## CONTENTS

**REVIEW**

- Molecular determinants of response to 5-fluorouracil-based chemotherapy in colorectal cancer: The undisputable role of micro-ribonucleic acids
  - Page 942

- Notch signalling pathway in development of cholangiocarcinoma
  - Rauff B, Malik A, Bhatti YA, Chudhary SA, Quadri I, Rafiq S
  - Page 957

**ORIGINAL ARTICLE**

**Basic Study**

- Identification of a nine-gene prognostic signature for gastric carcinoma using integrated bioinformatics analyses
  - Wu KZ, Xu XH, Zhan CP, Li J, Jiang JL
  - Page 975

**Retrospective Cohort Study**

- Prognostic significance of Borrmann type combined with vessel invasion status in advanced gastric cancer
  - Page 992

**Retrospective Study**

- Efficacy of uncovered self-expandable metallic stent for colorectal obstruction by extracolonic malignancy
  - Ahn JS, Hong SN, Chang DK, Kim YH, Kim ER
  - Page 1005

- γ-glutamyl transferase-to-platelet ratio based nomogram predicting overall survival of gallbladder carcinoma
  - Page 1014

- Clinical characteristics and outcome of primary hepatic neuroendocrine tumors after comprehensive therapy
  - Page 1031

- Oncological outcomes and predictors of radiofrequency ablation of colorectal cancer liver metastases
  - Wang CZ, Yan GX, Xin H, Liu ZY
  - Page 1044

**Observational Study**

- Methylation changes at the GNAS imprinted locus in pancreatic cystic neoplasms are important for the diagnosis of malignant cysts
  - Page 1056
CASE REPORT

1065  Response of human epidermal growth factor receptor 2-positive colorectal cancer to lapatinib monotherapy: A case report

   Guan JL, Liu JH, Wang Q, Cong YW, Chen YX, Huang KF, Huang ML, Huang L

1073  Colorectal cancer metastatic to the breast: A case report

ABOUT COVER
Editorial Board Member of World Journal of Gastrointestinal Oncology, Prof. Claudio Casella is Associate Professor of Surgery at the University of Brescia, Italy. He graduated from the University of Brescia Medical School in 1987. His post-graduate education culminated with a Digestive Surgery and Endoscopy Surgery degree in 1992. He is currently a general surgeon and oncology and endocrine surgeon specialist. His surgical track-record (in elective, urgent and emergency cases) covers all fields of general surgery, applying traditional and the latest minimally-invasive techniques. His scientific activity focuses on research of hormones and cancers, colorectal cancers, tumor markers in surgical oncology, and endocrine surgery, resulting in over 100 publications of scientific papers and communications in national and international journals. He participated in the International Study Group “Complications after Gastrectomy for Cancer”. (L-Editor: Filipodia)

AIMS AND SCOPE
The primary aim of World Journal of Gastrointestinal Oncology (WJGO, World J Gastrointest Oncol) is to provide scholars and readers from various fields of gastrointestinal oncology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

WJGO mainly publishes articles reporting research results and findings obtained in the field of gastrointestinal oncology and covering a wide range of topics including liver cell adenoma, gastric neoplasms, appendiceal neoplasms, biliary tract neoplasms, hepatocellular carcinoma, pancreatic carcinoma, cecal neoplasms, colonic neoplasms, colorectal neoplasms, duodenal neoplasms, esophageal neoplasms, gallbladder neoplasms, etc.

INDEXING/ABSTRACTING
The WJGO is now indexed in Science Citation Index Expanded (also known as SciSearch®), PubMed, and PubMed Central. The 2020 edition of Journal Citation Reports® cites the 2019 impact factor (IF) for WJGO as 2.898; IF without journal self cites: 2.880; 5-year IF: 3.316; Ranking: 143 among 244 journals in oncology; Quartile category: Q3; Ranking: 55 among 88 journals in gastroenterology and hepatology; and Quartile category: Q3.

RESPONSIBLE EDITORS FOR THIS ISSUE
Production Editor: Jia-Hui Li; Production Department Director: Xiang Li; Editorial Office Director: Jin-Lei Wang.

NAME OF JOURNAL
World Journal of Gastrointestinal Oncology

ISSN
ISSN 1948-5204 (online)

LAUNCH DATE
February 15, 2009

FREQUENCY
Monthly

EDITORS-IN-CHIEF
Rosa M Jimenez Rodriguez, Pashtoon Kasi, Monjur Ahmed

EDITORIAL BOARD MEMBERS

PUBLICATION DATE
September 15, 2020

COPYRIGHT
© 2020 Baishideng Publishing Group Inc

INSTRUCTIONS TO AUTHORS
https://www.wjgnet.com/bpg/gerinfo/204

GUIDELINES FOR ETHICS DOCUMENTS
https://www.wjgnet.com/bpg/GerInfo/287

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH
https://www.wjgnet.com/bpg/gerinfo/240

PUBLICATION ETHICS
https://www.wjgnet.com/bpg/GerInfo/288

PUBLICATION MISCONDUCT
https://www.wjgnet.com/bpg/gerinfo/208

ARTICLE PROCESSING CHARGE
https://www.wjgnet.com/bpg/gerinfo/242

STEPS FOR SUBMITTING MANUSCRIPTS
https://www.wjgnet.com/bpg/GerInfo/239

ONLINE SUBMISSION
https://www.itppublishing.com
Identification of a nine-gene prognostic signature for gastric carcinoma using integrated bioinformatics analyses

Kun-Zhe Wu, Xiao-Hua Xu, Cui-Ping Zhan, Jing Li, Jin-Lan Jiang

Abstract

BACKGROUND
Gastric carcinoma (GC) is one of the most aggressive primary digestive cancers. It has unsatisfactory therapeutic outcomes and is difficult to diagnose early.

AIM
To identify prognostic biomarkers for GC patients using comprehensive bioinformatics analyses.

METHODS
Differentially expressed genes (DEGs) were screened using gene expression data from The Cancer Genome Atlas and Gene Expression Omnibus databases for GC. Overlapping DEGs were analyzed using univariate and multivariate Cox regression analyses. A risk score model was then constructed and its prognostic value was validated utilizing an independent Gene Expression Omnibus dataset (GSE15459). Multiple databases were used to analyze each gene in the risk score model. High-risk score-associated pathways and therapeutic small molecule drugs were analyzed and predicted, respectively.

RESULTS
A total of 95 overlapping DEGs were found and a nine-gene signature (COL8A1, CTHRC1, COL5A2, AADAC, MAMDC2, SERPINE1, MAOA, COL1A2, and FNDC1) was constructed for the GC prognosis prediction. Receiver operating characteristic curve performance in the training dataset (The Cancer Genome Atlas-stomach adenocarcinoma) and validation dataset (GSE15459) demonstrated a robust...
prognostic value of the risk score model. Multiple database analyses for each gene provided evidence to further understand the nine-gene signature. Gene set enrichment analysis showed that the high-risk group was enriched in multiple cancer-related pathways. Moreover, several new small molecule drugs for potential treatment of GC were identified.

CONCLUSION
The nine-gene signature-derived risk score allows to predict GC prognosis and might prove useful for guiding therapeutic strategies for GC patients.

Key Words: Gastric carcinoma; Bioinformatic analysis; Prognosis; Overall survival; Differentially expressed genes

The nine-gene signature was identified to predict GC prognosis and prove potentially useful for guiding therapeutic strategies for GC patients.

INTRODUCTION
Gastric carcinoma (GC) is a lethal digestive malignant tumor that ranks as the fifth most commonly occurring cancer and the third cause of cancer-related death worldwide. In 2018, the global incidence and mortality of GC were estimated at 1033000 and 783000, respectively[1]. Despite advances in various therapeutic strategies, the 5-year survival rate for GC is still less than 30% and 70% of patients with GC are usually diagnosed at an advanced stage[2,3]. Therefore, it is necessary to search for a multiple-gene signature-derived model for predicting prognosis and accurately identifying anti-cancer targeted therapies to improve the prognostic stratification and personalized therapy for GC patients.

With the popularization and advancement of high throughput sequencing technologies, there is a real possibility of establishing multiple-gene signatures based on data integration and bioinformatics analysis in cancer research. The Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA) project provide invaluable resources for researchers worldwide to query gene expression and other functional genomics data[4,5]. For example, Zhao et al[6] constructed a five-gene signature based on data from TCGA databases that accurately predicted GC prognosis. Similarly, an 11-microRNA signature-derived risk score module was demonstrated to effectively predict prognosis in GC via a comprehensive genome-wide profiling analysis[7]. Therefore, it is necessary to identify genes that are significantly correlated with progression in GC patients and to further establish robust multiple-gene signatures, which could provide early diagnosis and optimized therapy for GC patients.

In the current study, GC gene expression data from TCGA and GEO datasets were first evaluated using a comprehensive bioinformatics analysis that filtered out overlapping differentially expressed genes (DEGs). Multivariate Cox regression was
applied to construct a nine-gene signature based on these identified DEGs to estimate prognosis and therapeutic outcomes in GC. The high-risk group was verified to be associated with tumor-associated pathways based on the nine-gene signature derived risk score model, which also identified promising small molecule drugs.

MATERIALS AND METHODS

GC patient data sets
The two independent GC microarray datasets GSE54129 (containing 111 GC and 21 non-cancerous samples) and GSE26899 (including 96 GC and 12 non-cancerous samples) were obtained from the GEO database and normalized using the robust multi-array average method\(^8\). Gene sequencing data and corresponding clinical information containing 375 GC samples and 32 non-cancerous samples were extracted from the TCGA-STAD (stomach adenocarcinoma) database. Subsequently, DEGs were filtrated out from the three-gene expression datasets. A flowchart of this study is showed in Figure 1.

Exploration of differentially expressed genes in GC
After standardization and log2 transformation of data from the original GEO datasets using the Affy package, the DEGs in GC were compared to normal gastric tissues and analyzed using the limma package in R software (version 3.2.1, https://www.r-project.org/)\(^9\). The \(|\log 2\text{FoldChange} (\log 2\text{FC})| \geq 1 \) and adjusted \(P\) value < 0.05 were defined as the cut-off criteria for identifying DEGs. In addition, the EdgeR package in R was used to explore DEGs for the RNA-Seq count from the TCGA database\(^10\). Data cut-off criteria were the same as described above. The upregulated/downregulated genes in the TCGA-GC cohort, GSE54129, and GSE26899 were overlapped to identify common and robust DEGs in GC.

Functional enrichment analysis
Gene ontology functional enrichment analysis was performed to expound potential biological processes, molecular functions, and cellular components for the common DEGs. Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis was performed to explore potential signaling pathways associated with overlapping genes, which might influence GC survival. All of the above analyses were performed by utilizing the online Database for Annotation, Visualization, and Integrated Discovery (DAVID, version 6.8, https://david.ncifcrf.gov/)\(^11\).

Establishment and validation of risk score model for prognosis
To further clarify the relationship between overlapping DEGs and the overall survival (OS) in GC patients, a univariate Cox proportional-hazards regression model in the TCGA-STAD cohort was utilized. Genes with a hazard ratio (HR) < 1 or > 1 were considered protective or risky, respectively. Subsequently, a multivariate Cox proportional-hazards regression analysis was performed to construct multiple DEG signatures. A risk score model was established using the Formula 1. In this equation, “coef,” represents the regression coefficient of gene \(X\) and “Expr,” is the expression value of gene \(X\) in the signature.

Gene expression data in the TCGA-STAD cohort were classified into high- and low-risk groups according to median cutoff of the risk score to evaluate the prognostic value of the risk score model. Survival differences between the two groups were compared using Kaplan-Meier (KM) survival analysis with the log-rank test. Reliability of the risk score model was assessed using the area under curve (AUC) of the receiver operating characteristic (ROC) curve.

Moreover, the reliability and prognostic value were validated using the ROC and Kaplan-Meier curves in an additional dataset GSE15459 containing 200 GC samples from the GEO database to explore whether the nine-gene signature functions as an independent prognostic factor.

Expression levels and survival analysis of nine genes in risk score model
The cBioPortal for the Cancer Genomics (http://www.cbioportal.org) database was utilized to verify a connection between genetic alterations and the nine genes. Then, Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/detail.php) was utilized to explore the expression of the nine genes at the transcriptional and translational levels, respectively. Furthermore, an OS analysis of
each gene in patients with GC was analyzed using the KM plotter database (http://kmplot.com/analysis/).

**Gene set enrichment analysis**
Gene set enrichment analysis (GSEA) (http://software.broadinstitute.org/gsea) was used to identify the promising signaling pathways for the high-risk group based on the risk score module\(^{12}\). \(P\) value < 0.05 and \(|\text{normalized enrichment score (NES)}| > 0.65\) were utilized to determine which functions to explore further.

**Identification of small molecule drug candidates**
The connectivity map (CMap) online database (http://www.broadinstitute.org) allows to investigate the interrelation among small molecule drugs, DNA microarray data, and diseases\(^{13}\). It was used to predict promising small molecule drugs involved in the overlapping DEGs from the GEO database and TCGA project that might be useful for treatment of GC.

**Statistical analysis**
Kaplan-Meier curves and log-rank method were utilized to validate the statistical criteria of observed differences in OS for low- and high-risk GC patients. The univariate and multivariate Cox proportional-hazards regression analyses were performed to estimate prognostic effects of independent genes and potential multiple-gene signatures. An ROC curve was used to evaluate the diagnostic performance of the nine-gene signature by calculating the AUC. All statistical analyses were performed using SPSS 20.0 (SPSS Inc., Chicago, IL, United States) and Prism 7.0 (GraphPad Software Inc., La Jolla, CA, United States) software. A \(P\) value < 0.05 was considered statistically significant.

**RESULTS**

**Identification of overlapping DEGs in GC**
Using the cut-off criteria, where \(P < 0.05\) and \(|\log2\text{FC}| > 1.0\), 1297 upregulated genes and 1165 downregulated genes in GSE54129, 331 upregulated genes and 173
downregulated in GSE26889, and 1034 highly expressed genes and 694 lowly expressed genes in the TCGA-STAD cohort were identified (Figure 2A). Furthermore, a total of 95 overlapping DEGs were screened out from the GEO microarray datasets and TCGA-STAD dataset, of which 59 were significantly upregulated and 36 were downregulated (Figure 2B). Hierarchical cluster heatmaps were used to explore DEG details between GC and non-cancerous tissues in each GC dataset (Figure 2C). Detailed information from the GEO datasets is shown in Supplementary Table 1.

**Gene functional enrichment analysis of overlapped genes**

Gene Ontology and Genes and Genomes Pathway analyses were used to further elucidate the potential biological function and promising signaling pathways of the overlapping genes in GC. The biological processes analysis indicated that the most genes were enriched during the cellular response to amino acid stimulus, cell chemotaxis, doxorubicin metabolic process, and extracellular matrix organization. The cellular components analysis showed that the genes were enriched in the extracellular space, extracellular region, and proteinaceous extracellular matrix. The molecular functions analysis indicated that the genes were enriched in the extracellular matrix structural constituent, oxidoreductase activity, and protease binding. Biological pathways were mainly enriched with chemical carcinogenesis, focal adhesion, drug metabolism-cytochrome P450, and PI3K/Akt signaling pathways (Figure 3).

**Identification of a nine-gene signature that predicts survival**

To determine promising biomarkers in connection with the prognosis of patients with GC, univariate Cox regression was performed to measure 95 overlapped genes in the TCGA-STAD cohort. A total of 46 genes ($P < 0.05$) were significantly correlated to OS in GC (Supplementary Table 2). These genes were then evaluated using multivariate Cox regression analysis.

Finally, a nine-gene signature ($COL8A1$, $CTHRC1$, $COL5A2$, $AADAC$, $MAMDC2$, $SERPINE1$, $MAOA$, $COL1A2$, and $FNDC1$) was constructed to assess the prognostic risk for each patient as follows: Risk score = $β_{COL8A1}ECOL8A1 + β_{CTHRC1}ECTHRC1 + β_{COL5A2}ECOL5A2 + β_{AADAC}EAADAC + β_{MAMDC2}EMAMDC2 + β_{SERPINE1}ESERPINE1 + β_{MAOA}EMAOA + β_{COL1A2}ECOL1A2 + β_{FNDC1}EFNDC1$ (Table 1), where "E" is the expression level of the genes obtained from multivariate Cox regression analysis based on the TCGA-STAD dataset.

Subsequently, patients with GC were classified into high- and low-risk groups according to the median risk score of the nine-gene signature. The ROC and KM curves were used to evaluate prognostic capacity of the nine-gene signature in GC. The AUC reached 0.751, suggesting that this nine-gene signature was relatively sensitive and specific in prognostic prediction for GC patients (Figure 4A). Moreover, results of the Kaplan–Meier curve for the two collectives indicated that patients in the low-risk group had a better OS than those in the high-risk group ($P < 0.001$; Figure 4B). Taken together, the results demonstrated that the nine-gene signature-derived risk score was significantly different for prognosis and OS between the two groups.

**Validation of the nine-gene signature**

To validate the repeatability and robustness of the nine-gene signature, an independent dataset GSE15459 was used as an external validation with the same formula. Patients in the dataset GSE15459 were divided into a high- or low-risk group with the same cutoff value as the training cohort. AUC for the nine-gene signature was calculated to be 0.682, which indicated that the model had a good prognostic capability for the survival of patients with GC in the testing collective (Figure 4C). Furthermore, in accordance with the training dataset, patients in the high-risk group had a significantly shorter OS than those in the low-risk group ($P = 0.011$; Figure 4D). These data further showed that the nine-gene signature could predict the prognosis of patients with GC.

**External validation of genetic alterations, expression levels, and survival analysis for nine genes**

Genetic alterations in the nine genes were analyzed by exploring 375 GC samples in the cbioPortal database. The results indicated that 158 (44%) samples had genetic alterations in the nine genes. Sequence mutations for each gene are shown in Figure 5A. Furthermore, expression levels for the nine genes were significantly different ($COL8A1$, $CTHRC1$, $COL5A2$, $SERPINE1$, $COL1A2$, and $FNDC1$ were upregulated and $AADAC$, $MAOA$, and $MAMDC2$ were downregulated) in GC tumor
Wu KZ et al. Prognosis prediction of gastric cancer patients

Table 1 Nine prognosis-associated genes for establishing the risk score system

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Description</th>
<th>Coef</th>
<th>HR</th>
<th>95%CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>COL8A1</td>
<td>Collagen type VIII alpha 1 chain</td>
<td>-0.39134</td>
<td>0.6761</td>
<td>0.4974-0.9191</td>
<td>0.01247*</td>
</tr>
<tr>
<td>CTHRC1</td>
<td>Collagen triple helix repeat containing 1</td>
<td>0.36470</td>
<td>1.4401</td>
<td>1.1432-1.8140</td>
<td>0.00196*</td>
</tr>
<tr>
<td>COL5A2</td>
<td>Collagen type V alpha 2 chain</td>
<td>0.45550</td>
<td>1.5770</td>
<td>0.9116-2.7278</td>
<td>0.10329</td>
</tr>
<tr>
<td>AADAC</td>
<td>Arylacetamide deacetylase</td>
<td>0.14823</td>
<td>1.1598</td>
<td>1.0437-1.2887</td>
<td>0.00585*</td>
</tr>
<tr>
<td>MAMDC2</td>
<td>MAM domain containing 2</td>
<td>0.21034</td>
<td>1.2341</td>
<td>1.0514-1.4485</td>
<td>0.01006*</td>
</tr>
<tr>
<td>SERPINE1</td>
<td>Serpin family E member 1</td>
<td>0.23183</td>
<td>1.2609</td>
<td>1.0583-1.5023</td>
<td>0.00949*</td>
</tr>
<tr>
<td>MAOA</td>
<td>Monoamine oxidase A</td>
<td>0.17086</td>
<td>1.1863</td>
<td>1.0151-1.3865</td>
<td>0.03172</td>
</tr>
<tr>
<td>COL1A2</td>
<td>Collagen type I alpha 2 chain</td>
<td>-0.54104</td>
<td>0.5821</td>
<td>0.3462-0.9790</td>
<td>0.04135*</td>
</tr>
<tr>
<td>FNDC1</td>
<td>Fibronectin type III domain containing 1</td>
<td>0.23264</td>
<td>1.2619</td>
<td>0.9808-1.6236</td>
<td>0.07041</td>
</tr>
</tbody>
</table>

*P < 0.05 was considered statistically significant. HR: Hazard ratio.

The initial step in this study was to identify the DEGs in GC using analysis of gene expression data from the TCGA and GEO datasets. These results showed that a total of tissues compared to non-cancerous tissues based on the GEPIA database (Figure 5B). KM plotter was used to study the prognostic performance of each gene in GC. The results identified that high COL8A1, CTHRC1, COL5A2, SERPINE1, COL1A2, MAMDC2, and FNDC1 expression was related to a worse prognosis, while high MAOA, and AADAC expression was related to a better prognosis in GC patients (Figure 6).

Gene set enrichment analysis of high-risk group

GSEA analysis was performed to explore potential signaling pathways associated with the high-risk group based on the nine-gene signature-derived risk score. The cut-off value was set at P < 0.05 and |enrichment score (ES)| > 0.65. Results showed that multiple tumor-associated pathways, such as angiogenesis, epithelial-mesenchymal transition, hedgehog signaling, Kirsten rat sarcoma viral oncogene homologue signaling, Notch signaling, and transforming growth factor (TGF)-β signaling, were enriched in the high-risk group GC patients (Figure 7).

Identification of related small molecule drugs

The nine-gene signature was further analyzed in the CMap database to predict potential small molecule drugs for GC. Ten small molecule drugs were revealed using the high connectivity score and P value < 0.05 (Table 2). A total of nine small molecule candidates were negatively correlated. The 3D conformers for the top five most significant candidates are shown in Supplementary Figure 1. All findings indicated that these drugs had potential therapeutic applications in GC.

DISCUSSION

Despite considerable development in the arena of various therapeutic GC strategies, including surgery, radiotherapy, chemotherapy, and targeted precise treatment, the OS of advanced GC patients has remained poor and the therapeutic effect is often unsatisfying. The current prognostic model established based on clinical prognostic factors, such as age, TNM stage, and pathology grade, is a routine predictive model for GC. However, because of the high GC heterogeneity, a conventional prognostic model cannot accurately predict the outcomes for GC patients. Multiple-gene assays, by contrast, are of great importance for precision medicine for GC patients[14]. Therefore, exploring potential molecular mechanisms and effective therapeutic targets is important for GC therapy and prevention. This study performed an integrated bioinformatics analysis to establish a nine-gene risk score model associated with prognosis and treatment response in GC patients. The high-risk group was identified to relate to tumor-associated signaling pathways based on the nine-gene signature-derived risk score and several novel small molecule drugs were discovered for potential GC treatment.

The initial step in this study was to identify the DEGs in GC using analysis of gene expression data from the TCGA and GEO datasets. These results showed that a total of...
95 overlapping DEGs were identified compared to normal gastric tissue. The common genes were further evaluated using functional enrichment analyses. The results indicated that common DEGs play a crucial important role in cancerous development. Subsequently, univariate and multivariate Cox regression analyses were performed to explore the relationship between DEGs and GC survival. A total of six genes (COL5A1, CTHRC1, COL5A2, SERPINE1, COL1A2, and FNDC1) were upregulated in GC and inversely correlated with OS (β > 0, HR > 1), whereas three genes (AADAC, MAMDC2, and MAOA) were downregulated and positively correlated with survival (β < 0, HR < 1). A novel multi-gene signature-derived risk score model was constructed using these nine DEGs. A comprehensive examination of the nine-gene signature prognostic value in the training (TCGA-STAD) and testing (GSE15459) datasets was carried out. ROC curve and Kaplan–Meier analysis performance in the training and validation datasets underscored the robust prognostic value of the risk score model. Among these nine genes, COL5A2, COL1A2, and COL8A1 are members of the collagen family, which is the main structural component of the extracellular matrix in tumors. COL5A2 encodes alpha 2 chain in type V collagen and is aberrantly expressed in ductal cancer in situ and invasive ductal cancer. It promotes the progression of cancer in situ to invasive cancer. Consistent with the present study, Hao et al. identified COL5A2 as a key gene in GC that serves as an oncogene associated with poor OS. Similarly, COL1A2 was reported to inhibit GC cell apoptosis and promote GC cell proliferation, invasion, and migration via the PI3k/Akt signaling pathway. COL5A2 was found to be upregulated and relevant to the poor clinical outcomes in multiple carcinomas, such as colon adenocarcinoma and bladder cancer. Its regulatory mechanism in GC remains unclear. CTHRC1 is a major glycosylated protein that has been demonstrated to be associated with cell proliferation, metastasis, and invasion via promoter demethylation and TGF-β1. It is also an independent prognostic predictor. SERPINE1, also known as PAI-1, can increase GC metastasis and promote peritoneal tumor growth and formation of bloody ascites in a mouse model of GC metastasis, which serves as an important prognostic gene in GC. FNDC1 is a principal component of the fibronectin structural domain that accelerates GC cell proliferation, differentiation, and metastasis via the epithelial-mesenchymal mechanism pathway. It also plays an important role in carcinogenesis in multiple cancers. AADAC is a microsomal serine esterase that mainly exists in the liver and gastrointestinal tract. Its main function is involved in drug hydrolysis, as well as triglyceride metabolism and Gilles de la Tourette syndrome. Liu et al. analyzed the prognostic genes in GC using bioinformatics analysis and found that ADAC is a significant tumor suppressor gene. MAOA degrades monoamine neurotransmitters and dietary amines and was also deemed to be a tumor suppressor in liver cancer, pancreatic cancer, and cholangiocarcinoma and a tumor promoter in prostate carcinoma, breast carcinoma, and non-small cell lung carcinoma. Its role in GC progression is poorly understood. There are few studies on the role of MAMDC2 in tumors. A meta-analysis study indicated that down-regulated MAMDC2 was related
Figure 2  Differentially expressed genes between gastric carcinoma and normal gastric tissues. A: Volcano plots visualizing the differentially expressed genes (DEGs) between gastric carcinoma and non-cancerous tissues in GSE54129, GSE26899, and The Cancer Genome Atlas datasets. Red dots represent significantly up-regulated genes; green dots represent significantly down-regulated genes; black dots represent non-differentially expressed genes. $P < 0.05$ and $|\log_2 \text{FC}| > 1.0$ were considered significant; B: Venn diagrams showing the upregulated overlapped DEGs (left) and downregulated overlapped DEGs (right) in three datasets; C: Heatmaps of the common genes in GSE54129, GSE26899, and The Cancer Genome Atlas datasets (top 50). The common genes include 59
upregulated genes and 36 downregulated genes. Each row represents the expression level of a gene, and each column represents a sample: Red for gastric carcinoma and blue for non-cancerous samples.

to a poor disease-free survival in breast carcinoma\cite{35}. Another study reported that miR-196a promotes head and neck squamous cell cancer migration, invasion, and adhesion to fibronectin via MAMDC2\cite{36}. The multiple database analysis of physiological and pathological functions for each gene in GC has provided important evidence helping to understand the prognostic and predictive capacity of the nine-gene signature.

GSEA analysis was utilized in order to provide a deeper insight into the molecular mechanisms for prognosis prediction of the nine-gene signature. Multiple cancer-associated signaling pathways were highlighted as a result, including angiogenesis, epithelial-mesenchymal transition, hedgehog signaling, Kirsten rat sarcoma viral oncogene homologue signaling, Notch signaling, and TGF-β signaling, which suggested that the nine-gene signature has predictive ability for prognosis and can reveal potential therapeutic targets in GC.

Therefore, the CMap database was utilized to explore promising small molecule drugs that have effective treatment response against GC. Levomepromazine, which belongs to antihistaminic compounds, is mainly used for treating breast cancer by binding to the translationally controlled tumor protein and induction of cell differentiation\cite{37}. Lasalocid is a carboxylic ionophore antibiotic produced by Streptomyces lasaliensis that is recognized as a choice for prostate cancer therapy because it increases cytotoxic apoptosis and cytoprotective autophagy\cite{38}. Trichostatin A is a histone deacetylase inhibitor that inhibits proliferation, migration, and invasion of GC cells and shows a good therapeutic effect in GC patients\cite{39,40}. Similarly, vorinostat is a histone deacetylase inhibitor approved by the United States Food and Drug Administration for cutaneous T-cell lymphoma. It is a promising therapeutic candidate in GC when combined with chemotherapeutic agents\cite{41}. Therefore, the present study suggested that these small molecule drugs could serve as novel therapeutic strategies for the high-risk GC group with a poor prognostic response.

One study reported that six genes related to GC prognosis based on DNA microarray data of 65 patients successfully prognosticated relapse in GC patients\cite{42}. A recent study identified a three-gene signature, which could predict GC survival using DNA microarray data of 129 GC patients\cite{43}. These studies were limited due to the small sample size or lack of suitable verification datasets, which limits the possibility of clinical application of the genes related to the GC prognosis. In the present study, a novel nine-gene signature was identified by examining the gene expression profile of 582 GC patients, which had a robustly effective prognostic capacity in GC.

However, this study includes some limitations. First, because the main sources of data in this study were downloaded from public databases that are constructed using available retrospective data, it is necessary to assess the probable utilization of molecular signatures for prognosis evaluation. Second, further studies including a greater number of GC patients are needed to verify the efficiency of the nine-gene signature in GC patients. A greater number of normal samples should also be included in the differential expression analyses. Moreover, multivariate Cox regression analysis was performed to obtain the expression level of multiple genes. More clinical events will be included to verify the prognosis effect of the nine-gene signature in further studies.

In conclusion, using a series of comprehensive bioinformatics analyses and validations, a novel nine-gene signature was constructed. The signature-derived risk score model had a robust prognostic capacity and therapeutic response in GC. Several small molecule drugs were identified to serve as potential therapeutic candidates for GC using bioinformatics. Further experimental studies are necessary to validate these findings and to elucidate the mechanisms for GC-related signaling pathways.
Figure 3  Functional analysis of differentially expressed genes. Significantly enriched gene ontology biological processes of differentially expressed genes in gastric carcinoma are shown. A: Biological Process; B: Cellular Components; C: Molecular Function; and D: Significantly Enriched Kyoto Encyclopedia of Genes and Genomes Pathways of DEGs in Gastric carcinoma. KEGG: Kyoto Encyclopedia of Genes and Genomes.
Figure 4  Performance of the risk score model in the training (the Cancer Genome Atlas-stomach adenocarcinoma) and validation (GSE15459) datasets. A: Receiving operating characteristic curve of the nine-gene signature in the training dataset (area under the curve = 0.751); B: Kaplan–Meier survival curve for gastric carcinoma patients in the training dataset (P < 0.001). C: Receiving operating characteristic curve of the nine-gene signature in the validation datasets (area under the curve = 0.682); D: Kaplan–Meier survival curve for gastric carcinoma patients in the validation dataset (P = 0.011). The blue curve represents low risk score group. The red curve represents high risk score group. AUC: Area under the curve.
Figure 5  Genetic alterations and expression of the nine prognostic genes in gastric carcinoma. A: Alteration proportion for the nine genes in 375 gastric carcinoma samples in the cBioPortal database; B: Gene expression levels of the nine genes between gastric carcinoma and normal gastric tissues in Gene
Expression Profiling Interactive Analysis database. Red represents $P < 0.05$.  

Figure 6  Prognostic value of the nine prognostic genes in gastric carcinoma. A: COL8A1 ($P < 0.001$); B: CTHRC1 ($P < 0.001$); C: COL5A2 ($P = 0.004$); D: AADAC ($P = 0.004$); E: MAMDC2 ($P < 0.001$); F: SERPINE1 ($P < 0.001$); G: MAOA ($P = 0.003$); H: COL1A2 ($P < 0.001$); I: MAOA ($P < 0.001$).
Figure 7  Promising signaling pathways identified by Gene Set Enrichment Analysis. Only six of the most significant tumor-associated pathways enriched in high-risk group gastric carcinoma patients are listed. A: Hedgehog signaling; B: Kirsten rat sarcoma viral oncogene homologue signaling; C: Notch signaling; D: TGF-β signaling; E: Angiogenesis; F: Epithelial-mesenchymal transition.
ARTICLE HIGHLIGHTS

Research background
With the popularization and advancement of high throughput sequencing technologies, there is a real possibility of establishing multiple-gene signatures based on data integration and bioinformatics analysis in cancer research. The present study aimed to identify prognostic biomarkers for gastric carcinoma (GC) patients using comprehensive bioinformatics analyses.

Research motivation
GC is one of the most aggressive primary digestive tumors. It has unsatisfactory therapeutic outcomes and is difficult to diagnose early. Therefore, it is necessary to search for a multiple-gene signature-derived model for predicting prognosis and accurately identifying anti-cancer targeted therapies to improve the prognostic stratification and personalized therapy for GC patients.

Research objectives
We aimed to explore the potential multiple-gene prognostic biomarkers and effective therapeutic targets for GC. In this study, we performed integrated bioinformatics analysis to establish a nine-gene risk score model (COL8A1, CTHRC1, COL5A2, SERPINE1, COL1A2, FNDC1, AADAC, MAOA, and MAMDC2) associated with prognosis and treatment response in GC patients. The nine-gene signature-derived risk score allows to predict GC prognosis and might prove useful for guiding therapeutic strategies for GC patients.

Research methods
Differentially expressed genes (DEGs) were screened using gene expression data from The Cancer Genome Atlas and GEO databases for GC. Overlapping DEGs were analyzed using univariate and multivariate Cox regression analyses. A risk score model was then constructed and signature prognostic values were validated utilizing an independent GEO dataset (GSE15459). CBioPortal, GEPIA, and KM-plotter databases were used to analyze each gene in the risk score model. Gene set enrichment analysis and the connectivity map database were used to predict high-risk score-associated pathways and therapeutic small molecule drugs, respectively.

Research results
A total of 95 overlapping DEGs were found and a nine-gene signature (COL8A1, CTHRC1, COL5A2, AADAC, MAMDC2, SERPINE1, MAOA, COL1A2, and FNDC1) was constructed for the GC prognosis prediction. Receiver operating characteristic curve performance in the training dataset (The Cancer Genome Atlas- stomach adenocarcinoma) and validation dataset (GSE15459) demonstrated a robust prognostic value of the risk score model. Multiple database analyses for each gene provided evidence to further understand the nine-gene signature. Gene set enrichment analysis showed that the high-risk group was enriched in multiple cancer-related pathways. Moreover, several new small molecule drugs for potential treatment of GC were identified.

Research conclusions
Using a series of comprehensive bioinformatics analyses and validations, a novel nine-gene signature was constructed. The signature-derived risk score model had a robust prognostic capacity and therapeutic response in GC. Several small molecule drugs were identified to serve as potential therapeutic candidates for GC using bioinformatics. Further experimental studies are necessary to validate these findings and to elucidate the mechanisms for GC-related signaling pathways.

Research perspectives
Multiple-gene assays are of great importance for precision medicine of GC patients. To further verify the prognostic capacity of the nine-gene signature, our future study may pay more attention to exploring the potential regulatory mechanisms how the nine-gene signature affects the development of GC.
REFERENCES


