Clinical Trial/Experimental Study

Medicine

OPEN

Identification of potential hub genes of gastric cancer

Xu-Dong Zhou, MD[a], Ya-Wei Qu, MD[b], Li Wang, MD[c], Fu-Hua Jia, MD[c], Peng Chen, MD[d], Yin-Pu Wang, MD[e], Hai-Feng Liu, MD[a,*]

Abstract

Background: Gastric cancer (GC) is a malignant tumor originated from gastric mucosa epithelium. It is the third leading cause of cancer mortality in China. The early symptoms are not obvious. When it is discovered, it has developed to the advanced stage, and the prognosis is poor. In order to screen for potential genes for GC development, this study obtained GSE118916 and GSE109476 from the gene expression omnibus (GEO) database for bioinformatics analysis.

Methods: First, GEO2R was used to identify differentially expressed genes (DEG) and the functional annotation of DEGs was performed by gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. The Search Tool for the Retrieval of Interacting Genes (STRING) tool was used to construct protein-protein interaction (PPI) network and the most important modules and hub genes were mined. Real time quantitative polymerase chain reaction assay was performed to verify the expression level of hub genes.

Results: A total of 139 DEGs were identified. The functional changes of DEGs are mainly concentrated in the cytoskeleton, extracellular matrix and collagen synthesis. Eleven genes were identified as core genes. Bioinformatics analysis shows that the core genes are mainly enriched in many processes related to cell adhesion and collagen.

Conclusion: In summary, the DEGs and hub genes found in this study may be potential diagnostic and therapeutic targets.

Abbreviations: BP = biological processes, DAVID = Database for Annotation, Visualization and Integrated Discovery, DEG = differentially expressed genes, GC = gastric cancer, GEO = gene expression omnibus, GO = gene ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, MCODE = molecular complex detection, MF = molecular function, PPI = protein–protein interaction, STRING = Search Tool for the Retrieval of Interacting Genes, TCGA = the cancer genome atlas.

Keywords: bioinformatic analysis, differentially expressed genes, gastric cancer, protein-protein interaction

1. Introduction

Gastric cancer (GC) is a malignant tumor originated from the gastric mucosal epithelium, mainly gastric adenocarcinoma. GC accounts for more than 95% of malignant tumors in the stomach and is one of the malignant tumors that seriously endanger human health. According to the results of the National Cancer Center of China in 2015, GC accounts for the third place in the mortality rate of malignant tumors in China.[1] The occurrence of GC is closely related to the adverse environment, lifestyle, dietary structure changes and Helicobacter pylori infection. Early GC symptoms are not obvious, some patients may have dyspepsia symptoms, and advanced GC may have upper abdominal pain, postprandial aggravation, poor appetite, anorexia, fatigue and weight loss. The common examination methods are gastroscopy and computed tomography, which are invasive and expensive.[2] When the patient has obvious symptoms, he is admitted to the hospital. The disease has developed to the advanced stage of GC, and the best surgical treatment time is lost. Except for Japan and South Korea, the 5-year survival rate of advanced GC in other countries and regions in the world is even less than 10%.[3] However, if GC can be diagnosed early, its 5-year survival rate will rise to 95%,[4] which means that the fundamental method for providing GC prognosis is early diagnosis and timely treatment. Currently, some serum biomarkers are used for screening early GC, such as CA19-9 and CEA, but these tumor markers are less sensitive and specific.[5] Therefore, to find a new effective biomarker for early GC, to further explore the pathogenesis of GC, to find potential diagnostic and therapeutic targets, to

This project was funded by the Capital Clinical Application Research-Research on the Diagnosis Value of Early Gastric Cancer and Precancerous Lesions by High-resolution Microendoscopy, No. Z141107002514099.

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

[a] The Clinical College of the General Hospital of Chinese People's Armed Police Forces, Anhui Medical University, Hefei, P.R. China, [b] Department of Gastroenterology, Third Medical Center of PLA General Hospital, Beijing, P.R. China, [c] Department of Gastroenterology, Huamei Hospital of China National University of Science and Technology, Ningbo, P.R. China, [d] Department of Ultrasound, Graduate School of Jinzhou Medical University, Jinzhou, P.R. China, [e] Department of Gastroenterology, Baoji Hospital Affiliated to Xi'an Jiaotong University, Baoji, P.R. China.

**Correspondence: Hai-Feng Liu, The Clinical College of the General Hospital of Chinese People's Armed Police Forces, Anhui Medical University, Hefei 230032, P.R. China (e-mail: haifengliu333@163.com).*

How to cite this article: Zhou X-D, Qu Y-W, Wang L, Jia F-H, Chen P, Wang Y-P, Liu H-F. Identification of potential hub genes of gastric cancer. Medicine 2022;101:41(e30741).

Received: 6 May 2022 / Received in final form: 24 August 2022 / Accepted: 25 August 2022

http://dx.doi.org/10.1097/MD.0000000000030741

achieve early detection, early diagnosis and targeted therapy, with significant clinical value and market Application prospect.

Bioinformatics is an emerging interdisciplinary subject that combines life sciences with computer science. It focuses on the collection, storage, processing, dissemination, analysis, and interpretation of biological information. The ability to process large amounts of complex biological data can be processed through the use of biological and informatics techniques. Microarray data information analysis technology has been widely used in the study of diseases such as tumors to explore the genetic correlation.[6,7] Microarray analysis technology can simultaneously acquire the expression information of tens of thousands of genes, and then explore the genomic changes related to the development of diseases. A large number of research and scholars[8,9] have used bioinformatics techniques to analyze differentially expressed genes (DEG) in tumor progression, and to study their roles in biological processes (BP), molecular functions (MF), and signaling pathways, and to elucidate the pathogenesis of diseases, so as to provide theoretical basis for early diagnosis and treatment.

In this study, bioinformatics technology was used to find the gene sequencing data of GC patients and normal people from gene expression omnibus (GEO). Two high-quality genetic data sets were extracted and analyzed for further analysis. Gene ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were performed by gene set enrichment analysis, and then important modules of the protein–protein interaction (PPI) network were screened. Using the genetic data of tumor patients and normal people in the sample, 73 gene sets and 11 significantly DEG molecules were found to be differentially expressed. These findings will enhance our understanding of the underlying mechanisms of GC and provide the basis for finding new diagnostic markers and targeted therapies.

2. Materials and Methods

2.1. Access to public data

GEO (http://www.ncbi.nlm.nih.gov/geo)[10] is an open high-throughput genomic database that includes microarrays, gene expression data and chips.

On November 20, 2019, the key words "(gastric cancer) AND gene expression" were set to detect the datasets, using a filter of "expression profiling by array" and "recent two years." There were 5 inclusion criteria: a sample number of more than 10 per dataset (samples of less than 10 were excluded), data from Homo sapiens (data from other species were excluded), a series entry type, expression profiling by array (data using methylation profiling by array were excluded), and a diagnosis of GC (data from other cancer diagnoses were excluded).

Two expression profile data sets (GSE118916 and GSE109476) were downloaded from the GEO database. The annotation platform for GSE118916 is GPL15207 platform, [PrimeView] Affymetrix Human Gene Expression Array. The GSE118916 data set is composed of 15 GC tissues and 15 stomach normal tissues. The annotation platform for GSE109476 is GPL24530platform, Arraystar Human LncRNA microarray V2.0 (Agilent-033010; custom-annotation; probe name version). The GSE109476 date set is composed of 5 GC tissues and 5 stomach normal tissues. All probe numbers are converted to gene symbols based on the annotation information in the platform.

2.2. Screening of DEGs via GEO2R

GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r)[11] is a system for online analysis of data in GEO. This tool system runs in the R language. It is accurate to say that it uses 2 R packages: GEOquery and limma. The former is used for data reading and the latter is used for calculation. The best thing about GEO2R is that is an online tool, easy and efficient to operate. GEO2R can perform a command to compare gene expression profiles between groups in order to identify DEGs between GC and stomach normal groups. In general, when the probe set has a corresponding gene symbol, the probe is considered valuable and will be retained. Statistically significant measure is P value <.01 and fold change >1.

2.3. Functional annotation of DEGs via GO and KEGG analysis

Database for Annotation, Visualization and Integrated Discovery (DAVID) (https://david.ncifcrf.gov/home.jsp) (version 6.8) is a bioinformatics database that integrates biological data and analytical tools.[12] KEGG (https://www.kegg.jp/) could help researcher to understand advanced functions and biological systems.[13] GO is an ontology widely used in bioinformatics, which covers 3 aspects of biology, including cellular components, MF and biological process.[14] In order to analyze the GO and pathway enrichment information of DEGs, the DAVID online tool was executed. Statistically significant measure is $P < .05$.

2.4. Construction and analysis of PPI network

Search Tool for the Retrieval of Interacting Genes (STRING; http://string-db.org) (version 10.5)[15] is a network that can be used to predict and track PPIs. Introducing DEGs into the tool makes intermolecular network analysis. The analysis of the interactions between different proteins can provide insights into the mechanisms of generation or development of GC. In this study, STRING database was used to construct PPI network with DEGs. The minimum required interaction score is that medium confidence > 0.4. Cytoscape (version 3.6.1) is an open visualization software that can be used to visualize PPI network.[16] Based on topological principles, the Molecular Complex Detection (MCODE) (version 1.5.1), a plug-in for Cytoscape, can mine tightly coupled regions from PPI. First, Cytoscape software plots the PPI network. Secondly, MCODE identifies the most important modules in the PPI network graph. The criteria of MCODE analysis is that node score cutoff = 0.2, degree cutoff = 2, Max depth = 100, MCODE scores > 5, and k-score = 2.

2.5. Mining and screening of core gene

The hub genes were selected with degrees ≥ 10. A network of the genes and their co-expression genes was analyzed using cBioPortal (http://www.cbioportal.org)[17,18] online platform. Hierarchical clustering of hub genes was constructed using UCSC Cancer Genomics Browser (http://genome-cancer.ucsc.edu).[19] The overall survival and disease-free survival analyses of hub genes were performed using Kaplan–Meier curve in cBioPortal.

2.6. RR-qPCR assay

A total of 10 GC participates were recruited. After surgery, 10 GC tumor samples from GC patients and 10 adjacent normal stomach tissues samples were obtained. The research conformed to the Declaration of Helsinki and was authorized by the Human Ethics and Research Ethics Committees of Third Medical Center of PLA General Hospital. The written informed consents were obtained from all participates.

Total RNA was extracted from 10 GC tumor samples and 10 adjacent normal stomach tissues samples by the RNAiso Plus (Trizol) kit (Thermofisher, Massachusetts, America), and reverse transcribed to cDNA. Real time quantitative polymerase chain reaction (RT-qPCR) was performed using a Light Cycler® 4800 System with specific primers for genes. Table 1 presents the primer sequences used in the experiments. The RQ values ($2^{-\Delta\Delta Ct}$, where Ct is the threshold cycle) of each sample were calculated,

and are presented as fold change in gene expression relative to the control group. GAPDH was used as an endogenous control.

The verification of hub genes expression and role on the overall survival of GC patients using the cancer genome atlas (TCGA) data

The gene expression dataset of GC in the TCGA was downloaded. There were a total of 580samples including 478 GC samples and 102 normal gastric samples. The IlluminaHiSeq UNC was selected as gene expression RNAseq in the research. In addition, the gene expression levels of hub genes between GC and normal gastric samples were compared using the one-way Anova.

Furthermore, effect of gene expression of hub genes on overall survival was analyzed by using the TCGA data.

2.7. Statistical analysis

Student's t test was used to determine the statistical significance when comparing the 2 groups. Statistical analysis was carried out using SPSS software version 21.0 (IBM Corp. Armonk, NY). Value of $P < .05$ were considered statistically significant.

Table 1

Primers and their sequences for PCR analysis

Primer	Sequence (5′–3′)
COL1A2-hF	AGGGAAGGTAGTAACAGTAG
COL1A2-hR	CCAGGATTACCCTATGAG
COL3A1-hF	TGAGCCTGGTAAGAATGG
COL3A1-hR	CCTGGAACACCTGGAATA
SPARC-hF	GGCTGGTCACATAGGTAC
SPARC-hR	GAGGGTTAAGCAAGGAAT
PCOLCE-hF	TCCTCCGTGCTGTGGTGT
PCOLCE-hR	GGTTCAGATCCCGTCCCT
COL8A1-hF	TGAACCAATCTGGCCTCC
COL8A1-hR	TTTGCTGCTAAGCCGTGA
SERPINH1-hF	CCTGAAGAATGGAGCAAA
SERPINH1-hR	AGGAGCGGAAAGGACACT
COL8A2-hF	GGAAACAAGAGCGATGACG
COL8A2-hR	CAGCGGTGAGAAGGGTGT
COL6A3-hF	CCCAGGAGTTCAAGACCA
COL6A3-hR	GAGGAGCCCAACACCATC
LAMA4-hF	CTGGACCTAACTGTGAAA
LAMA4-hR	GTATAAAGAATGGGCGAAA
LOXL1-hF	CGTTCACTGTAGCGTGTTT
LOXL1-hR	GTGCATCCTCTATGTCCCT
COL5A2-hF	AGCCAGGTTTGAGGAGCA
COL5A2-hR	GCAGCAATTAGTTGAGCC

PCR = polymerase chain reaction.

3. Results

3.1. Identification of DEGs in GC

One volcano plot presents the DEGs in the GSE118916 (Fig. 1A), and another volcano plot presents the DEGs in the GSE109476 (Fig. 1B). After standardization of the microarray results, DEGs (1768 in GSE118916, and 564 in GSE109476) were identified. The overlap among the 2 datasets contained 139 genes as shown in the Venn diagram (Fig. 1C), consisting of 189 downregulated genes and 84 upregulated genes between GC tissues and non-cancerous tissues.

3.2. KEGG and GO enrichment analyses of DEGs

To analyze the biological classification of DEGs, functional and pathway enrichment analyses were performed using DAVID. GO analysis results showed that changes in BP of DEGs were significantly enriched in collagen catabolic process, collagen fibril organization, extracellular matrix organization, integrin-mediated signaling pathway, cell adhesion and so on. Changes in MF were mainly enriched in collagen binding, growth factor binding, heparin binding, extracellular matrix structural constituent and so on (Table 1). Changes in cell component of DEGs were mainly enriched in extracellular matrix, proteinaceous extracellular matrix, collagen trimer, extracellular region and so on. The KEGG pathway analysis showed that all DEGs are mainly concentrated in ECM-receptor interaction, PI3K-Akt signaling pathway, Metabolism of xenobiotics by cytochrome P450, platelet activation, Gap junction, Protein digestion and absorption and Phagosome (Table 2).

3.3. PPI network construction and module analysis

The PPI network of DEGs was constructed (Fig. 2) and the most significant module was obtained using Cytoscape (Fig. 3). The functional analyses of genes involved in this module were analyzed using DAVID.

3.4. Hub gene selection and analysis

A total of 11 genes were identified as hub genes with degrees ≥10. The names, abbreviations and functions for these hub genes are shown in Table 3. A network of the hub genes and their co-expression genes was analyzed using cBioPortal online platform (Fig. 4A). Hierarchical clustering showed that the hub genes

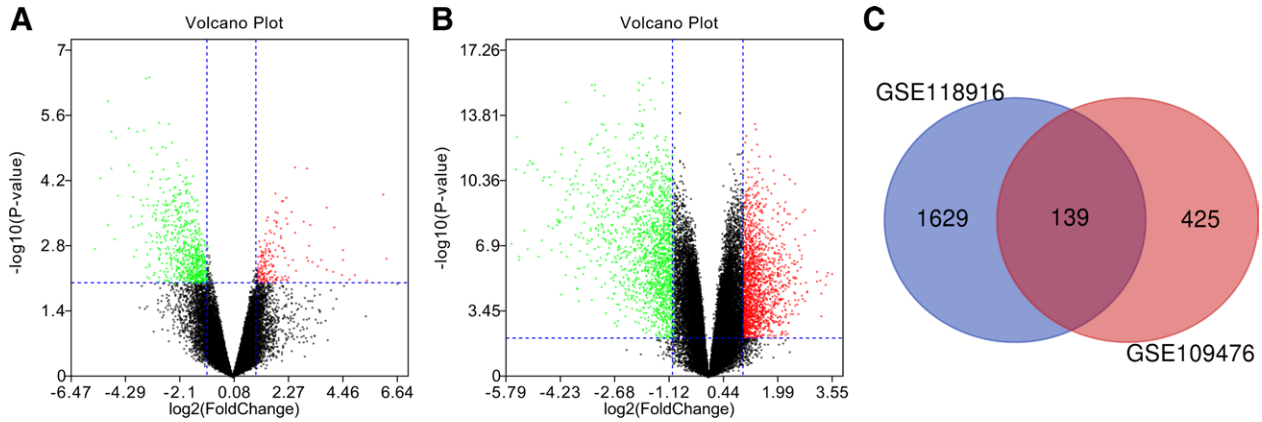

Figure 1. DEGs in GC. (A) One volcano plot presents the DEGs in the GSE118916. (B) another volcano plot presents the DEGs in the GSE109476. (C) Venn diagram, PPI network and the most significant module of DEGs. (A) DEGs were selected with a fold change > 1 and *P* value < .01 among the mRNA expression profiling sets GSE118916 and GSE109476. The 2 datasets showed an overlap of 139 genes. DEG = differentially expressed genes, GC = gastric cancer.

Table 2

GO and KEGG pathway enrichment analysis of DEGs in gastric cancer samples.

Term	Description	Count in gene set	*P* value
GO:0030574	Collagen catabolic process	8	5.49E-07
GO:0030199	Collagen fibril organization	6	1.11E-05
GO:0030198	Extracellular matrix organization	10	1.93E-05
GO:0007229	Integrin-mediated signaling pathway	7	1.09E-04
GO:0048593	Camera-type eye morphogenesis	3	.006
GO:0007263	Nitric oxide mediated signal transduction	3	.009
GO:0045926	Negative regulation of growth	3	.009
GO:0071294	Cellular response to zinc ion	3	.009
GO:0031012	Extracellular matrix	17	3.61E-10
GO:0005615	Extracellular space	29	2.23E-07
GO:0005578	Proteinaceous extracellular matrix	13	4.95E-07
GO:0005581	Collagen trimer	8	4.23E-06
GO:0005576	Extracellular region	28	2.18E-05
GO:0005788	Endoplasmic reticulum lumen	8	4.58E-04
hsa04512	ECM-receptor interaction	6	.001
hsa04151	PI3K-Akt signaling pathway	10	.003
hsa00980	Metabolism of xenobiotics by cytochrome P450	4	.029
hsa04611	Platelet activation	5	.030
hsa04540	Gap junction	4	.046
hsa04974	Protein digestion and absorption	4	.046
hsa04145	Phagosome	5	.048

DEGs = differentially expressed genes, GO = gene ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes.

could basically differentiate the GC samples from the non-cancerous samples (Fig. 4B). Subsequently, the overall survival analysis of the hub genes was performed using Kaplan–Meier curve. GC patients with COL1A2, COL3A1, SPARC, PCOLCE, COL8A1, SERPINH1, COL8A2, COL6A3, LAMA4, LOXL1, and COL5A2 alteration showed worse overall survival (Figs. 5 and 6).

3.5. Results of RT-qPCR analysis

According to the above expression analysis, COL1A2, COL3A1, SPARC, PCOLCE, COL8A1, SERPINH1, COL8A2, COL6A3, LAMA4, LOXL1, and COL5A2 were markedly up-regulated in GC tumor samples. As presented in Figure 7, the relative expression levels of COL1A2, COL3A1, SPARC, PCOLCE, COL8A1, SERPINH1, COL8A2, COL6A3, LAMA4, LOXL1, and COL5A2 were significantly higher in GC samples, compared with the normal stomach tissues groups. The result demonstrated that COL1A2, COL3A1, SPARC, PCOLCE, COL8A1, SERPINH1, COL8A2, COL6A3, LAMA4, LOXL1, and COL5A2 might be considered as biomarkers for GC.

3.6. The verification by TCGA

According to the above expression analysis, COL1A2, COL3A1, SPARC, PCOLCE, COL8A1, SERPINH1, COL8A2, COL6A3, LAMA4, LOXL1, and COL5A2 were significantly up-regulated in GC tumor samples compared with the normal gastric samples. After confirmation using TCGA data, these genes expression levels in GC samples were also significantly higher than the normal gastric samples (Fig. 8).

Overall survival analysis showed that GC patients with high expression levels of COL1A2, COL3A1, SPARC, PCOLCE, COL8A1, SERPINH1, COL8A2, COL6A3, LAMA4, LOXL1, and COL5A2 had poorer overall survival times than those with low expression levels ($P < .05$, Fig. 9).

4. Discussion

In 2018, there were more than 1 million new cases of GC in the world, and 783,000 deaths.[20] The most common sites of GC were gastric antrum (58%), cardia (20%), corpus (15%), whole stomach or most stomach (7%). GC can spread through direct spread, lymph node metastasis, hematogenous dissemination, and plant metastasis. At present, the treatment of GC is often treated by multiple means. The treatment may include partial gastrectomy or total gastrectomy, lymph node dissection and perioperative chemotherapy or postoperative radiotherapy and chemotherapy.[21–23] Patients may experience malnutrition, reduced immunity, and decreased quality of life during treatment. And it will bring a series of adverse reactions to patients, so that patients with GC not only suffer from physiologically great pain, but also psychologically bear tremendous pressure. After gastrectomy, the physiological function of patients will be seriously disturbed, and the body will also suffer from malnutrition, reflux esophagitis, absorption disorders and other adverse consequences.[24,25] On the other hand, since medicinal chemotherapy kills cancer cells and kills normal cells of the patient, it causes toxic effects and a series of adverse reactions, which cause serious damage to the patient's body and mind. The prognosis of patients is often associated with timely diagnosis and treatment, but there are large clinical heterogeneities in different individuals and tumor types. Therefore, it is of great clinical significance to further explore the pathogenesis of GC, to find early diagnostic markers, targeted therapeutic genes and molecules, and to achieve early diagnosis and individualized treatment according to different individuals and pathological types.

Bioinformatics technology has been widely used to find genetic molecules related to tumorigenesis and development, and to find genes and molecules that can be used as therapeutic targets. Cao et al found the PLEKHG1 molecule related to GC through this technology, and further confirmed the correlation between the gene and GC, suggesting that the molecule is a biomarker for diagnosis and prediction of outcome.[26] Wang et al found a molecule related to colorectal cancer proliferation and metastasis through bioinformatics technology, suggesting that it may serve as a potential therapeutic target.[27] In this study, DEGs between GC tissues and non-cancer tissues were obtained by analyzing 2 mRNA microarray data sets. A total of 139 DEGs were identified in 2 data sets. Bioinformatics analysis revealed high expression of COL1A2, COL3A1, SPARC, PCOLCE, COL8A1, SERPINH1, COL8A2, COL6A3, LAMA4, LOXL1, and COL5A2 in GC patients. At the same time, multiple gene

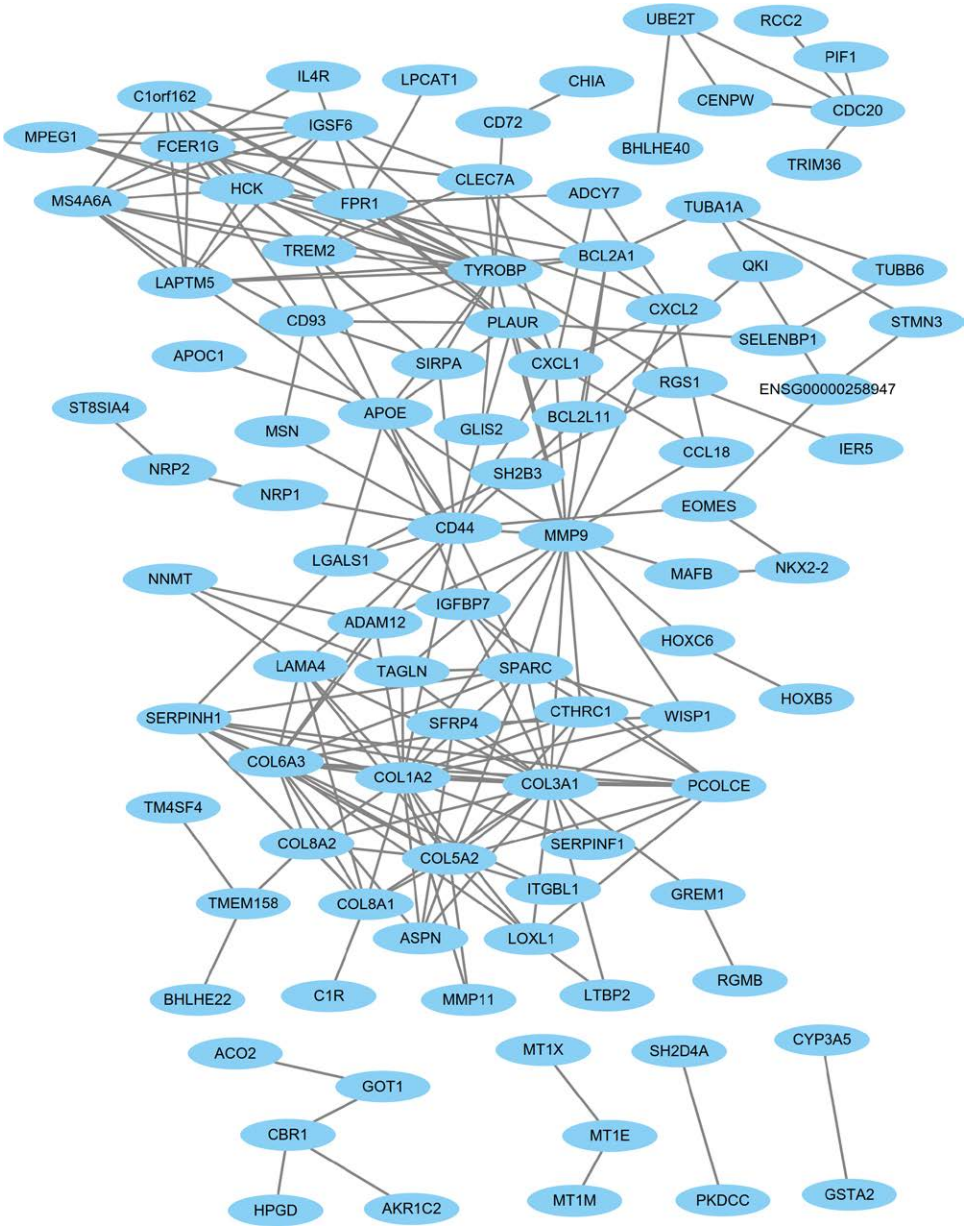

Figure 2. The PPI network of DEGs was constructed using Cytoscape. DEG = differentially expressed genes, PPI = protein–protein interaction.

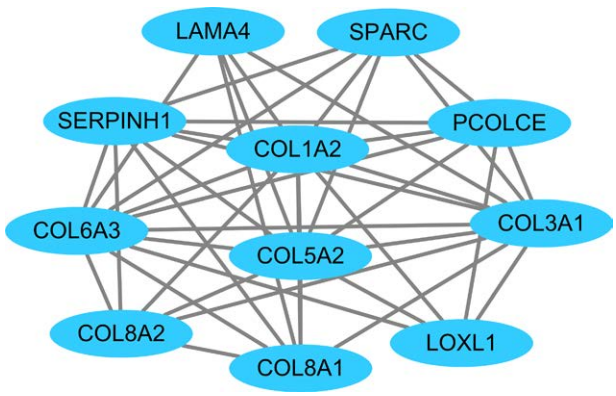

Figure 3. The most significant module was obtained from PPI network with 11 nodes. PPI = protein–protein interaction.

sets that were significantly up-regulated and down-regulated were found by GO analysis and KEGG analysis.

COL1A2 (Collagen Type I Alpha 2 Chain) is a member of the fibrocollagen family and encodes a pro-alpha 2 chain of type I collagen.[28] It acts to support the matrix structure, forming the interstitial part of most solid tumors, and regulates cell movement through interaction with the cytoskeleton. Studies have found that COL1A2 gene mainly affects cell proliferation, differentiation, adhesion and metastasis through extracellular matrix receptor interaction pathway and local adhesion pathway, mainly related to tumor invasion and metastasis.[29] Li et al found that the expression of COL1A2 in GC tissues was higher than that in adjacent normal tissues,[30] which was the same as the bioinformatics analysis in this study. Ponticos et al suggest that low expression of COL1A2 can inhibit the expression of TGF-B in cancer cells.[31] Since TGF-B contributes to the activation of PI3K signaling pathway, it is hypothesized that low expression of COL1A2 may inhibit the activation of

Table 3

Summaries for the function of 11 hub genes.

No.	Gene symbol	Full name	Function
1	COL1A2	Collagen Type I Alpha 2 Chain	Type I collagen is a member of group I collagen (fibrillar forming collagen). collagen type I, alpha 2, fibril forming, putative down-regulated c-Myc target gene, COL1A2.
2	COL3A1	Collagen Type III Alpha 1 Chain	Involved in regulation of cortical development. Is the major ligand of ADGRG1 in the developing brain and binding to ADGRG1 inhibits neuronal migration and activates the RhoA pathway by coupling ADGRG1 to GNA13 and possibly GNA12.
3	SPARC	Secreted Protein Acidic And Cysteine Rich	Appears to regulate cell growth through interactions with the extracellular matrix and cytokines. Binds calcium and copper, several types of collagen, albumin, thrombospondin, PDGF and cell membranes.
4	PCOLCE	Procollagen C-Endopeptidase Enhancer	Binds to the C-terminal propeptide of type I procollagen and enhances procollagen C-proteinase activity. C-terminal processed part of PCPE (CT-PCPE) may have an metalloproteinase inhibitory activity.
5	COL8A1	Collagen Type VIII Alpha 1 Chain	Macromolecular component of the subendothelium. Major component of the Descemet's membrane (basement membrane) of corneal endothelial cells. Also component of the endothelia of blood vessels.
6	SERPINH1	Serpin Family H Member 1	Binds specifically to collagen. Could be involved as a chaperone in the biosynthetic pathway of collagen.
7	COL8A2	Collagen Type VIII Alpha 2 Chain	Necessary for migration and proliferation of vascular smooth muscle cells and thus, has a potential role in the maintenance of vessel wall integrity and structure, in particular in atherogenesis.
8	COL6A3	Collagen Type VI Alpha 3 Chain	Collagen VI acts as a cell-binding protein. collagen type VI,alpha 3 (300kDa),microfibrillar,putative down-regulated c-Myc target gene, COL6A3
9	LAMA4	Laminin Subunit Alpha 4	Binding to cells via a high affinity receptor.
10	LOXL1	Lysyl Oxidase Like 1	Active on elastin and collagen substrates.
11	COL5A2	Collagen Type V Alpha 2 Chain	Type V collagen binds to DNA, heparan sulfate, thrombospondin, heparin, and insulin. Type V collagen is a key determinant in the assembly of tissue-specific matrices.

PI3K signaling pathway by down-regulating the expression of TGF-B in cancer cells, and promote the apoptosis of GC cells.[28] The high expression of COL1A2 can promote the proliferation, invasion and migration of GC, while the low expression of COL1A2 can inhibit the proliferation of GC cells, delay cell migration, and promote the apoptosis of GC cells. Therefore, COL1A2 can be a potential biomarker and therapeutic target.

SPARC (secreted protein acid and cysteine rich) is located in 5q33.1. It is a relative molecular mass of 32,000 nonstructural secreted extracellular matrix glycoprotein, it consists of a single polypeptide (285 amino acids), with the first 1981 U.S. TERMINE equal separation and purification of fetal bovine bone in humans.[32] It mediates the interaction of cell-microenvironment and has a wide range of biological effects in tumorigenesis, invasion, metastasis, angiogenesis and inflammation.[33] The study found that in some tumors with high metastatic characteristics, such as glioblastoma, melanoma, breast cancer and prostate cancer, SPARC can promote bone metastasis and epithelial-mesenchymal transition and promote tumor development, but as an anti-angiogenesis pancreatic cancer, colorectal cancer, gastric low metastatic tumors, pro-apoptotic, inhibition of cell proliferation and inhibition of cell cycle antitumor factor.[34,35] Its role in GC cells is highly controversial. Tsutomu et al found that the expression of SPARC mRNA in GC tissues was higher than that in the normal control group, and the prognosis of high SPARC expression was poor compared with low SPARC expression.[36] Chen et al also showed that in 140 ovarian cancer patients, high SPARC expression had a worse prognosis than low SPARC expression.[37] Chew et al and Liang et al reported that low SPARC expression was associated with poor long-term survival in 120 patients and 114 patients with colorectal cancer.[38,39] SPARC may play different roles in different stages of cancer and different stages of development of the same cancer. This study found that SPARC expression in GC tissues was higher than that in adjacent tissues, and the prognosis was poor.

SERPINH1 (Serpin Family H Member 1) is a member of the serine protease inhibitor H subfamily, also known as HPS47, a heat shock protein 47, and the coding gene is located in the 11q13.5 region of human chromosome 11. It is involved in BP such as collagen synthesis and endopeptidase activity, and can be used as a partner in the biosynthesis pathway of collagen.[40] SERPINH1 is closely associated with collagen-related diseases, including osteogenesis imperfecta, keloids, and fibrosis.[41,42] Qi et al found that SERPINH1 is highly expressed in renal clear cell carcinoma with poor prognosis.[43] Studies have reported that SERPINH1 is associated with the occurrence and development of glioma and cervical cancer, and is a possible therapeutic target.[44,45] Zhang et al found that SERPINH1 is up-regulated in GC,[46] and it is possible to promote tumor growth and invasion by regulating the extracellular matrix (ECM) network. This study found that high expression of SERPINH1 in GC tissues, poor prognosis in patients with low expression, can be a potential biomarker.

Our study identified 139 DEGs and 11 Hub genes that may be associated with the occurrence and development of GC. There are corresponding literatures indicating that COL1A2,[47,48] COL3A1,[49] SPARC,[50] SERPINH1,[51] COL6A3.[26] These genes are highly expressed in GC tissues, and the expression of LOXL1[52] is also related to distant metastasis of GC. However, the PCOLCE, COL8A2, COL8A1, and LAMA4 genes have not yet been documented to indicate their role in GC, and we subsequently recruited some patients. Relevant RT-qPCR experimental verification of these Hub genes is more indicative of the role of these genes in the development of GC than other studies.

Although the study conducted a rigorous bioinformatics analysis, a large number of clinical samples, animal experiments should be comprehensively verified to better understand the pathogenesis of primary colorectal cancer.

In summary, we identified 20 gene sets and 10 distinct DEGs from genetic samples from patients with GC and normal subjects through bioinformatics analysis. Hub genes in DEGs may provide new ideas and evidence for the diagnosis and targeted therapy of GC.

5. Conclusion

In conclusion, the present research aimed to identify DEGs which might be contained in the occurrence or development of GC. Finally, 139 DEGs and 11 hub genes were confirmed between GC tissues and normal tissues, which could be used as diagnostic and therapeutic biomarkers for GC. However, the biological functions of the all hub genes in GC require further researches.

Author contributions

Conceptualization: Xu-Dong Zhou.
Data curation: Ya-Wei Qu.
Formal analysis: Xu-Dong Zhou.
Investigation: Ya-Wei Qu, Fu-Hua Jia.

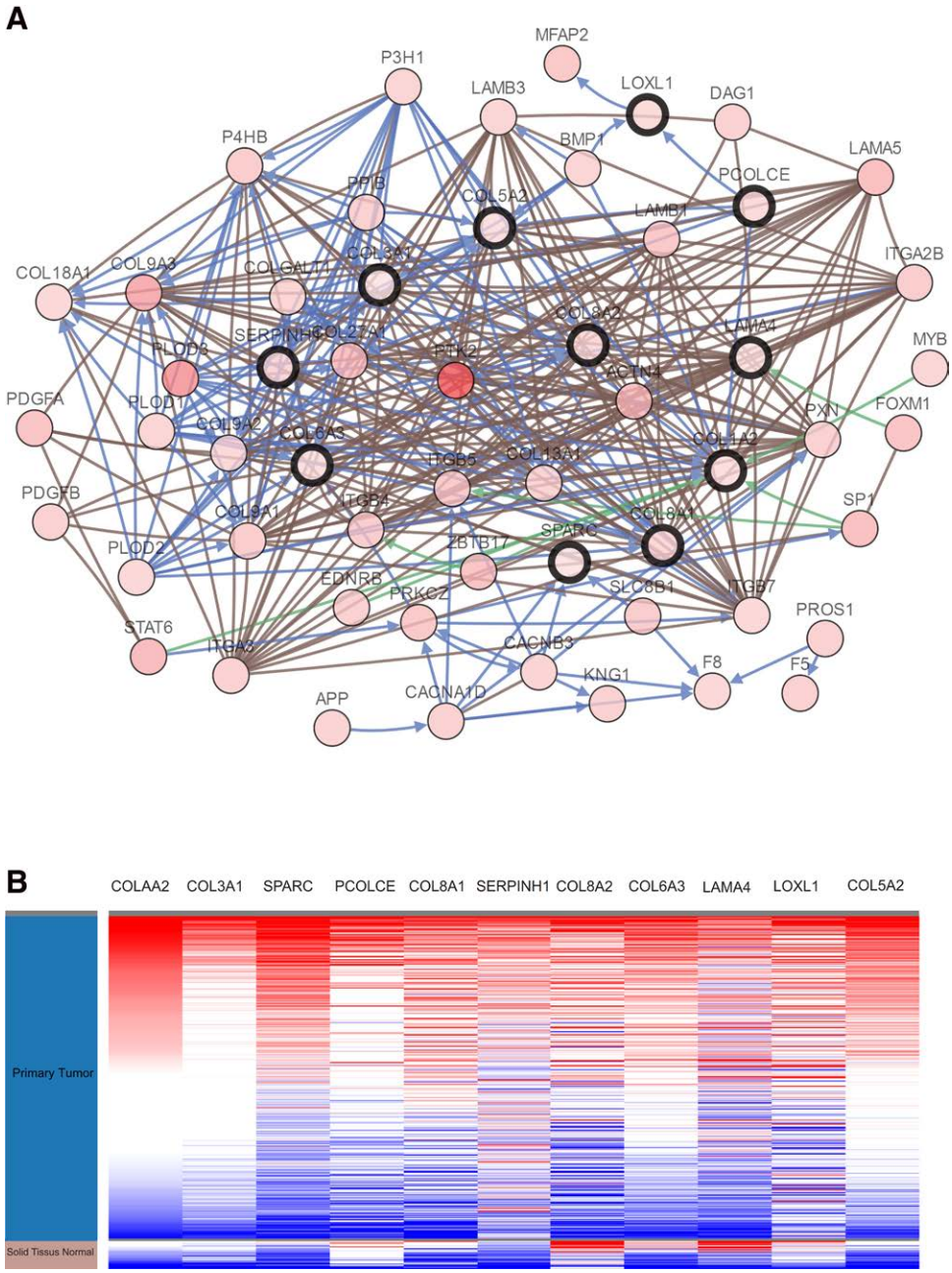

Figure 4. Interaction network and biological process analysis of the hub genes. (A) Hub genes and their co-expression genes were analyzed using cBioPortal. Nodes with bold black outline represent hub genes. Nodes with thin black outline represent the co-expression genes. (B) Hierarchical clustering of hub genes was constructed using UCSC. The samples under the pink bar are non-cancerous samples and the samples under the blue bar are GC samples. Upregulation of genes is marked in red; downregulation of genes is marked in blue. GC = gastric cancer.

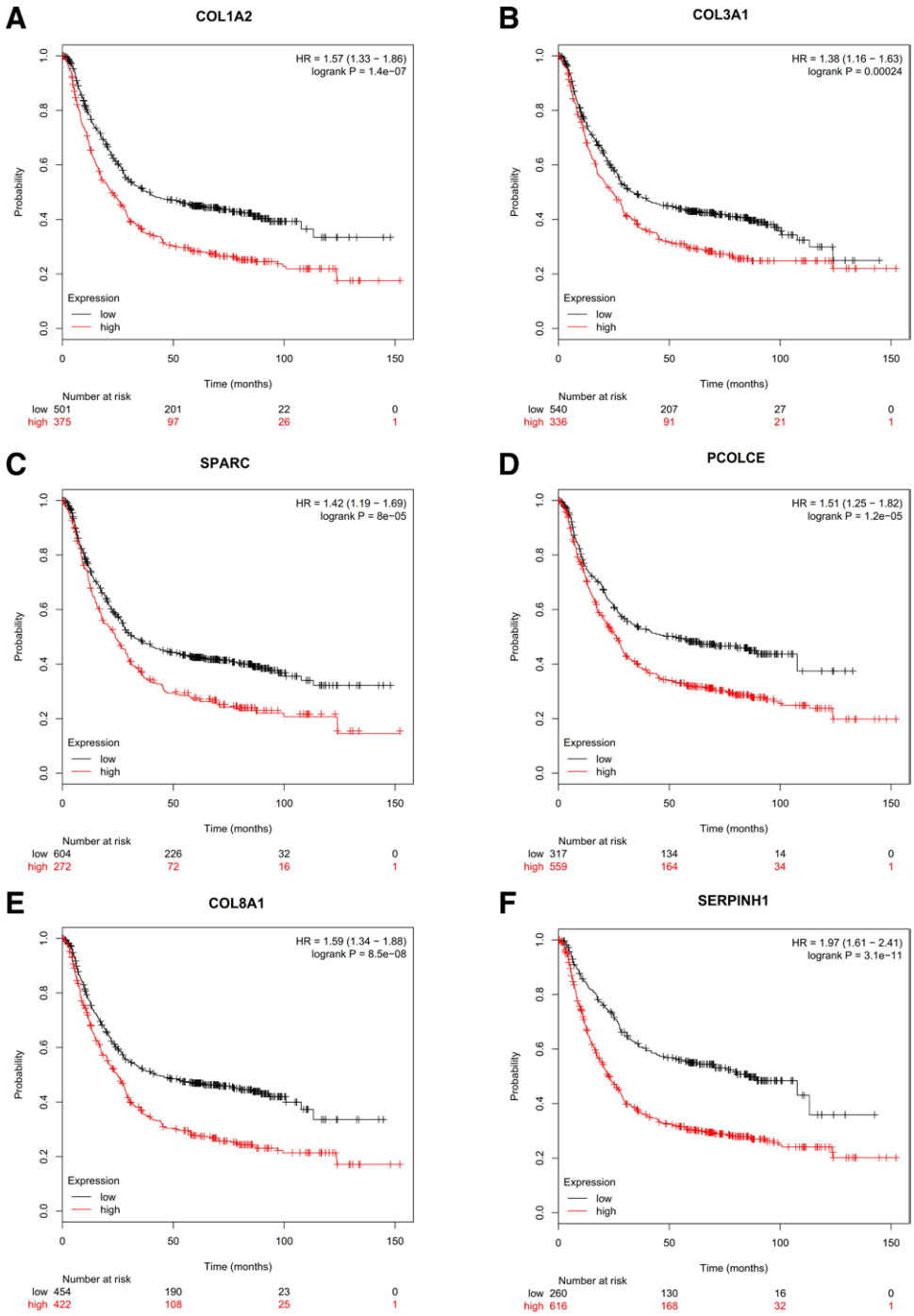

Figure 5. Overall survival analyses of hub genes (COL1A2, COL3A1, SPARC, PCOLCE, COL8A1, and SERPINH1). $P < .05$ was considered statistically significant.

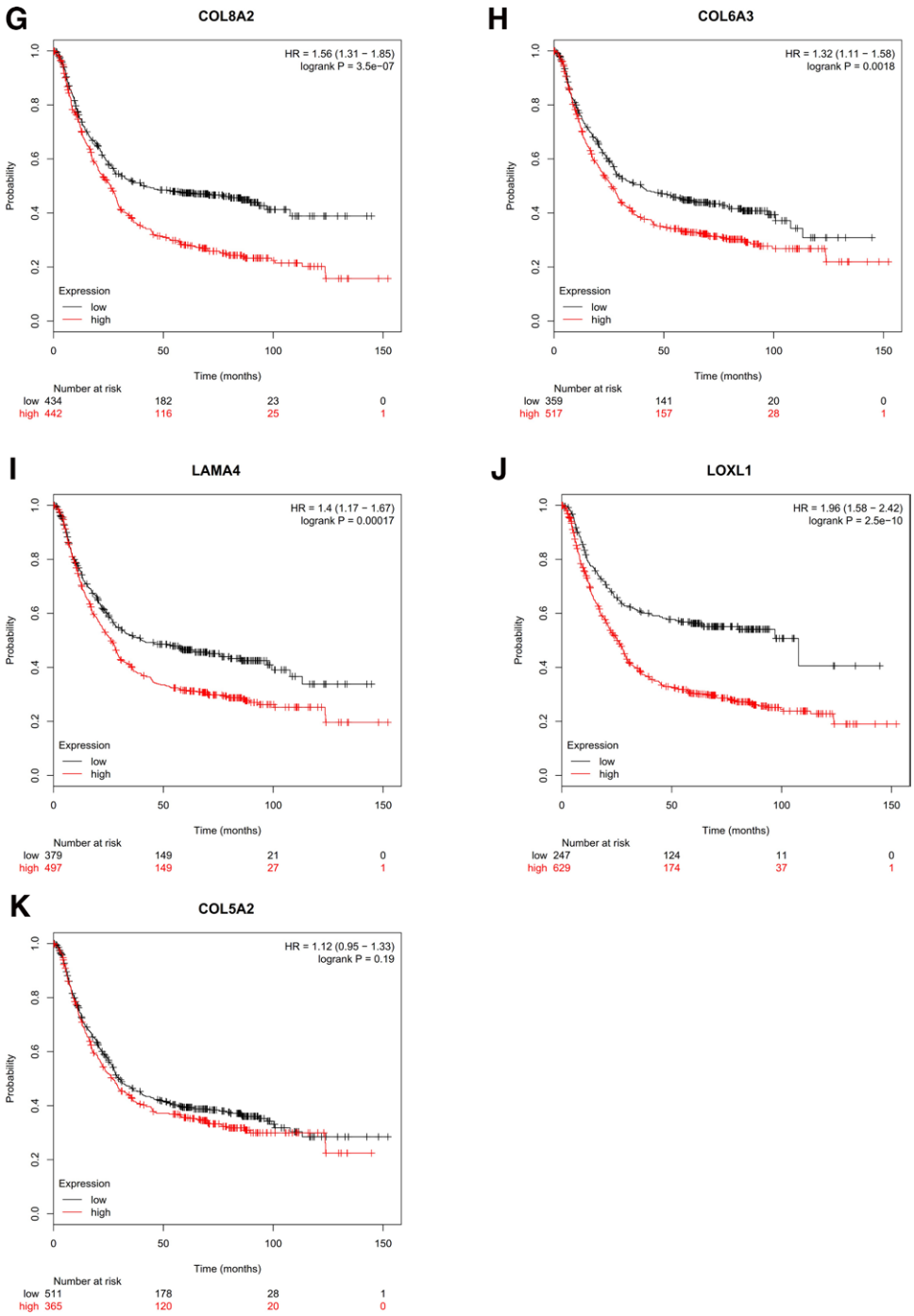

Figure 6. Overall survival analyses of hub genes (COL8A2, COL6A3, LAMA4, LOXL1, and COL5A2). $P < .05$ was considered statistically significant.

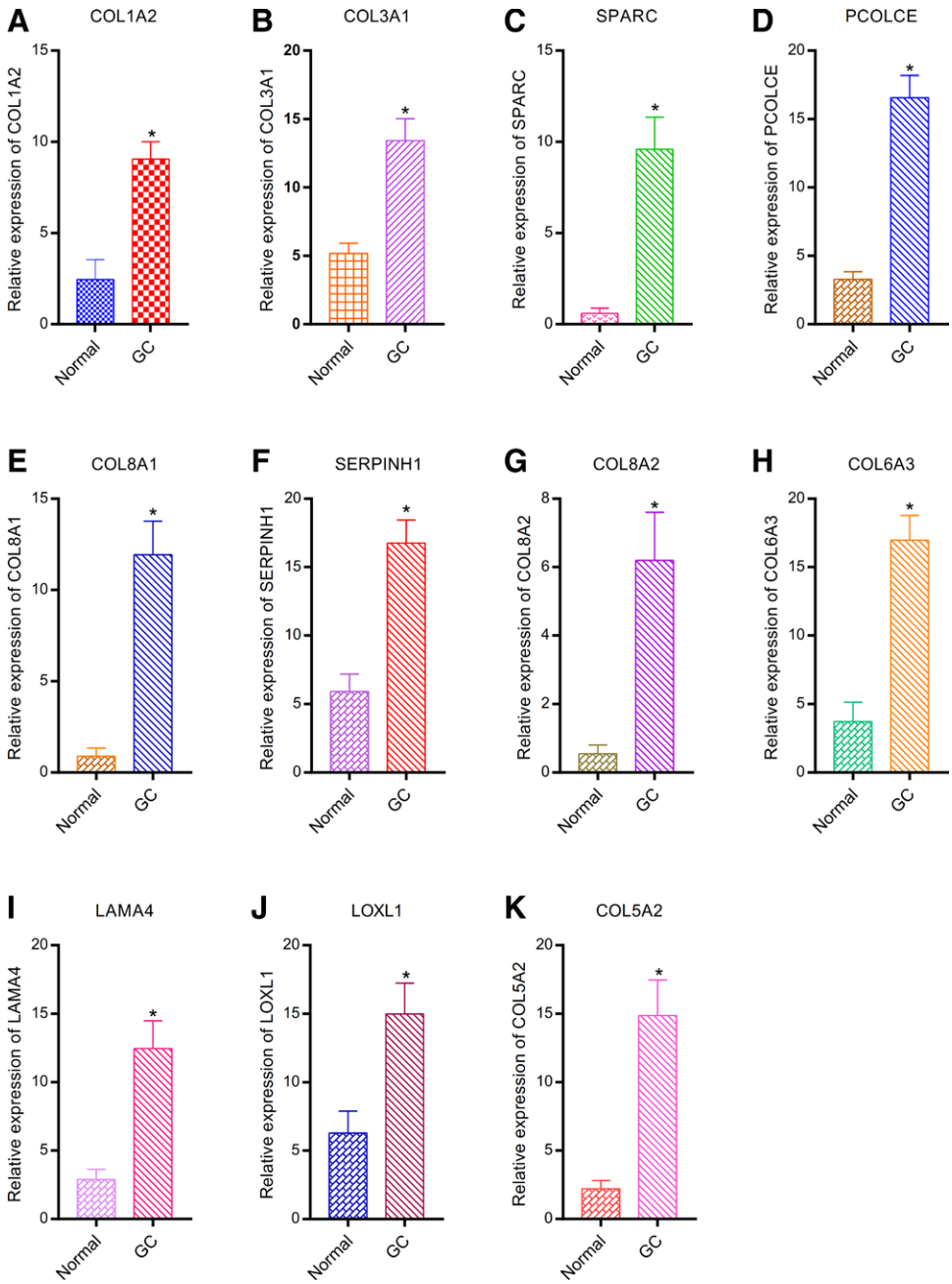

Figure 7. Relative expression of COL1A2, COL3A1, SPARC, PCOLCE, COL8A1, SERPINH1, COL8A2, COL6A3, LAMA4, LOXL1, and COL5A2 by RT-qPCR analysis. *$P < .05$, compared with normal stomach tissues. RT-qPCR = real time quantitative polymerase chain reaction.

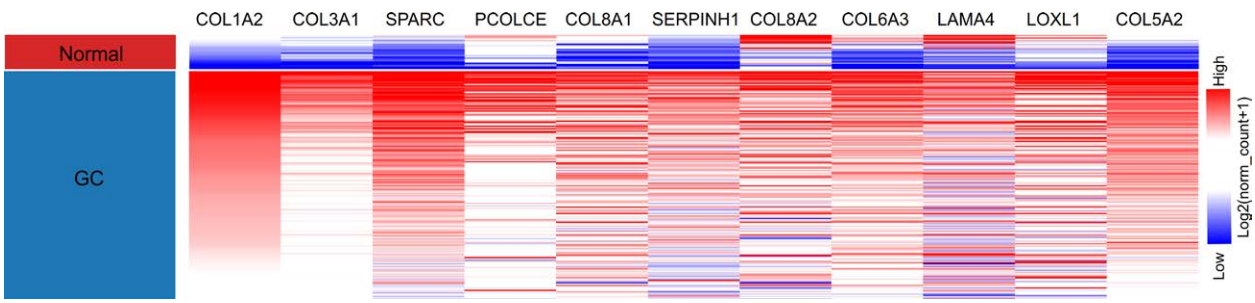

Figure 8. The confirmation of gene expression level using The Cancer Genome Atlas (TCGA) data. The genes expression levels of COL1A2, COL3A1, SPARC, PCOLCE, COL8A1, SERPINH1, COL8A2, COL6A3, LAMA4, LOXL1, and COL5A2 in GC samples were significantly higher than the normal gastric samples. GC = gastric cancer.

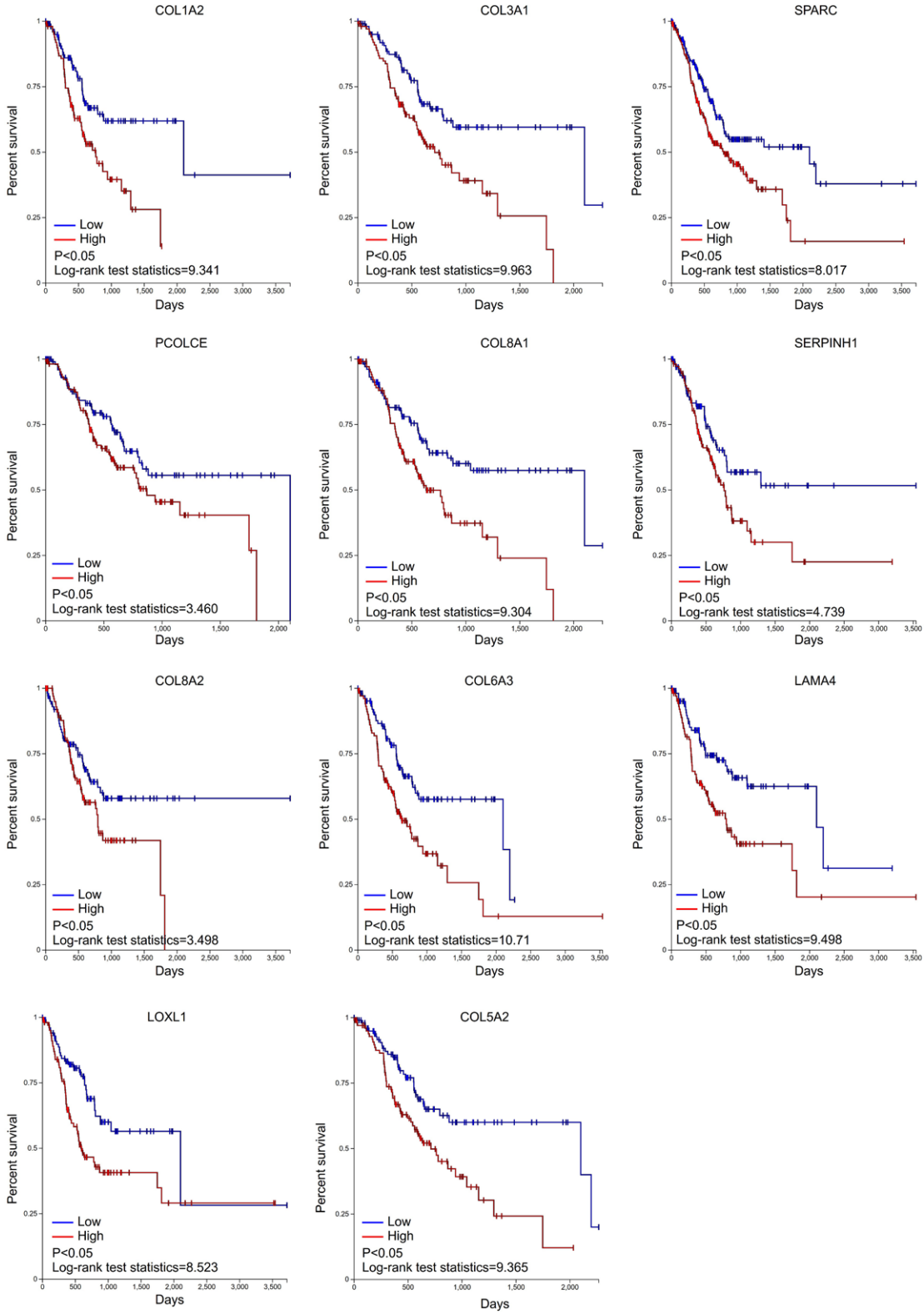

Figure 9. The effect of gene expression on overall survival by using the TCGA data. TCGA = The Cancer Genome Atlas.

Methodology: Fu-Hua Jia.
Resources: Xu-Dong Zhou.
Supervision: Peng Chen.
Validation: Peng Chen, Yin-Pu Wang.
Writing – original draft: Yin-Pu Wang.
Writing – review & editing: Hai-Feng Liu.

References

[1] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin. 2016;66:115–32.
[2] Ke D, Li H, Zhang Y, et al. The combination of circulating long noncoding RNAs AK001058, INHBA-AS1, MIR4435-2HG, and CEBPA-AS1 fragments in plasma serve as diagnostic markers for gastric cancer. Oncotarget. 2017;8:21516–25.
[3] Cunningham D, Okines AF, Ashley S. Capecitabine and oxaliplatin for advanced esophagogastric cancer. N Engl J Med. 2010;362:858–9.
[4] Parkin DM, Bray F, Ferlay J, et al. Global cancer statistics, 2002. CA Cancer J Clin. 2005;55:74–108.
[5] Shimada H, Noie T, Ohashi M, et al. Clinical significance of serum tumor markers for gastric cancer: a systematic review of literature by the Task Force of the Japanese Gastric Cancer Association. Gastric Cancer. 2014;17:26–33.
[6] Zhou Y, Liepe J, Sheng X, et al. GPU accelerated biochemical network simulation. Bioinformatics. 2011;27:874–6.
[7] Nobile MS, Cazzaniga P, Tangherloni A, et al. Graphics processing units in bioinformatics, computational biology and systems biology. Brief Bioinform. 2017;18:870–85.
[8] Li L, Lei Q, Zhang S, et al. Screening and identification of key biomarkers in hepatocellular carcinoma: evidence from bioinformatic analysis. Oncol Rep. 2017;38:2607–18.
[9] Milan T, Wilhelm BT. Mining cancer transcriptomes: bioinformatic tools and the remaining challenges. Mol Diagn Ther. 2017;21:249–58.
[10] Wang Z, AUID- O, Monteiro CD, et al. Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. Nat Commun. 2016;7:12846.
[11] Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013;41(Database issue):D991–5.
[12] Huang DW, Sherman BT, Tan Q, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 2007;8:R183.
[13] Kanehisa M. The KEGG database. Novartis Found Symp. 2002;247:91–101; discussion 101.
[14] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.
[15] Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(Database issue):D447–52.
[16] Su G, Morris JH, Demchak B, et al. Biological network exploration with Cytoscape 3. Curr Protoc Bioinformatics. 2014;47:8.13.1–24.
[17] Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6:pl1.
[18] Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2:401–4.
[19] Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. Genome Res. 2002;12:996–1006.
[20] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
[21] Macdonald JS, Smalley SR, Benedetti J, et al. Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction. N Engl J Med. 2001;345:725–30.
[22] Cunningham D, Allum WH, Stenning SP, et al. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. N Engl J Med. 2006;355:11–20.
[23] McCall MD, Graham PJ, Bathe OF. Quality of life: A critical outcome for all surgical treatments of gastric cancer. World J Gastroenterol. 2016;22:1101–13.
[24] Goh YM, Gillespie C, Couper G, et al. Quality of life after total and subtotal gastrectomy for gastric carcinoma. Surgeon. 2015;13:267–70.
[25] Ito Y, Yoshikawa T, Fujiwara M, et al. Quality of life and nutritional consequences after aboral pouch reconstruction following total gastrectomy for gastric cancer: randomized controlled trial CCG1101. Gastric Cancer. 2016;19:977–85.
[26] Cao W, Zhou D, Tang W, et al. Discovery of plasma messenger RNA as novel biomarker for gastric cancer identified through bioinformatics analysis and clinical validation. PeerJ. 2019;7:e7025.
[27] Wang X, Liu Z, Tong H, et al. Linc01194 acts as an oncogene in colorectal carcinoma and is associated with poor survival outcome. Cancer Manag Res. 2019;11:2349–62.
[28] Ao R. 0000-0001-7672-1793 AO, Guan L, Wang Y, Wang JN. Silencing of COL1A2, COL6A3, and THBS2 inhibits gastric cancer cell proliferation, migration, and invasion while promoting apoptosis through the PI3k-Akt signaling pathway. J Cell Biochem. 2018;119:4420–34.
[29] Haq F, Ahmed N, Qasim M. Comparative genomic analysis of collagen gene diversity. 3 Biotech. 2019;9:83.
[30] Rong L, Huang W, Tian S, et al. COL1A2 is a novel biomarker to improve clinical prediction in human gastric cancer: integrating bioinformatics and meta-analysis. Pathol Oncol Res. 2018;24:129–34.
[31] Ponticos M, Harvey C, Ikeda T, et al. JunB mediates enhancer/promoter activity of COL1A2 following TGF-beta induction. Nucleic Acids Res. 2009;37:5378–89.
[32] Termine JD, Kleinman HK, Whitson SW, et al. Osteonectin, a bone-specific protein linking mineral to collagen. Cell. 1981;26(1 Pt 1):99–105.
[33] Chlenski A, Cohn SL. Modulation of matrix remodeling by SPARC in neoplastic progression. Semin Cell Dev Biol. 2010;21:55–65.
[34] Feng J, Tang L. SPARC in tumor pathophysiology and as a potential therapeutic target. Curr Pharm Des. 2014;20:6182–90.
[35] Nagaraju GP, Dontula R, El-Rayes BF, et al. Molecular mechanisms underlying the divergent roles of SPARC in human carcinogenesis. Carcinogenesis. 2014;35:967–73.
[36] Sato T, Oshima T, Yamamoto N, et al. Clinical significance of SPARC gene expression in patients with gastric cancer. J Surg Oncol. 2013;108:364–8.
[37] Chen J, Wang M, Xi B, et al. SPARC is a key regulator of proliferation, apoptosis and invasion in human ovarian cancer. PLoS One. 2012;7:e42413.
[38] Chew A, Salama P, Robbshaw A, et al. SPARC, FOXP3, CD8 and CD45 correlation with disease recurrence and long-term disease-free survival in colorectal cancer. PLoS One. 2011;6:e22047.
[39] Liang JF, Wang HK, Xiao H, et al. Relationship and prognostic significance of SPARC and VEGF protein expression in colon cancer. J Exp Clin Cancer Res. 2010;29:71.
[40] Ito S, Nagata K. Biology of Hsp47 (Serpin H1), a collagen-specific molecular chaperone. Semin Cell Dev Biol. 2017;62:142–51.
[41] Friedman SL. Mechanisms of hepatic fibrogenesis. Gastroenterology. 2008;134:1655–69.
[42] Marini JC, Reich A, Smith SM. Osteogenesis imperfecta due to mutations in non-collagenous genes: lessons in the biology of bone formation. Curr Opin Pediatr. 2014;26:500–7.
[43] Qi Y, Zhang Y, Peng Z, et al. SERPINH1 overexpression in clear cell renal cell carcinoma: association with poor clinical outcome and its potential as a novel prognostic marker. J Cell Mol Med. 2018;22:1224–35.
[44] Wu ZB, Cai L, Lin SJ, et al. Heat shock protein 47 promotes glioma angiogenesis. Brain Pathol. 2016;26:31–42.
[45] Yamamoto N, Kinoshita T, Nohata N, et al. Tumor-suppressive microRNA-29a inhibits cancer cell migration and invasion via targeting HSP47 in cervical squamous cell carcinoma. Int J Oncol. 2013;43:1855–63.
[46] Zhang X, Yang JJ, Kim YS, et al. An 8-gene signature, including methylated and down-regulated glutathione peroxidase 3, of gastric cancer. Int J Oncol. 2010;36:405–14.
[47] Li J, Ding Y, Li A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. World J Surg Oncol. 2016;14:297.
[48] Li L, Zhu Z, Zhao Y, et al. FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics. Sci Rep. 2019;9:7827.
[49] Nie K, Shi L, Wen Y, et al. Identification of hub genes correlated with the pathogenesis and prognosis of gastric cancer via bioinformatics methods. Minerva Med. 2020;111:213–25.
[50] Liao P, Li W, Liu R, et al. Genome-scale analysis identifies SERPINE1 and SPARC as diagnostic and prognostic biomarkers in gastric cancer. Onco Targets Ther. 2018;11:6969–80.
[51] Tian S, Peng P, Li J, et al. SERPINH1 regulates EMT and gastric cancer metastasis via the Wnt/beta-catenin signaling pathway. Aging (Albany NY). 2020;12:3574–93.
[52] Kasashima H, Yashiro M, Okuno T, et al. Significance of the Lysyl oxidase members Lysyl oxidase like 1, 3, and 4 in gastric cancer. Digestion. 2018;98:238–48.