MINIREVIEWS

4734 Inflammatory myofibroblastic tumor of the distal common bile duct: Literature review with focus on pathological examination
Cordier F, Hoorens A, Ferdinande L, Van Dorpe J, Creytens D

4740 Probiotics and autoprobiotics for treatment of Helicobacter pylori infection
Baryshnikova NV, Ilina AS, Ermolenko EI, Uspsenskiy YP, Suvorov AN

4752 Plant-based diet and its effect on coronary artery disease: A narrative review

ORIGINAL ARTICLE
Clinical and Translational Research

4763 Identification of survival-associated biomarkers based on three datasets by bioinformatics analysis in gastric cancer

4788 High expression of autophagy-related gene EIF4EBP1 could promote tamoxifen resistance and predict poor prognosis in breast cancer

4800 Delineation of fatty acid metabolism in gastric cancer: Therapeutic implications

4814 Mechanical analysis of the femoral neck dynamic intersection system with different nail angles and clinical applications
Wang Y, Ma JX, Bai HH, Lu B, Sun L, Jin HZ, Ma XL

Retrospective Cohort Study

4824 Development and validation of a predictive model for spinal fracture risk in osteoporosis patients
Lin XM, Shi ZC

Retrospective Study

4833 Risk prediction model for distinguishing Gram-positive from Gram-negative bacteremia based on age and cytokine levels: A retrospective study
Zhang W, Chen T, Chen HJ, Chen N, Xing ZJ, Fu XY

4843 Sudden death in the southern region of Saudi Arabia: A retrospective study
Al-Emam AMA, Dajam A, Alraiji M, Alfaifi W, Al-Shraim M, Helaly AM
## Contents

Thrice Monthly Volume 11 Number 20 July 16, 2023

### Diagnostic value of preoperative examination for evaluating margin status in breast cancer
4852
Liu P, Zhao Y, Rong DD, Li KF, Wang YJ, Zhao J, Kang H

### Prospective Study
4865
Defining the awareness and attitude of the clinicians through pharmacovigilance in Turkey
Aydin OC, Aydin S, Guney HZ

### Predictive value of the trans-perineal three-dimensional ultrasound measurement of the pubic arch angle for vaginal delivery
4874
Liang ZW, Gao WL

### CASE REPORT
4883
Microwave ablation of solitary T1N0M0 papillary thyroid carcinoma: A case report
Dionísio T, Lajut L, Sousa F, Violante L, Sousa P

4890
Acute spinal subdural haematoma complicating a posterior spinal instrumented fusion for congenital scoliosis: A case report
Michon du Marais G, Tabard-Fougère A, Dayer R

4897
Subacute osteomyelitis due to *Staphylococcus caprae* in a teenager: A case report and review of the literature
Vazquez O, De Marco G, Gavira N, Habre C, Bartuez M, Steiger CN, Dayer R, Ceroni D

4903
ABCB4 gene mutation-associated cirrhosis with systemic amyloidosis: A case report
Cheng N, Qin YJ, Zhang Q, Li H

4912
Metagenomic next-generation sequencing in the diagnosis of neurocysticercosis: A case report
Xu WB, Fu JJ, Yuan XJ, Xian QJ, Zhang LJ, Song PP, You ZQ, Wang CT, Zhao QG, Pang F

4920
Drug-coated balloons for treating *de novo* lesions in large coronary vessels: A case report
Zhang ZQ, Qin YR, Yin M, Chen XH, Chen L, Liang WY, Wei XQ

4926
Pretreatment with a modified St. Thomas’ solution in patients with severe upper limb injuries: Four case reports
Sun ZY, Li LY, Xing JX, Tong LC, Li Y

4932
Unexpected diffuse lung lesions in a patient with pulmonary alveolar proteinosis: A case report
Jian L, Zhao QQ

4937
Contrast-induced ischemic colitis following coronary angiography: A case report
Qiu H, Li WP

4944
Posterior pedicle screw fixation combined with local steroid injections for treating axial eosinophilic granulomas and atlantoaxial dislocation: A case report
Tu CQ, Chen ZD, Yao XT, Jiang YJ, Zhang BF, Lin B

4956
Antithrombin III deficiency in a patient with recurrent venous thromboembolism: A case report
Luo JQ, Mao SS, Chen JY, Ke XY, Zhu YF, Huang W, Sun HM, Liu ZJ
<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>4961</td>
<td>Laryngospasm as an uncommon presentation in a patient with anti-N-methyl-D-aspartate receptor encephalitis: A case report</td>
<td>Wang L, Su HJ, Song GJ</td>
</tr>
</tbody>
</table>
ABOUT COVER
Editorial Board Member of *World Journal of Clinical Cases*, Kengo Moriyama, MD, PhD, Associate Professor, Department of Clinical Health Science, Tokai University School of Medicine, Tokai University Hachioji Hospital, Hachioji 1838, Tokyo, Japan. osaru3moving@gmail.com

AIMS AND SCOPE
The primary aim of *World Journal of Clinical Cases* (*WJCC, World J Clin Cases*) is to provide scholars and readers from various fields of clinical medicine with a platform to publish high-quality clinical research articles and communicate their research findings online.

*WJCC* mainly publishes articles reporting research results and findings obtained in the field of clinical medicine and covering a wide range of topics, including case control studies, retrospective cohort studies, retrospective studies, clinical trials studies, observational studies, prospective studies, randomized controlled trials, randomized clinical trials, systematic reviews, meta-analysis, and case reports.

INDEXING/ABSTRACTING
The *WJCC* is now abstracted and indexed in Science Citation Index Expanded (SCIE, also known as SciSearch®), Journal Citation Reports/Science Edition, Current Contents®/Clinical Medicine, PubMed, PubMed Central, Reference Citation Analysis, China National Knowledge Infrastructure, China Science and Technology Journal Database, and Superstar Journals Database. The 2023 Edition of Journal Citation Reports® cites the 2022 impact factor (IF) for *WJCC* as 1.1; IF without journal self cites: 1.1; 5-year IF: 1.3; Journal Citation Indicator: 0.26; Ranking: 133 among 167 journals in medicine, general and internal; and Quartile category: Q4.

RESPONSIBLE EDITORS FOR THIS ISSUE
Production Editor: Hao-Ge Yu; Production Department Director: Xu Guo; Editorial Office Director: Jin-Lai Wang.

NAME OF JOURNAL
*World Journal of Clinical Cases*

ISSN
ISSN 2307-8960 (online)

LAUNCH DATE
April 16, 2013

FREQUENCY
Thrice Monthly

EDITORS-IN-CHIEF
Bao-Gan Peng, Jerzy Tadeusz Chudek, George Kontogeorgos, Maurizio Serati, Ja Hyeon Ku

EDITORIAL BOARD MEMBERS
https://www.wjgnet.com/2307-8960/editorialboard.htm

PUBLICATION DATE
July 16, 2023

COPYRIGHT
© 2023 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.f6publishing.com
Identification of survival-associated biomarkers based on three datasets by bioinformatics analysis in gastric cancer

Long-Kuan Yin, Hua-Yan Yuan, Jian-Jun Liu, Xiu-Lian Xu, Wei Wang, Xiang-Yu Bai, Pan Wang

BACKGROUND
Gastric cancer (GC) is one of the most common malignant tumors with poor prognosis in terms of advanced stage. However, the survival-associated biomarkers for GC remains unclear.

AIM
To investigate the potential biomarkers of the prognosis of patients with GC, so as to provide new methods and strategies for the treatment of GC.

METHODS
RNA sequencing data from The Cancer Genome Atlas (TCGA) database of STAD tumors, and microarray data from Gene Expression Omnibus (GEO) database (GSE19826, GSE79973 and GSE29998) were obtained. The differentially expressed genes (DEGs) between GC patients and health people were picked out using R software (x64 4.1.3). The intersections were underwent between the above obtained co-expression of differential genes (co-DEGs) and the DEGs of GC from Gene Expression Profiling Interactive Analysis database, and Gene Ontology (GO) analysis, Kyoto Encyclopedia of Gene and Genome (KEGG) pathway analysis, Gene Set Enrichment Analysis (GSEA), Protein-protein Interaction (PPI) analysis and Kaplan-Meier Plotter survival analysis were performed on these DEGs. Using Immunohistochemistry (IHC) database of Human Protein Atlas (HPA), we verified the candidate Hub genes.

RESULTS
With DEGs analysis, there were 334 co-DEGs, including 133 up-regulated genes and 201 down-regulated genes. GO enrichment analysis showed that the co-DEGs...
were involved in biological process, cell composition and molecular function pathways. KEGG enrichment analysis suggested the co-DEGs pathways were mainly enriched in ECM-receptor interaction, protein digestion and absorption pathways, etc. GSEA pathway analysis showed that co-DEGs mainly concentrated in cell cycle progression, mitotic cell cycle and cell cycle pathways, etc. PPI analysis showed 84 nodes and 654 edges for the co-DEGs. The survival analysis illustrated 11 Hub genes with notable significance for prognosis of patients were screened. Furtherly, using IHC database of HPA, we confirmed the above candidate Hub genes, and 10 Hub genes that associated with prognosis of GC were identified, namely BGN, CEP55, COL1A2, COL4A1, FZD2, MAOA, PDGFRB, SPARC, TIMP1 and VCAN.

CONCLUSION
The 10 Hub genes may be the potential biomarkers for predicting the prognosis of GC, which can provide new strategies and methods for the diagnosis and treatment of GC.

Key Words: Gastric cancer; Survival-associated biomarkers; Bioinformatics analysis; Hub genes

©The Author(s) 2023. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: Gastric cancer (GC) is one of the most common leading cause of death worldwide. The cases with advanced GC usually have poor prognosis. To date, the prognostic biomarkers of GC remain unclear. In this article, we investigated the co-expression of differential genes (co-DEGs) between GC tissues and normal tissues based on the data from Gene Expression Omnibus, Gene Expression Profiling Interactive Analysis and The Cancer Genome Atlas. By using bioinformatics analysis, the signal pathways of co-DEGs involvement in GC were identified, and 10 Hub biomarkers for the survival of GC were screened.

INTRODUCTION
Gastric cancer (GC) is one of the most common malignant tumors with high morbidity and mortality[1-3]. It is known that the occurrence of GC is the result of multiple factors. The genetic factors, dietary habits and Helicobacter pylori infection play a very important role in the occurrence and development of GC[4-6]. The most common screening methods for GC are gastroscopy and pathological examination, which can effectively improve the detection rate of early GC[7,8]. However, those patients, who with advanced GC, have poor treatment effect and poor prognosis[9].

The development of GC is a complex pathological process involving changes of various genes and pathways[10]. Previous studies have shown that some changes were happened between the GC tissues and normal tissues, especially the expression of the genes[11]. The biomarkers measured in different stages of GC are helpful as indicators of early diagnosis, routine screening, postoperative monitoring or pharmacological response to a therapeutic intervention[12]. Therefore, exploring the survival related biomarkers of GC may provide more approaches for the treatment of GC, improving the overall survival time of the patients. To date, the survival related biomarkers for GC remains unclear. Currently, the development of high-throughput sequencing technology has generated a large number of functional genomic data[13], making it possible to reveal the survival related biomarkers of GC by analyzing the differential gene expression data between the GC tissues and the normal tissues. In recent years, bioinformatics is widely used to analyze the genomic and proteomic data of tumors, and to reveal the function of gene products at the molecular level for cancer[14].

In this study, bioinformatics strategy was used to obtain data from Gene Expression Omnibus (GEO), Gene Expression Profiling Interactive Analysis (GEPIA) and The Cancer Genome Atlas (TCGA). In briefly, the software of Gene Expression Profiling Interactive Analysis (GEPIA), R software (x64.1.3), STRING, Kaplan-Meier plotter and Human Protein Atlas (HPA), were performed to analyze and integrate the mRNA expression data of GC tissues and adjacent tissues or normal gastric tissues to explore the molecular functions (MF) of differential genes and signal pathways of GC. Finally, we successfully screened the genes of 10 key biomarkers for the survival of GC. The present study analyzed the high-throughput data of multi database and multi-chip datasets, which could more accurately reveal the potential prognosis biomarkers of GC. The detailed analysis workflow is as shown in Figure 1.
Yin LK et al. Bioinformatics analysis of survival-associated biomarkers of GC

Figure 1 Flowchart for research into bioinformatics data from Gene Expression Omnibus, The Cancer Genome Atlas and Gene Expression Profiling Interactive Analysis. GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; GEPIA: Gene Expression Profiling Interactive Analysis; DEGs: Differentially expressed genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Gene and Genome; GSEA: Gene Set Enrichment Analysis.

MATERIALS AND METHODS

Data sets download
GDC TCGA Stomach Cancer (STAD) related datasets were downloaded from UCSC Xena (https://xenabrowser.net/) database[15], including 373 GC tissue samples and 32 normal gastric tissue samples. R software (x64 4.1.3) was used to download and analyze datasets of gene expression profiles from GEO database (GSE19826, GSE79973 and GSE29998) with Tidyverse and query packages, including 72 GC tissue samples and 74 normal gastric tissue samples.

Differentially expressed genes screening
Based on TCGA data sets, the Tidyverse and DESeq2R packages of software (x64 4.1.3) were used for differential gene expression analysis. RNA sequencing data from normal and tumor tissue samples were extracted for analysis. Volcano map was drawn to show the folding changes and P values of differentially expressed genes (DEGs) (|LogFC| ≥ 1, adjusted P value < 0.05). R software dealt with GEO data sets (GSE19826, GSE79973 and GSE29998, respectively), then the ID was converted, and the adjust P value < 0.05 and |LogFC| ≥ 1 were set as the cut off criterion. Subsequently, a Venn
diagram method was used to screen out co-expression of differential genes (co-DEGs). Of these co-DEGs, only protein-coding genes were further analyzed.

**Functional enrichment analysis of co-DEGs**

GEPIA\[16] is a newly developed interactive web server for analyzing the RNA sequencing expression data from the TCGA and the GTEx projects. The GEPIA data sets were analyzed using the stat packet of R software to predict the potential functions of the co-DEGs. The functional analyses of Gene Ontology (GO) analysis, Kyoto Encyclopedia of Gene and Genome (KEGG) pathway analysis and Gene Set Enrichment Analysis (GSEA) pathway analysis were performed on the co-DEGs by using R software. \( P < 0.05 \) was considered statistically significant.

**Protein-protein interaction network analysis of co-DEGs**

Protein-protein interaction (PPI) Analysis of the identified DEGs was constructed by STRING 11.5. The STRING Database\[17\] (https://cn.string-db.org/) is a database of known and predicted PPIs. Using STRING 11.5, the interaction network between the above co-DEGs and the related genes was presented by setting the maximum confidence at 0.9. Furthermore, and Cytoscape_v3.9.1 software was carried out for analyzing and mapping.

**Survival analysis of co-DEGs**

Using Boxplot functions of GEPIA database, we set \(| \log 2 FC | \geq 1, P-\text{value cut off} < 0.05 \) as the cut off criterion. The tumor-related data from STAD database were selected to match the normal gastric tissue data of TCGA and GTEx, the genes with significant differences were screened out, and box-plot was performed for the screened Hub genes.

**Verifying the Hub genes again**

Kaplan-Meier plotter\[18\] was able to assess the association between 30 K gene (mRNA, microRNA, protein) expression and survival in 25 K + samples from 21 tumor types, including breast cancer, ovarian cancer, lung cancer, GC, etc. We used Kaplan-Meier plotter online tool to perform visualization analysis of GC related database again, and verified the above results.

**Immunohistochemical analysis to determine the final Hub genes**

Due to the highly specific characteristics of antigen and antibody binding, immunohistochemistry (IHC) can reveal the relative distribution and abundance of proteins. Then, through HPA\[19\] (https://www.proteinatlas.org), currently the largest and most comprehensive Human tissue Protein spatial database, we can more intuitively observe the difference of Hub gene expression between normal stomach tissue and GC tissue. \( \chi^2 \) test was used to compare the difference between the two groups, and \( P < 0.05 \) is statistically significant.

## RESULTS

### Identifying DEGs

The TCGA STAD counts dataset was processed by using tidyverse and DESeq2 packages of R software (x64 4.1.3), and DEGs was screened (\(| \log 2 FC | \geq 1, P-\text{value cut off} < 0.05 \) as the cut off criterion). Our results showed that there were 2133 up-regulated genes and 2349 down-regulated genes. R software (x64 4.1.3) was used to process data sets (GSE19826, GSE79973 and GSE29998), and then ID conversion was performed. Limma package was used to process the above three data sets respectively, and the screening criteria were \(| \log 2 FC | \geq 1 \) and \( P-\text{value} < 0.05 \). There were 2202 up-regulated genes and 2700 down-regulated genes in GSE19826, 665 up-regulated genes and 1507 down-regulated genes in GSE79973, 4346 up-regulated genes and 3002 down-regulated genes in GSE29998, respectively. Heat maps were drawn by Pheatmap package of R software (x64 4.1.3) (Figures 2A, C and E), and volcano maps were drawn by ggplot2 package of R software (x64 4.1.3) (Figures 2B, D and F).

The TCGA STAD FPKM data set was processed by tidyverse and Pheatmap packages of R software (x64 4.1.3) combined with DEGs, and the heatmap was drawn (Figure 2G). Tidyverse, GGploT2 packages and DEGs were used to draw the volcano map (Figure 2H).

Venn Diagram package of R software (x64 4.1.3) was used to make Venn Diagram of DEGs in GEO and TCGA datasets. A total of 334 DEGs were obtained, including 133 up-regulated genes and 201 down-regulated genes (Figures 3A and B).

### GO functional enrichment analysis of co-DEGs

GO functional enrichment analysis of co-DEGs showed that the GO annotation of co-DEGs was divided into three parts: Biological process (BP), cell composition (CC) and MF, and a diagram of which was shown in Figures 4A-E and Table 1 (list of top 5 GO pathways). Arraying the ascending order of \( P \) value \(( P < 0.05 \)), our results revealed that the GO pathways of DEGs were enriched as the follows: Collagen fibrillary tissue, cell division, extracellular matrix organization, skeletal system development and copper ion detoxification, etc., were enriched in BP; extracellular space, extracellular matrix, extracellular region, collagen trimer and centromere, etc., were enriched in CC; extracellular matrix structural components, extracellular matrix structural components that give tensile strength, platelet-derived growth factor binding, zinc ion binding, creatine kinase activity and so on, were enriched in MF (details in Supplementary Tables 1 and 2).
**Table 1 Gene ontology enrichment pathways of co-expression of differential genes (top 5)**

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>Count in gene set</th>
<th>(P) value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GOTERM_BP_DIRECT</strong></td>
<td>GO:0005199: Collagen fibril organization</td>
<td>20</td>
<td>7.76E-08</td>
<td>CENPW, UBE2C, RCC2, CDC47, KIF14, NCPG, NDC80, CDC25B, CDC20, CCNB1, PRC1, NUF2, CDK1, NEK2, KIF2C, BUB1, MAD2L1, SPC25</td>
</tr>
<tr>
<td></td>
<td>GO:0051001: Extracellular matrix organization</td>
<td>13</td>
<td>2.67E-07</td>
<td>OLFML2B, MMP7, MMPI, TNFRSF11B, COL3A1, ADAMTS2, COL1A2, COL5A1, COL4A1, COL5A2, COL6A1, COL10A1</td>
</tr>
<tr>
<td></td>
<td>GO:0001001: Skeletal system development</td>
<td>12</td>
<td>3.61E-07</td>
<td>TEAD4, COMP, COL3A1, VCAN, PKDCC, COL1A2, CDH11, COL5A2, COL10A1, TNFRSF11B, HOXA13, HOXC10</td>
</tr>
<tr>
<td></td>
<td>GO:0010273: Detoxification of copper ion</td>
<td>6</td>
<td>7.19E-07</td>
<td>MT2A, MT1M, MT1G, MT1H, MT1X, MT1E</td>
</tr>
<tr>
<td></td>
<td><strong>GOTERM_CC_DIRECT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GO:0005615: Extracellular space</td>
<td>61</td>
<td>6.57E-14</td>
<td>PIGR, SPARC, OLFML2B, CXCL12, COL1A1, CXCL17, COMP, VMO1, PLAU, CA2, FAM3B, COL10A1, SOSTDC1, CPXMI, TIMP1, CPA2, CHA1, MMPI, GPX3, GKN1, BGN, GKN2, PGC, GP, SFRP4, ALDH1A1, VCAN, COL4A1, SCGB2A1, SFRPS1, ANOS1, COL3A5, COL1A1, TF2, TF1, CELA3B, CPB1, TNFRSF11B, LR8, SCUBE2, SELENPB1, CST2, CHAD, SPP1, SERPINH1, CKB, APOE, WNT2, CTHRC1, LING01, ANGPT2, CKM, SULF1, KLK11, ATP4A, COL3A1, COL1A2, COL5A1, FAP, COL5A2, ADA</td>
</tr>
<tr>
<td></td>
<td>GO:0001012: Extracellular matrix</td>
<td>21</td>
<td>1.06E-11</td>
<td>LING01, OLFML2B, MMP7, MMPI, BGN, TNFRSF11B, COMP, COL3A1, ADAMTS2, VCAN, COL1A2, COL5A1, COL4A1, COL5A2, CHAD, ANOS1, COL4A5, COL10A1, TIMP1, APOE</td>
</tr>
<tr>
<td></td>
<td>GO:0005576: Extracellular region</td>
<td>57</td>
<td>5.21E-10</td>
<td>SPARC, OLFML2B, CXCL8, PSGA, COL1A1, AQP4, COMP, LIPF, ADAMTS2, PLAU, FAM3B, COL10B1, OLR1, TIMP1, CPA2, CHA1, MMPI, GPX3, MMPI, GKN1, MAMDRC2, BGN, PGC, SFRP4, VCAN, FCN1C1, COL4A1, NRG4, ANOS1, COL3A5, COL1A1, TF2, TF1, LYVE, PKDCX, TNFRSF11B, THY1, THBS2, PLAZ7G, SCUBE2, PTPRZ1, SPP1, APOE, METTL7A, WNT2, GPRIHP1, CTHRC1, ANGPT2, B3GAT1, OCA2B, COL3A1, AKR1B10, COL1A2, COL3A1, QPCT, COL5A2, APOC1, CNTN3</td>
</tr>
<tr>
<td></td>
<td>GO:0005581: Collagen trimmer</td>
<td>12</td>
<td>7.15E-09</td>
<td>COL3A1, COL1A2, COL5A1, COL4A1, MMP1, COL1A2, COL5A2, SERPINH1, COL10A1, COL4A5, TIMP1, CTHRC1</td>
</tr>
<tr>
<td></td>
<td>GO:0000776: Kinetochore</td>
<td>11</td>
<td>4.25E-06</td>
<td>CENPW, NUF2, HJURP, KIF2C, CENPFN, NEK2, CENPFA, BUB1, NDC80, MAD2L1, SPC25</td>
</tr>
<tr>
<td></td>
<td><strong>GOTERM_MF_DIRECT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GO:0005201: Extracellular matrix structural constituent</td>
<td>14</td>
<td>7.32E-09</td>
<td>SPARC, BGN, THBS2, COMP, COL3A1, COL1A2, COL5A1, COL4A1, COL5A2, ANOS1, COL4A5, COL6A1, COL10A1, CTHRC1</td>
</tr>
<tr>
<td></td>
<td>GO:0005020: Extracellular matrix structural constituent conferring tensile strength</td>
<td>9</td>
<td>1.97E-08</td>
<td>COL3A1, COL1A2, COL5A1, COL4A1, COL10A1, COL5A2, COL6A1, COL8A1</td>
</tr>
<tr>
<td></td>
<td>GO:0048407: Platelet-derived growth factor binding</td>
<td>5</td>
<td>5.29E-06</td>
<td>PDGFRB, COL3A1, COL1A2, COL5A1, COL4A1</td>
</tr>
<tr>
<td></td>
<td>GO:0008270: Zinc ion binding</td>
<td>23</td>
<td>4.29E-04</td>
<td>CPA2, TRIM50, CPB1, MMPI, ADH1C, MMP1, MT1M, ESRRB, MT1X, ESRRG, MYRIP, ADH7, ADAMTS2, MT2A, CA2, QPCT, MT1G, ZNF385B, CPXMI, MT1H, TIMP1, ADA, MTIE</td>
</tr>
<tr>
<td></td>
<td>GO:0004111: Creatine kinase activity</td>
<td>3</td>
<td>0.001918971</td>
<td>CKMT2, CKM, CKB</td>
</tr>
</tbody>
</table>

**KEGG functional enrichment analysis of co-DEGs**

Using KEGG functional enrichment analysis, the results of pathways were arranged in ascending order of \(P\) value, and the \(P < 0.05\) is the cutoff value (as shown in Figures 4F-J). As shown in Table 2 (list of top 5 KEGG pathways), our findings suggested that co-DEGs pathways were mainly enriched in ECM-receptor interaction, protein digestion and absorption, gastric acid secretion, mineral absorption and cell cycle pathways, etc. (detailed in Supplementary Tables 1 and 2).

**GSEA functional enrichment analysis of co-DEGs**

The results of pathways were arranged in ascending order of \(P\) value (as shown in Figure 4k, \(P < 0.05\)). As shown in
Figure 2 Identification of differentially expressed genes from The Cancer Genome Atlas and Gene Expression Omnibus Datasets. A: The heat map of GSE19826 data set, with 2202 up-regulated genes and 2700 down-regulated genes. The blue normal samples, and the red are tumor samples; B: The volcanic map of GSE19826 dataset; C: The heat map of GSE79973 dataset, with 665 up-regulated genes and 1507 down-regulated genes. The blue are normal samples, and the red are tumor samples; D: The volcanic map of GSE79973 dataset; E: The heat map of GSE29998 dataset, with 4346 up-regulated genes and 3002 down-regulated genes. The blue are normal samples, and the red are tumor samples; F: The volcanic map of GSE29998 data set; G: The heat map of The Cancer Genome Atlas (TCGA) data set, with 2133 up-regulated genes and 2349 down-regulated genes. The blue are normal samples, and the red are tumor samples; H: The volcanic map of TCGA dataset. In B, D, F and H, the red dots are down-regulated genes and the blue dots are up-regulated genes. TCGA: The Cancer Genome Atlas.

Table 3 (list of top 5 GSEA pathways), GSEA pathway analysis illustrated that co-DEGs mainly concentrated in cell cycle progression, mitotic cell cycle, cell cycle, organelle fission and mitosis pathways, etc. (detailed in Supplementary Tables 1 and 2).

**PPI analysis of co-DEGs**
STRING11.5 was conducted for PPI analysis on the above-mentioned differential expressed genes. The maximum confidence was set to 0.9, and the isolated genes without interaction were deleted (Figure 5A). Cytoscape_v3.9.1 software was used for further analyzing and mapping, showing 84 nodes and 654 edges (Figure 5B).

**Survival analysis of Hub genes**
The Boxplot tool of GEPIA database was used for analyzing the above 84 selected gene nodes, and there was significant
Yin LK et al. Bioinformatics analysis of survival-associated biomarkers of GC

Table 2 Kyoto Encyclopedia of Gene and Genome enrichment pathways of co-expression of differential genes (top 5)

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>Count in gene set</th>
<th>P value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG_pathway</td>
<td>hsa04974: Protein digestion and absorption</td>
<td>13</td>
<td>2.23E-08</td>
<td>CPA2, CELA3B, CPB1, COL12A1, COL3A1, COL1A2, SLC7A8, COL5A1, COL4A1, COL5A2, COL4A5, COL8A1, COL10A1</td>
</tr>
<tr>
<td></td>
<td>hsa04512: ECM-receptor interaction</td>
<td>8</td>
<td>2.57E-04</td>
<td>COMP, COL1A2, COL4A1, ITGA2, CHAD, SPP1, COL4A5, THBS2</td>
</tr>
<tr>
<td></td>
<td>hsa04971: Gastric acid secretion</td>
<td>7</td>
<td>7.46E-04</td>
<td>ATP4B, ATP4A, KCNE2, CCK8B, CA2, KCNJ15, KCNJ16</td>
</tr>
<tr>
<td></td>
<td>hsa04978: Mineral absorption</td>
<td>6</td>
<td>0.001635962</td>
<td>MT2A, MT1M, MT1G, MT1H, MT1X, MT1E</td>
</tr>
<tr>
<td></td>
<td>hsa04110: Cell cycle</td>
<td>8</td>
<td>0.00218248</td>
<td>CDC20, CCNB2, CCNB1, ORC1, CDK1, BUB1, CDC25B, MAD2L1</td>
</tr>
</tbody>
</table>

KEGG: Kyoto Encyclopedia of Gene and Genome.

Table 3 Gene Set Enrichment Analysis enrichment pathways of co-expression of differential genes (top 5)

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
<th>Set size</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO</td>
<td>GO_CELL_CYCLE_PROCESS</td>
<td>41</td>
<td>1.00E-10</td>
</tr>
<tr>
<td></td>
<td>GO_MITOTIC_CELL_CYCLE</td>
<td>37</td>
<td>1.40E-10</td>
</tr>
<tr>
<td></td>
<td>GO_CELL_CYCLE</td>
<td>44</td>
<td>1.49E-10</td>
</tr>
<tr>
<td></td>
<td>GO_ORGANELLE_FISSON</td>
<td>24</td>
<td>9.65E-08</td>
</tr>
<tr>
<td></td>
<td>GO_MITOTIC_NUCLEAR_DIVISION</td>
<td>19</td>
<td>1.20E-07</td>
</tr>
<tr>
<td></td>
<td>GO_CELL_CYCLE_PHASE_TRANSITION</td>
<td>21</td>
<td>1.82E-07</td>
</tr>
</tbody>
</table>

GO: Gene Ontology.

Figure 3 The Venn analysis of differentially expressed genes from Gene Expression Omnibus and The Cancer Genome Atlas datasets. A: The Venn diagram of GSE19826 (red) 2202 up-regulated genes, GSE79973 (blue) 665 up-regulated genes, GSE29998 (green) 4346 up-regulated genes, The Cancer Genome Atlas (TCGA)-Stad (purple) 2133 up-regulated genes, among which 133 genes intersect; B: The Venn diagram of GSE19826 (red) 2700 down-regulated genes, GSE79973 (blue) 1507 down-regulated genes, GSE29998 (green) 3002 down-regulated genes, TCGA-Stad (purple) 2349 down-regulated genes, among which 201 genes intersect. TCGA: The Cancer Genome Atlas.

expression difference between the tumor tissues and normal tissues (Figure 6). Then, GEPIA’s survival tool was used for visual analysis of the selected genes [the cutoff value is Logrank P < 0.05 and hazard ratio (HR) < 0.05]. Our results suggested that 12 genes with notable significance for prognosis of patients were screened, including CEP55, COL1A2, COL3A1, gpihbp1, Vcan, TIMP1, SPARC, PDGFRb, MAOA, fzd2, COL4A1 and BGN. Visualized analysis of survival curve showed that CEP55 was the protective factor (Logrank P < 0.05, HR < 1), COL1A2, COL3A1, gpihbp1, VCAN, TIMP1, SPARC, PDGFRB, MAOA, FZD2, COL4A1 and BGN were the risk factors (Logrank P < 0.05, HR > 1) (Figure 7).
Yin LK et al. Bioinformatics analysis of survival-associated biomarkers of GC

A

B

Gene ratio

Count

Yin LK et al. Bioinformatics analysis of survival-associated biomarkers of GC

WJCC | https://www.wjgnet.com

4771 | July 16, 2023 | Volume 11 | Issue 20
Yin LK et al. Bioinformatics analysis of survival-associated biomarkers of GC

C

Size

Category

- Chromosome segregation
- External encapsulating structure organization
- Extracellular matrix organization
- Extracellular structure organization
- Nuclear division

D

GO Terms

- Chromosome segregation
- Extracellular matrix organization
- Extracellular structure organization
- External encapsulating structure organization
- Nuclear division
- Mitotic sister chromatid segregation
- Sister chromatid segregation

Log FC

-1 3
Yin LK et al. Bioinformatics analysis of survival-associated biomarkers of GC

G

- Protein digestion and absorption
- ECM–receptor interaction
- Gastric acid secretion
- Mineral absorption
- Cell cycle
- Drug metabolism – cytochrome P450
- Focal adhesion
- Metabolism of xenobiotics by cytochrome P450
- Tyrosine metabolism
- AGE–RAGE signaling pathway in diabetic complications
- Progesterone-mediated oocyte maturation
- Histidine metabolism
- PI3K–Akt signaling pathway
- Arginine and proline metabolism
- Collecting duct acid secretion
- Human papillomavirus infection

H

GO Terms
- Protein digestion and absorption
- ECM–receptor interaction
- Gastric acid secretion
- Mineral absorption
- Cell cycle
- Drug metabolism – cytochrome P450
- Focal adhesion
- Metabolism of xenobiotics by cytochrome P450
- Tyrosine metabolism

[Bar chart and pie chart visualizations of gene ontology terms with color-coded log fold change values ranging from -3 to 3]
Survival analysis validation of Hub gene

For the above results, Kaplan-Meier plotter online network was carried out for visual analysis and verification. Excluding COL3A1 (Logrank \(P > 0.05\)), 11 key genes, including CEP55, COL1A2, GPIHBP1, VCAN, TIMP1, SPARC, PDGFRB, MAOA, FZD2, COL4A1, and BGN, were obtained (Logrank \(P < 0.05\)). The results confirmed that CEP55 was a protective factor (Logrank \(P < 0.05, \text{HR} < 1\); COL1A2, GPIHBP1, VCAN, TIMP1, SPARC, PDGFRB, MAOA, FZD2, COL4A1, and BGN were risk factors (Logrank \(P < 0.05, \text{HR} > 1\)) (Figure 8).

Immunohistochemical analysis and final identification of Hub genes

Compared with normal gastric tissues, the expressed proteins of 10 genes of BGN, CEP55, COL1A2, COL4A1, FZD2, MAOA, PDGFRB, SPARC, TIMP1, and VCAN were up-regulated in GC tissues by HPA IHC database (\(P < 0.05\)) (Figure 9). Therefore, in this study, 10 Hub genes related to the prognosis of GC were finally screened.

DISCUSSION

Based on GC gene expression data from GEO and TCGA databases, in this study, bioinformatics was conducted to screen differential expressed genes with relation to survival prognosis of GC. A total of 334 DEGs were analyzed by GO, KEGG and GSEA enrichment, respectively. By GO analysis (Figure 4), these DEGs were found to be enriched in BP, CC and MF. In KEGG analysis, ECM-receptor interaction, protein digestion and absorption, gastric acid secretion, mineral absorption, cell cycle and other signal pathways were enriched. GSEA pathway analysis showed that these DEGs were mainly concentrated in cell cycle progression, mitotic cell cycle, cell cycle, organelle fission, mitosis and other signaling pathways (Table 3). The GEPIA database was then used to verify the differences in expression of these key genes between tumors and normal tissues. The PPI network of these DEGs was analyzed and constructed by STRING11.5, and the credibility was set as the highest: 0.9. Eighty-four Hub genes were screened. GEPIA and Kaplan-Meier plotter identified 11 Hub genes (CEP55, COL1A2, COL4A1, FZD2, MAOA, PDGFRB, SPARC, TIMP1, and VCAN) that were associated with GC prognosis. The combination of three databases (GEO, TCGA and GEPIA) makes our results more credible, which is a prominent feature of this study. Subsequently, 10 key genes involved in the prognosis of GC, including BGN, TIMP1, VCAN, COL1A2, COL4A1, FZD2, MAOA, PDGFRB, SPARC, and CEP55, were screened by HPA immunohistochemical analysis.

The results of this study showed that BGN and VCAN genes, encoding multifunctional proteoglycans, which were highly expressed in GC tissues and associated with poor prognosis of patients. In this article, by GO, KEGG and GSEA analysis, we found that the two genes (BGN and VCAN) are mainly enriched in BP and MF such as extracellular matrix, extracellular matrix structural components, extracellular space and glycosaminoglycan binding (shown in Supplementary Tables 1 and 2). These BPs and MFs may be involved in the occurrence and development of GC, but the specific mechanism is still unclear, and further experimental research is needed. Previous studies have found that BGN may have important value in predicting the overall survival and tumor immune infiltration of GC\[20\]. BGN is not only a potential...
Figure 5 The protein-protein interaction analysis of co-expression of differential genes. A: The protein-protein interaction network diagram of co-expression of differential genes (co-DEGs) constructed by STRING11.5; B: 84 nodes and 654 edges of co-DEGs was revealed by Cytoscape software. *P < 0.05.

As shown in Supplementary Tables 1 and 2, these findings also revealed that TIMP1 and MAOA genes, encoding enzyme proteins, predicted poor survival for patients; TIMP1 mainly participates in the degradation of extracellular matrix, promotes cell proliferation, and have anti-apoptosis function in tumors. Previous studies have found that TIMP1 is positively correlated with the pathological N stage of GC, which may inhibit the growth and metastasis of GC cells through mir-6745-TIMP1 axis[26,27], and Chemerin receptor antagonists down-regulate the expression of TIMP1 and TIMP2 through chemokine-like receptor-1 and G-protein coupled receptor 1 pathways, reducing the metastatic and invasive ability of GC cells[28]. Thus, TIMP1 could promote the progression of GC and shorten the survival period of
Yin LK et al. Bioinformatics analysis of survival-associated biomarkers of GC
patients. MAOA, encoding a mitochondrial enzyme, catalyzes the oxidative deamination of amines. The loss or decrease of MAOA expression can be used as a marker to monitor the immunotherapeutic effect of GC[29]; MAOA can also facilitate the proliferation and metastatic ability of gastric tumor cells by regulating mitochondrial function and aerobic glycolysis[30]. In the present study, we found that MAOA may affect the molecular functions of cells, such as protein binding, flavin adenine dinucleotide binding, oxidoreductase and so on. Therefore, it might affect the survival and prognosis of patients with GC by regulating the metabolic function of cells.

In addition, the results of this article suggested that two high-expressed genes of FZD2 and PDGFRB, encoding receptor proteins, were related to the prognosis of patients. It has been reported that a protein encoded by FZD2 is involved in binding to β-catenin typical signaling pathways and participates in regulation β-catenin dependent pathways. At present, only few literatures indicate that FZD2 may play a key role in the occurrence and development of GC[31]. Our results demonstrated that FZD2 may act on: Atypical Wnt signaling pathway, classical Wnt signaling pathway, Wnt protein binding and other BPs. Previous studies have reported that PDGFRB is related to immune cell infiltration in GC, and may serve as a potential prognosis biomarker for GC[32]. PDGFRB is significantly correlated with the malignant phenotype of tumors, and the high expression of PDGFRB significantly reduced the overall survival of patients with GC[33,34]. In this study, we found that PDGFRB might promote tumor angiogenesis, cell proliferation and cell migration, and inhibit the aging and apoptosis of tumor cells, increasing the possibility of GC metastasis (details in Supplementary Tables 1 and 2).

The other four genes (COL1A2, COL4A1, SPARC and CEP55) are related to the prognosis of GC, which mainly encode collagen, acidic matrix related protein and centrosome protein of GC cells. The previous studies have found that collagen-encoding genes COL1A1 and P4HA3 may be related to the prognosis of GC[35]. At the present study, we found that COL4A1 and COL1A2 may have some effects on extracellular matrix and its interaction with receptors, protein binding, protein digestion and absorption, and regulate PI3K Akt signaling pathway. For the CEP55, some researchers have reported that it participates in promoting the malignant biological behavior of GC cells[36]. The expression of CEP55 in GC tissues is elevated, and CEP55 can also be a potential therapeutic target for GC[37]. The high expression of SPARC is associated with poor prognosis and shorter overall survival in patients with GC[38-41]. Our results suggested that CEP55 and SPARC could affect cell mitosis, cytokinesis and extracellular matrix structure, but its mechanisms are still remaining unclear (details in Supplementary Tables 1 and 2).

To sum up, this Bioinformatics present that BGN, CEP55, COL1A2, COL4A1, FZD2, MAOA, PDGFRB, SPARC, TIMP1 and VCAN, identified as the prognosis of GC, are involved in the occurrence and development of GC, and thus affect the survival and prognosis of patients by themselves or the encoded proteins. In this study, the 10 Hub genes obtained by comprehensive analysis of multiple databases and datasets may be used as survival biomarkers. Nevertheless, there are still many shortcomings in this study, for example, our data is currently limited to the online database, the selection of data may be biased or incomplete, and the relevant molecular mechanism needs further experimental research to verify.
Figure 7 The prognosis of 12 key genes associated with gastric cancer screened by Gene Expression Profiling Interactive Analysis survival analysis tool, showing CEP55 was the protective factor (Logrank P < 0.05, hazard ratio < 1).
CONCLUSION

In this study, we analyzed the gene expression profiles and sequencing data of GC tissues and adjacent or normal gastric tissues to explore the pathogenesis of GC using bioinformatics, investigated the signal pathways of co-DEGs involved in GC, and identified the 10 Hub genes correlated with the prognosis of patients with GC. The 10 key genes obtained through the analysis of multiple databases and datasets may be used as objective and reliable biomarkers for the survival analysis of patients. In addition, these genes or their encoded proteins can also be used as potential therapeutic targets for GC, improving the survival time of patients with GC. However, the mechanisms of 10 Hub genes in GC is still unclear, which needs further confirmation through molecular biology and clinical experiments.
ARTICLE HIGHLIGHTS

Research background
Gastric cancer (GC) is one of the most common malignant tumors, and its pathogenesis and biomarkers are still unclear.

Research motivation
The present study for the first time investigated the 10 Hub genes as the potential biomarkers of the prognosis of patients using bioinformatics.

Research objectives
The aims of this study are to explore the potential biomarkers of the prognosis of patients with GC, so as to provide new strategies for the treatment of GC.

Research methods
In this study, bioinformatics strategy was used to obtain Datasets from The Cancer Genome Atlas, Gene Expression Omnibus and Gene Expression Profiling Interactive Analysis. The software of R software, STRING, Kaplan-Meier plotter and Human Protein Atlas, were performed to analyze and integrate the mRNA datasets, respectively.

Research results
The signal pathways of the involvement of the co-expression of differential genes in GC were screened out, and the 10 Hub genes, including BGN, CEP55, COL1A2, COL4A1, FZD2, MAOA, PDGFRB, SPARC, TIMP1 and VCAN, were associated with prognosis of GC and identified as the potential prognostic biomarkers of GC.

Research conclusions
The 10 key genes obtained through the analysis of multiple datasets may be used as objective and reliable biomarkers for the survival analysis of patients. In addition, these genes or their encoded proteins can also be as potential therapeutic targets for GC, improving the survival time of patients with GC.

Research perspectives
The mechanisms of 10 Hub genes in GC is still unclear, which needs further confirmation through molecular biology and clinical experiments.

FOOTNOTES

Author contributions: Yin LK and Yuan HY contributed equally to this work. Yin LK, Yuan HY, Liu JJ and Wang P contributed to data collection and manuscript drafting; Yin LK, Xu XL, Wang W, Bai XY and Wang P prepared for the figures and tables; and all authors have approved the final manuscript.

Institutional review board statement: All of the data of this paper are from the public database of TCGA, GEO and GEPIA, and there is no
Yin LK et al. Bioinformatics analysis of survival-associated biomarkers of GC

ethical statement needed to be declared for this manuscript.

Clinical trial registration statement: All of the data of this paper are from the public database of TCGA, GEO and GEPIA, and no clinical trial registration needed for this manuscript.

Informed consent statement: All of the data of this paper are from the public database of TCGA, GEO and GEPIA, and no informed consent form documents are needed to be signed by any patients.

Conflict-of-interest statement: All the authors report no relevant conflicts of interest for this article.

Data sharing statement: The data supporting the results of this study are available from GEO database (GSE19826, GSE79973 and GSE29998).

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: https://creativecommons.org/Licenses/by-nc/4.0/

Country/Territory of origin: China

ORCID number: Long-Kuan Yin 0000-0001-5974-5443; Hua-Yan Yuan 0000-0001-5983-3849; Jian-Jun Liu 0000-0003-1206-6480; Xiu-Lian Xu 0000-0002-4177-6919; Wei Wang 0000-0002-6035-9525; Xiang-Yu Bai 0000-0003-0259-1994; Pan Wang 0000-0003-2805-8346.

S-Editor: Wang JJ
L-Editor: A
P-Editor: Cai YX

REFERENCES


Yin LK et al. Bioinformatics analysis of survival-associated biomarkers of GC


Ding YL, Sun SF, Zhao GL, COL5A2 as a potential clinical biomarker for gastric cancer and renal metastasis. Medicine (Baltimore) 2021; 100: e24561 [PMID: 33067786 DOI: 10.1097/MD.0000000000024561]


