MEDICAL SCIENCE MONITOR

DATABASE ANALYSIS

e-ISSN 1643-3750
© Med Sci Monit, 2021; 27: e929104
DOI: 10.12659/MSM.929104

Received: 2020.10.11
Accepted: 2020.11.27
Available online: 2020.12.10
Published: 2021.02.14

Identification of Potential Biomarkers Associated with Prognosis in Gastric Cancer via Bioinformatics Analysis

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABCDEF 1 Dong Li
ACEG 2 Yi Yin
BDF 2 Muqun He
BDF 2 Jianfeng Wang

1 Cancer Institute, Fujian Cancer Hospital and Fujian Medical University Cancer Hospital, Fuzhou, Fujian, P.R. China
2 Department of Medical Oncology, Fujian Cancer Hospital and Fujian Medical University Cancer Hospital, Fuzhou, Fujian, P.R. China

Corresponding Authors: Dong Li, e-mail: jerryldong@163.com, Yi Yin, e-mail: elvof1@163.com
Source of support: This research was supported by startup funding for scientific research from Fujian Medical University (No. 2017XQ1216)

Background: Gastric cancer (GC) is one of the leading causes of cancer-related mortality worldwide. We aimed to identify differentially expressed genes (DEGs) and their potential mechanisms associated with the prognosis of GC patients.

Material/Methods: This study was based on gene profiling information for 37 paired samples of GC and adjacent normal tissues from the GSE118916, GSE79973, and GSE19826 datasets in the Gene Expression Omnibus database. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were used to investigate the biological role of the DEGs. The protein–protein interaction (PPI) network was constructed by Cytoscape, and the Kaplan-Meier plotter was used for prognostic analysis.

Results: We identified 119 DEGs, including 21 upregulated and 98 downregulated genes, in GC. The 21 upregulated genes were mainly enriched in extracellular matrix-receptor interaction, focal adhesion, and transforming growth factor-β signaling, while the 98 downregulated genes were significantly associated with gastric acid secretion, retinol metabolism, and metabolism of xenobiotics by cytochrome P450. Thirty hub DEGs were obtained for further analysis. Twenty-five of the 30 hub DEGs were significantly associated with the prognosis of GC, and 21 of the 25 hub DEGs showed consistent expression trends within the 3 profile datasets. KEGG reanalysis of these 21 hub DEGs showed that *COL1A1*, *COL1A2*, *COL2A1*, *COL11A1*, *THBS2*, and *SPP1* were mainly enriched in the extracellular matrix-receptor interaction pathways.

Conclusions: We identified 6 genes that were significantly related to the prognosis of GC patients. These genes and pathways could serve as potential prognostic markers and be used to develop treatments for GC patients.

Keywords: Gene Expression Profiling • Prognosis • Stomach Neoplasms • Tumor Markers, Biological

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/929104

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Background

Gastric cancer (GC) is the fourth most common cancer in the world and the second in China, and the second leading cause of cancer-related death worldwide and in China [1,2]. In China, there were approximately 679 100 new cases of GC and 498 000 GC-related deaths in 2015 [3]. Based on histopathological classification, GC is separated into 3 main groups, adenocarcinoma, undifferentiated carcinoma, and signet ring-cell carcinoma, and more than 95% of cases are gastric adenocarcinomas. Most GC patients are asymptomatic in the early stages, and GC is frequently not diagnosed until the disease has already progressed to an advanced stage. Due to the difficulty of early diagnosis, rapid spread, and distant metastasis of GC, the prognosis of GC remains poor. Therefore, novel reliable prognostic biomarkers need to be identified to understand the molecular mechanisms of GC and to improve therapeutic outcomes.

Microarray analysis has been used for more than 10 years as a reliable technique to probe differentially expressed genes (DEGs) and identify potential clinical biomarkers. There are a large number of valuable datasets in public databases for exploring new research questions. In this study, we analyzed the differential expression patterns between GC tumor tissues and matched normal tissues to explore the hub genes and key pathways associated with GC prognosis. We used bioinformatics methods to investigate differential gene expression and to conduct functional enrichment analyses to identify potential molecular mechanisms in GC.

Material and Methods

Microarray Data Information

We obtained GSE118916, GSE79973, and GSE19826 mRNA expression profiles from the Gene Expression Omnibus (GEO) database (*http://www.ncbi.nlm.nih.gov/geo*), a free public microarray/gene profile database [4-7]. The GSE118916, GSE79973, and GSE19826 datasets were respectively based on 15, 10, and 12 pairs of GC tissues and adjacent nontumor (normal) tissues, with gene expression profiles generated by the Affymetrix Human Gene Expression Array platform. The diagnosis of GC was independently confirmed for all patients by 2 pathologists using the criteria provided by the American Joint Committee on Cancer (AJCC) [8].

Identification of DEGs

Common DEGs from the 3 datasets were screened using the GEO2R online tool and Venn diagram software. DEGs were identified with |logFC| >2 and adjusted *P* value <0.05 as the cutoff criteria by GEO2R online tools. Venn software online was used to detect common DEGs among the 3 datasets. The DEGs with logFC <0 or logFC >0 were considered downregulated or upregulated genes, respectively.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Analysis

The DAVID database (*https://david.ncifcrf.gov/*) [9] was used to carry out Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) enrichment analysis of identified DEGs, including those associated with molecular functions, cellular components, and biological processes [10]. GO terms and KEGG pathways with $P<0.05$ were selected as significant.

Construction of Protein–protein Interaction Network and Module Analysis

Protein–protein interaction (PPI) information was obtained through an online STRING analysis (*http://string-db.org*) [11]. Cytoscape software for additional analyses was used to identify hub DEGs and to construct the PPI network. The plug-in MCODE of Cytoscape was applied to detect significant modules in the PPI network. The cutoff criteria were set with degree cutoff=2, node score cutoff=0.2, maximum depth=100, and k-core=2 [12].

Validation of Hub DEGs

The online Kaplan-Meier plotter database (*https://kmplot.com/analysis/*) was used to assess the prognostic value of hub DEGs among GC patients ($P<0.05$) [13]. There were 875 GC patients recruited for survival analysis. Hazard ratios, their 95% confidence intervals, and log-rank *P* values were calculated. The GEPIA server (*http://gepia.cancer-pku.cn/*) was used to further validate the expression of hub DEGs in GC tissues and normal tissues ($P<0.05$) [14].

Results

Identification of DEGs in GC

We extracted 350, 579, and 340 DEGs from 37 paired GC and normal gastric tissues from GSE118916, GSE79973, and GSE19826, respectively. We identified 119 common DEGs in the 3 datasets by using online Venn diagram software. These DEGs included 21 upregulated genes (logFC >0) and 98 downregulated genes (logFC <0) (**Table 1, Figure 1**).

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 1. All 119 commonly differentially expressed genes (DEGs; 21 upregulated and 98 downregulated genes) in 3 profile datasets between gastric cancer tissues and adjacent normal tissues.

DEGs	Genes name
Up-regulated	SFRP4, FNDC1, COL11A1, CEMIP, CTHRC1, IGF2BP3, SULF1, SERPINH1, FAP, RARRES1, THBS2, INHBA, CLDN4, COL1A2, SPP1, COL1A1, COL8A1, CLDN1, CLDN7, THBS4, COL10A1
Down-regulated	GPR155, GREM2, LIPF, ETNPPL, GSTA1, PDIA2, GUCA2B, FBP2, AKR1B10, CNTD1, ANXA10, TCN1, SLC2A12, FAM150B, SLC16A7, MUC6, ZNF385B, CYP2C18, FGA, ALDH3A1, ATP4A, FUT9, UGT2B15, KRT20, KIAA1324, AZGP1, GKN1, CTSE, ADGRG2, COL2A1, RDH12, C16orf89, GHRL, GIF, ALDH1A1, DUOX2, CA2, LTF, FOLR1, GATA5, ATP4B, AQP4, SULT1C2, CHGA, CAPN9, SLC26A9, CHIA, ESRRG, AKR7A3, APLP1, REG1A, ADTRP, IRX3, SSTR1, ACER2, MT1G, CPA2, C6orf58, DPCR1, VSIG1, PGC, FAM3B, SLC1A2, MYRIP, KCNE2, SOSTDC1, PDILT, CA9, VSIG2, CYP2C9, PIK3C2G, SIGLEC11, TMED6, SST, SH3GL2, TFF2, ALDOB, SCNN1G, CKMT2, CCKBR, HPGD, PSAPL1, DNER, PSCA, CWH43, KCNJ16, KCNJ15, SLC26A7, TFF1, SCGB2A1, CLDN18, HDC, RASSF6, FBXL13, CLIC6, GKN2, CXCL17, TRIM50

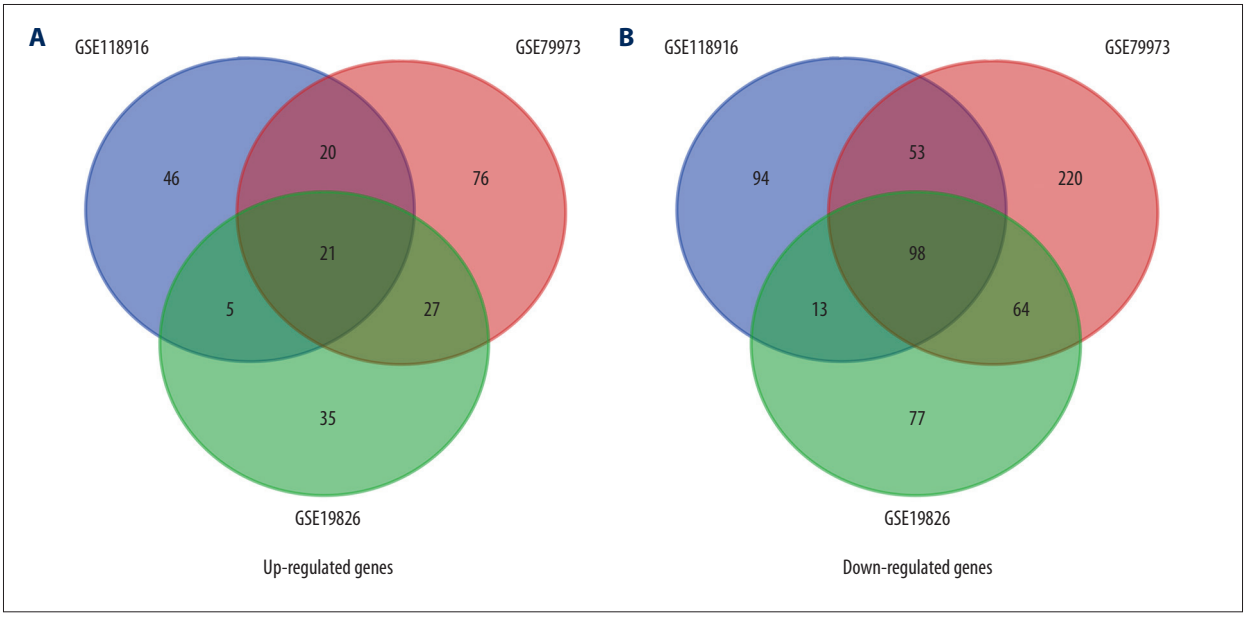

Figure 1. Venn diagrams of all screened differentially expressed genes (DEGs) identified from 3 gene expression profiles (GSE118916, GSE79973, and GSE19826). (**A**) Twenty-one upregulated genes. (**B**) Ninety-eight downregulated genes.

Functional Enrichment Analysis of DEGs

Functional enrichment analysis of the 119 DEGs was conducted using DAVID software. The DEGs were divided into 3 functional groups: biological processes, cellular components, and molecular functions. The GO analysis showed that for biological processes, upregulated DEGs were mainly related to collagen fibril organization, endodermal cell differentiation, negative regulation of angiogenesis, collagen biosynthetic process, skin morphogenesis, and protein heterotrimerization, while the downregulated DEGs were related to digestion, cellular aldehyde metabolic process, gastric acid secretion, potassium ion import, xenobiotic metabolic process, and oxidation-reduction process. For cellular components, upregulated DEGs were involved in collagen trimer, extracellular space, proteinaceous extracellular matrix (ECM), collagen type I trimer, bicellular tight junction, and apicolateral plasma membrane, and the downregulated DEGs were associated with extracellular space, extracellular exosome, basolateral plasma membrane, integral component of plasma membrane, apical plasma membrane, and extracellular region. For molecular function, upregulated DEGs were particularly enriched in ECM structural constituents and structural molecule activity, and downregulated DEGs were involved in inward-rectifier potassium channel activity, benzaldehyde dehydrogenase (NAD^+) activity, hydrogen-potassium exchange ATPase activity, identical protein binding, chloride channel activity, and retinal dehydrogenase activity (**Table 2**).

The KEGG analysis results provided by DAVID software are shown in **Table 3**. The upregulated DEGs were mainly associated

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 2. Gene Ontology (GO) term enrichment analysis of differentially expressed genes (DEGs) in gastric cancer.

Expression	Category	Term	Count	%	*P*-value	FDR
Up-regulated	GOTERM_BP_DIRECT	GO: 0030199~collagen fibril organization	4	11.49	6.43E-06	0.007663
	GOTERM_BP_DIRECT	GO: 0035987~endodermal cell differentiation	3	8.62	4.66E-04	0.553934
	GOTERM_BP_DIRECT	GO: 0016525~negative regulation of angiogenesis	3	8.62	0.001384	1.636565
	GOTERM_BP_DIRECT	GO: 0032964~collagen biosynthetic process	2	5.75	0.005185	6.005438
	GOTERM_BP_DIRECT	GO: 0043589~skin morphogenesis	2	5.75	0.009057	10.272816
	GOTERM_BP_DIRECT	GO: 0070208~protein heterotrimerization	2	5.75	0.010345	11.652002
	GOTERM_CC_DIRECT	GO: 0005581~collagen trimer	5	14.37	2.69E-07	2.44E-04
	GOTERM_CC_DIRECT	GO: 0005615~extracellular space	10	28.74	4.16E-07	3.78E-04
	GOTERM_CC_DIRECT	GO: 0005578~proteinaceous extracellular matrix	4	11.49	0.001308	1.182888
	GOTERM_CC_DIRECT	GO: 0005584~collagen type I trimer	2	5.75	0.002728	2.453896
	GOTERM_CC_DIRECT	GO: 0005923~bicellular tight junction	3	8.62	0.004148	3.709096
	GOTERM_CC_DIRECT	GO: 0016327~apicolateral plasma membrane	2	5.75	0.016264	13.853770
	GOTERM_MF_DIRECT	GO: 0005201~extracellular matrix structural constituent	3	8.62	0.001010	0.806984
	GOTERM_MF_DIRECT	GO: 0005198~structural molecule activity	3	8.62	0.012211	9.385704
Down-regulated	GOTERM_BP_DIRECT	GO: 0007586~digestion	11	8.70	1.33E-12	1.92E-09
	GOTERM_BP_DIRECT	GO: 0006081~cellular aldehyde metabolic process	4	3.16	2.46E-05	0.035612
	GOTERM_BP_DIRECT	GO: 0001696~gastric acid secretion	3	2.37	4.30E-04	0.619734
	GOTERM_BP_DIRECT	GO: 0010107~potassium ion import	4	3.16	4.57E-04	0.659491
	GOTERM_BP_DIRECT	GO: 0006805~xenobiotic metabolic process	5	3.95	8.48E-04	1.219547
	GOTERM_BP_DIRECT	GO: 0055114~oxidation-reduction process	11	8.70	0.001358	1.947012
	GOTERM_CC_DIRECT	GO: 0005615~extracellular space	26	20.56	1.19E-08	1.32E-05
	GOTERM_CC_DIRECT	GO: 0070062~extracellular exosome	30	23.73	1.24E-04	0.137838
	GOTERM_CC_DIRECT	GO: 0016323~basolateral plasma membrane	7	5.54	3.39E-04	0.375425
	GOTERM_CC_DIRECT	GO: 0005887~integral component of plasma membrane	15	11.86	0.013478	13.981885
	GOTERM_CC_DIRECT	GO: 0016324~apical plasma membrane	6	4.75	0.017388	17.691052
	GOTERM_CC_DIRECT	GO: 0005576~extracellular region	16	12.65	0.017544	17.836177
	GOTERM_MF_DIRECT	GO: 0005242~inward rectifier potassium channel activity	3	2.37	0.004085	5.024893
	GOTERM_MF_DIRECT	GO: 0018479~benzaldehyde dehydrogenase (NAD+) activity	2	1.58	0.009574	11.410763
	GOTERM_MF_DIRECT	GO: 0008900~hydrogen: potassium-exchanging ATPase activity	2	1.58	0.014327	16.618617
	GOTERM_MF_DIRECT	GO: 0042802~identical protein binding	9	7.12	0.027002	29.160325
	GOTERM_MF_DIRECT	GO: 0005254~chloride channel activity	3	2.37	0.027701	29.799387
	GOTERM_MF_DIRECT	GO: 0001758~retinal dehydrogenase activity	2	1.58	0.033114	34.565617

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 3. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of differentially expressed genes (DEGs) in gastric cancer.

Expression	Pathway ID	Name	Count	%	P-value	Genes
Up-regulated	hsa04512	ECM-receptor interaction	6	2.87	2.58E-07	COL1A2, COL1A1, THBS2, COL11A1, THBS4, SPP1
	hsa04510	Focal adhesion	6	2.87	1.97E-05	COL1A2, COL1A1, THBS2, COL11A1, THBS4, SPP1
	hsa04350	TGF-beta signaling pathway	3	1.44	0.011911	INHBA, THBS2, THBS4
	hsa04670	Leukocyte transendothelial migration	3	1.44	0.021274	CLDN7, CLDN4, CLDN1
	hsa04514	Cell adhesion molecules (CAMs)	3	1.44	0.026257	CLDN7, CLDN4, CLDN1
	hsa04530	Tight junction	3	1.44	0.027005	CLDN7, CLDN4, CLDN1
Down-regulated	hsa04971	Gastric acid secretion	9	0.07	2.05E-08	KCNJ16, KCNJ15, CCKBR, ATP4A, ATP4B, SLC26A7, KCNE2, CA2, SST
	hsa00830	Retinol metabolism	5	0.04	8.30E-04	ALDH1A1, RDH12, CYP2C9, CYP2C18, UGT2B15
	hsa00980	Metabolism of xenobiotics by cytochrome P450	5	0.04	0.001431	GSTA1, CYP2C9, AKR7A3, UGT2B15, ALDH3A1
	hsa05204	Chemical carcinogenesis	5	0.04	0.001910	GSTA1, CYP2C9, CYP2C18, UGT2B15, ALDH3A1
	hsa00982	Drug metabolism – cytochrome P450	4	0.03	0.010356	GSTA1, CYP2C9, UGT2B15, ALDH3A1
	hsa01100	Metabolic pathways	16	0.13	0.010455	ETNPPL, PIK3C2G, FUT9, CYP2C9, CYP2C18, ACER2, ALDOB, FBP2, ALDH3A1, RDH12, ALDH1A1, AKR1B10, CKMT2, HDC, UGT2B15, LIPF
	hsa04966	Collecting duct acid secretion	3	0.02	0.013808	ATP4A, ATP4B, CA2
	hsa00051	Fructose and mannose metabolism	3	0.02	0.019104	AKR1B10, ALDOB, FBP2

with the signaling pathways in ECM-receptor interaction, focal adhesion, transforming growth factor-β signaling pathway, leukocyte transendothelial migration, cell adhesion molecules, and tight junctions, while the downregulated DEGs were significantly enriched in gastric acid secretion, retinol metabolism, metabolism of xenobiotics by cytochrome P450, chemical carcinogenesis, drug metabolism-cytochrome P450, metabolic pathways, collecting duct acid secretion, and fructose and mannose metabolism.

PPI Network and Cluster Analysis

STRING and Cytoscape were used to construct the PPI network and conduct further explorations. A total of 119 DEGs (21 upregulated and 98 downregulated genes) were filtered into the PPI network, which contained 90 nodes and 162 edges (**Figure 2A**) and excluded 29 DEGs. We applied Cytoscape MCODE for further analysis to obtain 30 hub nodes among 90 nodes and 75 edges (**Figure 2B**). The 30 hub DEGs included 13 upregulated and 17 downregulated genes.

Survival Analysis and Validation of Hub DEGs

To further analyze the effect of hub DEGs in GC, the Kaplan-Meier plotter was used to identify 30 hub DEGs associated with the prognosis of 875 GC patients. The results demonstrated 25 hub DEGs were significantly associated with the prognosis of GC patients, while *CLDN1*, *GKN1*, *HDC*, *GIF*, and *FAP* were not statistically significant ($P>0.05$, **Table 4, Figure 3**). The online GEPIA software was used to validate the expression of 25 hub DEGs between GC tissues and normal tissues. Twenty-one of the 25 hub DEGs showed significantly different expression and consistent expression trends in GSE118916, GSE79973, and GSE19826 (**Table 5, Figure 4**). A total of 11 out of 21 hub

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

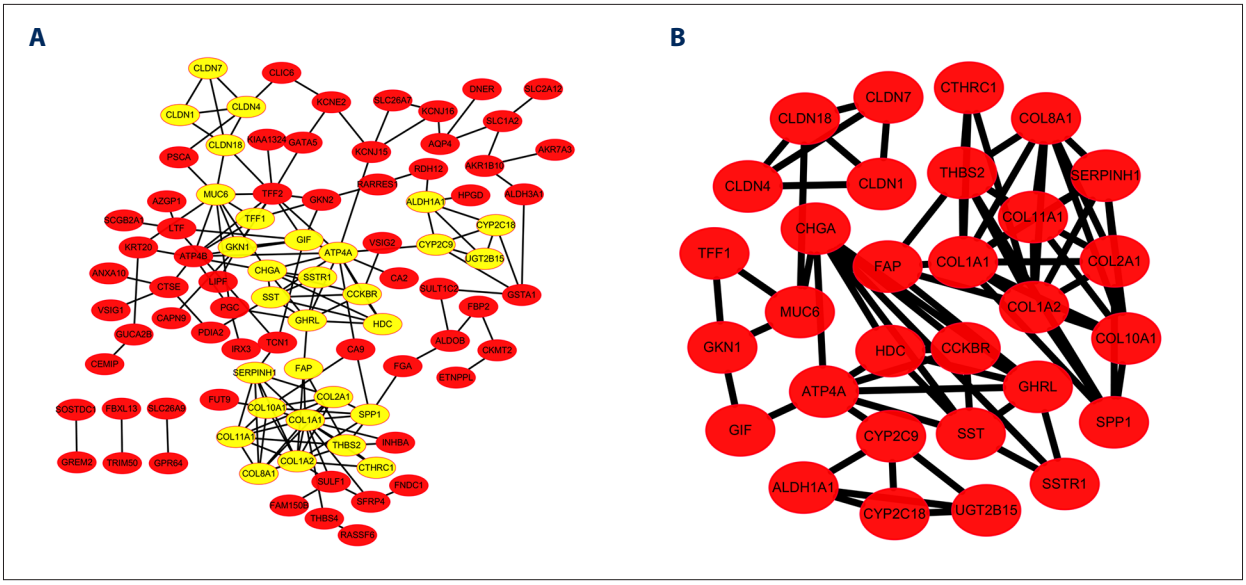

Figure 2. Differentially expressed genes (DEGs) protein–protein interaction (PPI) network analysis. (**A**) DEGs in PPI network complex by STRING and Cytoscape, which demonstrated 90 nodes and 162 edges and excluded 29 DEGs. (**B**) Module identified by Cytoscape MCODE plug-in. The nodes represent proteins; the edges represent protein interactions.

upregulated genes identified in the present study were also overexpressed ($P<0.05$), while the other 10 hub downregulated genes were also downregulated in the GEO dataset.

Pathway Enrichment Reanalysis and Stage Analysis of Hub DEGs

Twenty-one hub DEGs were reanalyzed by DAVID software to identify strongly associated pathways. The results showed that *COL1A1*, *COL1A2*, *COL2A1*, *COL11A1*, *THBS2*, and *SPP1* were mainly enriched in the ECM-receptor interaction or focal adhesion pathways (**Table 6**). We used the online GEPIA software to validate the expression of these 6 hub DEGs in different GC stages. Statistical analysis showed significant differential expression of *COL1A1*, *COL1A2*, *COL11A1*, and *THBS2*, while *COL2A1* and *SPP1* expression was not significantly different across different GC stages (**Figure 5**).

Discussion

Although surgery, chemotherapy, radiation therapy, immunotherapy, and other treatment methods have improved in GC, the current state of treatment and prognosis for GC patients remains unsatisfactory due to the difficulty of early diagnosis. Many patients are not diagnosed until GC is at an advanced or inoperable stage. To identify more effective biomarkers of prognosis in GC, we analyzed the gene expression profiles (GSE118916, GSE79973, and GSE19826) of 37 paired GC and adjacent normal tissues from the GEO public database. We identified 119 DEGs (21 upregulated and 98 downregulated genes) by GEO2R and Venn diagram online software. To better understand the interactions among DEGs, we further analyzed gene functional enrichment using DAVID software. Then, we constructed the PPI network via the STRING online database and Cytoscape software. Thirty hub DEGs were screened by Cytoscape MCODE. Subsequently, we conducted survival analysis using the Kaplan-Meier plotter to study the relationship between 30 hub DEGs and GC prognosis. The results showed 25 of the 30 hub DEGs were significantly associated with GC prognosis ($P<0.05$). To validate the expression of these 25 hub DEGs, we conducted a GEPIA analysis, which showed that 21 of these DEGs had significantly different expression between GC tissues and normal tissues and expression trends were consistent among datasets from the GEO database. Reanalysis of the 21 hub DEGs by KEGG pathway enrichment revealed that *COL1A1*, *COL1A2*, *COL2A1*, *COL11A1*, *THBS2*, and *SPP1* were mainly enriched in ECM-receptor interaction or the focal adhesion pathway and may be new effective biomarkers for the prognosis of GC patients.

Type I collagen is abundant in bone, cornea, dermis, and tendon. It consists of a heterotrimer of 2 chains of collagen type I alpha 1 (COL1A1) and 1 chain of collagen type I alpha 2 (COL1A2). Studies have demonstrated that abnormal expression of COL1A1 and COL1A2 is associated with tumor invasion and progression [15,16]. COL1A2 was found to be upregulated in colorectal cancer [17] and breast cancer [18]. COL1A2 was also found to be downregulated in bladder cancer, and it was suggested that inactivation of *COL1A2* through CpG hypermethylation may contribute to proliferation and migration activity of bladder cancer [16]. Moreover, *COL1A2* was identified as an

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 4. Survival analysis of the 30 hub differentially expressed genes (DEGs). Twenty-five hub DEGs were significantly correlated with the survival of patients with gastric cancer (*P*<0.05) while 5 hub DEGs were not significant (*P*>0.05).

Category	Genes
P<0.05	CTHRC1, COL10A1, COL2A1, COL8A1, MUC6, CLDN7, CLDN4, CYP2C18, COL11A1, ALDH1A1, GHRL, THBS2, CYP2C9, UGT2B15, SERPINH1, ATP4A, CHGA, SSTR1, SST, COL1A1, COL1A2, CCKBR, SPP1, CLDN18, TFF1
P>0.05	GKN1, FAP, CLDN1, HDC, GIF

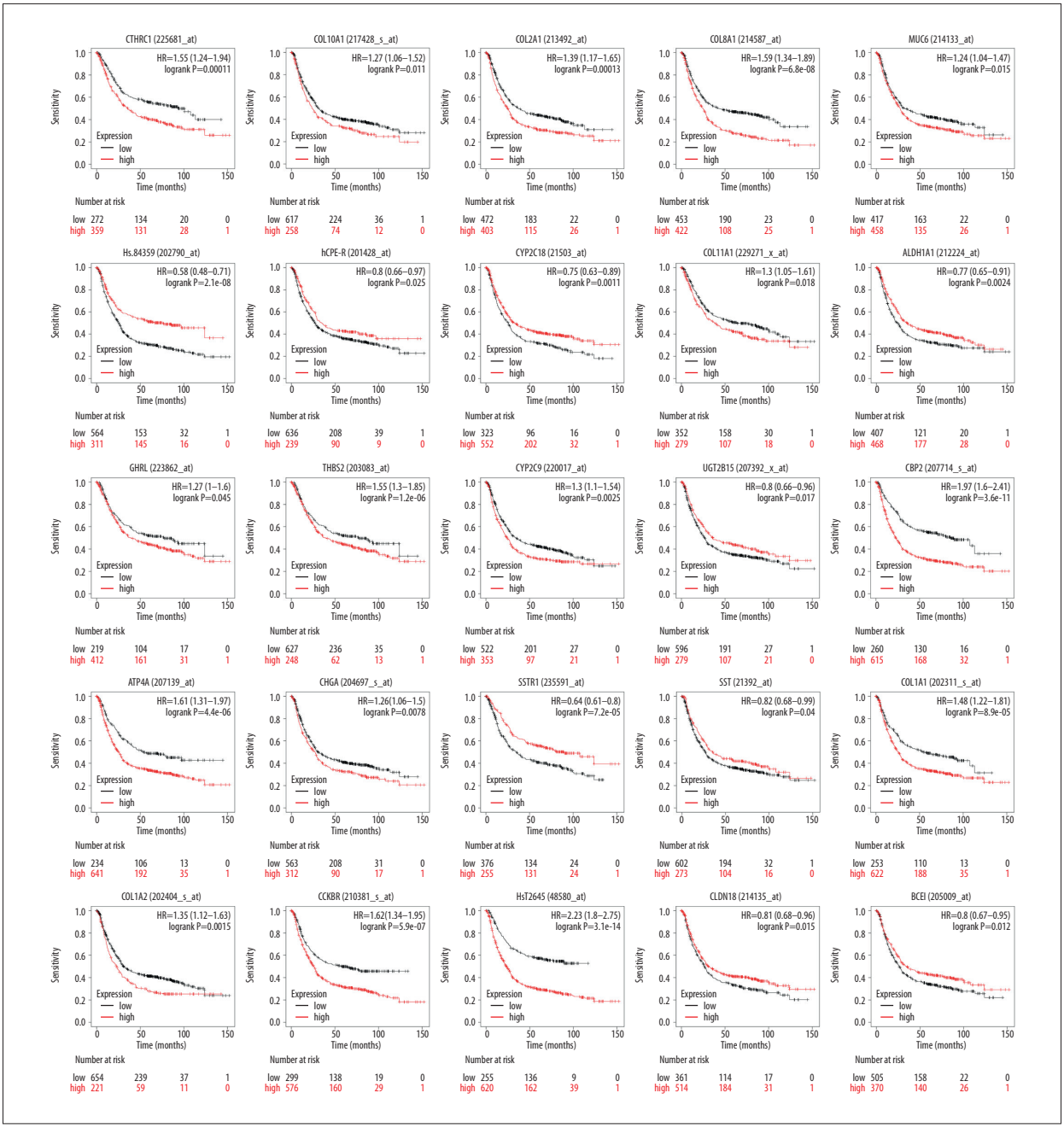

Figure 3. Overall survival analysis of 30 hub differentially expressed genes (DEGs). Twenty-five of 30 hub DEGs were significantly correlated with the survival of gastric cancer (GC) patients (*P*<0.05). Hs.84359 meant *CLDN7*; hCPE-R meant *CLDN4*; CBP2 meant *SERPINH1*; HsT2645 meant *SPP1*; and BCEI meant *TFF1*.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 5. Expression validation of 21 hub differentially expressed genes (DEGs; 11 upregulated and 10 downregulated genes).

Category	Genes
Up-regulated	COL11A1, CTHRC1, SERPINH1, THBS2, CLDN4, COL1A2, SPP1, COL1A1, COL8A1, CLDN7, COL10A1,
Down-regulated	MUC6, ATP4A, UGT2B15, COL2A1, GHRL, ALDH1A1, CHGA, CYP2C9, SST, CCKBR

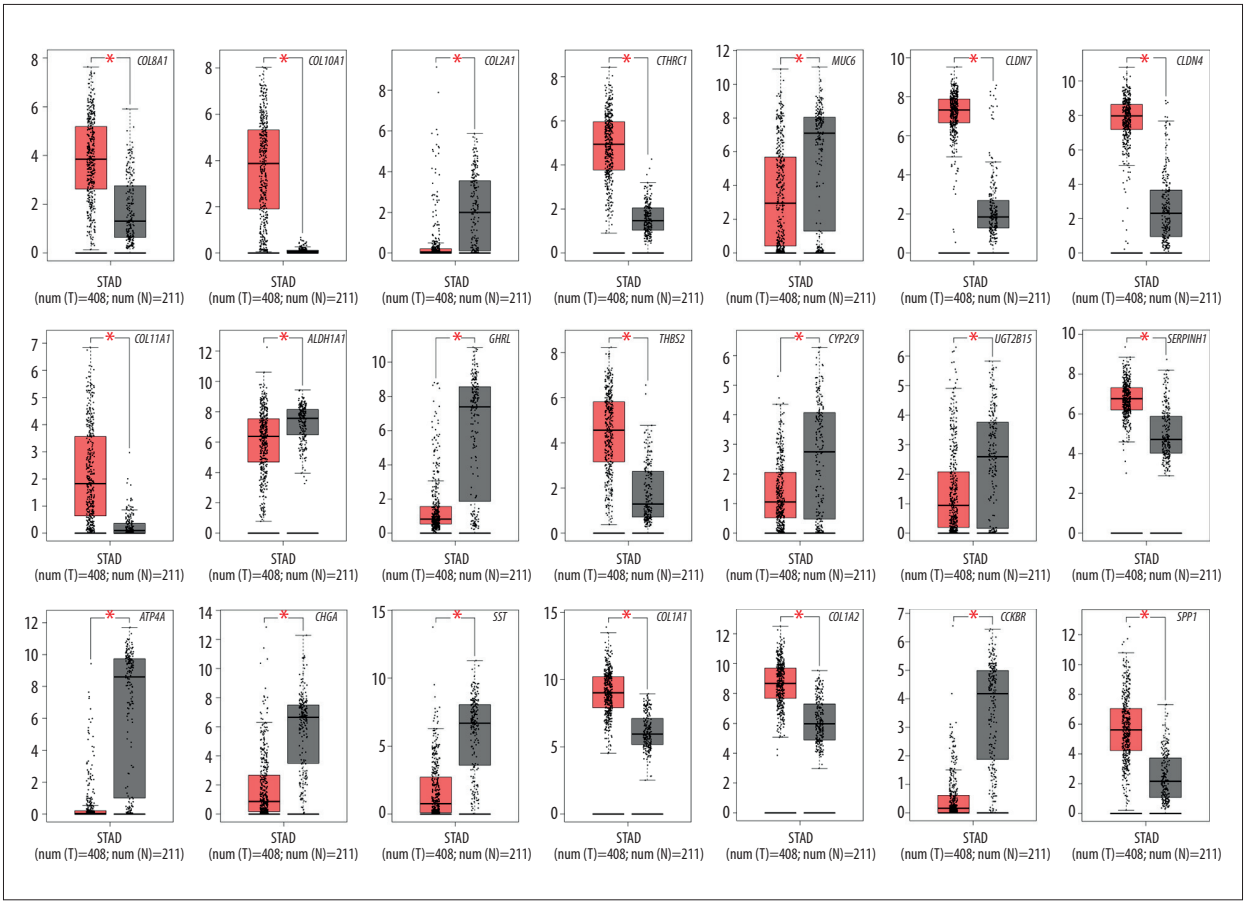

Figure 4. Validation of the expression of 25 hub differentially expressed genes (DEGs) by GEPIA website in gastric cancer (GC) tissues and normal tissues. The red box indicates tumor samples, and the gray box indicates normal samples.

Table 6. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway reanalysis of 21 hub differentially expressed genes (DEGs).

Pathway ID	Name	Count	%	*P*-value	Genes
hsa04512	ECM-receptor interaction	6	2.84	7.89E-07	COL1A2, COL2A1, COL1A1, THBS2, COL11A1, SPP1
hsa04510	Focal adhesion	6	2.84	5.80E-05	COL1A2, COL2A1, COL1A1, THBS2, COL11A1, SPP1
hsa00830	Retinol metabolism	3	1.42	0.006825	ALDH1A1, CYP2C9, UGT2B15

upregulated DEG in GC, which was consistent with our results, but *COL1A2* expression was not observed to have prognostic value in GC [19,20]. In contrast, *COL1A2* expression was associated with GC prognosis in our study. Li et al [21] reported that COL1A1 was notably elevated in patients' ascites of epithelial ovarian cancer compared with normal peritoneal fluids, and it promoted migration and invasion by ovarian cancer cells. COL1A1 had reduced expression in endometrial carcinoma

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

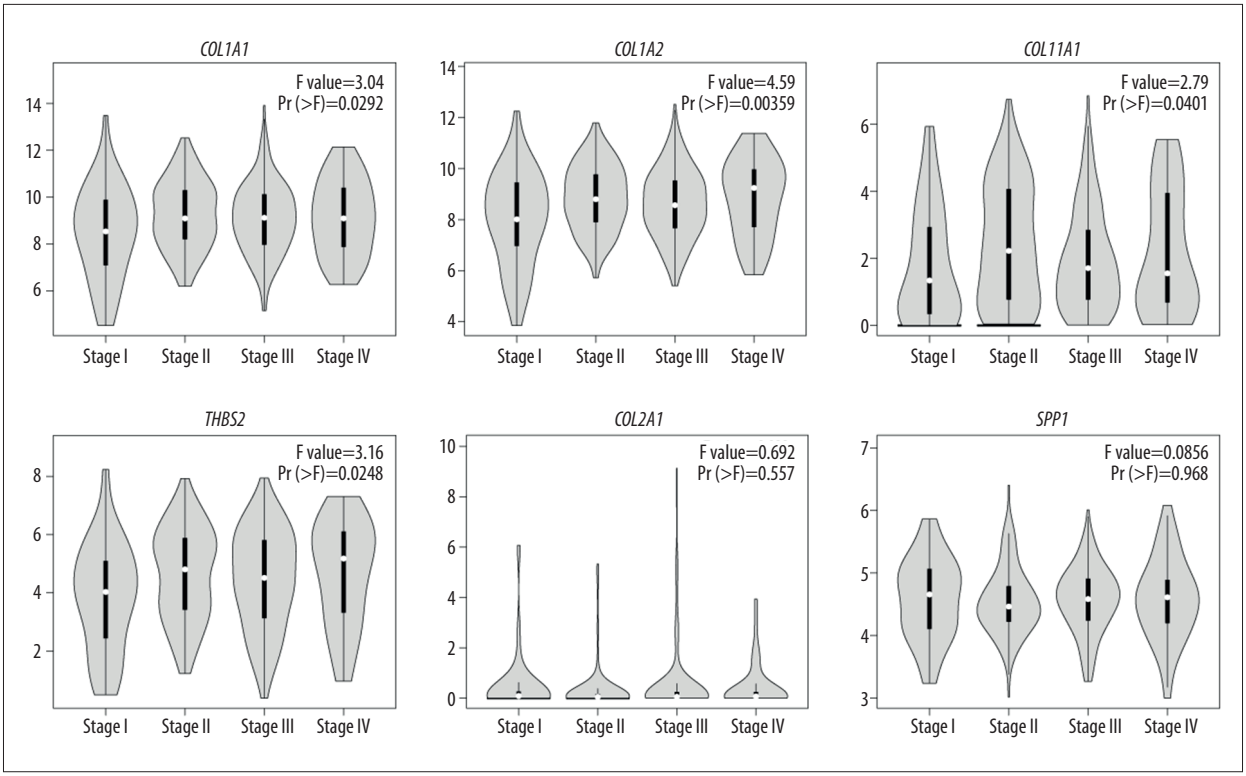

Figure 5. Pathological stage plot of gastric cancer (GC) hub differentially expressed genes (DEGs). *COL1A1*, *COL1A2*, *COL11A1*, and *THBS2* showed significant differences, while *COL2A1* and *SPP1* were not significantly different at various stages.

and hepatocellular carcinoma tumor tissues compared with normal tissue and affected the prognosis of patients [22,23].

COL2A1 is a fibrillar collagen found in cartilage and the vitreous humor of the eye [24]. Many studies have reported that *COL2A1* gene polymorphisms are associated with genetic diseases, such as COL2A1 rhegmatogenous retinal detachment, the most common type of retinal detachment [25,26]. However, few studies have reported differential expression of COL2A1 in association with human cancer. A 7-gene prognostic signature (*FBN1*, *MMP1*, *PLAU*, *SPARC*, *COL1A2*, *COL2A1*, and *ATP4A*) was proposed using integrated bioinformatics methods, and it may provide potential biomarkers for both prognosis and new therapeutic targets in GC [20].

Collagen type XI alpha 1 chain (COL11A1) was found to be upregulated in various cancers, including colorectal [27], pancreatic [28], ovarian cancer [29], and gastric cancer. COL11A1 could be used as a diagnostic indicator between premalignant and malignant lesions in the stomach [30,31]. Recently, studies have demonstrated that COL11A1 suppression in HGC-27 cells significantly inhibited proliferation, migration, and invasion in vitro. COL11A1 might be an oncogene in GC, which may regulate multiple genes responsible for cell growth and/or invasion and may be a potential therapeutic target in future investigation [32].

Thrombospondin 2 (THBS2) belongs to the thrombospondin family and mediates cell-cell and cell-matrix interactions. Studies suggest that the THBS2 protein may be involved in cell adhesion and migration and possibly function as a potential inhibitor of tumor growth and angiogenesis. The expression and prognosis of THBS2 have been investigated in breast cancer [33], ovarian cancer [34], and lung cancer. THBS2 was markedly overexpressed in a number of datasets of non-small-cell lung carcinoma, including lung adenocarcinoma. THBS2 may play a double role in the lung adenocarcinoma progression, including antiangiogenic and oncogenic functions. Overexpression of THBS2 was associated with poorer survival in lung adenocarcinoma patients [35]. THBS2 expression was remarkably related with the TNM stage, AJCC stages, and clinical outcomes ($P<0.05$), and may be a strong prognostic indicator in colorectal cancer [36]. THBS2 was a highly specific, independent diagnostic marker in patients with pancreatic cancer [37].

The protein encoded by secreted phosphoprotein 1 (*SPP1*, osteopontin), a matricellular protein, is involved in several pathophysiological processes including acute brain injury, cancer progression, and metastasis. After acute intracerebral hemorrhage, plasma SPP1 concentrations were significantly higher in patients than in controls and were strongly associated with National Institutes of Health Stroke Scale scores and hematoma

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

volumes. Plasma SPP1 could be a useful prognostic biomarker in intracerebral hemorrhage [38]. Previous studies have shown that SPP1 is upregulated in several human cancers and plays a potential role in tumor formation. Higher SPP1 cytoplasmic expression was associated with a significantly lower recurrence rate in colorectal cancer patients, and it was highly correlated with tumor grade, tumor invasion, and distant metastasis [39]. SPP1 overexpression was demonstrated to be significantly correlated with progression-free survival and poor overall survival in Chinese and Japanese patients with esophageal squamous cell carcinoma [40]. Elevated SPP1 expression may be a prognostic risk factor in different cancers.

Our studies demonstrated that 6 genes (*COL1A1*, *COL1A2*, *COL2A1*, *COL11A1*, *THBS2*, and *SPP1*) were particularly enriched in ECM-receptor interaction or focal adhesion pathways. Qiu et al [41] also identified *COL1A1*, *COL1A2*, *THBS2*, and *SPP1* as being significantly overexpressed in GC tissues, and *COL1A2* and *THBS2* were associated with significantly reduced survival time in GC patients. The ECM mixture serves an important role in tissue and organ morphogenesis and the maintenance of cell and tissue structure and function. The receptor interaction or cell-matrix adhesions lead to direct or indirect control of cellular activities such as adhesion, migration, differentiation, proliferation, and apoptosis. Four of the 6 genes found to be enriched in the current study are collagen genes. Collagen is the major constituent of the tumor ECM component, and degradation of collagen could play a key role in the invasion of the surrounding tissues by tumor cells [42]. Recently, several collagen genes were found to be closely related to tumor invasion and metastasis. *COL1A1* and *COL1A2* were upregulated in colorectal cancer and also associated with invasion and progression of other tumors, such as hypopharyngeal squamous cell carcinoma and bladder cancer [15,16]. Gao et al [31] reported that 7 collagen genes had elevated expression in GC, including *COL1A2* and *COL11A1* identified in our study. Previous studies have reported that the above 6 genes were correlated with progression and prognosis of different types of cancer, but few studies have shown that these 6 hub genes and ECM-receptor interaction or focal adhesion pathways play a vital role in GC and the prognosis of GC patients.

The current study has several strengths. First, different bioinformatics analysis tools were used for cross-validation and reanalysis, and they verified the study results. Second, several hub DEGs, especially collagen genes, have been reported to be significantly associated with the prognosis of GC patients in other studies, and this previous research supports our results. However, the present study also has some limitations. The online Kaplan-Meier plotter database was only used to assess the prognostic value. We did not have more information for prognosis from associated factors, such as age, sex, and treatment, and we may have missed some valuable information. In addition, our findings need to be confirmed through molecular biology studies.

Conclusions

Our studies identified 6 hub DEGs (*COL1A1*, *COL1A2*, *COL2A1*, *COL11A1*, *THBS2*, and *SPP1*) in GC tissues and adjacent normal tissues. These 6 genes, particularly collagen genes, are mainly enriched in the ECM-receptor interaction or focal adhesion pathway and affected the prognosis of GC patients. These genes and pathways could serve as potential prognostic markers and be utilized in the development of treatment for GC patients.

Acknowledgments

We thank all those who participated in this study.

Conflict of Interest

None.

References:

1. Sitarz R, Skierucha M, Mielko J, et al. Gastric cancer: Epidemiology, prevention, classification, and treatment. Cancer Manag Res, 2018;10:239-48
2. Van Cutsem E, Sagaert X, Topal B, et al. Gastric cancer. Lancet, 2016;388(10060):2654-64
3. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. Cancer J Clin, 2016;66(2):115-32
4. 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(1):7827
5. Jin Y, He J, Du J, et al. Overexpression of HS6ST2 is associated with poor prognosis in patients with gastric cancer. Oncol Lett, 2017;14(5):6191-97
6. He J, Jin Y, Chen Y, et al. Downregulation of ALDOB is associated with poor prognosis of patients with gastric cancer. Onco Targets Ther, 2016;9:6099-109
7. Wang Q, Wen Y-G, Li D-P, et al. Upregulated INHBA expression is associated with poor survival in gastric cancer. Med Oncol, 2012;29(1):77-83
8. Washington K. 7th Edition of the AJCC cancer staging manual: Stomach. Ann Surg Oncol, 2010;17(12):3077-79
9. Huang Da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc, 2009;4(1):44-57
10. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res, 2000;28(1):27-30
11. 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:D447-52
12. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res, 2003;13(11):2498-504
13. Szász AM, Lánczky A, Nagy Á, et al. Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. Oncotarget, 2016;7(31):49322-33

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

14. Tang Z, Li C, Kang B, et al. GEPIA. A web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res, 2017;45(W1):W98-102
15. Lin P, Tian P, Pang J, et al. Clinical significance of COL1A1 and COL1A2 expression levels in hypopharyngeal squamous cell carcinoma. Oncol Lett, 2020;20(1):803-9
16. Mori K, Enokida H, Kagara I, et al. CpG hypermethylation of collagen type I alpha 2 contributes to proliferation and migration activity of human bladder cancer. Int J Clin Oncol, 2009;34(6):1593-602
17. Zou X, Feng B, Dong T, et al. Up-regulation of type I collagen during tumorigenesis of colorectal cancer revealed by quantitative proteomic analysis. J Proteomics, 2013;94:473-85
18. Lin J, Goldstein L, Nesbit A, et al. Influence of hormone receptor status on spinal metastatic lesions in patients with breast cancer. World Neurosurg, 2016;85:42-48
19. Zhuo C, Li X, Zhuang H, et al. Elevated THBS2, COL1A2, and SPP1 expression levels as predictors of gastric cancer prognosis. Cell Physiol Biochem, 2016;40(6):1316-24
20. Wang J, Gao P, Song Y, et al. Prognostic value of gastric cancer-associated gene signatures: Evidence based on a meta-analysis using integrated bioinformatics methods. J Cell Mol Med, 2018;22(11):5743-47
21. Li M, Wang J, Wang C, et al. Microenvironment remodeled by tumor and stromal cells elevates fibroblast-derived COL1A1 and facilitates ovarian cancer metastasis. Exp Cell Res, 2020;394(1):112153
22. Zhu Y, Shi L, Chen P, et al. Identification of six candidate genes for endometrial carcinoma by bioinformatics analysis. World J Surg Oncol, 2020;18(1):161
23. Hayashi M, Nomoto S, Hishida M, et al. Identification of the collagen type 1 α 1 gene (COL1A1) as a candidate survival-related factor associated with hepatocellular carcinoma. BMC Cancer, 2014;14:108
24. Cheah KS, Stoker NG, Griffin JR, et al. Identification and characterization of the human type II collagen gene (COL2A1). Proc Natl Acad Sci USA, 1985;82(9):2555-59
25. Okazaki S, Meguro A, Ideta R, et al. Common variants in the *COL2A1* gene are associated with lattice degeneration of the retina in a Japanese population. Mol Vis, 2019;25:843-50
26. Go SL, Maugeri A, Mulder JJ, et al. Autosomal dominant rhegmatogenous retinal detachment associated with an Arg453Ter mutation in the COL2A1 gene. Invest Ophthalmol Vis Sci, 2003;44(9):4035-43
27. Fischer H, Stenling R, Rubio C, Lindblom A. Colorectal carcinogenesis is associated with stromal expression of COL11A1 and COL5A2. Carcinogenesis, 2001;22(6):875-78
28. Badea L, Herlea V, Dima SO, et al. Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. Hepatogastroenterology, 2008;55(88):2016-27
29. Wu YH, Chang TH, Huang YF, et al. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. Oncogene, 2014;33(26):3432-40
30. Zhao Y, Zhou T, Li A, et al. A potential role of collagens expression in distinguishing between premalignant and malignant lesions in stomach. Anat Rec (Hoboken), 2009;292(5):692-700
31. Gao X, Zhong S, Tong Y, et al. Alteration and prognostic values of collagen gene expression in patients with gastric cancer under different treatments. Pathol Res Pract, 2020;216(3):152831
32. Li A, Li J, Lin J, et al. COL11A1 is overexpressed in gastric cancer tissues and regulates proliferation, migration and invasion of HGC-27 gastric cancer cells in vitro. Oncol Rep, 2017;37(1):333-40
33. Perez-Rivas LG, Jerez JM, Fernandez-De Sousa CE, et al. Serum protein levels following surgery in breast cancer patients: A protein microarray approach. Int J Clin Oncol, 2012;41(6):2200-6
34. Wang M, Wang J, Liu J, et al. Systematic prediction of key genes for ovarian cancer by co-expression network analysis. J Cell Mol Med, 2020;24(11):6298-307
35. Weng TY, Wang CY, Hung YH, et al. Differential expression pattern of THBS1 and THBS2 in lung cancer: Clinical outcome and a systematic-analysis of microarray databases. PLoS One, 2016;11(8):e0161007
36. Tian Q, Liu Y, Zhang Y, et al. THBS2 is a biomarker for AJCC stages and a strong prognostic indicator in colorectal cancer. J BUON, 2018;23(5):1331-36
37. Peng HY, Chang MC, Hu CM, et al. Thrombospondin-2 is a highly specific diagnostic marker and is associated with prognosis in pancreatic cancer. Ann Surg Oncol, 2019;26(3):807-14
38. Li HJ, Han NN, Nan Y, et al. Plasma osteopontin acts as a prognostic marker in acute intracerebral hemorrhage patients. Clin Chim Acta, 2020;500:208-12
39. Assidi M, Gomaa W. Prognostic value of osteopontin (SPP1) in colorectal carcinoma requires a personalized molecular approach. Tumour Biol, 2019;41(9):1010428319863627
40. Wang Y, Lu Y, Xu W, et al. Prognostic value of osteopontin expression in esophageal squamous cell carcinoma: A meta-analysis. Pathol Res Pract, 2019;215(10):152571
41. Qiu J, Sun M, Wang Y, Chen B. Identification of hub genes and pathways in gastric adenocarcinoma based on bioinformatics analysis. Med Sci Monit, 2020;26:e920261
42. Yin Y, Zhao Y, Li AQ, Si JM. Collagen: A possible prediction mark for gastric cancer. Med Hypotheses, 2009;72(2):163-65

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]