Development and validation of focal adhesion-related genes signature in gastric cancer

Guanghui Zhao  
Sun Yat-sen University Cancer Center

Tianqi Luo  
Sun Yat-sen University Cancer Center

Zexian Liu  
Sun Yat-sen University Cancer Center

Jianjun Li (lijj@sysucc.org.cn)  
Sun Yat-sen University Cancer Center

Article

Keywords: gastric cancer, TCGA, focal adhesion, prognosis, GEO

Posted Date: September 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2033724/v1

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

Background

This study aims to construct a focal adhesion-related genes-based prognostic signature (FAS) to accurately predict the overall survival (OS) of patients with gastric cancer (GC) and to identify key prognostic genes related to GC.

Results

The gene expression data and corresponding clinical characteristics of GC patients were obtained from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA). Subsequently, the GEO dataset was randomly distributed into training and test cohorts. The TCGA dataset was used to validate the external cohort. The least absolute shrinkage and selection operator (Lasso) Cox regression was used to detect OS-related genes in the GEO cohort. A risk score model was established according to the screened genes. A nomogram, based on the clinical characteristics and risk score, was generated to predict the prognosis of GC patients. A time-dependent receiver operating characteristic (ROC) and calibration curve were applied to assess our newly formed model. The patients were grouped into a high- or low-risk group depending on the risk score. Low-risk patients exhibited higher OS than high-risk patients (entire cohort: p < 0.001; train cohort: p < 0.001, test cohort: p < 0.001). This study found that a high-risk score was associated with the circulatory system process and high infiltration of macrophages, CD44, and HLA-DMB.

Conclusions

The generated model based on the genetic characteristics of the focal adhesion prognostic gene can aid in the prognosis of GC patients in the future.

Introduction

Gastric cancer (GC) is among the world’s most common malignant solid tumors and the fifth leading type of global cancer, with 7.7% of all cancer-related deaths, second only to lung and liver cancers (Sung et al. 2021). Due to the uncommon symptomology of early GC, most patients are only diagnosed in the advanced stages of disease (Sitarz et al. 2018). Therefore, understanding the heterogenetic nature of GC to accurately evaluate the prognosis, personalized clinical diagnosis, and treatment. Focal adhesion is a physical connection between the extracellular matrix and actin cytoskeleton, consisting of multiprotein complexes, such as integrins, coflin proteins, and focal adhesion kinase (FAK) (Paluch, Aspalter, and Sixt 2016). Focal adhesion plays an important role in regulating cell adhesion, mechanical induction, cell differentiation, distant metastasis, and chemoradiotherapeutic resistance of tumor cells (Landowski et al. 2003, Damiano et al. 1999). Focal adhesion up-regulates B3 and FAK expressions in GC, facilitating
cancer cells to resist fluorouracil, leading to failed treatment (Ngabire et al. 2020). Focal adhesion protein also promotes GC cell invasion by enhancing cell proliferation (Shen et al. 2013). Inhibition of the focal adhesion signal pathway can potentially treat GC effectively.

This study investigates the relationship between focal adhesion-related genes (FARGs) and corresponding clinical characteristics of GC patients in the Gene Expression Omnibus (GEO) \((n = 684)\) and The Cancer Genome Atlas (TCGA) \((n = 330)\) databases. According to the univariate Cox regression analysis, significant genes associated with prognosis were identified, and a dependable FARG signature was constructed based on Lasso analysis. Receiver operating characteristic (ROC) and Kaplan-Meier (KM) analyses were used to evaluate the signature. Then, a nomogram model was constructed based on the FARGs-signature and corresponding clinical characteristics. The accuracy of our newly developed nomogram that predicts the prognosis of GC patients was validated using a TCGA external validation cohort. In summary, this study highlighted the crucial role of focal adhesion-related genes signature (FAS) and developed a nomogram that predicts OS in GC patients.

**Results**

**Patient characteristics and establishment of FAS**

As illustrated in Fig. 1, upon exclusion of cases with a survival time of fewer than 30 days and normal cases, 684 samples were collated in the four GEO (Home - GEO - NCBI (nih.gov)) datasets (GSE13861, GSE26942, GSE29272, and GSE62254). These cases were randomly divided into a training (478) or test cohort (206) in a 7:3 ratio. Table 1 summarizes the patient clinical characteristics. The “limma” package extracted genes associated with focal adhesion in the GEO database. In the training cohort, the univariate Cox and LASSO regression analyses were conducted to screen 11 genes associated with GC patient OS, as depicted in Figs. 2A and B.
### Table 1
Clinicopathological characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Training cohort</th>
<th>Test cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High-risk</td>
<td>Low-risk</td>
</tr>
<tr>
<td>N</td>
<td>244</td>
<td>234</td>
</tr>
<tr>
<td>Risk score(median)</td>
<td>14.61</td>
<td>13.96</td>
</tr>
<tr>
<td>Age(median)</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>69</td>
</tr>
<tr>
<td>Stage</td>
<td>I</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>26</td>
</tr>
<tr>
<td>Overall survival</td>
<td>Alive</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>Dead</td>
<td>39</td>
</tr>
<tr>
<td>Survival time(median)</td>
<td>804</td>
<td>1807</td>
</tr>
</tbody>
</table>

The risk score of each sample was calculated as follows:

\[
\text{Risk score} = \text{CMOP} \times 0.2496 + \text{IGF1R} \times 0.7161 + \text{ITGB5} \times 1.6124 + \text{LAMA4} \times 2.203 + \text{MYLK} \times 0.1340 + \text{THBS4} \times 0.1248 + \text{TNN} \times 0.7119 + \text{VEGFB} \times 1.0125 + \text{VWF} \times 0.8667.
\]

The patients were divided into a high- or low-risk group based on the optimal cut-off value of the risk score in the Survminer R package. The distribution of the risk scores and the survival statuses of patients in the training cohort are displayed in Figs. 3A–C. KM analysis determines the differences between the two groups in the training cohort (Fig. 4A). Lastly, time-dependent ROC exhibits the prognostic values of our signature (Fig. 4D).

**Validation and evaluation of the prognostic gene signature**

Internal and external validations on the test cohort using TCGA datasets were conducted. Consistent with the training cohort results, OS was lower in high-risk patients than in low-risk patients (Figure 4B, \(p<0.001\) in test cohort; Figure 4C, \(p \leq 0.01\) in TCGA cohort). The area under the ROC curve demonstrated that the signature could precisely predict GC prognosis (Figures 4E and F). The AUCs for the internal and external validation cohorts were 0.704 and 0.598 at three years, respectively, and 0.657 and 0.667 at five years,
respectively. The PCA and t-SNE analyses were conducted to determine the accuracy of the signature. Both PCA and t-SNE plots revealed that the high- and low-risk groups had different directions in the training, test, and external cohorts (Figures 5A–F).

**FAS is an independent predictor of GC**

The Cox regression analysis demonstrates the relationship between the risk scores acquired from the prognostic model with other clinical parameters. Based on the uni- and multivariate regression analyses, stage ($P < 0.001$, HR = 3.5; $P<0.001$, HR=3.164, respectively) and risk score ($P < 0.001$, HR = 3.43; $P<0.001$, HR=3.253, respectively) were independent OS prognostic factors in the training cohort (Figures 6A and B). The test cohort was validated, whereby both stage ($P < 0.001$, HR = 3.012; $P<0.001$, HR=2.539, respectively) and risk score ($P < 0.001$, HR = 2.328; $P=0.002$, HR=2.061, respectively; Figures 6C and D) were demonstrated to be independent risk factors for OS in GC patients.

**Subgroup analysis of the prognostic value of FAS**

The training cohort was further divided into subgroups to investigate the prognostic value of the developed model among different patient populations based on their clinical characteristics and estimate OS between high- and low-risk groups. The KM analysis revealed that the risk score could distinguish differences between various subgroups, such as age and gender (Figures 7A–F). Similarly, the test cohort demonstrated differences in age and gender between the high- and low-risk groups. However, there are no differences in the stage subgroup, which may be related to our study’s relatively small sample size (Sup Figures 1A–F).

The relationship between risk score and corresponding clinical characteristics of GC patients ($n = 684$) in the GEO dataset was analyzed. The risk scores of stage I + II patients were significantly lower than those of stage III + IV patients ($P=1.1e-09$, Figure 8A). Similar results were shown in the age subgroup ($P=0.019$, Figure 8B). No significant relationship between the risk score and gender ($P=0.92$, Figure 8C).

**Generation of a prognostic nomogram that predicts OS in GC patients**

To predict the prognosis of GC patients accurately, a nomogram was generated to predict the 1-, 3-, and 5-year OS rates of GC patients, based on the uni- and multi-variate regression analyses (Figure 9). Additionally, ROC analysis revealed that the sensitivity of the nomogram was higher than other clinicopathological features in the training, internal validation, external validation, or entire cohort (Figures 10A–H). The calibration plots for the training, internal validation, and entire cohort were in agreement between the actual OS and the predicted from the nomogram (Figures 11A–D).

**GSEA**

To elucidate the different gene functional and signal pathways enrichment between the high- and low-risk score groups, a Gene Set Enrichment Analysis (GSEA) was conducted on the
“c2.cp.kegg_v7.4.symbols.gmt” and “c5.go.v7.4.symbols.gmt” gene sets. The top five pathways and gene functions in the high- and low-risk groups are displayed in Figures 12A–D.

**Immune Cells Infiltration and Immune-Related Pathways**

TME regulates tumor treatment resistance and is associated with tumor occurrence, development, and metastasis. It consists of the tumor, immune and stromal cells, and myriad cytokines. Alterations in the TME (such as changes in the immune cell components) enhance tumor progression. To analyze the distribution characteristics of immune cells in the GC TME and to investigate the interaction mode between GC tumor and immune cells, the ssGSEA tool was used to predict 16 common immune cells and 13 relative components of immune-related functions based on the GC gene expression profile data.

Patients in the low-risk group had high ratios of B cell, T regulator, and follicular helper T cells than in the high-risk group (Figure 13A). Moreover, the low-risk group displayed higher antigen-presenting cell (APC) co-inhibition, CC chemokine receptor (CCR) cytolytic activity, inflammation promotion, and T cell inhibition than patients with a high-risk score (Figure 13B). Additionally, we demonstrated that the expression levels of CD200, CD28, CD40, CD40LG, CD44, CD86, LAIR1, NRP1, TNFRSF4, TNFRSF8, TNFSF18, TNFSF4, and VTCN1 in the high-risk group were higher than those in the low-risk group (Figure 13C). The finding demonstrates that the immune microenvironment is partially associated with the OS prognosis of GC patients with a high level of FARGs.

**Correlation of m6A expression**

N6-methyladenosine (m6A) is the most abundant RNA modification in eukaryotic cells (Yue, Liu, and He 2015). Extensive RNA processing and metabolism research revealed that m6A is a key contributor to cancer development. m6A is a potential prognostic marker involved in multiple aspects of cancer treatment (Ma et al. 2019). To assess the relationship between m6A expression and our GC prognostic signature, the levels of 13 m6A genes in different GC samples were estimated. It is found that an elevated expression of FTO, METTL3, RBM15, YTHDC1, and YTHDF1 genes in the high- versus low-risk group (Figure 13D).

**Correlation between TME subcomponents and the FARGs risk score and outcome of GC patients**

TME consists of diverse immune and stromal cells linked to disease development, prognosis, and treatment outcome. Based on our ESTIMATE algorithm, TME was separated and scored into stromal, immune, and estimate subcomponents to investigate potential relationships between this study’s risk scores and TME. A high immune or matrix score indicates a high proportion of the immune or matrix components in the TME. The ESTIMATE score is the sum of the immune and stromal scores, indicating the combined proportion of these two components in TME. The Pearson’s correlation analysis indicated that the stromal and ESTIMATE scores were positively correlated with the risk score of the entire cohort (r=0.6718, p<2.2e-16; r=0.4148, p<2.2e-16, respectively, Figures 14A and C). There is no correlation between the immune and risk scores (P=0.484, Figure 14B).
To further clarify the survival rate of different GC sub-populations, the entire cohort was grouped into median immune, median stromal, and median ESTIMATE scores as cut-off points. As illustrated in Figures 14D and F, patients with elevated stromal and ESTIMATE scores exhibited worse OS than those with low stromal and ESTIMATE scores (P<0.001, P<0.001, respectively). The survival rate between patients with high and low immune scores (P=0.83) was similar (Figure 14E). A heatmap was generated to depict the immune cell scores in the high- and low-risk groups, as illustrated in Figure 15.

**Drug sensitivity prediction**

The correlation between drug Z score and genes was analyzed, and the top 16 significant drug-gene pairs are displayed in Figure 16. The overall 246 were statistically different in Supplement Figure 2. JNJ-38877605, Staurosporine, and XAV-939 depicted the most positive correlation with FDA gene expression. On the contrary, By-Product of CUDC-305, Palbocic, and Oxaliplatin was negatively correlated with FDA gene expression.

**Validation of a FAD-Based Prognostic Model in a clinical sample**

The protein expression of focal adhesion related-genes was analyzed using the Human Protein Atlas (HPA) database. Seven of the nine focal adhesion related-genes (CM0P, IGF1R, ITGB5, LAMA4, MYLK, THBS4, TNN, VEGFB, and VWF) showed elevated expression in GC (Figure 17). The clinical data of STAD in the KM-plotter database were investigated to observe the prognosis of patients with different FA-relation gene expressions. Patients with a high FA relation gene expression had a better OS (Figure 18) than those with a low FAD-related gene expression, excluding the THBS1.

**Discussion**

GC patients showed symptoms in the early stages of the disease. Most of them miss the surgery opportunity due to local or distant metastasis at the time of diagnosis. However, multiple factors could easily regulate a single clinical factor or gene feature. Thus, a single factor or gene is not a reliable prognostic marker. With the rapid development of high-throughput sequencing technology, this study systematically investigates a class of gene sets related to patient survival prediction. The formation and turnover of focal adhesion are critical to tumor cell migration and progression (Eke and Cordes 2015, Wu et al. 2021, Maziveyi et al. 2018). Therefore, it is crucial to evaluate the prognosis of FARGs in GC patients.

This study used the bioinformatics approach and the publicly available TCGA and GEO databases to screen for FARGs. A risk score was assigned to each case to predict their prognosis, and it was found that high-risk cases exhibited a worse prognosis than low-risk ones. Similar results were found in patients of different genders, ages, and stages. A nomogram was created by combining risk scores and corresponding clinical characteristics to demonstrate the accuracy of the generated prognostic model in predicting 3- and 5-year survival in GC patients.
The focal adhesion and GC prognosis-related genes exhibited high predictive value and accurately individualized GC patient information.

Prior studies described the relationship between the screened genes and cancer, particularly GC. COMP was significantly higher in colon cancer and is strongly associated with cellular appreciation and tumor progression(Nfonsam et al. 2020). IGF1R signaling regulates the biological process of GC by increasing β-catenin activation, epithelial-mesenchymal transition, and cell proliferation(Xu, Zhou, et al. 2017). Yang et al. demonstrated that ITGB5 expression reduction could inhibit cell proliferation using CRISPRa and CRISPRi Technologies(Yang et al. 2021). ITGB5 also promoted lymph node metastasis in colorectal cancer patients(Capriotti et al. 2020). LAMA4 induces resistance to cisplatin by activating androgen receptors in GC, failing the treatment in GC patients(Peng et al. 2020). Han et al. demonstrated that LAMC1 knockdown inhibits GC cell proliferation, migration, invasion, and the Warburg effect by suppressing AKT and MEK/ERK pathways(Han, Jiang, and Fan 2021). MYLK is heavily involved in GC metastasis via its AR-V12-mediated regulation(Xia et al. 2019). Both extracellular matrix proteins THBS1 and THBS4 strongly regulate key tumor cell processes, such as proliferation, attachment, adhesion, and migration. Elevated THBS1 and THBS4 expressions may be closely associated with the rising tumor grading and prediction of poor prognosis in GC patients(Zhang et al. 2021, Chen et al. 2019).

The biological functions that hold greater relevance to high-risk patients were identified using GO analysis. Axon development, blood vessel morphogenesis, circulatory system process, collagen fibril organization, and epithelial cell proliferation were high in high-risk patients. High-risk patients undergo biological processes, such as forming new cells, cell division, and proliferation, indicating vigorous division and proliferation of tumor cells in this patient population. This finding was consistent with the previous clinical conclusion that the prognosis of patients in the high-risk group is generally poor. The KEGG functional pathways in high- and low-risk GC patients were investigated using the risk model established by our 11-gene signature. The five major pathways are generally associated with cancer development. Namely, axon guidance, cell adhesion molecules, complement and coagulation cascades, ECM receptor interaction, and focal adhesion were high in the high-risk patients(Jurcak et al. 2019, Francavilla, Maddaluno, and Cavallaro 2009, Zhang et al. 2020, Yang et al. 2020). In addition, a larger population of macrophages and neutrophils were found in high-risk patients. Tumor-associated macrophages or neutrophils are typically associated with poor prognosis in GC patients(Gambardella et al. 2020, Li et al. 2019). At the primary site, macrophages are present at all stages of tumor progression and are closely associated with the invasion of tumor cells(Nielsen and Schmid 2017).

The APC co-inhibitory effect and cytolytic activity score were lower in the high-risk group than in the low-risk group. The weakening antitumor immunity in high-risk patients may lead to a poor prognosis. Research on tumor m6A has become a hot spot recently. The levels of FTO, METTL3, and YTHDC1 in the high-risk group were remarkably higher than in the low-risk group. FTO plays a critical role in the progression and metastasis of GC associated with low differentiation, lymph node metastasis, TNM stage, and poor prognosis, making it an important molecular marker for monitoring GC(Xu, Shao, et al. 2017). Yue et al. demonstrated that an enhanced expression of METTL3 is positively related to the poor
prognosis of GC patients. It is also revealed that METTL3 benefits the process of epithelial-mesenchymal transformation and metastasis in vivo (Yue et al. 2019). One study analyzed various biological information from different human cancer databases and found that YTHDF1 mutations are present in approximately 7% of GC patients and that high YTHDF1 levels are associated with augmented cancer proliferation, invasiveness, and poor patient OS (Pi et al. 2021).

The official HPA website further validated the target gene expression in cancer and normal tissues. KM survival curve also verified the difference in gene expression in gastric cancer. The sensitive drugs closely related to gene expression were predicted to further explore our key genes’ therapeutic effect on tumors. Positive correlation refers to the high expression of these genes in GC, directly proportional to drug sensitivity. The negative correlation refers to the high gene expression in GC, which may affect drug efficacy.

Currently, this is the first prognostic model involving focal adhesion in GC patients. This study developed and validated this model as an excellent predictor of GC patient OS prognosis. In addition, our model provides further information concerning immune infiltration, immune checkpoint indicators, and pathway enrichment in various subgroups.

Although the model was validated in various aspects, it still faces certain limitations. First, the data was downloaded from the TCGA and GEO databases, and the significance of the model in other cohorts needs further validation. In addition, more investigation is needed to determine if the genes in the model act synergistically to influence GC patient prognosis. This study investigated the prognostic relevance of the focal adhesion gene in GC patients using retrospective analysis, and its predictive ability needs to be tested in prospective studies. Unlike traditional biological research methods, this method was based on a large dataset and possessed the advantages of enhanced efficiency, flexibility, and pertinence. With the continuous development of sequencing technology, this model can likely hold great potential in clinical application.

**Conclusions**

We developed a FAS (COMP, IGF1R, ITGB5, LAMA4, LAMC1, MYLK, THBS1, THBS4, TNN, VEGFB, and VWF) model to distinguish the clinical features of GC patients and to predict GC patient prognosis better. Our study provides a valuable basis and direction for subsequent basic experimental and clinical research involving GC.

**Materials And Methods**

**Data collection**

The clinicopathological information and corresponding gene expression data of GC patients were obtained from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). A total of 684 cases (GSE13861, GSE29272, GSE62254, and GSE26942) were examined in the entire cohort. 330 TCGA-STAD samples and
their corresponding clinicopathological data were extracted from the Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) to be used for the external validation cohort. A list of focal adhesion genes was downloaded from the MSigDB database (https://www.gsea-msigdb.org/gsea/msigdb) to aid this analysis. Our study based on public source, there are no ethical issues and other conflicts of interests.

**Data processing**

Gene symbols of each gene matrix file were extracted based on the corresponding platform file used by the Perl software. The batch effect was impaired using the Empirical Bayes method (“sva” package) among the series. Lastly, the entire cohort was randomly grouped into either training or test cohort in a 7:3 ratio.

**FARGs signature construction and validation.**

In total, 199 FARGs were selected from the MSigDB database. Univariate Cox, LASSO regression, and multivariate Cox analyses were performed using the “glmnet” and “survival” R package to identify the relationship between the FARGs and GC prognosis and develop a GC prognostic signature. The risk score for each sample was calculated using FARGs expression associated with prognosis, and its corresponding correlation coefficient, as follows: \[ \text{Risk Score} = \sum_{i=1}^{n} (\text{Coei} \times \text{Expri}) \]

where, Coei represented coefficients, and Expri represented the expression of each FARGs.

Each sample was divided into a high- or low-risk group based on the risk score. The KM survival and time-dependent ROC curves were used to validate the accuracy of the signature.

**Establishment and evaluation of a FAS-based nomogram model to predict OS of GC patients**

A prognostic nomogram was established to effectively identify the OS of GC patients based on the focal adhesion risk score, patient’s gender, age, and the American Joint Committee on Cancer (AJCC) stage. The calibration and ROC curves were used to assess the accuracy of nomograms in the 1-, 3-, and 5-year OS of GC patients.

**Gene set enrichment analysis**

The GSEA software (version 4.1.0) was obtained from the GSEA website (http://www.gseamsigdb.org/gsea/index.jsp) to identify the functional enrichment pathways regulated by the FARGs signature. Additionally, the “c2.cp.kegg.v7.4.symbols.gmt” and “c5.go.v7.4.symbols.gmt” gene sets were extracted from the molecular signatures database and used as the target enrichment sets for GSEA analysis.

**Calculations of the immune, stromal, and estimate scores**

The estimation R package was applied to predict each GC sample’s immune and stromal component scores in the tumor microenvironment (TME). The calculation method was estimated and displayed in
the form of three scores: immune, stromal, and estimated. If the risk score positively correlates with the immune, stromal, and overall scores, it produces a higher risk score and a greater proportion of corresponding TME components. The “survive” and “survminer” software packages in the R software were employed to analyze the TME score. A total of 684 samples with detailed survival times and statuses were then divided into a high- or low evaluation group according to the median values of the immune, stromal, and estimate scores for survival analysis.

**Single-sample gene set enrichment analysis (ssGSEA)**

The single sample gene set enrichment analysis (ssGSEA) in the “gsva” R package quantifies the infiltration statuses of 16 immune cells and the activities of 13 immune-related pathways in the high- and low-risk groups.

**Drug Sensitivity Analysis**

The drug sensitivity analysis was conducted using cellminer (https://discover.nci.nih.gov/SclcCellMinerCDB/) Database data, screening FDA-approved and clinical trial data, and analyzing the relationship between FDA genes expression level and drug sensitivity. Spearman's correlation analysis was conducted to determine the correlation using R software, and the top 16 drugs were selected.

**Validation of a FAD-Based Prognostic Model in a clinical sample**

The Human Protein Atlas (HPA) dataset was used to evaluate the protein expression levels of the focal adhesion related-genes. The accuracy of the data was further validated following the analysis of the clinical data of STAD in the KM-plotter database (https://kmplot.com/).

**Statistical Analyses**

All statistical analyses were performed using R software (version 4.0.0), and KM survival analysis assessed the differential OS durations between the high- and low-risk groups. P < 0.05 was set as the significance threshold.

**Declarations**

**Acknowledgments**

We thank Home for Researchers editorial team (www.home-for-researchers.com) for language editing service. We express sincere gratitude to the TCGA, GEO, GSEA databases for the availability of data.

**Funding**

There is no funding to report.
Data availability statement

The raw data of this study are derived from the TCGA (https://portal.gdc.cancer.gov/). The dataset GSE13861, GSE26942, GSE29272, and GSE62254 for this study can be found in the GEO database (https://www.ncbi.nlm.nih.gov/geo). All data generated or analysed during this study are included in this published article.

Author contributions

Designation: J.L; Data collection: T.L; Drawing: G.Z; Original draft writing: G.Z, J.L and Z.L supervised the overall workflow and revised the manuscript and guaranteed of the study. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References


Figures

Figure 1

The flow chart showing the scheme of our study on focal adhesion prognostic signatures in GC
Figure 2

(A-B) Determination of the number of factors by the LASSO analysis
Figure 3

Risk score distribution (A), survival status (B), and 11 FARGs expression profiles (C) for patients in high-risk and low-risk groups in training cohort.
Figure 4

(A) Differences in survival between high-risk and low-risk groups of the training cohort (p <0.001).
Validation of the 11-FARG prognostic signature in the testing cohort and the external cohort. Differences in survival between high-risk and low-risk groups of the test cohort (p <0.001) (B) and external cohort (p =0.01) (C). (D)-(F) Time-dependent ROC analysis for the FARGs signature in GC. (D) Time-dependent ROC analysis of the FARGs signature in the training cohort. (E) Time-dependent ROC analysis of the FARGs signature in the test cohort. (F) Time-dependent ROC analysis of the FARGs signature in the external cohort.
Figure 5

(A) The PCA plot in the training cohort. (B) The t-SNE plot in the training cohort. (C) The PCA plot in the test cohort. (D) The t-SNE plot in the test cohort. (E) The PCA plot in the external cohort. (F) The t-SNE plot in the external cohort.
Figure 6

Univariate and multivariate Cox regression analysis showed the relationship between age, gender, stage, risk score, and overall survival, and indicated that risk score could be used as an independent prognostic factor for training cohort (A-B), test cohort (C-D).
Figure 7

The high-risk group in training cohort showed a poor prognosis than the low-risk group in different clinical stratification like age (A-B), gender (C-D), stage (E-F)
Figure 8

Stratified analysis of the prognostic signature in the training cohort. The relationships between the FARGs prognostic signature and stage (A), age (B), gender (C).
Figure 9

Nomogram for the prediction of 1-, 3-, and 5-year survival probability in patients GC
Figure 10

(A-D) 3-year time-dependent ROC analysis for nomogram in GC. (A) Time-dependent ROC analysis of the nomogram in the training cohort. (B) Time-dependent ROC analysis of nomogram in the test cohort. (C) Time-dependent ROC analysis of the nomogram in the entire cohort. (D) Time-dependent ROC analysis of the nomogram in the external cohort. (E-H) 5-year time-dependent ROC analysis for nomogram in GC. (E) Time-dependent ROC analysis of the nomogram in the training cohort. (F) Time-dependent ROC analysis of nomogram in the test cohort. (G) Time-dependent ROC analysis of the nomogram in the entire cohort. (H) Time-dependent ROC analysis of the nomogram in the external cohort.
Figure 11

The calibration plot for internal validation and external validation of the nomogram. (A) The calibration plot for training cohort. (B) The calibration plot for test cohort. (C) The calibration plot for entire cohort. (D) The calibration plot for external cohort.
Figure 12

Exploration of biological function. (A). Significantly enriched GO pathways and KEGG pathways in the entire cohort by GSEA. (A) The GSEA analysis for GO pathway in the low-risk group. (B) The GSEA analysis for GO pathway in the high-risk group. (C) The GSEA analysis for KEGG pathway in the low-risk group. (D) The GSEA analysis for GO pathway in the high-risk group.
Figure 13

(A) The immune cell between high-risk and low-risk groups in entire cohort; ***p<0.05, **p<0.01, *P <0.001;
(B) The immune related function between high-risk and low-risk groups in entire cohort; ***p<0.05, **p<0.01, *P <0.001;
(C) The checkpoint differences between high-risk and low-risk groups in entire cohort; ***p<0.05, **p<0.01, *P <0.001;
(D) The m6A expression differences between high-risk and low-risk groups in entire cohort. ***p<0.05, **p<0.01, *P <0.001.
Figure 14

Association of TME subcomponents with FARGs risk score and outcome in patients with GC. (A) Scatter plots depicting the high positive correlation between stromal score and FARGs risk score in human GC samples. Pearson’s correlation coefficient is shown in the graphs ($P < 2.2.0 \times 10^{-6}$). (B) Scatter plots depicting the no correlation between immune score and FARGs risk score in human GC samples. Pearson’s correlation coefficient is shown in the graphs ($P=0.4$). (C) Scatter plots depicting the high positive correlation between ESTIMATE score and FARGs risk score in human GC samples. Pearson’s correlation coefficient is shown in the graphs ($P < 2.2.0 \times 10^{-6}$). (D) Kaplan–Meier curves for overall survival of 684 GC patients according to stromal score. Log-rank test, $p<0.001$. (E) Kaplan–Meier curves for overall survival of 684 GC patients according to immune score. Log-rank test, $p=0.83$. (F) Kaplan–Meier curves for overall survival of 684 GC patients according to ESTIMATE score. Log-rank test, $p<0.001$. 
Figure 15

Heatmap revealing the scores of immune cells in the high-risk and low-risk groups.
Figure 16

Correlations between FDA gene expression and drug sensitivity. The figure shows the top 16 significant drug-gene pairs with significant correlation. X-axis: gene expression; y-axis: drug sensitivity Z scores.
Figure 17

Expression of FDA gene in human protein map library.
Figure 18

Univariate survival analysis of the FDA genes using Kaplan-Meier curves.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementTable1.docx
- supplementTable2.docx
- Supplementfigurelegend.docx
- supplementfigure1.jpg