A risk score model with five long non-coding RNAs for predicting prognosis in gastric cancer: an integrated analysis combing TCGA and GEO dataset

Yiguo Wu [Equal first author, 1], **Junping Deng** [Equal first author, 2], **Shuhui Lai** [1], **Yujuan You** [Corresp., 3], **Jing Wu** [Corresp. 4]

[1] Department of Medicine, Nanchang University, Nan Chang, China

[2] Department of General Surgery, The First Affiliated Hospital of Nanchang University, Nan Chang, China

[3] Department of Anesthesiology, The Second Affiliated Hospital of Nanchang University, Nan Chang, China

[4] Shenzhen Prevention and Treatment Center for Occupational Diseases, Shen Zhen, China

Corresponding Authors: Yujuan You, Jing Wu
Email address: 506737972@qq.com, 446346807@qq.com

Background. Gastric cancer (GC) is one of the most common carcinomas of the digestive tract, and the prognosis for these patients may be poor. There are evidence that some long non-coding RNAs (lncRNAs) could predict the prognosis of gastric cancer. However, few lncRNA signatures have been used to predict the prognosis of the cancer. We herein aimed at constructing a risk score model combining with lncRNAs to predict the prognosis of gastric cancer and providing some new potential therapeutic targets in the future. **Methods.** We performed bayesian analysis and survival analysis to identify differential expressed lncRNAs that had significantly different survival times by using gastric cancer patient expression profile data from The Cancer Genome Atlas (TCGA). We then established a formula including five lncRNAs to predict prognosis in GC patients. In addition, to verified the prognostic value of this risk score model, two independent the Gene Expression Omnibus (GEO) datasets (GSE62254 (N=300) and GSE 15459 (N=200)) were employed to act as validation groups. **Results.** Based on the character of five-lncRNA, high or low risk subgroups can be divided among GC patients. The prognostic value of the five-lncRNA signature was confirmed in both TCGA dataset and the other two independent GEO datasets. Furthermore, stratification analysis found that the prognostic value of this risk model was independent in GC patients with Ⅱ-Ⅳ stage. Moreover, we constructed a nomogram model combining the clinical factors and five lncRNAs to heighten the accuracy of prognostic prediction. Enrichment analysis based on Kyoto Encyclopedia of Genes and Genomes (KEGG) suggested that five lncRNAs may be touched upon multiple cancer occurrence and progress-related pathways. **Conclusion.** Our results showed that the risk score model combining five-lncRNA signature predicts prognosis of GC patients well especially in stage II-IV and may provide potential therapeutic targets in

future.

A risk score model with five long non-coding RNAs for predicting prognosis in gastric cancer: an integrated analysis combing TCGA and GEO dataset

Yiguo Wu [Equal first author, 1], Junping Deng [Equal first author, 2], Shuhui Lai[1], Yujuan You[3], Jing Wu[4]
[1] Department of Medicine, Nanchang University, Nanchang City, Jiangxi Province, China
[2] Department of General Surgery, The First Affiliated Hospital of Nanchang University, Nanchang City, Jiangxi Province, China
[3]Department of Anesthesiology, The Second Affiliated Hospital of Nanchang University, Nanchang City, Jiangxi Province, China
[4]Department of Health Surveillance, Shenzhen Prevention and Treatment Center for Occupational Diseases, Shenzhen City, Guangdong Province, China

Corresponding Author:
Yujuan You
No. 1 Minde Road, Nanchang City, Jiangxi Province，330006, China
Email address: 506737972@qq.com
Jing Wu
No.2019 Buxin Road, Luohu District, Shenzhen City, Guangdong Province, 518020, China
Email address:446346807@qq.com

Abstract

Background. Gastric cancer (GC) is one of the most common carcinomas of the digestive tract, and the prognosis for these patients may be poor. There are evidence that some long non-coding RNAs (lncRNAs) could predict the prognosis of gastric cancer. However, few lncRNA signatures have been used to predict the prognosis of the cancer. We herein aimed at constructing a risk score model combining with lncRNAs to predict the prognosis of gastric cancer and providing some new potential therapeutic targets in the future.

Methods. We performed bayesian analysis and survival analysis to identify differential expressed lncRNAs that had significantly different survival times by using gastric cancer patient expression profile data from The Cancer Genome Atlas (TCGA). We then established a formula including five lncRNAs to predict prognosis in GC patients. In addition, to verified the prognostic value of this risk score model, two independent the Gene Expression Omnibus (GEO) datasets (GSE62254 (N=300) and GSE15459 (N=200)) were employed to act as validation groups.

Results. Based on the character of five-lncRNA, high or low risk subgroups can be divided among GC patients. The prognostic value of the five-lncRNA signature was confirmed in both TCGA dataset and the other two independent GEO datasets. Furthermore, stratification analysis

found that the prognostic value of this risk model was independent in GC patients with Ⅱ-Ⅳ stage. Moreover, we constructed a nomogram model combining the clinical factors and five lncRNAs to heighten the accuracy of prognostic prediction. Enrichment analysis based on Kyoto Encyclopedia of Genes and Genomes (KEGG) suggested that five lncRNAs may be touched upon multiple cancer occurrence and progress-related pathways.

Conclusion. Our results showed that the risk score model combining five-lncRNA signature predicts prognosis of GC patients well especially in stage II-IV and may provide potential therapeutic targets in future.

Introduction

Gastric cancer (GC) is one of the most common carcinomas of the digestive tract, and is especially prevalent in Asian countries. It is estimated that about 679,100 individuals in china were diagnosed with gastric cancer in 2015, and almost 498,000 died from the condition in that year (Saka et al., 2011; Chen et al., 2016).The standard therapies for gastric cancer are surgery and chemotherapy. However, most patients with advanced gastric cancer will have recurrence of the malignancy, and metastasis, after treatment, resulting in a poor prognosis. Despite considerable research into therapies for gastric cancer, the prospects of survival of GC patients remain bleak (Saka et al., 2011). The identification of gastric cancer patients with poor survival prognoses and the administration of effective treatment as early as possible are the keys to improving survival times. The investigation of potential therapeutic and prognostic biomarkers for gastric cancer is of considerable importance.

Long non-coding RNAs (lncRNAs) are RNAs of 200 nucleotides or more than with no or limited protein-coding potential. There is considerable evidence that lncRNAs play key roles in the initiation and developments of tumors. For example, lncRNA-ATB disorders have been shown to contribute to cancer cell proliferation, migration, invasion, and drug resistance in tumors and to prompt epithelial-mesenchymal transition (EMT) through competitive bounding to miRNAs(Li et al., 2017; Balas & Johnson, 2018). Some researchers have suggested that lncRNAs could act as new prognostic biomarkers in cancers. These potential biomarkers include CCAT2(Yu et al., 2017), HOXB-AS3(Huang et al., 2017) and ASLNC07322(Li et al., 2019) in colon cancer. A large number of lncRNAs closely related to the prognosis of gastric cancer have been identified, including MEG3(Wei & Wang, 2017), SNHG7(Wang et al., 2017), and DANCR(Mao et al., 2017). Risk score models have also been constructed to predict the prognosis of human tumors. In non-small cell lung cancer, differences in prognosis could be identified by their the 8-lncRNA signature (Miao et al., 2019). However, the identification of lncRNA related to prognosis in patients with gastric cancer remains in its early stages and additional research is necessary.

In this study, we analyzed data from 450 GC patients from The Cancer Genome Atlas (TCGA) database according to their risk score, in order to identify differentially expressed lncRNAs for the prediction of prognoses. Two independent Gene Expression Omnibus (GEO) datasets were employed to validate the selected-lncRNA. We explored the accuracy of prediction of five lncRNAs in different clinical subgroups, using the lncRNA data in combination with the clinical

characteristics of the patients. We constructed a nomogram model combining the clinical factors and five lncRNAs to increase the accuracy of prognostic prediction. Finally, we performed a pathway enrichment analysis to understand the potential functions of these lncRNAs in GC.

Materials & Methods

Preparation of GC datasets

We acquired a training dataset of gastric cancer samples from TCGA, comprised of 450 samples and 14147 LncRNAs (case: normal = 414:36). 450 samples were included to perform differential expression analysis. After that, excluding 6 cases with missing OS prognostic information, a total of 408 cases were recruited for further univariate Cox proportional hazards regression analysis and subsequent analysis in the training set. The microarray data for the validation set, and the survival data of the patients are publicly available at the GEO with accession numbers GSE62254 (N=300; 1397 LncRNAs) and GSE15459 (N=200; 1397 LncRNAs).

Normalization of GEO data

Because of the differented expression profiles of the two GEO datasets (GSE62254, and GSE15459), we performed quantile normalization on the original data and downloaded it as a probe-level CEL file. Affymetrix U133 Plus2.0 was used as the probe matching platform. We downloaded the data from Affymetrix website (http://www.affymetrix.com), and a total of 2986 lncRNA-specific probes were included.

Creation of an lncRNAs-based risk model from the test cohort

LncRNAs which were differentially expressed between GC and non-cancerous gastric tissue in the TCGA dataset were identified using Bayesian analysis using the limma R package of the R statistical computing environment ((log2|fold change| >1 and adjusted $P < 0.05$)), and the adjusted P was used to reduce false positives(Deng et al., 2019; Zeng et al., 2019). The candidate LncRNAs were analyzed using a univariate Cox proportional hazards regression analysis ($p<0.01$). The cutoff values of lncRNA expression were determined as the median of all expression values in the Cox survival analysis. We identified 278 lncRNAs with statistically significant differences. After identifying the lncRNAs common to both the TCGA and GEO(GSE62254) datasets, multivariate Cox hazards analyses was performed to identify independent prognostic lncRNAs. Finally, an lncRNAs-based risk model was created from a linear combination of the expression levels of these lncRNAs, multiplied by the regression coefficients obtained from the multivariate Cox hazard analyses.

Validation of the lncRNA-based model for prognostic prediction

We calculated the risk scores of each case, and the median score was used as the cutoff value to classify the patients into two risk score groups including high risk subgroup and low risk subgroup. Kaplan-Meier analysis was applied to differentiate between the survival of the two groups. Time-dependent receiver operating characteristic (ROC) curves were constructed to assess the model. Two GEO datasets were employed for validation of the lncRNAs-based model for the prediction of prognosis. Cox hazards analyses were conducted to estimate the hazard ratio

of this risk score model with a 95% confidence interval, To further evaluate the predictive value of the model for each clinic subgroup. The clinical subgroups were determined by gender, TNM stage, histologic grade, race, and age. Finally, a nomogram combining the risk score model with the clinical factors was constructed using the RMS package in R. We calculated the concordance index (C-index) and plotted a calibration curve to determine its predictive accuracy and discriminatory capacity.

Potential functions of the five lncRNAs

To understand the potential functions of the five lncRNAs, which appeared to be discriminatory, we performed linear regression analysis of the relationship between the lncRNAs and the protein-coding genes in the TCGA dataset. The screening criteria for the protein-coding genes was that these genes were positively associated with at least one lncRNA (Pearson coefficient > 0.4). After identifying the candidate genes, aberrantly activated signaling pathways were screened out using the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis Web-based Gene Set Analysis Toolkit (http://www.webgestalt.org/), one of the popular software tools for functional enrichment analysis related to KEGG pathways(Yang et al., 2019; Wang et al., 2013),.

Statistical analysis

R software (version 3.6.1) was used for the statistical analyses. Bayesian analysis was carried out using the limma R. Univariate and multivariate Cox proportional hazards regression analysis were conducted to identify the prognosis-related lncRNAs. The survival R package was used for Kaplan–Meier survival analysis. A time-dependent ROC curve was constructed to assess the specificity and sensitivity of the risk model. The Review Manager software (version 5.3) was used to plot the forest plot. Chi-square tests were applied for comparison of the rate of reoccurrence and death between the high-risk and low-risk. Pearson's linear regression analysis was used to explore the relationship between the lncRNAs and the protein-coding genes.

Results

Identification of five prognostic lncRNAs

After downloading the raw data from the TCGA database, the samples which included clinical and prognostic information were included in the study as the training cohort. We performed Bayesian analysis (log2|fold change| >1 and adjusted $P < 0.05$) and univariate Cox proportional hazard regression analysis ($p<0.01$) to identify survival-related lncRNAs. A total of 278 lncRNAs were further analyzed (Table S1). To validate the predictive accuracy, we intersected the lncRNAs selected from the TCGA database with the GEO validation set. Thirty-seven shared lncRNAs were found to be present in both the 278 lncRNAs and the validation dataset (GSE62254) (Table S1). After multivariable Cox proportional hazards regression analyses (Table S3)), we identified five lncRNAs as independent prognostic factors for gastric cancer: LINC00205, TRHDE-AS1, OVAAL, LINC00106, MIR100HG (Table 1). The expression of the five lncRNAs in gastric cancer patients was plotted as volcano and heat maps (Fig. 1A-B).

Survival curves were also plotted based on the overall survival (OS) and disease-free survival (DFS) of these 408 patients (Fig. 1C-D).

Creation of a lncRNAs-based risk model from the test cohort

According to the schematic workflow of the present study (Table 2), using the coefficients of five lncRNAs identified by multivariable Cox hazards analyses, we created a risk-score formula as follows: risk score = (0.249092 × expression level of LINC00205) + (0.182045 × expression level of TRHDE-AS1) + (0.271169 × expression level of OVAAL) + (−0.20794 × expression level of LINC00106) + (0.502539 × expression level of MIR100HG). Among the five lncRNAs, a negative coefficient indicates a protective factor, such as LINC00106. The remaining four lncRNAs with positive coefficients, LINC00205, TRHDE-AS1, OVAAL and MIR100HG, were risk factors. The risk scores of each patient in the test cohort were calculated (Table S2). The risk score in the TCGA ranged from -2.086959745 to 2.270305234. The patients in the test cohort were divided into two subgroups: high risk (n = 204); and low risk group (n = 204), with the median score (-0.001085) was used as the cut-off value. We performed Kaplan-Meier survival analysis to assess the effect of the lncRNAs-based model on the OS and DFS for GC in the test cohort (Fig. 2A-B). Our results indicated that the high-risk group had a significantly worse prognosis than the low risk group for both OS and DFS, and the P value were 1 × 10 ^ (-6) and 6 × 10 ^ (-6), respectively. The scatter plots for the death and recurrence incidence of GC patients are shown in (Fig. 2C-F). The rates of both death and recurrence for GC cases in the high risk group were significantly higher than low-risk group （$P < 0.001$）. Finally, in order to more accurately evaluate the prognostic value of the five lncRNAs signatures using the risk score model, we performed time-dependent ROC analysis using the 1-4 years cut-off of OS and the 1-2 years cut-off of DFS as the ROC ending points (Fig. 2G-H) (Fig. S1A-D). The area under the ROC curve (AUC) for the 4-year cut-off of OS and the 2-year cut-off of DFS was 0.734 and 0.692, respectively, and have a highest predictive value among those years, suggesting that this model was for the valuable prediction survival GC patients (Fig. 2G-H).

Validation of the lncRNAs-based model for prognostic prediction in independent cohorts

To assess the prognostic significance of this novel prognostic model involving five signatures in GC patients, we used the other two independent validation sets from the GEO database. We calculated the risk score using the formula given above. The GC patients in the GSE62254 (validation group-1, n = 300) and GSE15459 (validation group-2, n = 200) datasets were divided into high-risk and low-risk groups as well. Because of lack of DFS data in GSE15459, we only calculated the OS of the patients. The cases in the high-risk validation subgroups had a poorer OS than those in the low-risk group (log-rank test P = 0.009 and 0.02, respectively) (Fig. 3A-B). The scatter plots for death events are shown in (Fig. 3C-D). The incidence of death and reoccurrence in GC patients in the high risk group was significantly higher than low risk group （$P < 0.001$）. The AUC of the two validation cohort in four-year cut-off OS was 0.622 and 0.610, respectively (Fig. 3E-F). The ROC curve of the two validation cohort in 1-3 year cut-off OS was showed in (Fig. S3A-F). Furthermore, we verified the risk model in DFS of the

GSE62254 dataset (Fig. S2A-D). Our results further confirmed the value of this risk score model.

The lncRNAs-based model had a favorable prognostic prediction in stage II, stage III, and stage IV patients

To further investigate the performance of the lncRNAs-based model, stratified Kaplan-Meier survival analysis for OS in the training group was performed. This analysis was based on the AJCC TNM stage: I, II, III, or IV (Fig. 4A-D). The five-lncRNA signature showed good predictive value for OS in subgroups stage II (P = 0.008), stage III (P = 0.02)and stage IV (P = 0.01). Otherwise, but not for stage I (P = 0.3).

To estimate the hazard ratio of each subgroup of patients as defined by gender, TNM stage, histologic grade, race and age(> =or < fifty years) (Table 3), the risk score model was used to divide the patients into two risk groups using median cut-off value. Forest plots are shown in (Fig. 5). The risk score model involving the five-lncRNA signature had a relatively good prognostic value in the clinic subgroups of gender, histologic grade and age (> =or < fifty years). To improve the prognostic value of this model, we combined the clinical factors with the risk score model to construct a nomogram model to predict prognosis. The nomogram model and nomogram calibration curve are shown in (Fig. 6A-B). To evaluate the effect of the nomogram model, we also calculated C-index. The C-index for predicting the four year OS of GC patients was 0.69668, indicating that it is valuable for predicting prognosis.

Potential functions of the five lncRNAs

In order to investigate the functions of the five lncRNAs in GC, we calculated the Pearson correlations between the five lncRNA signatures and 19605 protein-coding genes in the TCGA dataset. A total of 3069 genes (Table S4) were positively correlated with at least one lncRNA (Pearson's coefficient> 0.4) (Fig. 7A), and were further selected for KEGG pathway enrichment analyses. Ranked by –logP value (Q value), we selected the top 10 pathways to draw bubble plot (Fig. 7B) (Zeng et al., 2019; Deng et al., 2019). For biological processes, the co-expressed genes were mainly enriched in pathways involved in cancer, such as the Focal adhesion pathway, the cGMP−PKG signaling pathway and Calcium signaling pathway. This finding indicates that these five lncRNAs may be related to the regulation of the initiation and progress of tumors.

Discussion

In this study, we identified a potential signature involving five lncRNAs which are differentially expressed in tumor and normal tissues, and which may be valuable for the prediction of prognosis in GC. The prognostic performance of the risk score model involving the five lncRNAs was verified by both the TCGA dataset and GEO datasets. Stratified analysis suggested that the risk score model was valuable for the prediction of prognosis in GC patients with stage II to IV. To enhance the predictive accuracy of the model, we combined clinical parameters with the five-lncRNA signature to construct a nomogram model and confirmed its performance using a calibration curve and C index.

Gastric cancer is a common malignancy in the digestive system (Siegel et al., 2019). Despite continuous improvement in treatment, the five-year survival rate of patients with advanced gastric cancer still hovers at 20% (Min et al., 2019; Misawa et al., 2019). Therefore, early diagnosis, early identification of high-risk patients and the implementation of effective treatment measures as early as possible are key to improving survival times. It is also important to develop novel prognostic indicators for GC. Over the past few decades, a large amount of research evidence has showed that protein-encoding genes(Ghoorun et al., 2019; Luo et al., 2019) and microRNAs(Li et al., 2020; Zhou et al., 2019), play vital roles in the occurrence and development of various tumors, and could predict the prognosis as well. A number of nomogram models involving clinical factors have been constructed to predict the prognosis of GC. For example, Yue et al (Yu & Zhang, 2019)used tumor size and tumor site, as independent prognostic factors, to construct OS nomograms for predicting outcome in GC patients, and the C-index for this model indicated that the model was able to predict the prognosis of GC patients in OS. Recently, more lncRNAs related to the prognosis of gastric cancer have been discovered, but prognostic prediction models involving lncRNAs still lack a unified conclusion so far. We present a nomogram including clinical factors and a five-lncRNA signature which may be of value for the prediction of prognosis in GC patients.

As a result, it is urgent to explore new biomarkers to improve the assessment of diagnosis and prognosis of GC patients due to the limitations of the AJCC TNM staging system and some related socring systems. Many lncRNAs have been identified, of which only a small proportion has been functionally annotated recently. However, there is evidence to indicate that lncRNAs, acting either as carcinogenes or tumor suppressors, participate in the tumorigenesis and development of various tumors by regulating the processes of chromatin remodeling, transcription and post-transcriptional modification(Bartonicek et al., 2016; Iyer et al., 2015), and therefore may be valuable for tumor diagnosis and prognosis. Some studies have found that gastric cancer-related lncRNAs are involved in biological behaviors such as the proliferation, migration, invasion, and autophagy of gastric cancer cells, affecting the initiation and prognosis of GC (Mao et al., 2017; Wei & Wang, 2017). For example, lncRNA MEG3 appears to inhibit the proliferation, metastasis and prognosis of GC through up regulation of the p53 expression, a key tumor suppressor (Wei & Wang, 2017). We identified five lncRNAs lncRNAs(LINC00205, TRHDE-AS1, OVAAL, LINC00106, and MIR100HG) as predictors of GC prognosis, and developed a risk score model. Kaplan-Meier analysis suggested that this model is valuable for predicting prognosis in GC patients. We used two independent GEO datasets as validation datasets. Our results confirmed that the risk score model was stable and performed well in predicting the prognosis of GC.

Among the five lncRNAs, including LINC00205, TRHDE-AS1, OVAAL and MIR100HG, acted as risk factors for GC patients, otherwise, the LINC00106 was a protective factor. Except for LINC00205 and MIR100HG, the other three lncRNAs have been less reported in the literatures. Furthermore, except for LINC00106. In this study, LINC00205, TRHDE-AS1, OVAAL and MIR100HG were identified as potentially prognostic biomarkers in GC for the first

time. Consistent with our result, it has previously been reported that high expression of LINC00106 is indicative of prolonged overall survival in GC (Qi et al., 2020). Nevertheless, the function of this lncRNA in gastric cancer and its specific mechanism needs further study. Interestingly, in hepatocellular carcinoma (HCC), the expression of LINC00205, a tumor suppressor, has been positively associated with OS and recurrence-free survival by a comprehensive genome-wide analysis (Cui et al., 2017). A study showed that, as a competing endogenous RNA with lower expression level levels in tumor tissues, LINC00205 may negatively regulate the progression of HCC via the miR-184/EPHX1 axis (Long et al., 2019), While another research has indicated that LINC00205, can act as a oncogene, promoting proliferation, migration and invasion of HCC cells by targeting miR-122-5p(Zhang et al., 2019). Moreover, LINC00205 appeared to act as a protective factor in pancreatic cancer survival [HR = 0.58, p (Log rank) = 0.0091](Giulietti et al., 2018). The reported role and therefore value for prognostic prediction of LINC00205 in various cancers shows significant discrepancies. These discrepancies might be associated with the specificities of different tumors. It has been reported that up-regulation of the TRHDE-AS1 inhibits the growth of lung carcinoma through competitive combination with miRNA-103/KLF4 axis(Zhuan et al., 2019). One study found that OVVAL is highly expressed in colon cancer and melanoma, and further experimental results showed that OVAAL promotes the proliferation of cancer cells via dual mechanisms controlling RAF/MEK/ERK signaling and p27-mediated cell senescence(Sang et al., 2018). The lncRNA MIR100HG has been studied as a oncogene in acute megakaryoblastic leukemia(Emmrich et al., 2014), laryngeal squamous cell carcinoma(Huang et al., 2019), and for its role in mediating cetuximab resistance via Wnt/β-catenin signaling(Lu et al., 2017) in colorectal cancer. Although the roles of these lncRNAs in cancer need further elucidate, our results may provide a novel approach for the study of gastric cancer.

To further investigate the functions of the five lncRNAs in gastric cancer, we performed a pathway enrichment analysis. The pathways in which the genes are enriched are involved in regulation of cancer, including the pathway in cancer, cGMP−PKG signaling pathway, Calcium signaling pathway, and Focal adhesion pathway etc. This finding suggests that the five lncRNAs may play important roles in tumor occurrence and development in GC patients. There is evidence that lncRNA can promote tumorigenesis through the cGMP−PKG signaling pathway. For example, overexpression of SRRM2-AS accelerated angiogenesis in nasopharyngeal carcinoma via cGMP-PKG signaling pathway(Chen et al., 2019). It has been reported the Calcium signaling pathway mainly involved in metabolic diseases and heart diseases over the past years(Berridge, 2016; Dewenter et al., 2017). A latest research showed that Calcium signaling pathway was associated with cancer cell survival, but more details on how it affects are still to be studied(Reczek & Chandel, 2018). Focal adhesion are special sites where integrin receptors aggregated in cells interact with extracellular matrix and intracellular actin skeleton(Burridge, 2017), and it plays a critical role in tumor invasion and migration(Shen et al., 2018). There are evidence that knock down of the Linc01060 could promote the pancreatic cancer progression via Vinculin-Mediated Focal Adhesion pathway Turnover(Shi et al., 2018).

However, whether the LncRNA can mediate the progress of GC through Focal Adhesion pathway is less reported. In short, lncRNA may participate in the genesis and development of various tumors through the above pathways.

A previous study reported a 24-lncRNA signature could predict outcome of GC patients. The signature was identified using lncRNA expression profiles of GC from GEO (Zhu et al., 2016), and it provide a new perspective on the identification of novel potential targets treatment of GC. However, due to the limitated amount of data in the GEO dataset, the lncRNAs identified in this study may not represent the complete population of lncRNAs underlying GC. In our study, we took full advantage of TCGA and GEO data to comprehensively investigate potentially prognostic lncRNAs. To evaluate predictive performance of this five-lncRNA signature, we determined the end point of a ROC curve based on the cut-off value of the OS and DFS curve rather than the survival outcome, which may be able to more accurately evaluate the performance of the model. In order to improve the accuracy of the five-LncRNA prognosis model, we combined it with clinically relevant prognostic factors to develop a nomogram model which could predict OS of GC patients.

In our study, we identified the prognostic lncRNAs by mining the expression profiles available online. The raw data of GEO was standardized using Affymetrix U133 platform which is a common commercial probe technology. Then, we integrated the data both the TCGA and GEO datasets to draw a conclusion. Integrated analysis has been proved to be an effective approach for multiple datasets with different platforms and detection times using R package(Zhang et al., 2019), and is widely applied for bioinformatic analysis(Li et al., 2018). Integration of several available datasets to improve the number of the cases could promote the results reliability(Ma et al., 2017). However, because the characteristics of patients in different data sets may be inconsistent, although we did not analyze it, such as the different TNM stage, the different distribution of age, and different races, this might inevitably lead to a bias in conclusion. Moreover, owing to dataset in the TCGA has much more lncRNAs than the GEO, intersection of different datasets may omit potential prognostic lncRNAs inevitably. We took full advantage of TCGA and GEO data to comprehensively investigate potentially prognostic lncRNAs in general. Secondly, because of the lack of DFS data in one GEO validation group, we used only one validation cohorts to verify the prognostic value of the five-lncRNA signature for the DFS of the GC patients. Thirdly, due to the limitated amount of this tumor data and researches about these lncRNAs so far, experimental research into these lncRNAs is highly needed to further understand these functions in GC in the future.

Conclusions

We established a risk score model including five lncRNAs to predict GC patients' OS and DFS, particularly in those with Ⅱ-Ⅳ stage. Our findings also provided evidence of developing effective prognostic biomarkers for GC patients and potential therapeutic targets in the future.

Acknowledgements

At the point of finishing this paper, we would thank to Jie Peng for technical assistance with the data analysis.

References

Balas, M.M., and Johnson, A.M. 2018. Exploring the mechanisms behind long noncoding RNAs and cancer. *Noncoding RNA Res* 3:108-117. 10.1016/j.ncrna.2018.03.001

Bartonicek, N., Maag, J.L., and Dinger, M.E. 2016. Long noncoding RNAs in cancer: mechanisms of action and technological advancements. *Molecular Cancer* 15:43. 10.1186/s12943-016-0530-6

Berridge, M.J. 2016. The Inositol Trisphosphate/Calcium Signaling Pathway in Health and Disease. *PHYSIOLOGICAL REVIEWS* 96:1261-1296. 10.1152/physrev.00006.2016

Burridge, K. 2017. Focal adhesions: a personal perspective on a half century of progress. *FEBS Journal* 284:3355-3361. 10.1111/febs.14195

Chen, S., Lv, L., Zhan, Z., Wang, X., You, Z., Luo, X., and You, H. 2019. Silencing of long noncoding RNA SRRM2-AS exerts suppressive effects on angiogenesis in nasopharyngeal carcinoma via activating MYLK-mediated cGMP-PKG signaling pathway. *JOURNAL OF CELLULAR PHYSIOLOGY*. 10.1002/jcp.29382

Chen, W., Zheng, R., Baade, P.D., Zhang, S., Zeng, H., Bray, F., Jemal, A., Yu, X.Q., and He, J. 2016. Cancer statistics in China, 2015. *CA: A Cancer Journal for Clinicians* 66:115-132. 10.3322/caac.21338

Cui, H., Zhang, Y., Zhang, Q., Chen, W., Zhao, H., and Liang, J. 2017. A comprehensive genome-wide analysis of long noncoding RNA expression profile in hepatocellular carcinoma. *Cancer Med* 6:2932-2941. 10.1002/cam4.1180

Deng, J., Xu, Y., and Wang, G. 2019. Identification of Potential Crucial Genes and Key Pathways in Breast Cancer Using Bioinformatic Analysis. *Frontiers in Genetics* 10. 10.3389/fgene.2019.00695

Dewenter, M., von der Lieth, A., Katus, H.A., and Backs, J. 2017. Calcium Signaling and Transcriptional Regulation in Cardiomyocytes. *CIRCULATION RESEARCH* 121:1000-1020. 10.1161/CIRCRESAHA.117.310355

Emmrich, S., Streltsov, A., Schmidt, F., Thangapandi, V.R., Reinhardt, D., and Klusmann, J.H. 2014. LincRNAs MONC and MIR100HG act as oncogenes in acute megakaryoblastic leukemia. *Molecular Cancer* 13:171. 10.1186/1476-4598-13-171

Ghoorun, R.A., Wu, X.H., Chen, H.L., Ren, D.L., and Wu, X.B. 2019. Prognostic Significance of FKBP14 in Gastric Cancer. *Onco Targets Ther* 12:11567-11577. 10.2147/OTT.S221943

Giulietti, M., Righetti, A., Principato, G., and Piva, F. 2018. LncRNA co-expression network analysis reveals novel biomarkers for pancreatic cancer. *CARCINOGENESIS* 39:1016-1025. 10.1093/carcin/bgy069

Huang, J.Z., Chen, M., Chen, Gao, X.C., Zhu, S., Huang, H., Hu, M., Zhu, H., and Yan, G.R. 2017. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *MOLECULAR CELL* 68:171-184. 10.1016/j.molcel.2017.09.015

Huang, Y., Zhang, C., and Zhou, Y. 2019. LncRNA MIR100HG promotes cancer cell proliferation, migration and invasion in laryngeal squamous cell carcinoma through the downregulation of miR-204-5p. *Onco Targets Ther* 12:2967-2973. 10.2147/OTT.S202528

Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S., Poliakov, A., Cao, X., Dhanasekaran, S.M., Wu, Y.M., Robinson, D.R., Beer, D.G., Feng, F.Y., Iyer, H.K., and Chinnaiyan, A.M. 2015. The landscape of long noncoding RNAs in the human transcriptome. *NATURE GENETICS* 47:199-208. 10.1038/ng.3192

Li, J., Li, Z., Zheng, W., Li, X., Wang, Z., Cui, Y., and Jiang, X. 2017. LncRNA-ATB: An indispensable cancer-related long noncoding RNA. *Cell Prolif* 50. 10.1111/cpr.12381

Li, J.P., Zhang, H.M., Liu, M.J., Xiang, Y., Li, H., Huang, F., Li, H.H., Dai, Z.T., Gu, C.J., Liao, X.H., and Zhang, T.C. 2020. miR-133a-3p/FOXP3 axis regulates cell proliferation and autophagy in gastric cancer. *JOURNAL OF CELLULAR BIOCHEMISTRY*. 10.1002/jcb.29613

Li, X., Lv, X., Li, Z., Li, C., Li, X., Xiao, J., Liu, B., Yang, H., and Zhang, Y. 2019. Long Noncoding RNA ASLNC07322 Functions in VEGF-C Expression Regulated by Smad4 during Colon Cancer Metastasis. *Mol Ther Nucleic Acids* 18:851-862. 10.1016/j.omtn.2019.10.012

Li, Y., Gu, J., Xu, F., Zhu, Q., Ge, D., and Lu, C. 2018. Transcriptomic and functional network features of lung squamous cell carcinoma through integrative analysis of GEO and TCGA data. *Scientific Reports* 8. 10.1038/s41598-018-34160-w

Long, X., Li, Q., Zhi, L.J., Li, J.M., and Wang, Z.Y. 2019. LINC00205 modulates the expression of EPHX1 through the inhibition of miR‐184 in hepatocellular carcinoma as a ceRNA. *JOURNAL OF CELLULAR PHYSIOLOGY* 235:3013-3021. 10.1002/jcp.29206

Lu, Y., Zhao, X., Liu, Q., Li, C., Graves-Deal, R., Cao, Z., Singh, B., Franklin, J.L., Wang, J., Hu, H., Wei, T., Yang, M., Yeatman, T.J., Lee, E., Saito-Diaz, K., Hinger, S., Patton, J.G., Chung, C.H., Emmrich, S., Klusmann, J.H., Fan, D., and Coffey, R.J. 2017. lncRNA MIR100HG-derived miR-100 and miR-125b mediate cetuximab

resistance via Wnt/beta-catenin signaling. *NATURE MEDICINE* 23:1331-1341. 10.1038/nm.4424
Luo, X., He, Y., Tang, H., Cao, Y., Gao, M., Liu, B., and Hu, Z. 2019. Effects of HER2 on the invasion and migration of gastric cancer. *American Journal of Translational Research* 11:7604-7613.
Ma, T., Liang, F., Oesterreich, S., and Tseng, G.C. 2017. A Joint Bayesian Model for Integrating Microarray and RNA Sequencing Transcriptomic Data. *JOURNAL OF COMPUTATIONAL BIOLOGY* 24:647-662. 10.1089/cmb.2017.0056
Mao, Z., Li, H., Du B, Cui, K., Xing, Y., Zhao, X., and Zai, S. 2017. LncRNA DANCR promotes migration and invasion through suppression of lncRNA-LET in gastric cancer cells. *Biosci Rep* 37. 10.1042/BSR20171070
Miao, R., Ge, C., Zhang, X., He, Y., Ma, X., Xiang, X., Gu, J., Fu, Y., Qu, K., Liu, C., Wu, Q., and Lin, T. 2019. Combined eight-long noncoding RNA signature: a new risk score predicting prognosis in elderly non-small cell lung cancer patients. *Aging (Albany NY)* 11:467-479. 10.18632/aging.101752
Min, S.H., Won, Y., Kim, G., Lee, Y., Park, Y.S., Ahn, S.H., Park, D.J., and Kim, H.H. 2019. 15-year experience of laparoscopic gastrectomy in advanced gastric cancer: analysis on short-term and long-term oncologic outcome. *SURGICAL ENDOSCOPY AND OTHER INTERVENTIONAL TECHNIQUES*. 10.1007/s00464-019-07292-x
Misawa, K., Mochizuki, Y., Sakai, M., Teramoto, H., Morimoto, D., Nakayama, H., Tanaka, N., Matsui, T., Ito, Y., Ito, S., Tanaka, K., Uemura, K., Morita, S., and Kodera, Y. 2019. Randomized clinical trial of extensive intraoperative peritoneal lavage versus standard treatment for resectable advanced gastric cancer (CCOG 1102 trial). *Br J Surg* 106:1602-1610. 10.1002/bjs.11303
Qi, M., Yu, B., Yu, H., and Li, F. 2020. Integrated analysis of a ceRNA network reveals potential prognostic lncRNAs in gastric cancer. *Cancer Medicine*. 10.1002/cam4.2760
Reczek, C.R., and Chandel, N.S. 2018. ROS Promotes Cancer Cell Survival through Calcium Signaling. *CANCER CELL* 33:949-951. 10.1016/j.ccell.2018.05.010
Saka, M., Morita, S., Fukagawa, T., and Katai, H. 2011. Present and future status of gastric cancer surgery. *JAPANESE JOURNAL OF CLINICAL ONCOLOGY* 41:307-313. 10.1093/jjco/hyq240
Sang, B., Zhang, Y.Y., Guo, S.T., Kong, L.F., Cheng, Q., Liu, G.Z., Thorne, R.F., Zhang, X.D., Jin, L., and Wu, M. 2018. Dual functions for OVAAL in initiation of RAF/MEK/ERK prosurvival signals and evasion of p27-mediated cellular senescence. *Proceedings of the National Academy of Sciences* 115:E11661-E11670. 10.1073/pnas.1805950115
Shen, J., Cao, B., Wang, Y., Ma, C., Zeng, Z., Liu, L., Li, X., Tao, D., Gong, J., and Xie, D. 2018. Hippo component YAP promotes focal adhesion and tumour aggressiveness via transcriptionally activating THBS1/FAK signalling in breast cancer. *J Exp Clin Cancer Res* 37:175. 10.1186/s13046-018-0850-z
Shi, X., Guo, X., Li, X., Wang, M., and Qin, R. 2018. Loss of Linc01060 induces pancreatic cancer progression through vinculin-mediated focal adhesion turnover. *CANCER LETTERS* 433:76-85. 10.1016/j.canlet.2018.06.015
Siegel, R.L., Miller, K.D., and Jemal, A. 2019. Cancer statistics, 2019. *CA Cancer J Clin* 69:7-34. 10.3322/caac.21551
Wang, J., Duncan, D., Shi, Z., and Zhang, B. 2013. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *NUCLEIC ACIDS RESEARCH* 41:W77-W83. 10.1093/nar/gkt439
Wang, M.W., Liu, J., Liu, Q., Xu, Q.H., Li, T.F., Jin, S., and Xia, T.S. 2017. LncRNA SNHG7 promotes the proliferation and inhibits apoptosis of gastric cancer cells by repressing the P15 and P16 expression. *Eur Rev Med Pharmacol Sci* 21:4613-4622.
Wei, G.H., and Wang, X. 2017. lncRNA MEG3 inhibit proliferation and metastasis of gastric cancer via p53 signaling pathway. *Eur Rev Med Pharmacol Sci* 21:3850-3856.
Yang, M., Qin, X., Qin, G., and Zheng, X. 2019. The role of IRAK1 in breast cancer patients treated with neoadjuvant chemotherapy. *Onco Targets Ther* 12:2171-2180. 10.2147/OTT.S185662
Yu, C., and Zhang, Y. 2019. Development and validation of prognostic nomogram for young patients with gastric cancer. *Annals of Translational Medicine* 7:641. 10.21037/atm.2019.10.77
Yu, Y., Nangia-Makker, P., Farhana, L., and Majumdar, A. 2017. A novel mechanism of lncRNA and miRNA interaction: CCAT2 regulates miR-145 expression by suppressing its maturation process in colon cancer cells. *Molecular Cancer* 16:155. 10.1186/s12943-017-0725-5
Zeng, J., Lu, W., Liang, L., Chen, G., Lan, H., Liang, X., and Zhu, X. 2019. Prognosis of clear cell renal cell carcinoma (ccRCC) based on a six-lncRNA-based risk score: an investigation based on RNA-sequencing data. *Journal of Translational Medicine* 17. 10.1186/s12967-019-2032-y
Zhang, L., Wang, Y., Sun, J., Ma, H., and Guo, C. 2019. LINC00205 promotes proliferation, migration and invasion of HCC cells by targeting miR-122-5p. *Pathology - Research and Practice* 215:152515. 10.1016/j.prp.2019.152515
Zhang, X., Zhang, W., Jiang, Y., Liu, K., Ran, L., and Song, F. 2019. Identification of functional lncRNAs in gastric cancer by integrative analysis of GEO and TCGA data. *JOURNAL OF CELLULAR BIOCHEMISTRY* 120:17898-

17911. 10.1002/jcb.29058
Zhou, H.Y., Wu, C.Q., and Bi, E.X. 2019. MiR-96-5p inhibition induces cell apoptosis in gastric adenocarcinoma. *World J Gastroenterol* 25:6823-6834. 10.3748/wjg.v25.i47.6823
Zhu, X., Tian, X., Yu, C., Shen, C., Yan, T., Hong, J., Wang, Z., Fang, J., and Chen, H. 2016. A long non-coding RNA signature to improve prognosis prediction of gastric cancer. *Molecular Cancer* 15. 10.1186/s12943-016-0544-0
Zhuan, B., Lu, Y., Chen, Q., Zhao, X., Li, P., Yuan, Q., and Yang, Z. 2019. Overexpression of the long noncoding RNA TRHDE‐AS1 inhibits the progression of lung cancer via the miRNA‐103/KLF4 axis. *JOURNAL OF CELLULAR BIOCHEMISTRY*. 10.1002/jcb.29029

Figure 1

The expression information of five lncRNAs, overall survival and disease free survival in gastric cancer patients in the TCGA dataset.

(A) Volcano plot with blue dots indicating five lncRNAs expression levels which is significantly different between tumor and normal tissue based on the criteria of an absolute log2 fold change (FC)>1 and adjusted $P < 0.05$. (B) Heatmap of the five-lncRNA expression profile of the 414 patients in the TCGA dataset. Among five lncRNAs, MIR100HG and TRHDE-AS1 have a similar expression in 414 patients in the TCGA dataset, otherwise the other three lncRNAs do as well. (C-D) Kaplan-Meier analysis of patients' overall survival and disease-free survival in the TCGA dataset.

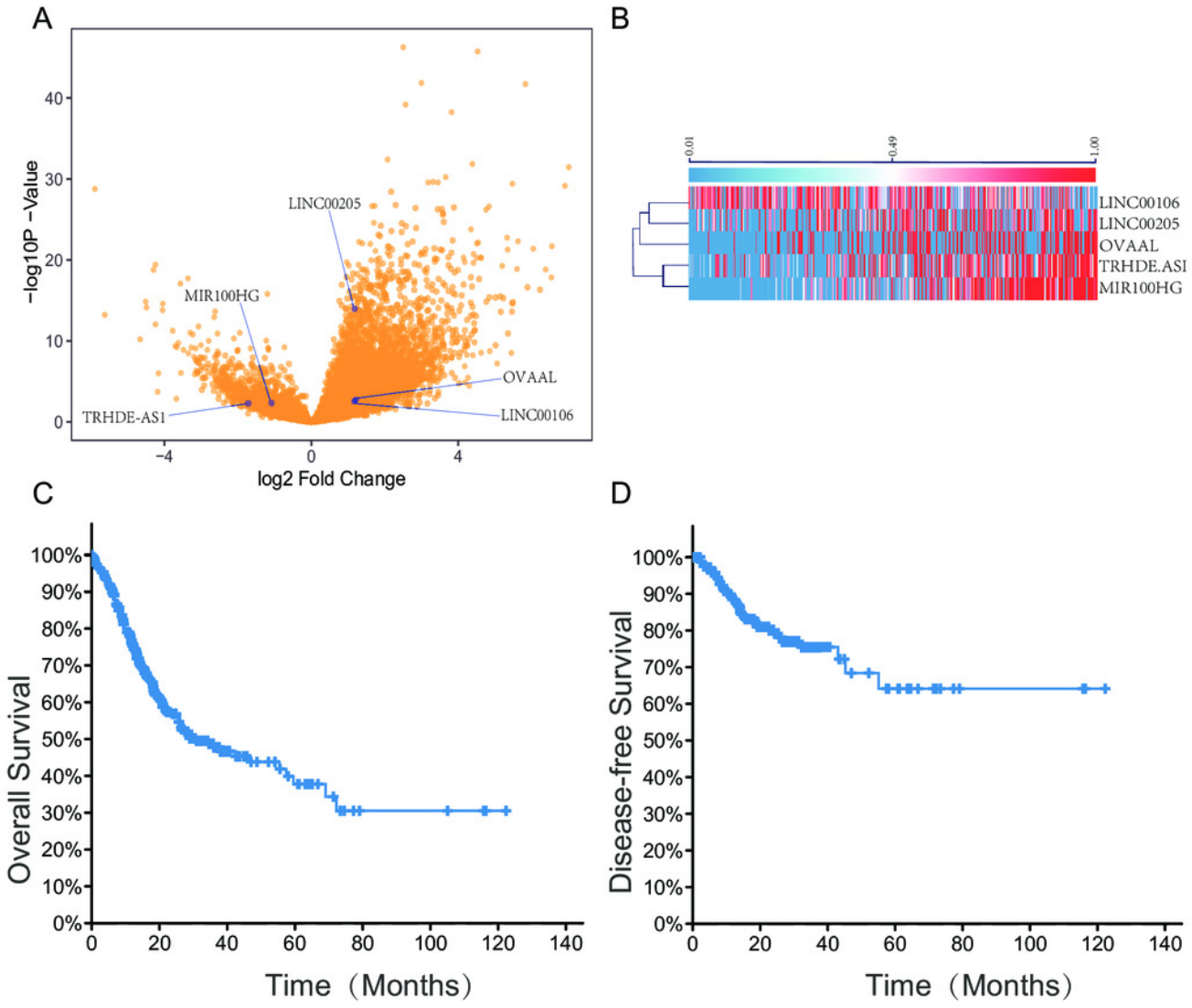

A
−log10P −Value
log2 Fold Change
LINC00205
MIR100HG
OVAAL
TRHDE-AS1
LINC00106
B
LINC00106
LINC00205
OVAAL
TRHDE.ASI
MIR100HG
C
Overall Survival
Time（Months）
D
Disease-free Survival
Time（Months）

Figure 2

The prognostic value of five-lncRNA signature in training group.

The Prognostic value of five-lncRNA signature in training group. (A-B) Kaplan-Meier analysis of patients' overall survival and disease-free survival in the high-risk (n = 204) and low-risk (n = 204) subgroups of the training set. (C)The scatter plot of five-lncRNA-based risk score distribution for patient survival status. (D) The percentage of patient survival status in the high-risk and low-risk subgroups of the training set. (E) The five-lncRNA-based risk score distribution for patient recurrence. (F) The percentage of patient recurrence in the high-risk and low-risk subgroups of the training set. (G-H) The time-dependent ROC analysis of the risk score for prediction the 4-year cut-off OS and 2-year cut-off DFS of the training set. The area under the curve was calculated for ROC curves. ***P<0.001.

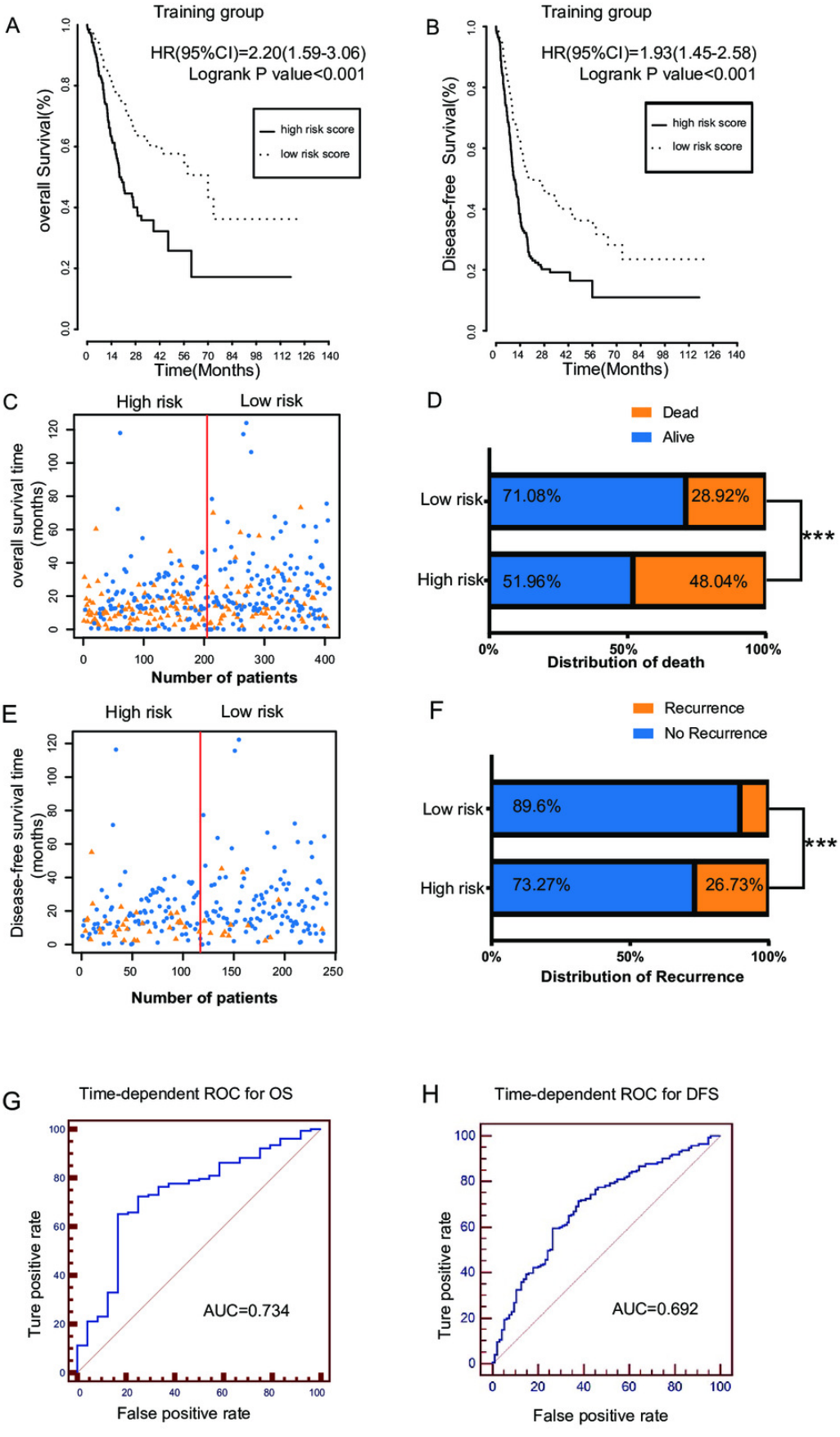

A
Training group
HR(95%CI)=2.20(1.59-3.06)
Logrank P value<0.001
high risk score
low risk score
overall Survival(%)
Time(Months)
B
Training group
HR(95%CI)=1.93(1.45-2.58)
Logrank P value<0.001
high risk score
low risk score
Disease-free Survival(%)
Time(Months)
C
High risk
Low risk
overall survival time (months)
Number of patients
D
Dead
Alive
Low risk
71.08%
28.92%
High risk
51.96%
48.04%
***
0%
50%
100%
Distribution of death
E
High risk
Low risk
Disease-free survival time (months)
Number of patients
F
Recurrence
No Recurrence
Low risk
89.6%
High risk
73.27%
26.73%
***
0%
50%
100%
Distribution of Recurrence
G
Time-dependent ROC for OS
Ture positive rate
AUC=0.734
False positive rate
H
Time-dependent ROC for DFS
Ture positive rate
AUC=0.692
False positive rate

Figure 3

The prognostic value of five-lncRNA signature in two independent GEO validation groups.

(A-B) Kaplan-Meier analysis of predicting overall survival of GC patients based on the high-risk and low-risk subgroups in two independent validation groups(GSE62254 and GSE15459). (C-D)The scatter plot of five-lncRNA-based risk score distribution for patient survival status in two independent validation groups.(E-F) The time-independent ROC analysis of the risk score for prediction the 4-year cut-off OS of the two independent validation groups. The area under the curve was calculated for ROC curves.

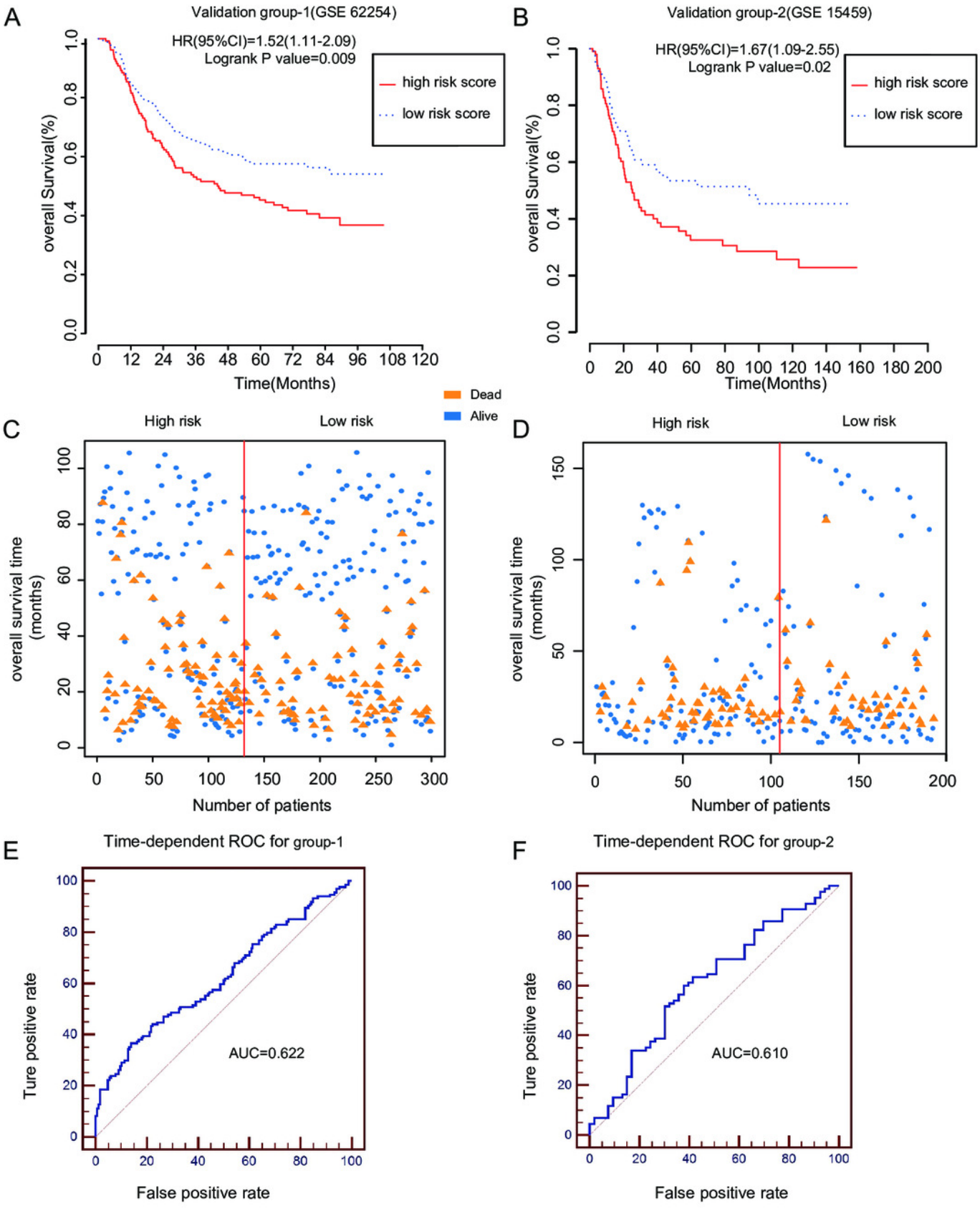

A
Validation group-1(GSE 62254)
HR(95%CI)=1.52(1.11-2.09)
Logrank P value=0.009
high risk score
low risk score
overall Survival(%)
Time(Months)
B
Validation group-2(GSE 15459)
HR(95%CI)=1.67(1.09-2.55)
Logrank P value=0.02
high risk score
low risk score
overall Survival(%)
Time(Months)
C
High risk
Low risk
Dead
Alive
overall survival time (months)
Number of patients
D
High risk
Low risk
overall survival time (months)
Number of patients
E
Time-dependent ROC for group-1
AUC=0.622
Ture positive rate
False positive rate
F
Time-dependent ROC for group-2
AUC=0.610
Ture positive rate
False positive rate

Figure 4

The prognostic value of five-lncRNA signature in subgroups according to the TNM stage.

(A-D) Kaplan-Meier analysis of the overall survival of GC patients with stageI,stageII,stageIIIand stageIV,respectively.

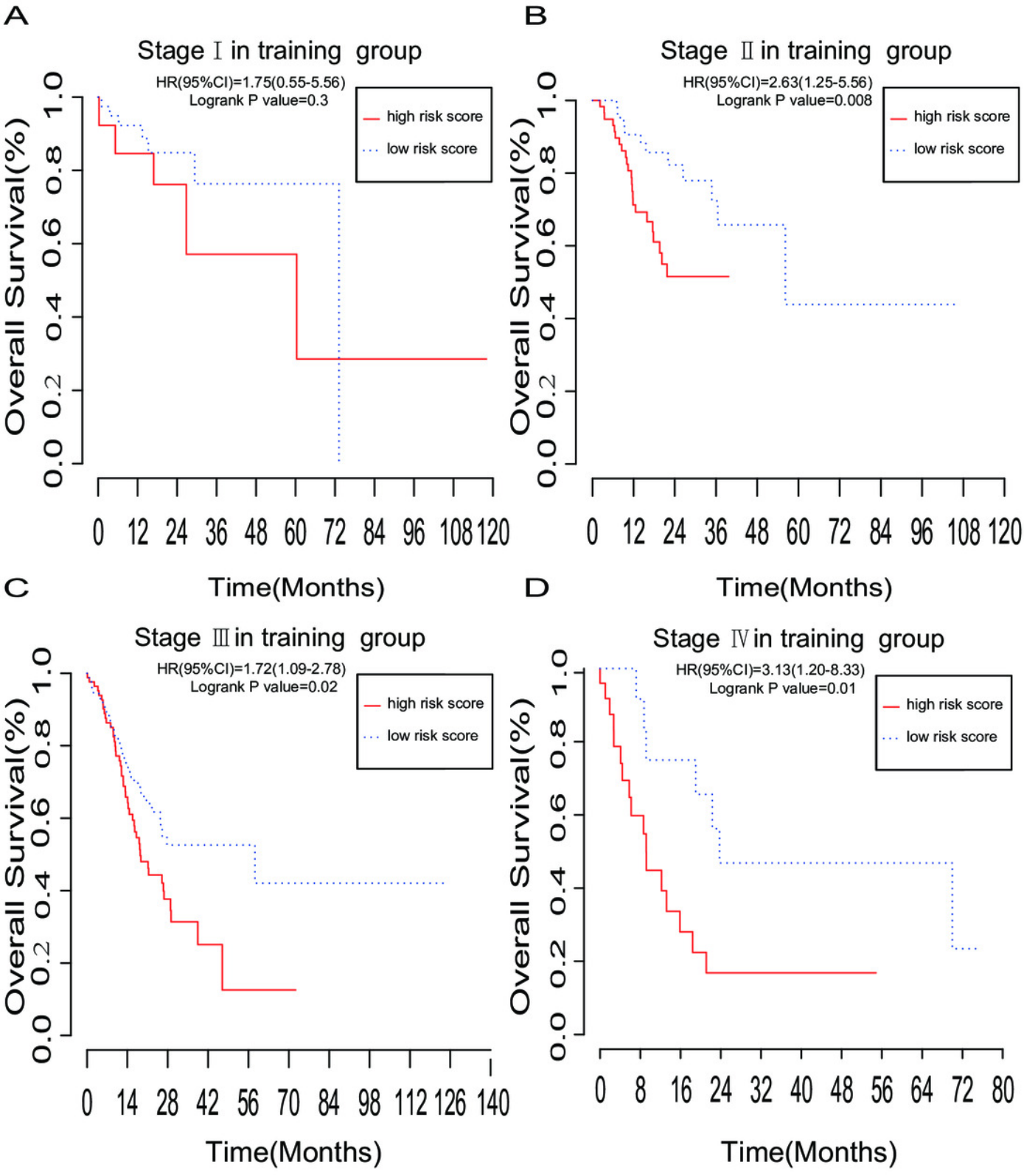

A
Stage Ⅰ in training group
HR(95%CI)=1.75(0.55-5.56)
Logrank P value=0.3
high risk score
low risk score
Overall Survival(%)
0.0 0.2 0.4 0.6 0.8 1.0
0 12 24 36 48 60 72 84 96 108 120
Time(Months)
B
Stage Ⅱ in training group
HR(95%CI)=2.63(1.25-5.56)
Logrank P value=0.008
high risk score
low risk score
Overall Survival(%)
0.0 0.2 0.4 0.6 0.8 1.0
0 12 24 36 48 60 72 84 96 108 120
Time(Months)
C
Stage Ⅲ in training group
HR(95%CI)=1.72(1.09-2.78)
Logrank P value=0.02
high risk score
low risk score
Overall Survival(%)
0.0 0.2 0.4 0.6 0.8 1.0
0 14 28 42 56 70 84 98 112 126 140
Time(Months)
D
Stage Ⅳ in training group
HR(95%CI)=3.13(1.20-8.33)
Logrank P value=0.01
high risk score
low risk score
Overall Survival(%)
0.0 0.2 0.4 0.6 0.8 1.0
0 8 16 24 32 40 48 56 64 72 80
Time(Months)

Figure 5

Forest plot to evaluate prognostic value of five-lncRNA signature in subgroups divided by clinical factors.

Forest plot for clinic subgroup

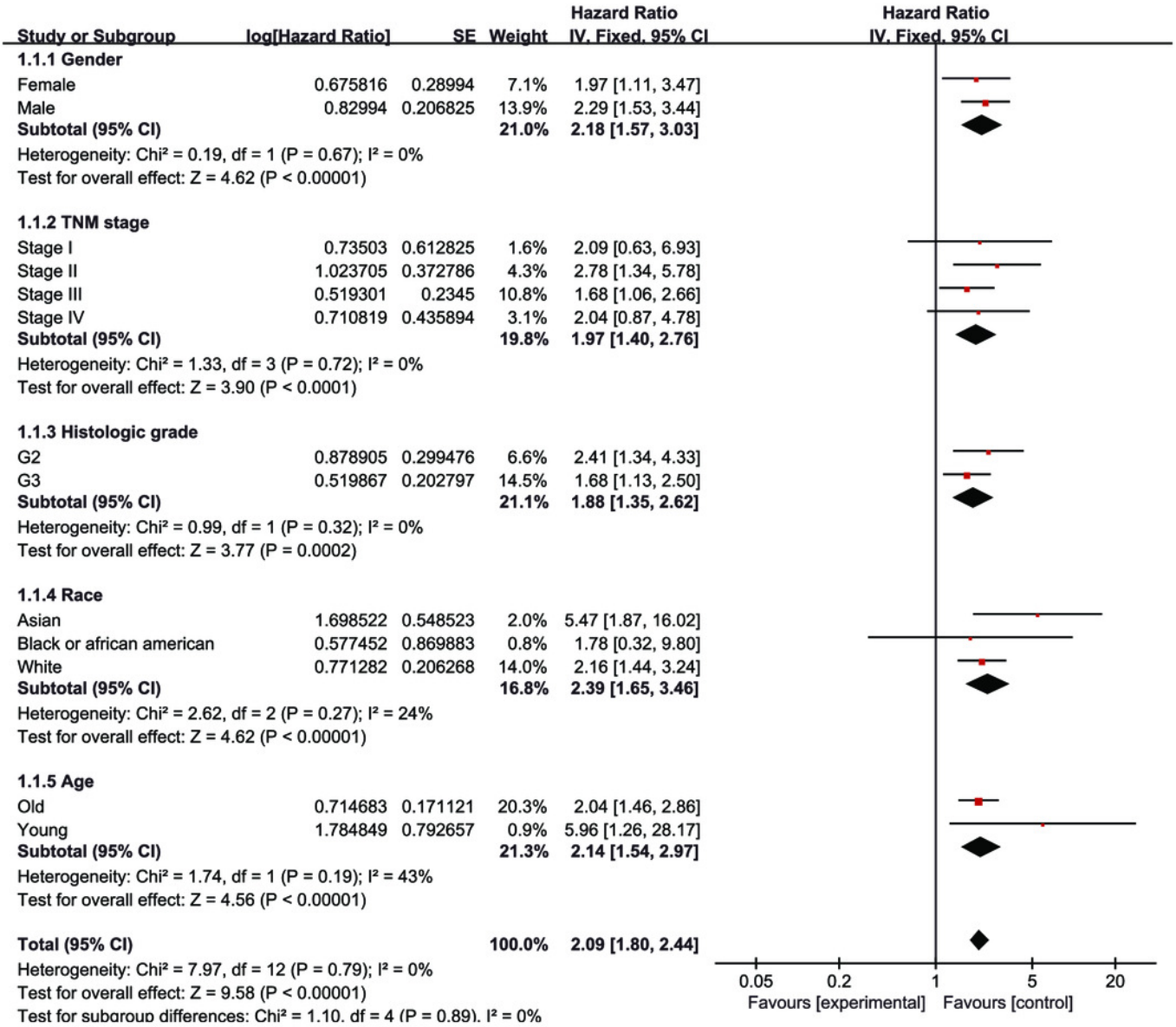
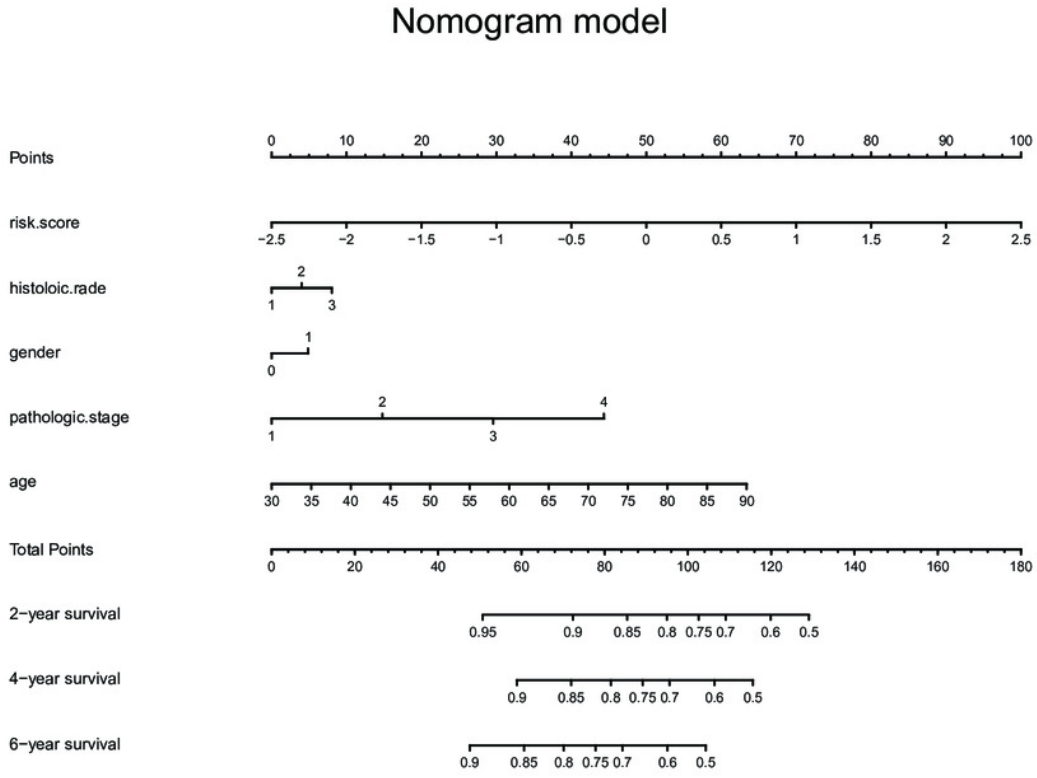


Figure 6

The prognostic value of a nomogram model combining five-lncRNA signature with the clinical factors.

(A) A nomogram model combining five-lncRNA signature with the clinical factors for predicting the 4-year OS of GC patients. (B) The nomogram calibration curve to evaluate the prediction of 4-year OS of GC patients. The C index of this model was also calculated.

A


Nomogram model
Points
0 10 20 30 40 50 60 70 80 90 100
risk.score
−2.5 −2 −1.5 −1 −0.5 0 0.5 1 1.5 2 2.5
histoloic.rade
2
1 3
gender
1
0
pathologic.stage
2 4
1 3
age
30 35 40 45 50 55 60 65 70 75 80 85 90
Total Points
0 20 40 60 80 100 120 140 160 180
2−year survival
0.95 0.9 0.85 0.8 0.75 0.7 0.6 0.5
4−year survival
0.9 0.85 0.8 0.75 0.7 0.6 0.5
6−year survival
0.9 0.85 0.8 0.75 0.7 0.6 0.5

B

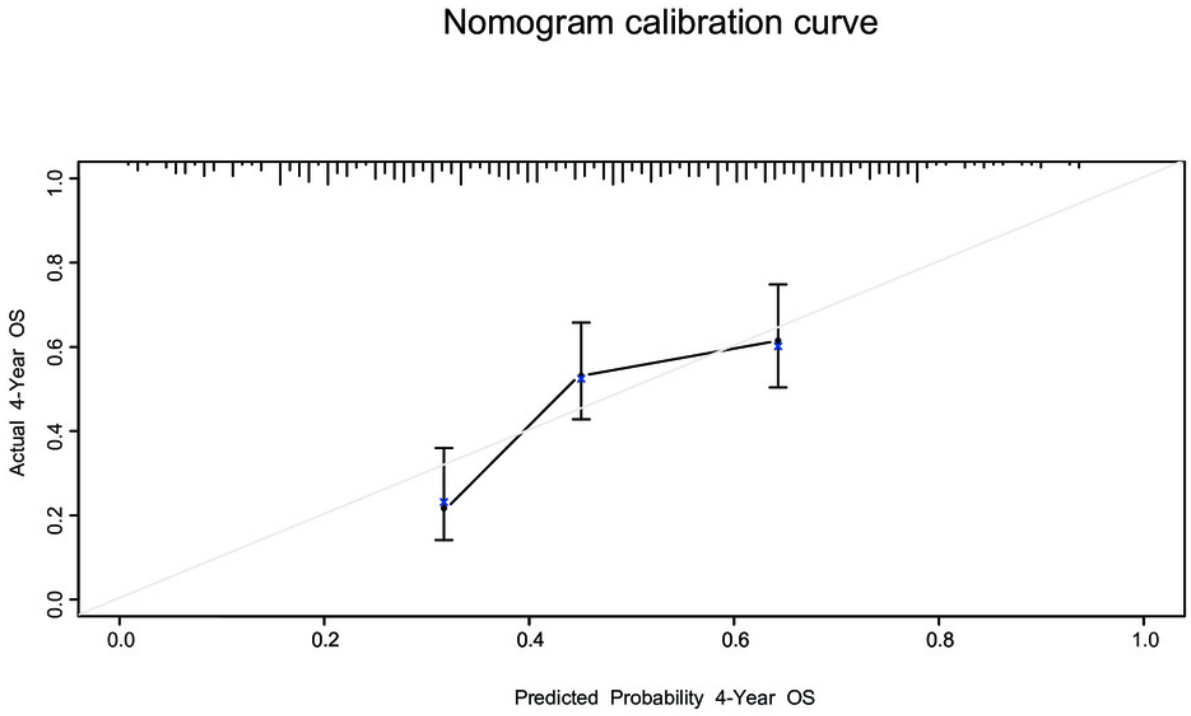
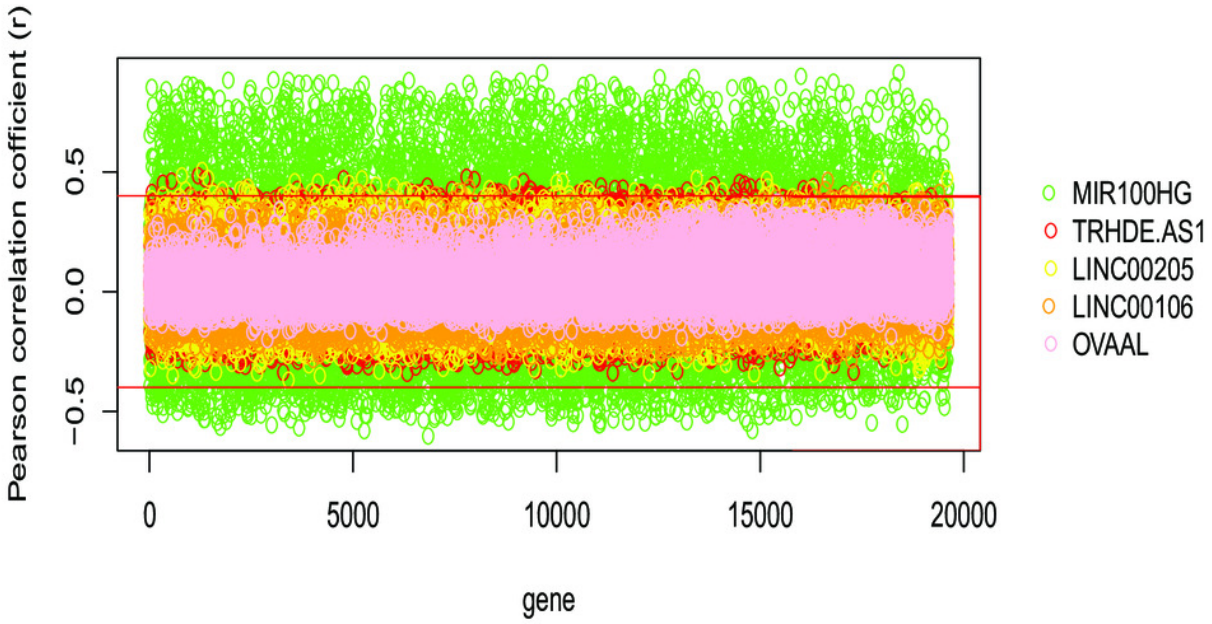

Nomogram calibration curve
Actual 4-Year OS
0.0 0.2 0.4 0.6 0.8 1.0
Predicted Probability 4-Year OS
0.0 0.2 0.4 0.6 0.8 1.0

Figure 7

Potential functions of the five lncRNAs

(A) The Pearson correlation coefficient between 19605 protein-coding genes and five lncRNAs in TCGA dataset. (B) The functional enrichment bubble map of pathways by KEGG pathway analysis. Bubble size represents the number of gene enriched in the pathway.

A


Pearson correlation cofficient (r)
0.5
0.0
−0.5
0
5000
10000
15000
20000
gene
MIR100HG
TRHDE.AS1
LINC00205
LINC00106
OVAAL

B

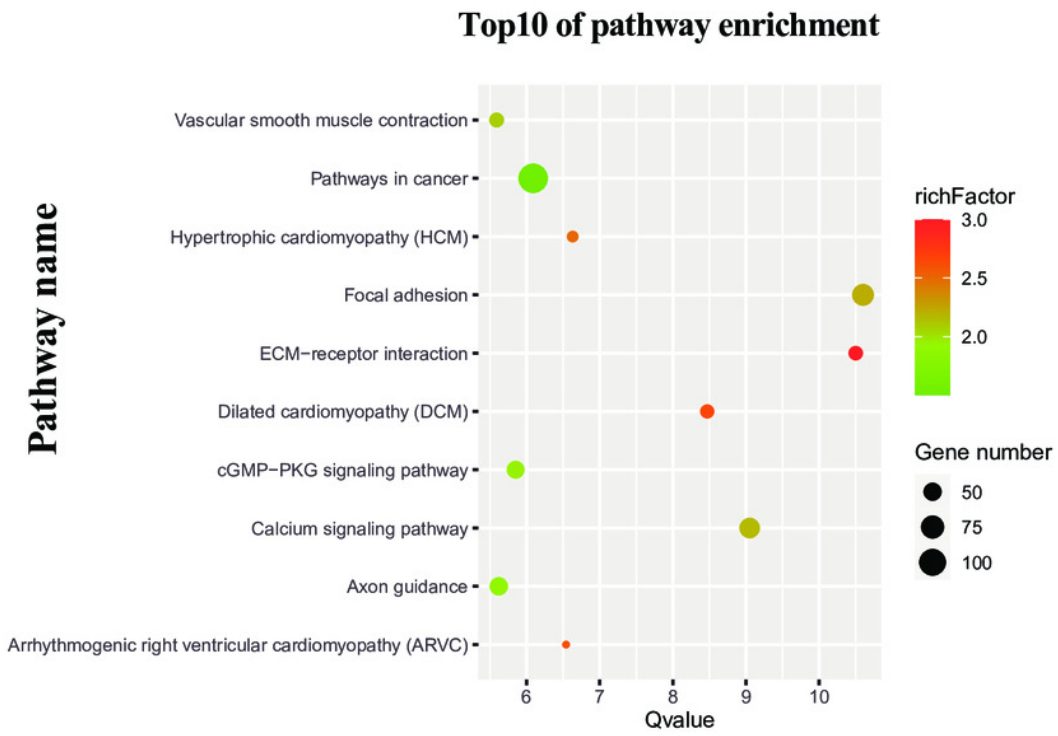

Top10 of pathway enrichment
Pathway name
Vascular smooth muscle contraction
Pathways in cancer
Hypertrophic cardiomyopathy (HCM)
Focal adhesion
ECM−receptor interaction
Dilated cardiomyopathy (DCM)
cGMP−PKG signaling pathway
Calcium signaling pathway
Axon guidance
Arrhythmogenic right ventricular cardiomyopathy (ARVC)
6
7
8
9
10
Qvalue
richFactor
3.0
2.5
2.0
Gene number
50
75
100

Table 1(on next page)

Five lncRNAs significantly associated with prognosis of GC patients in the training group.

Derived from the multivariable Cox proportional hazards regression analysis in the training group.

Gene name	Ensemble ID	Chr.	Coordinate	Coefficient	Hazard ratio	P value
LINC00205	ENSG00000223768.1	21	45288052-45297354	0.249092	1.373451497	0.047216345
TRHDE-AS1	ENSG00000236333.3	12	72253507-72273509	0.182045	1.846654514	0.000109193
OVAAL	ENSG00000236719.2	1	180558974-180566518	0.271169	1.880897277	0.0000744
LINC00106	ENSG00000236871.6	X&Y	1397025-1399412	-0.207942	0.624972486	0.003469142
MIR100HG	ENSG00000255248.6	11	122028329-122422871	0.502539	1.396343319	0.036829012

1

Table 2(on next page)

The schematic workflow of the present study

1

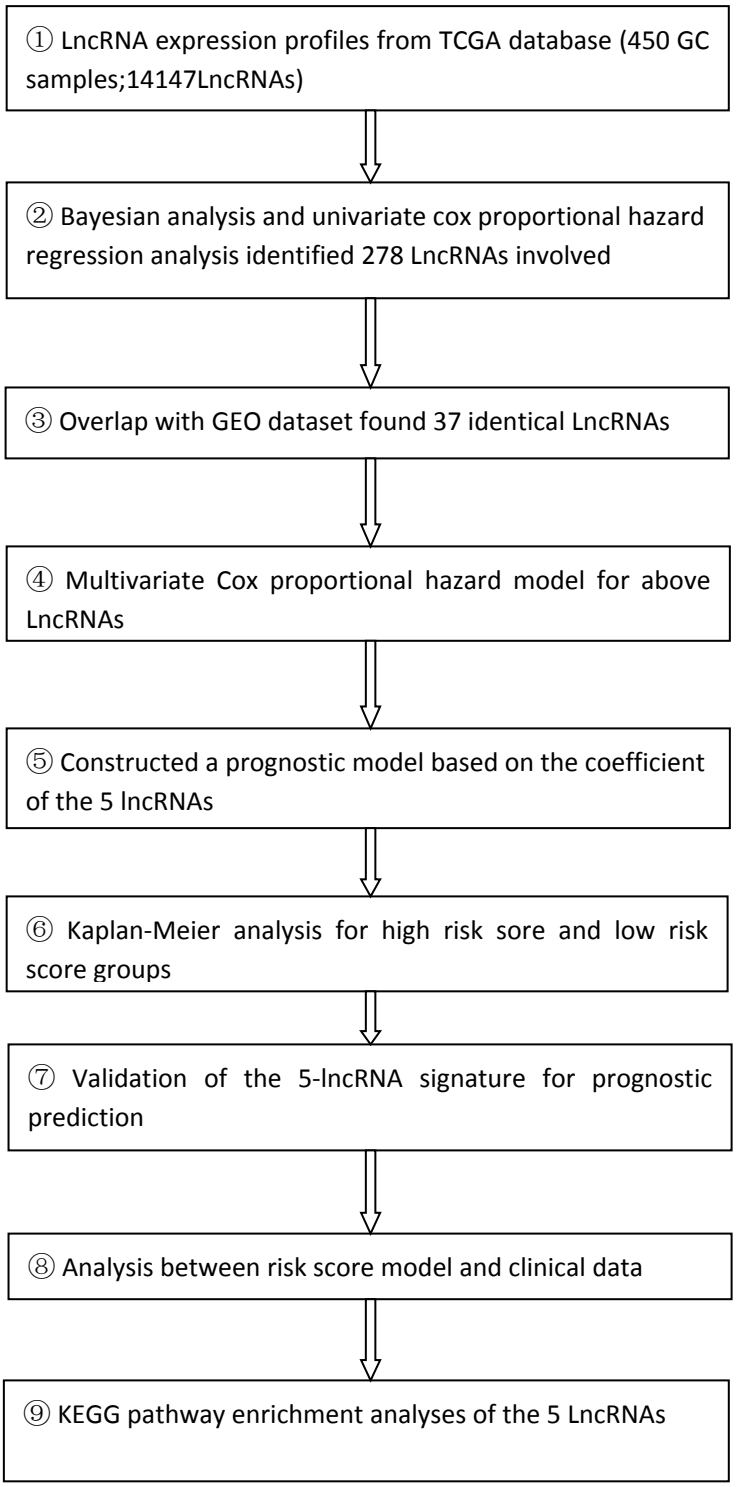

① LncRNA expression profiles from TCGA database (450 GC samples;14147LncRNAs)
② Bayesian analysis and univariate cox proportional hazard regression analysis identified 278 LncRNAs involved
③ Overlap with GEO dataset found 37 identical LncRNAs
④ Multivariate Cox proportional hazard model for above LncRNAs
⑤ Constructed a prognostic model based on the coefficient of the 5 lncRNAs
⑥ Kaplan-Meier analysis for high risk sore and low risk score groups
⑦ Validation of the 5-lncRNA signature for prognostic prediction
⑧ Analysis between risk score model and clinical data
⑨ KEGG pathway enrichment analyses of the 5 LncRNAs

Table 3(on next page)

The association between five-lncRNA signature and OS of GC patients in training group.

Abbreviations: HR, Hazard ratio; 95%CI, 95% confidence interval.

	Number (High Risk score/Low Risk score)	HR (95%CI)	P value
Total	204/204	2.09 (1.80, 2.44)	0.000001
Gender			
Male	129/134	2.29 (1.53, 3.44)	0.00002
Female	75/70	1.97 (1.11, 3.47)	0.01
Histologic grade			
G2	47/97	2.41 (1.34, 4.33)	0.0006
G3	146/97	1.68 (1.13, 2.50)	0.02
Race			
Asian	44/41	5.47 (1.87, 16.02)	0.001
Black or african american	4/8	1.78 (0.32, 9.80)	0.6
White	138/120	2.16 (1.44, 3.24)	0.0003
Age			
Old (>=50 years old)	186/191	2.04 (1.46, 2.86)	0.00001
Young (<50 years old)	18/13	5.96 (1.26, 28.17)	0.008
TNM stage			
Stage I	14/41	2.09 (0.63, 6.93)	0.3
Stage II	62/58	2.78 (1.34, 5.78)	0.008
Stage III	87/77	1.68 (1.06, 2.66)	0.02
Stage IV	25/16	2.04 (0.87, 4.78)	0.01

1