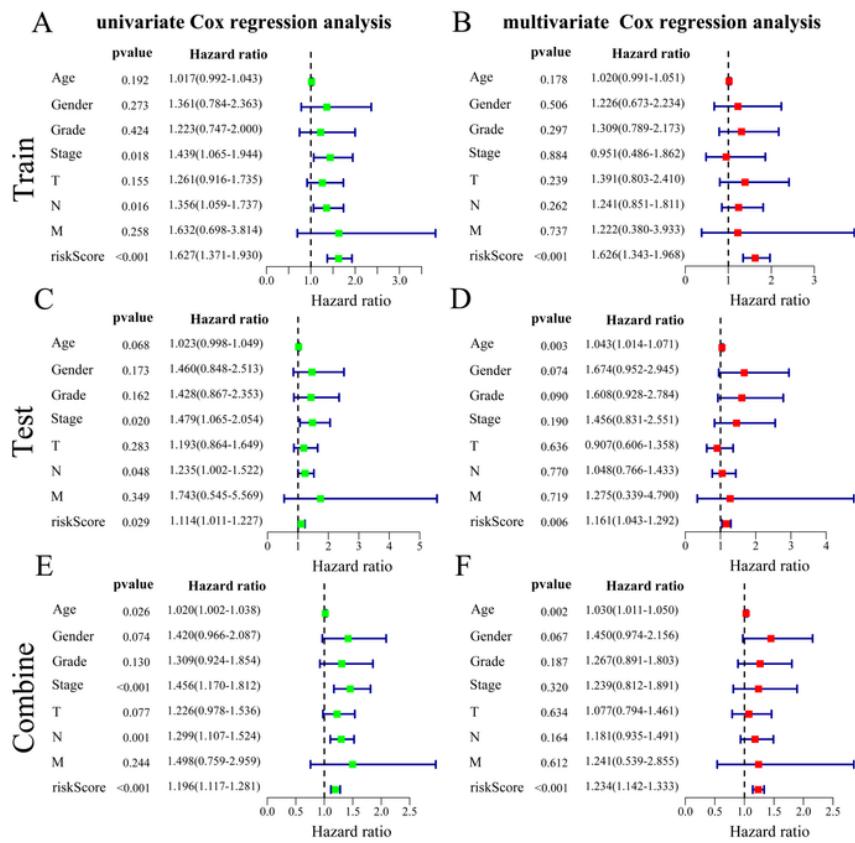


Figure 6

The nomogram model for the prediction of GC patient survival. A The nomogram model based on clinical factors and risk score. B Calibration curve of predicted probabilities. C ROC analysis of the predicted nomogram

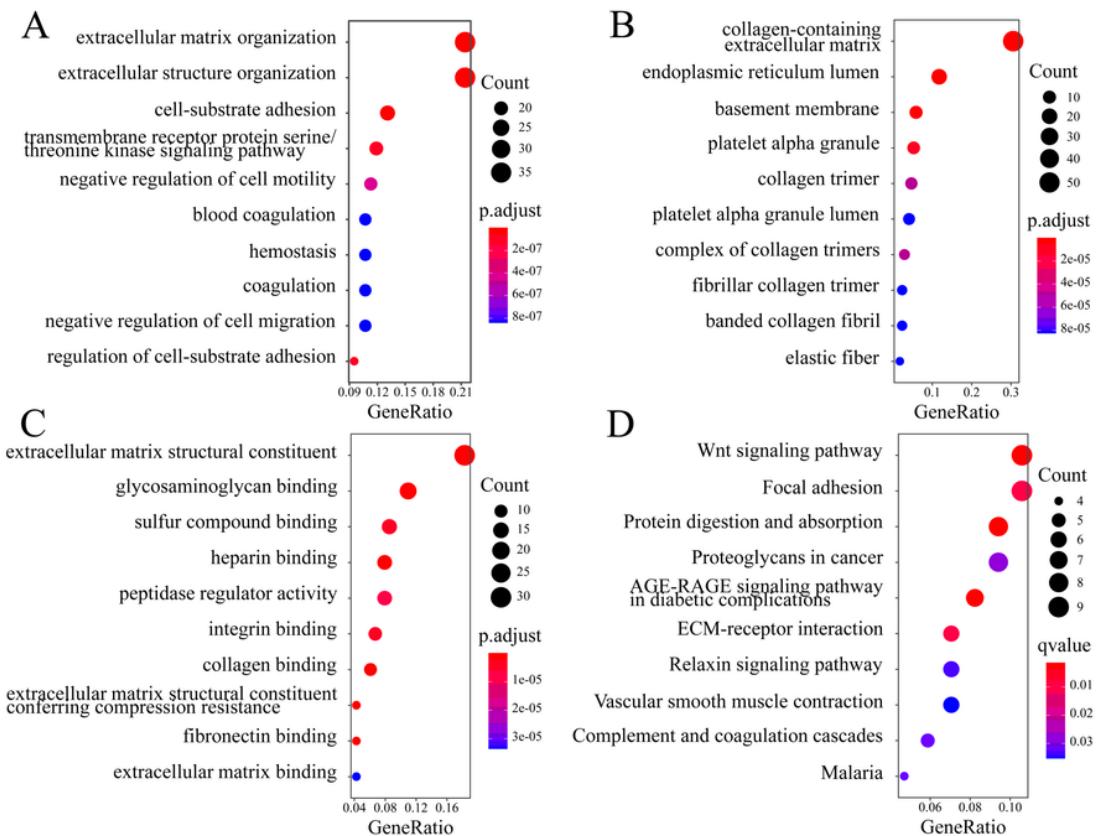


Figure 7

GO and KEGG enrichment analysis results. A biological process analysis. B Cellular component analysis. C Molecular function analysis. D KEGG pathway analysis

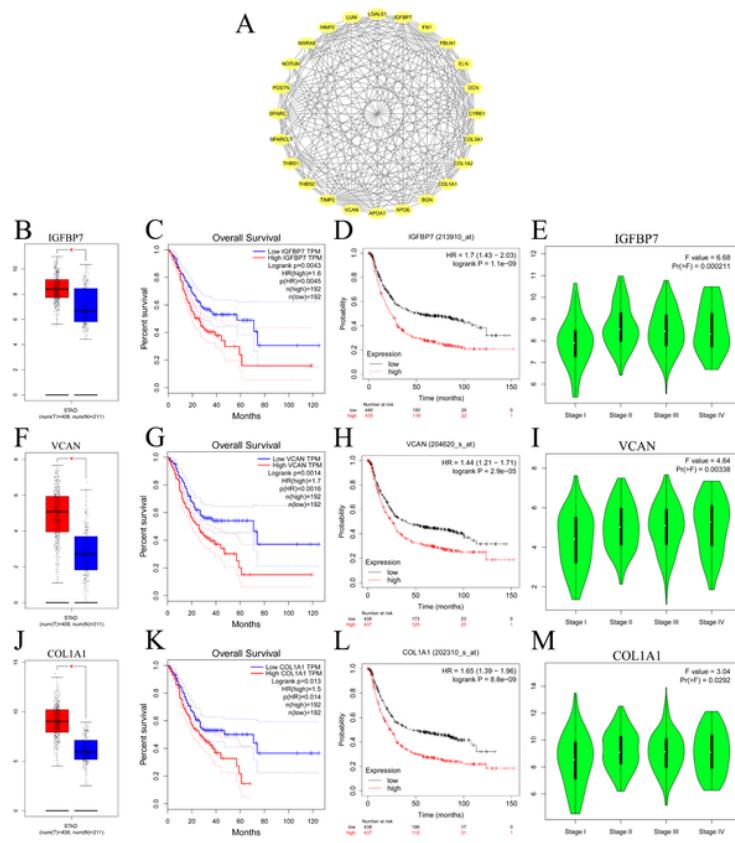


Figure 8

Establishment of the PPI network and validation of target genes. A The interaction of genes in PPI network. B , F , J Expression of IGFBP7, VCAN, COL1A1 in the GEPIA database. C , G , K) Analysis of survival according to IGFBP7, VCAN, and COL1A1 expression in the GEPIA database. D, H, L Analysis of survival according to IGFBP7, VCAN, and COL1A1 expression in the K-M plotter database. E, I, M Expression of IGFBP7, VCAN, and COL1A1 in different tumor stages according to the GEPIA database

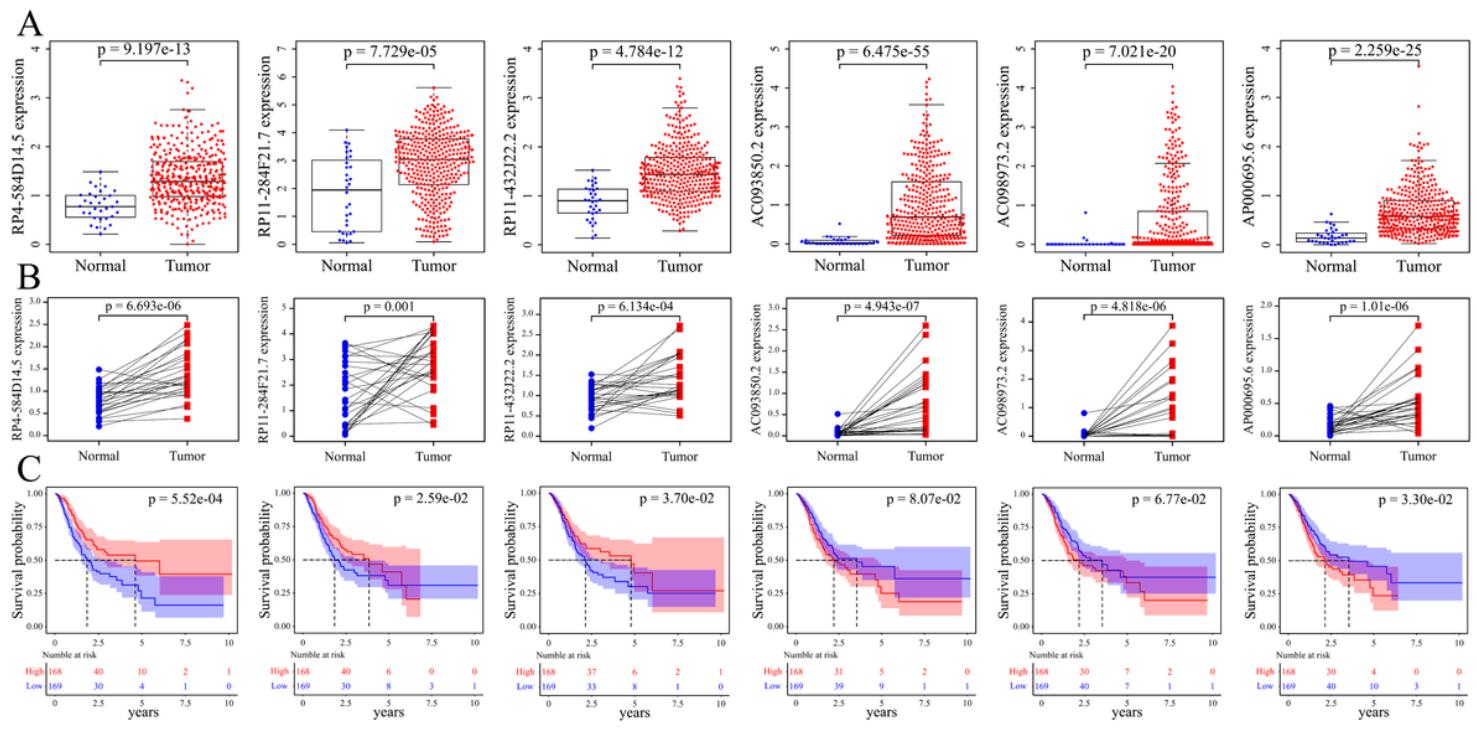


Figure 9

The expression and prognosis of the six lncRNAs in TCGA database. A The expression of six lncRNAs in cancer tissues and nontumor tissues. B The expression of six lncRNAs in 27 pairs of GC tissues and paracancerous tissues. C K-M analysis of survival according to the expression of the prognostic lncRNA

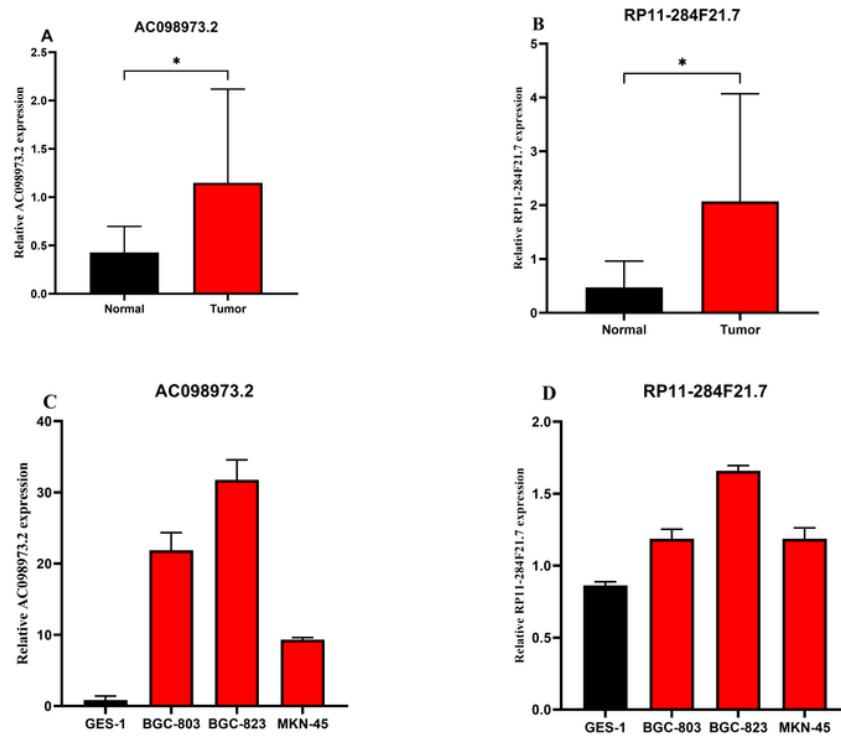


Figure 10

qRT-PCR results for the two lncRNAs in tissue samples and cells. (A-B) Relative expression of AC098973.2 and RP11-284F21.7 in 9 pairs of cancer tissues and paracancerous tissues ($P < 0.05$). (C-D) Relative expression of AC098973.2 and RP11-284F21.7 in GC cell lines and GES-1 ($P < 0.05$)