MEDICAL SCIENCE MONITOR

LAB/IN VITRO RESEARCH

e-ISSN 1643-3750
© Med Sci Monit, 2020; 26: e920261
DOI: 10.12659/MSM.920261

Received: 2019.09.23
Accepted: 2019.11.27
Available online: 2020.01.21
Published: 2020.02.14

Identification of Hub Genes and Pathways in Gastric Adenocarcinoma Based on 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 Jieping Qiu
BD 1 Mengyu Sun
BC 1 Yaoqun Wang
ADG 2 Bo Chen

1 Department of Clinical Medicine, The First Clinical College, Anhui Medical University, Hefei, Anhui, P.R. China
2 Department of Gastrointestinal Surgery Center, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, P.R. China

Corresponding Author: Bo Chen, e-mail: chenbo831116@163.com
Source of support: National Natural Science Foundation (NO. 81602425); Anhui Provincial Teaching Research Project (NO. 2016jyxm0529); National Innovation and Entrepreneurship Project for College Students (NO. 201710366009)

Background: Gastric adenocarcinoma accounts for 95% of all gastric malignant tumors. The purpose of this research was to identify differentially expressed genes (DEGs) of gastric adenocarcinoma by use of bioinformatics methods.

Material/Methods: The gene microarray datasets of GSE103236, GSE79973, and GSE29998 were imported from the GEO database, containing 70 gastric adenocarcinoma samples and 68 matched normal samples. Gene ontology (GO) and KEGG analysis were applied to screened DEGs; Cytoscape software was used for constructing protein-protein interaction (PPI) networks and to perform module analysis of the DEGs. UALCAN was used for prognostic analysis.

Results: We identified 2909 upregulated DEGs (uDEGs) and 7106 downregulated DEGs (dDEGs) of gastric adenocarcinoma. The GO analysis showed uDEGs were enriched in skeletal system development, cell adhesion, and biological adhesion. KEGG pathway analysis showed uDEGs were enriched in ECM-receptor interaction, focal adhesion, and Cytokine-cytokine receptor interaction. The top 10 hub genes – COL1A1, COL3A1, COL1A2, BGN, COL5A2, THBS2, TIMP1, SPP1, PDGFRB, and COL4A1 – were distinguished from the PPI network. These 10 hub genes were shown to be significantly upregulated in gastric adenocarcinoma tissues in GEPIA. Prognostic analysis of the 10 hub genes via UALCAN showed that the upregulated expression of COL3A1, COL1A2, BGN, and THBS2 significantly reduced the survival time of gastric adenocarcinoma patients. Module analysis revealed that gastric adenocarcinoma was related to 2 pathways: including focal adhesion signaling and ECM-receptor interaction.

Conclusions: This research distinguished hub genes and relevant signal pathways, which contributes to our understanding of the molecular mechanisms, and could be used as diagnostic indicators and therapeutic biomarkers for gastric adenocarcinoma.

MeSH Keywords: Prognosis • Stomach Neoplasms • Tumor Markers, Biological

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/920261

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 a common malignant disease with a mortality rate of about 10% [1], which does a great harm to global health. Gastric adenocarcinoma (GAC) is the most common pathological type of gastric cancer, accounting for 95% of gastric malignant tumors [2], and it is characterized by easy invasion and metastasis [3]. Most GC patients are diagnosed in advanced stages, which is the major reason for its poor prognosis [4]. Although multimodal therapy, including surgery, chemotherapy, radiotherapy, and targeted therapy, has recently improved, the 5-year overall survival rate of patients with terminal GC is still less than 20% [5], and it can be as high as 90% if GC is detected in the early stage [6]. Accordingly, the early diagnosis and treatment of GAC is crucial.

Studies have shown that many biochemical molecular markers are involved in the occurrence and development of tumors and can be used for early screening of tumors. However, many markers are highly expressed in various types of tumors and do not have good specificity [7]. Therefore, it is necessary to further explore new and specific diagnostic markers of gastric adenocarcinoma as an auxiliary detection project for early diagnosis. Recently, bioinformatics has become a promising and effective tool for screening significant genetic or epigenetic variations that occur in carcinogenesis and determine the diagnosis and prognosis of cancer [8]. Various bioinformatics databases, such as the GEO database, provide opportunities for data mining for gene expression profiles of cancer.

In this study, we imported 3 gastric adenocarcinoma datasets from the GEO database. We screened differentially expressed genes (DEGs) by comparing the gene expression between gastric adenocarcinoma samples and paired normal mucosa samples. Then, function annotations and signal pathway analysis of DEGs were performed using Gene ontology (GO) and KEGG signal pathway enrichment analysis in the DAVID database. Subsequently, to study the mechanism of occurrence and development of GAC at the molecular level, we used UALCAN for prognosis analysis and GEPIA for verification of the mRNA expression level, which may provide valuable insights for diagnosis, targeted drug research, and prognosis evaluation of GAC.

Material and Methods

Datasets

The Gene Expression Omnibus database (GEO, *http://www.ncbi.nlm.nih.gov/geo*) is a communal functional genic database including array-based and sequence-based data, and is available to users free of charge. The gene expression datasets of GSE103236 [9], GSE79973 [10], and GSE29998 [11] were acquired from the GEO database. The 3 datasets selected in this experiment all met 3 criteria: (1) samples from human gastric tissue; (2) with case-control group; and (3) sample number ≥18, and only for the pathological type of GAC. GSE103236 was based on the GPL4133 platform (Agilent-014850 Whole Human Genome Microarray 4x44K G4112F). GSE79973 was based on the GPL570 platform ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array). GSE29998 was based on the GPL6947 platform (Illumina HumanHT-12 V3.0 expression BeadChip). GSE103236 contains 19 samples, including 10 gastric adenocarcinoma samples and 9 matched normal mucosa samples. GSE79973 contains 20 samples, including 10 gastric adenocarcinoma samples and 10 matched normal mucosa samples. GSE29998 contains 99 samples, including 50 gastric adenocarcinoma samples and 49 matched normal mucosa samples.

Data processing

GEO2R (*http://www.ncbi.nlm.nih.gov/geo/geo2r/*) is an online tool with which different groups of samples from the GEO database can be compared to identify DEGs [12]. The data were divided into a gastric adenocarcinoma group and a normal group for further analysis by GEO2R. The benchmark adj. $p<0.05$ and $|\log2FC|>1$ were determined as the cutoff values for statistical analysis of each dataset, and the intersecting parts of the 3 datasets were determined by use of the online tool Draw Venn diagram (bioinformatics. psb. ugent. be/webtools/Venn/).

Gene ontology (GO) and KEGG signal pathway analysis of DEGs

The GO (*http://www.geneontology.org*) database [13] can provide functional classification for genomic data, including biological processes (BP), cellular component (CC), and molecular function (MF). GO analysis is a widely used annotating tool of genes and genic productions. The Kyoto Encyclopedia of Genes and Genomes (KEGG, *http://www.genome.ad.jp/kegg/*) database [14] is a networked website designed for genic function analysis, exegesis, and visualizing. The Database for Annotation, Visualization and Integrated Discovery (DAVID, *http://david.abcc.ncifcrf.gov/*) [15] is an online tool for genic functional classification, which can be applied for gene analysis to assess the biological function of genes. In this research, GO enrichment analysis and KEGG pathway analysis were applied using the DAVID website to study the functions of DEGs. $p<0.05$ was set as the cutoff point for statistical significance.

Integration of protein-protein interaction (PPI) network and module analysis

The Search Tool for the Retrieval of Interacting Genes (STRING, *http://string.embl.de/*) [16] is a biological database designed

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

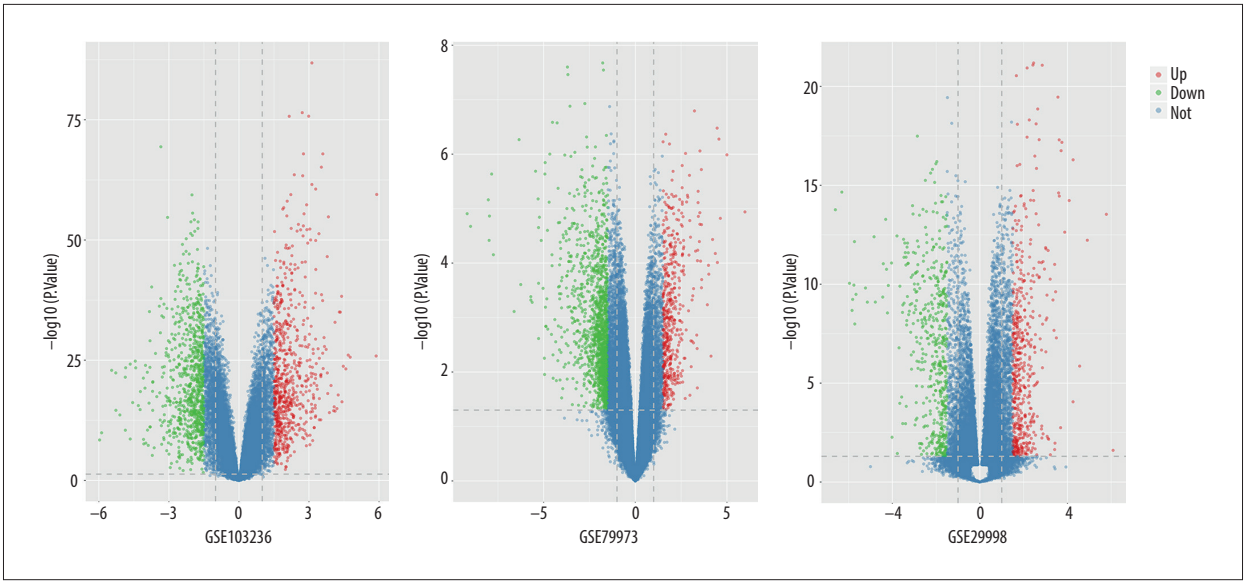


Figure 1. Volcano plot of all significant DEGs. Volcano plot of all significant DEGs, including a total of 2909 uDEGs and 7106 dDEGs. Red color means uDEGs, green color represents dDEGs, and blue color represents genes that are not significantly different in expression. The criterion: |foldchange|>1, p<0.05 is determined as the cutoff value.

for predicting PPI networks. The DEGs were imported to STRING to assess the interactive relationships, and a confidence score >0.9 was considered as significant. Then, we used Cytoscape [17], which biological graph visualization software that can construct comprehensive models of biologic molecular interaction. The Molecular Complex Detection (MCODE), a pluggable unit of Cytoscape, was applied for screening the modules of the PPI network. The benchmarks were determined as: degree cutoff=2, node score cutoff=0.2, k-core=4, and maximum depth=100. The KEGG signal pathway enrichment analysis was reapplied to DEGs located in the modules to study their major functions.

Expression levels and prognostic analysis of hub genes

GEPIA (Gene Expression Profiling Interactive Analysis) [18] is a well-known platform that can be used to analyze differences in the mRNA expression levels of a specific gene in specific cancers between cancerous tissues and paired normal tissues. We used GEPIA to study mRNA expression levels of hub genes in GAC and paired normal tissues. UALCAN (*http://ualcan.path.uab.edu*) [19] was used to assess the prognosis of hub genes. For each gene, cancer patients were automatically separated into high-expression and low-expression groups in accordance with the expression value of RNA, and the difference p<0.05 was regarded as significant.

Results

Screening of DEGs

In the aggregate, 70 gastric adenocarcinoma samples and 68 matched normal mucosa samples from 3 datasets were analyzed. In view of the GEO2R analysis, using the adj. p<0.05 and |log2FC|>1 criteria, 2909 upregulated DEGs (uDEGs) and 7106 downregulated DEGs (dDEGs) were screened in GAC tissues compared with normal tissues (Figure 1). A total of 250 genes were collected from all 3 datasets, including 92 uDEGs (Figure 2, Table 1) and 158 dDEGs (Figure 2, Table 1).

GO term enrichment analysis

GO analysis outcomes showed that for biological process (BP), uDEGs were markedly enriched in skeletal system development, cell adhesion, and biological adhesion (Figure 3, Table 2); the dDEGs are mainly in ion transport, homeostatic process, and chemical homeostatic (Figure 3, Table 2). For molecular function (MF), the uDEGs are enriched in structural molecule activity, extracellular matrix (ECM) structural constituent, and growth factor binding (Figure 3, Table 2); and the dDEGs are enriched in channel activity, passive transmembrane transporter activity, and substrate specific channel activity (Figure 3, Table 2). Cellular component (CC) analysis revealed that uDEGs are concentrated in extracellular region, extracellular region part, and proteinaceous extract (Figure 3, Table 2); and dDEGs are concentrated in extracellular region, plasma membrane part, and extracellular region part (Figure 3, Table 2).

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

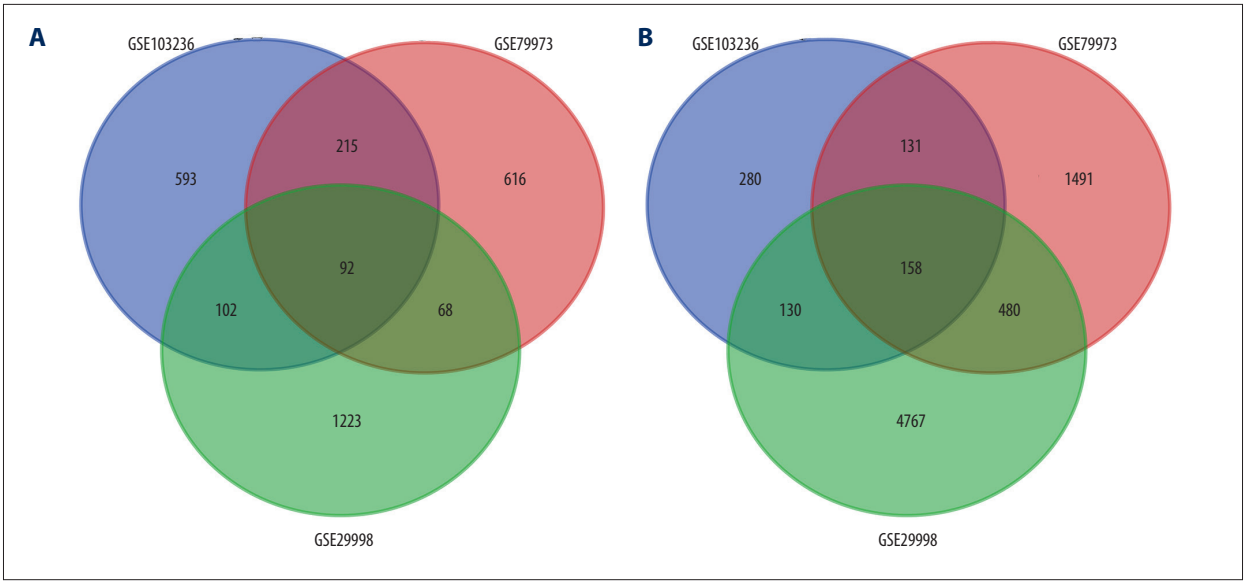


Figure 2. (**A, B**) Venn diagram of all screened DEGs. Venn diagram shows: uDEGs shared by GSE103236, GSE79973, and GSE29998 microarrays. A total of 92 uDEGs and 158 dDEGs were identified in the intersections.

Table 1. DEGs in gastric adenocarcinoma shared in 3 microarrays.

DEGS	Total	Elements
uDEGs	92	IGF2BP3, CDC25B, C5AR1, ZMYND15, COL1A1, GDPD5, ANLN, CHRNA5, FNDC1, COL18A1, PRRX1, PDGFRB, COL5A2, KIF4A, THY1, ASCL2, ANTXR1, SPP1, WNT5A, OLR1, MSR1, MELK, CDH11, TIMP1, BGN, COL8A1, TEAD4, ECT2, MMP11, KRT80, DDX31, FSCN1, SRPX2, WNT2, LRP8, CEMIP, BMP1, DIO2, ARPC1B, MFAP2, WISP1, VMO1, COL4A1, SLC1A3, SULF1, CLDN1, COL11A1, TREM1, COL1A2, APOC1, COL12A1, ESM1, ARHGAP11A, PLAU, RFC3, TGM2, OSMR, FOXC1, CHEK1, TNFRSF11B, VSNL1, IGFBP7, CST1, RCC2, LEF1, IL13RA2, LZTS1, SPHK1, KIF2C, AGPAT4, BUB1, TNFRSF12A, TROAP, ANGPT2, COL3A1, TMEM158, SERPINH1, FAP, INHBA, CDCA3, SLC5A6, CKAP2, THBS2, OLFML2B, S100A10, COL6A3, CSF2RA, HSD11B1, PMEPA1, CTHRC1, GAD1, NCAPG
dDEGs	158	EPB41L4B, COL4A5, ZNF385B, NRG4, IRX3, KLF4, NDRG2, IGFBP2, SLC9A2, SCNN1B, ESRRB, HTR4, F13A1, CNTD1, ADHFE1, CELA3B, OSBPL7, ADGRG2, MYOC, GIF, GPER1, HBB, GNG7, COL4A6, ERO1B, SOSTDC1, SBSPON, TMED6, ARHGAP24, CKMT2, KCNJ16, SLC26A7, ADAMTSL1, SYT4, PTGER3, ATP4A, DNASE1L3, CKB, MAL, AQP4, ESRRG, STOX2, CPA2, OXCT1, ABCA8, TTLL7, AXDND1, FXYD4, PER3, SHMT1, ETNPPL, DPT, PACRG, CAPN13, SLC5A5, FGA, NTN1, LGI1, ATP4B, EPM2A, SLC2A4, ADH1A, KLF15, SCUBE2, GPX3, KCNMB2, MAMDC2, SELENBP1, GPAT3, CLIC6, HAPLN1, TRIM50, B3GAT1, MS4A2, BHMT, GHRL, PNOC, GPRC5C, KCNJ13, DGKD, BMP6, GC, ALDH6A1, PDGFD, HTR1E, KIT, ADH1C, TOX, GCNT2, SST, PLCXD3, XYLT2, SLC6A16, CWH43, PDK4, GFRA2, GPR155, GREM2, NR3C2, PLA2G1B, MUC6, LINGO2, RNASE1, MYZAP, GUCA1C, AKR1C1, ACACB, PNPLA7, FXYD1, BAALC, PPP2R3A, BMP5, CKM, FBXL13, COBLL1, RGMB, FBP2, FAM150B, HIF3A, GSTA4, FGG, CHGA, RAB26, CAPN9, MT1M, SGSM1, ASPA, SCGN, SULT2A1, GAMT, CDHR3, CCKBR, ADRB2, GRIA3, SCGB2A1, SLC7A8, DUOX1, SORCS1, ARHGEF37, SCARA5, PACSIN1, LIFR, FNDC5, GLUL, CHIA, METTL7A, FAM189A2, SSTR1, PAIP2B, ACER2, ADH1B, MYRIP, KCNE2, PPP1R3C, TCEA3, PDE8B, SIGLEC11, KBTBD12

KEGG signal pathway analysis

The most remarkably enriched pathways of uDEGs and dDEGs identified by KEGG analysis are shown in Table 3. The uDEGs are enriched in focal adhesion, ECM-receptor interaction, and cytokine-cytokine receptor interaction, while the dDEGs are enriched in pathways in arginine and proline metabolism, as well as glycine, serine, and threonine metabolism.

PPI network construction, module analysis and hub genes determination

The interaction between DEGs was calculated using the STRING database, and 250 DEGs differently expressed in all 3 data sets were imported into Cytoscape software for visualization. PPI network involves 143 nodes and 578 edges (Figure 4). The top 10 genes in connectivity ranking in the PPI network

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

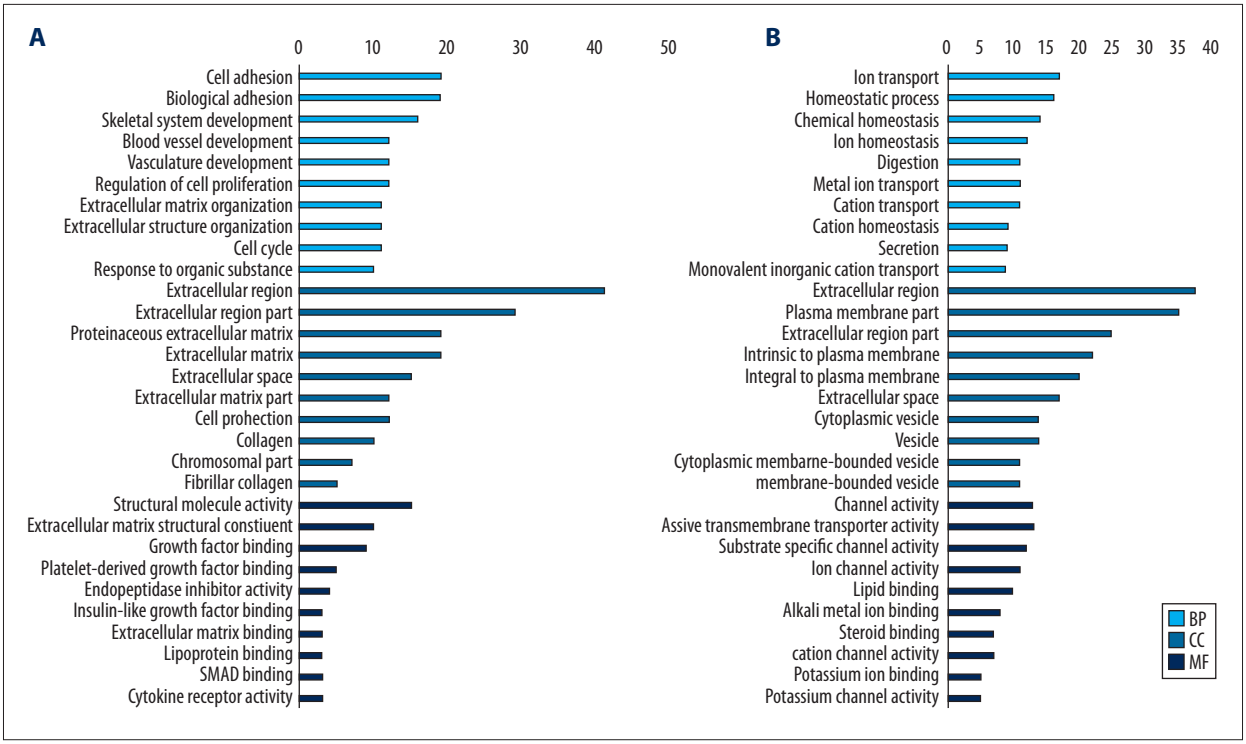


Figure 3. Gene ontology analysis of DEGs related to gastric adenocarcinoma. The x-axis stands for the number of DEGs, and the vertical axis stands for GO terms, (**A**) The top 10 enriched biological processes (BP), cellular component (CC), and molecular function (MF) of 92 uDEGs. (**B**) The top 10 enriched BP, CC, and MF for 158 dDEGs.

were selected as hub genes. The results showed that COL1A1 ranked highest among all DEGs, with 34 degree, followed by COL3A1, COL1A2, BGN, COL5A2, THBS2, TIMP1, SPP1, PDGFRB, and COL4A1 (Table 4).

The module analysis of 143 nodes showed that the most important module with higher score involves 15 nodes and 143 edges (Figure 4). All 15 nodes are all upregulated genes, which suggests the vital role of uDEGs in GAC. KEGG signal pathway analysis of the 15 genes showed that they mainly participated in 2 pathways: ECM-receptor interaction and focal adhesion. It is noteworthy that 8 of the 15 genes in the module (COL4A1, COL6A3, COL3A1, COL1A2, COL1A1, COL11A1, COL4A6, and THBS2) are involved in both pathways.

Expression levels and prognostic analysis of hub genes

GEPIA database showed that all 10 hub genes are upregulated in GAC (Figure 5). To assess the prognostic value of 10 hub genes, we used UALCAN for prognostic analysis. The results of the prognostic analysis showed the upregulated expression of COL3A1, COL1A2, BGN, and THBS2 significantly reduce the survival time of GAC patients (Figure 6).

Discussion

Gastric cancer is a leading cause of death. Early diagnosis and treatment are essential to prolong the survival time of GC patients. GAC is the most common type of gastric cancer. Therefore, it is crucial to further explore the predictive indicators and therapeutic targets of GAC. Recently, with the rapid development of bioinformatics, DNA microarray is increasingly applied to explore the early diagnosis, treatment, and prognosis of cancer [20]. Therefore, the present study explored the potential target genes and pathways of GAC by use of bioinformatics methods.

In this study, 2909 uDEGs and 7106 dDEGs were identified from the GSE103236, GSE79973, and GSE29998 datasets downloaded from the GEO database, among which, 92 uDEGs and 158 dDEGs were significantly expressed in all 3 datasets. To further define the role of these DEGs in gastric adenocarcinoma, we performed a series of bioinformatics and prognostic analysis of these DEGs.

GO analysis revealed uDEGs are highly involved in cell adhesion, biological adhesion, and skeletal system development, while the dDEGs are mainly in ion transport, homeostatic process, and chemical homeostasis. Studies [21] have shown that the decrease of cell adhesion is a key step in the metastasis of

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 analysis of DEGs related to gastric adenocarcinoma.

Expression	Category	Term	Count	%	P value	FDR
Upregulated	GOTERM_BP_FAT	GO: 0001501~skeletal system development	19	3.140495868	6.25E-17	1.67E-13
	GOTERM_BP_FAT	GO: 0007155~cell adhesion	20	3.305785124	4.39E-12	6.55E-09
	GOTERM_BP_FAT	GO: 0022610~biological adhesion	20	3.305785124	4.50E-12	6.72E-09
	GOTERM_BP_FAT	GO: 0030199~collagen fibril organization	8	1.32231405	1.19E-11	1.78E-08
	GOTERM_BP_FAT	GO: 0030198~extracellular matrix organization	8	1.32231405	1.30E-07	1.95E-04
	GOTERM_BP_FAT	GO: 0001503~ossification	8	1.32231405	2.60E-07	3.89E-04
	GOTERM_BP_FAT	GO: 0060348~bone development	8	1.32231405	4.12E-07	6.16E-04
	GOTERM_BP_FAT	GO: 0043062~extracellular structure organization	8	1.32231405	2.75E-06	0.0041046
	GOTERM_BP_FAT	GO: 0032963~collagen metabolic process	5	0.826446281	3.71E-06	0.005542186
	GOTERM_BP_FAT	GO: 0043588~skin development	5	0.826446281	4.29E-06	0.006410681
	GOTERM_CC_FAT	GO: 0005578~proteinaceous extracellular matrix	25	4.132231405	1.86E-24	2.10E-21
	GOTERM_CC_FAT	GO: 0031012~extracellular matrix	25	4.132231405	1.14E-23	1.28E-20
	GOTERM_CC_FAT	GO: 0005576~extracellular region	40	6.611570248	6.86E-20	7.75E-17
	GOTERM_CC_FAT	GO: 0044421~extracellular region part	30	4.958677686	8.27E-19	9.34E-16
	GOTERM_CC_FAT	GO: 0044420~extracellular matrix part	14	2.314049587	1.64E-15	1.88E-12
	GOTERM_CC_FAT	GO: 0005581~collagen	9	1.487603306	1.49E-12	1.68E-09
	GOTERM_CC_FAT	GO: 0005583~fibrillar collagen	5	0.826446281	1.48E-07	1.67E-04
	GOTERM_CC_FAT	GO: 0005604~basement membrane	6	0.991735537	2.04E-05	0.023013251
	GOTERM_CC_FAT	GO: 0031093~platelet alpha granule lumen	5	0.826446281	2.76E-05	0.031185256
	GOTERM_CC_FAT	GO: 0060205~cytoplasmic membrane-bounded vesicle lumen	5	0.826446281	3.67E-05	0.041403378
	GOTERM_MF_FAT	GO: 0005201~extracellular matrix structural constituent	11	1.818181818	5.06E-13	5.90E-10
	GOTERM_MF_FAT	GO: 0005198~structural molecule activity	14	2.314049587	3.15E-07	3.67E-04
	GOTERM_MF_FAT	GO: 0005518~collagen binding	5	0.826446281	8.88E-06	0.010355681
	GOTERM_MF_FAT	GO: 0005539~glycosaminoglycan binding	7	1.157024793	1.19E-05	0.013923194
	GOTERM_MF_FAT	GO: 0001871~pattern binding	7	1.157024793	2.05E-05	0.023961179
	GOTERM_MF_FAT	GO: 0030247~polysaccharide binding	7	1.157024793	2.05E-05	0.023961179
	GOTERM_MF_FAT	GO: 0008201~heparin binding	6	0.991735537	3.72E-05	0.043401787
	GOTERM_MF_FAT	GO: 0050840~extracellular matrix binding	4	0.661157025	1.30E-04	0.151895387
	GOTERM_MF_FAT	GO: 0005509~calcium ion binding	12	1.983471074	4.20E-04	0.488103889
	GOTERM_MF_FAT	GO: 0008237~metallopeptidase activity	6	0.991735537	5.52E-04	0.641829759

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 continued. Gene ontology analysis of DEGs related to gastric adenocarcinoma.

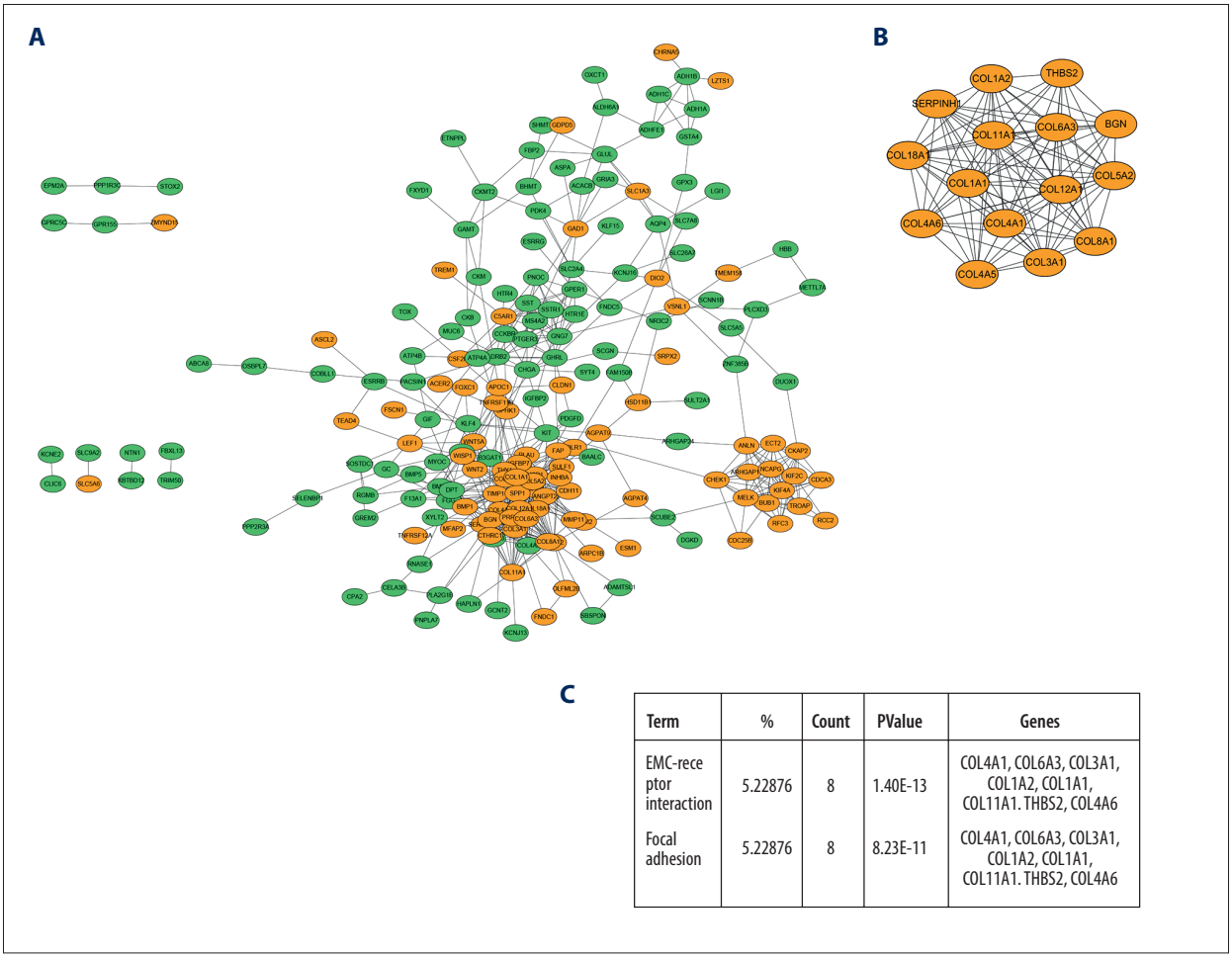
Expression	Category	Term	Count	%	P value	FDR
Downregulated	GOTERM_BP_FAT	GO: 0007586~digestion	12	1.038062284	7.45E-12	1.09E-08
	GOTERM_BP_FAT	GO: 0055114~oxidation reduction	15	1.297577855	3.07E-05	0.045029161
	GOTERM_BP_FAT	GO: 0006081~cellular aldehyde metabolic process	4	0.346020761	6.53E-04	0.952399247
	GOTERM_BP_FAT	GO: 0030001~metal ion transport	10	0.865051903	0.00213818	3.08666015
	GOTERM_BP_FAT	GO: 0006812~cation transport	10	0.865051903	0.006666616	9.333089916
	GOTERM_BP_FAT	GO: 0006811~ion transport	12	1.038062284	0.007065784	9.865306976
	GOTERM_BP_FAT	GO: 0046903~secretion	7	0.605536332	0.010249221	14.00683938
	GOTERM_BP_FAT	GO: 0015672~monovalent inorganic cation transport	7	0.605536332	0.01337189	17.89646714
	GOTERM_BP_FAT	GO: 0006813~potassium ion transport	5	0.432525952	0.016821807	22.00279704
	GOTERM_BP_FAT	GO: 0022600~digestive system process	3	0.259515571	0.018374485	23.78773935
	GOTERM_CC_FAT	GO: 0005576~extracellular region	31	2.6816609	1.26E-05	0.014253173
	GOTERM_CC_FAT	GO: 0045177~apical part of cell	7	0.605536332	0.001371031	1.545706714
	GOTERM_CC_FAT	GO: 0016324~apical plasma membrane	6	0.519031142	0.002113034	2.373127171
	GOTERM_CC_FAT	GO: 0005624~membrane fraction	11	0.951557093	0.047644803	42.55171216
	GOTERM_MF_FAT	GO: 0031420~alkali metal ion binding	7	0.605536332	0.004013875	5.198242704
	GOTERM_MF_FAT	GO: 0004033~aldo-keto reductase activity	3	0.259515571	0.004305787	5.566366984
	GOTERM_MF_FAT	GO: 0004198~calcium-dependent cysteine-type endopeptidase activity	3	0.259515571	0.004899806	6.31139254
	GOTERM_MF_FAT	GO: 0008289~lipid binding	9	0.778546713	0.009796998	12.2496158
	GOTERM_MF_FAT	GO: 0016620~oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	3	0.259515571	0.010025001	12.51741656
	GOTERM_MF_FAT	GO: 0030955~potassium ion binding	5	0.432525952	0.010277041	12.81257072
	GOTERM_MF_FAT	GO: 0008233~peptidase activity	10	0.865051903	0.013517076	16.52574449
	GOTERM_MF_FAT	GO: 0015267~channel activity	8	0.692041522	0.019316845	22.80964632
	GOTERM_MF_FAT	GO: 0022803~passive transmembrane transporter activity	8	0.692041522	0.01954695	23.04969236
	GOTERM_MF_FAT	GO: 0008900~hydrogen: potassium-exchanging ATPase activity	2	0.173010381	0.019742308	23.25294859

cancer, which agrees with our GO analysis results. For MF, the uDEGs are markedly enriched in ECM structural constituent, structural molecule activity, and growth factor binding, while the dDEGs were enriched in channel activity, passive transmembrane transporter activity, and substrate specific channel activity. GO CC analysis revealed that uDEGs were concentrated in extracellular region part, proteinaceous extract, and extracellular region, while dDEGs were concentrated in extracellular region, plasma membrane part, and extracellular region part. The role of ECM and collagen binding in development and

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. KEGG pathway analysis of DEGs associated with gastric adenocarcinoma.

Expression	Term	Count	%	P value	FDR
Upregulated	hsa04512: ECM-receptor interaction	11	1.818181818	2.66E-13	2.14E-10
	hsa04510: Focal adhesion	11	1.818181818	1.80E-09	1.44E-06
	hsa04350: TGF-beta signaling pathway	4	0.661157025	0.005149174	4.056689024
Downregulated	hsa00982: Drug metabolism	8	0.692041522	1.46E-07	1.41E-04
	hsa00830: Retinol metabolism	7	0.605536332	1.38E-06	0.001338048
	hsa00980: Metabolism of xenobiotics by cytochrome P450	7	0.605536332	2.60E-06	0.002518197
	hsa00010: Glycolysis/Gluconeogenesis	4	0.346020761	0.007818033	7.311759835
	hsa00591: Linoleic acid metabolism	3	0.259515571	0.015552893	14.070282
	hsa00350: Tyrosine metabolism	3	0.259515571	0.036348334	30.10544025



Term	%	Count	PValue	Genes
EMC-receptor interaction	5.22876	8	1.40E-13	COL4A1, COL6A3, COL3A1, COL1A2, COL1A1, COL11A1. THBS2, COL4A6
Focal adhesion	5.22876	8	8.23E-11	COL4A1, COL6A3, COL3A1, COL1A2, COL1A1, COL11A1. THBS2, COL4A6

Figure 4. PPI network construction, module analysis, and pathway enrichment analysis. Protein-protein interaction network for products of DEGs. A total of 143 nodes and 419 interaction associations were identified. (**A**) The nodes mean proteins; the edges mean the interactions of proteins; green circles meant dDEGs and orange circles meant uDEGs. (**B**) Module analysis based on Cytoscape software. (**C**) KEGG pathway enrichment analysis of DEGs in the module.

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. Connectivity and regulation of the top 10 hub genes.

Gene symbol	Gene title	Connectivity	Regulation
COL1A1	Collagen type I alpha 1 chain	34	Up
COL3A1	Collagen type III alpha 1 chain	30	Up
COL1A2	Collagen type I alpha 2 chain	29	Up
BGN	Biglycan	29	Up
COL5A2	Collagen type V alpha 2 chain	23	Up
THBS2	Thrombospondin 2	23	Up
TIMP1	TIMP metallopeptidase inhibitor 1	21	Up
SPP1	Secreted phosphoprotein 1	20	Up
PDGFRB	Platelet-derived growth factor receptor beta	20	Up
COL4A1	Collagen type IV alpha 1 chain	19	Up

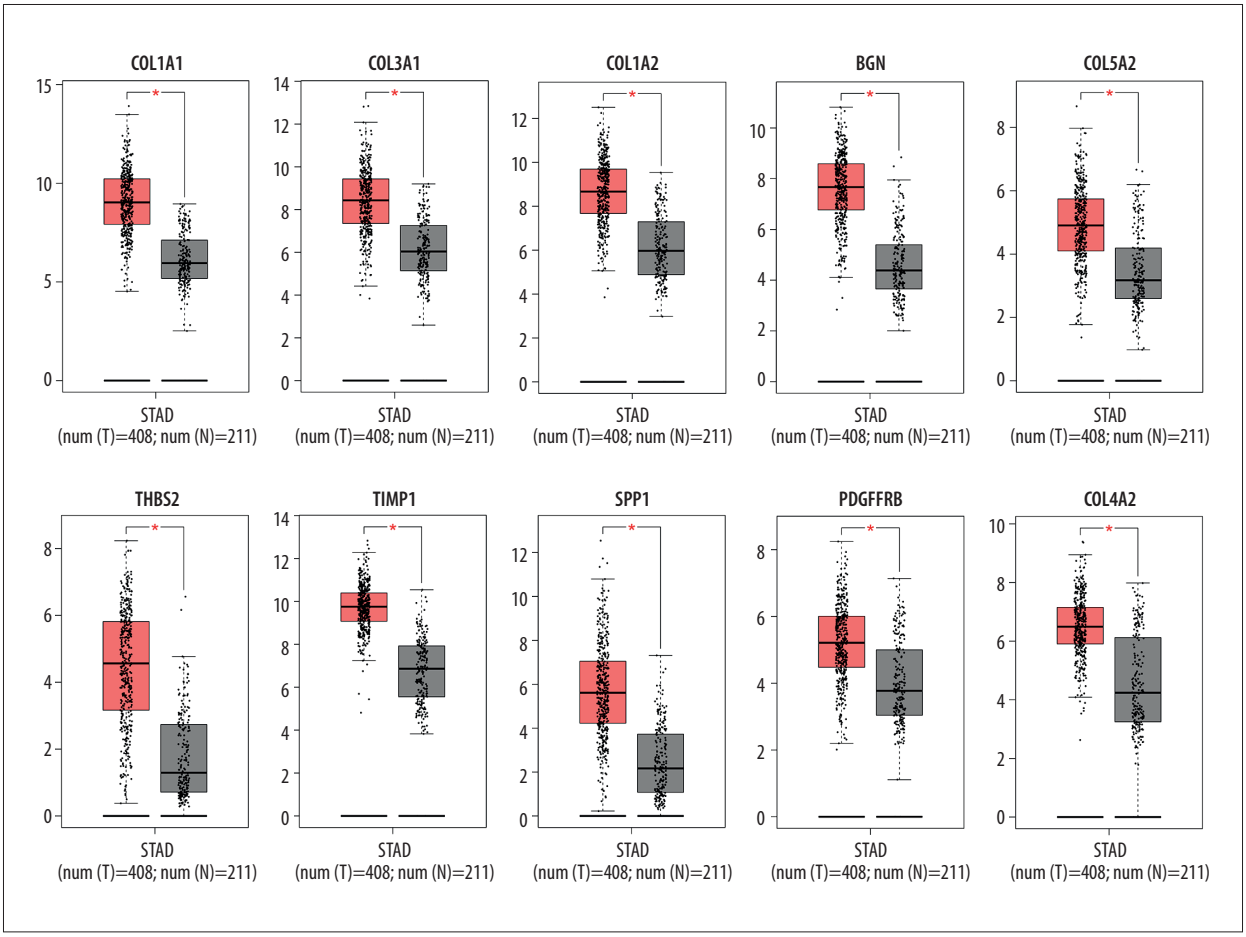


Figure 5. Validation of the gene expression levels of COL1A1, COL3A1, COL1A2, BGN, COL5A2, THBS2, TIMP1, SPP1, PDGFRB, and COL4A1 between GAC and normal gastric tissues in the GEPIA database. They are significantly upregulated in GAC compared with normal tissues (P<0.01). The red * represents P<0.01.

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

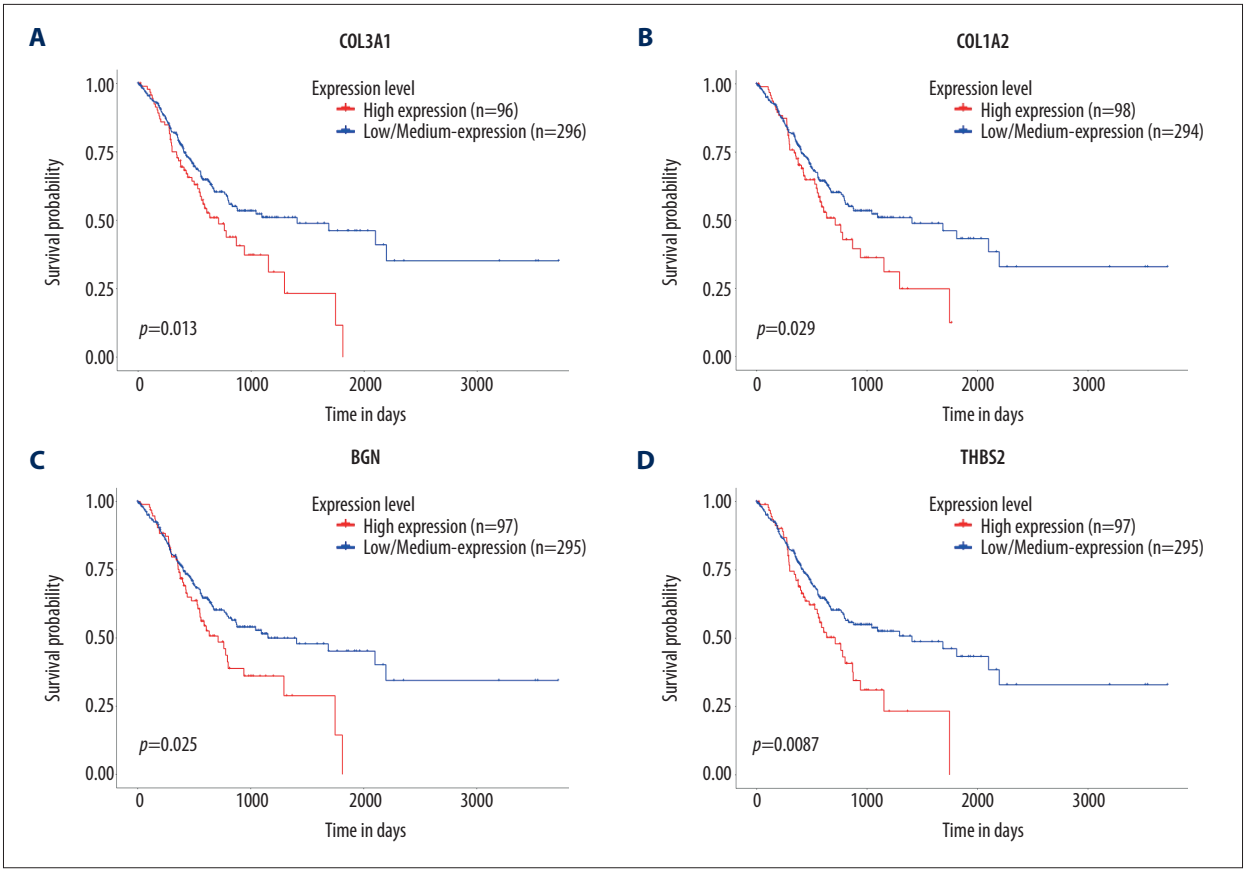


Figure 6. UALCAN overall survival analysis plot of the top 10 hub genes expressed in gastric adenocarcinoma patient samples and 4 DEGs among the top 10 hub genes that are significantly related to the survival of gastric adenocarcinoma patients ($P<0.05$). (**A**) COL3A1; (**B**) COL1A2; (**C**) BGN; (**D**) THBS2.

progression of tumors has been confirmed in some previous studies [22,23], which agrees with results of the present study.

To better understanding the relationships and interactions between these DEGs, we used Cytoscape software to construct a PPI network of DEGs-encoded proteins, and screened out 10 hub genes with high degrees. The order of degree from high to low was COL1A1, COL3A1, COL1A2, BGN, COL5A2, THBS2, TIMP1, SPP1, PDGFRB, and COL4A1. COL1A1, COL1A2, COL3A1, COL5A2, and COL4A1, which belong to the collagen (COL) family, are the top 10 hub genes, which suggests that the collagen gene is likely to be a potential target for gastric adenocarcinoma. Collagen is the main protein in bone and teeth, and is involved in the adhesion of tumor cells, gap junction, and formation of extracellular matrix (ECM) [24]. COL1A1 is the major component of type I collagen. Some studies have shown that mir-129-5p stops the invasion and proliferation of gastric cancer cells by inhibiting COL1A1[25]. Ma et al. [26] found silencing the collagen gene inhibits tumor proliferation and metastasis. Our study also found the uDEGs are enriched in cell adhesion and biological adhesion at the BP level, which suggests that DEGs belonging to the COL family play a vital role in invasion and metastasis of tumor cells. Studies [27,28] showed that COL1A2 is highly expressed in colorectal cancer and medulloblastoma. Research has revealed the high expression of COL3A1 is independently associated with the low survival rate of colorectal carcinoma [29]. However, the relationship between COL3A1 and gastric adenocarcinoma has not been studied. Zhang et al. [30] found that the high expression of COL4A1 is closely related to the depth of invasion, TNM stage, and lymph node metastasis. Makito et al. [32] demonstrated that COL4A1 can promote invasive ability and invasive growth pattern by activating the AKT pathway and upregulating epithelial-mesenchymal transition. Zhao et al. [32] also used bioinformatics methods show that COL5A2 is a key factor in gastric cancer, but there is no laboratory evidence to prove that COL5A2 is involved in gastric adenocarcinoma.

ECM is a protein compound that plays an indispensable role in cell migration and cancer development [33]. BGN, as an integral part of ECM, is considered to be a pathway for malignant tumor cells to acquire migration and invasiveness [34]. Studies have shown that the expression of BGN in GC is notably upregulated, and correlated with depth of tumor invasion

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

and TNF staging [35]. Thromboreactive protein (THBS) is an extracellular glycoprotein that plays roles in cell matrix and intercellular interactions [36]. Studies have shown that high THBS2 expression is correlated with low proliferation rate of gastric cancer cells [37]. Tissue inhibitor matrix metalloproteinase-1 (TIMP-1) is classified into the family of tissue inhibitors of metalloproteinases, and the proteins encoded by TIMP-1 are considered to be the key biofactors in the invasion and metastasis of tumors [38]. Wang et al. [39] showed that the expression level of TIMP1 in peripheral blood was associated with the stage of cancer, and the upregulation of TIMP1 may be an adverse prognostic factor for recurrence of gastric cancer. SPP1 is an ECM-related protein that has carcinogenic and anti-tumor effects [40]. Li et al. [41] also identified SPP1 as a prognostic pivotal gene in gastric cancer by bioinformatics. Sharvesh et al. [42] found that SPP1 is highly expressed in gastric cancer tissues compared with normal adjacent tissues, and its expression increased with the depth of tumor invasion. Platelet-derived growth factor receptors (PDGFRs) can induce activation of intracellular signal transduction pathways, which can promote cell proliferation, metastasis, and invasion [43]. Chen et al. [44] identified PDGFRB as a candidate gene for gastric cancer by constructing a gene co-expression network, which is consistent with the results of our study. It has also been affirmed that PDGFRB is upregulated in gastric cancer tissues, and its high expression is positively correlated with poor prognosis of gastric cancer patients [45]. These results suggest that BGN, THBS2, TIMP1, SPP1, and PDGFRB are key factors in GAC.

Module analysis from the PPI network showed that gastric adenocarcinoma is closely related to focal adhesion and ECM-receptor interaction. Focal adhesion is a complex, dynamic process involving the driving activity of actin cytoskeleton and the participation of specific receptors and signal transduction [46]. Studies have found that focal adhesions are intensely involved in multiple key pathways of tumor migration and metastasis [47]. Research by Lu et al. showed that abnormal ECM can promote the growth and metastasis of tumors by directly promoting cell metastasis on the one hand, and indirectly by promoting the formation of tumor microvessels on the other hand [48]. It is noteworthy that 8 of the 15 genes in the module (COL4A1, COL6A3, COL3A1, COL1A2, COL1A1, COL11A1, COL4A6, and THBS2) are involved in both pathways, and most of them belong to the COL family, which strengthens the findings of the role of the COL family in gastric adenocarcinoma.

To study the expression levels and prognostic value of 10 hub genes, we used GEPIA database and UALCAN for expression validation and prognostic analysis. The GEPIA database showed all the 10 hub genes are upregulated in GAC compared to normal gastric tissue. The results of the prognostic analysis showed that the upregulated expression of COL3A1, COL1A2, BGN, and THBS2 significantly reduced the survival time of GAC patients. Therefore, COL3A1, COL1A2, BGN, and THBS2 appear to be ideal prognostic indicators for gastric adenocarcinoma.

In sum, we identified DEGs and performed GO analysis, pathway enrichment analysis, and PPI network construction to understand their roles in gastric adenocarcinoma. In addition, we identified COL3A1, COL1A2, BGN, and THBS2 as hub genes and evaluated their prognostic value. This study provided evidence for early diagnosis and prognostic evaluation of gastric adenocarcinoma at the molecular level, but these findings need to be confirmed by subsequent laboratory studies.

Conclusions

In this study, we used bioinformatics to predict the DEGs of gastric adenocarcinoma and its enriched pathways and screened and evaluated some hub genes to provide some ideas and references for the early diagnosis and treatment of gastric adenocarcinoma at the molecular level. However, the limitation of our research lies in the lack of laboratory evidence. Therefore, further laboratory studies are needed to validate these findings.

Acknowledgements

We express our cordial thanks to all those who participated in this study.

Conflict of interest

None.

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

References:

1. Yao L, Shi W, Gu J: Micro-RNA 205-5p is involved in the progression of gastric cancer and targets phosphatase and tensin homolog (PTEN) in SGC-7901 human gastric cancer cells. Med Sci Monit, 2019; 25: 6367–77
2. Ferlay J, Shin HR, Bray F et al: Estimates of worldwide burden of cancer in 2008: globocan 2008. Int J Cancer, 2010; 127(12): 2893–917
3. Wu P-L, He Y-F, Yao H-H et al: Martrilin-3 (MATN3) overexpression in gastric adenocarcinoma and its prognostic significance. Med Sci Monit, 2018; 24: 348–55
4. Zhang Z, Zhu X: Clinical significance of lysophosphatidic acid receptor-2 (LPA2) and Krüppel-like factor 5 (KLF5) protein expression detected by tissue microarray in gastric adenocarcinoma. Med Sci Monit, 2019; 25: 4705–15
5. Ahn S, Park DY: Practical points in gastric pathology. Arch Pathol Lab Med, 2016; 140(5): 397–405
6. Beeharry MK: New blood markers detection technology: A leap in the diagnosis of gastric cancer. World J Gastroenterol, 2016; 22(3): 1202–12
7. Blee TK, Gray NK, Brook M: Modulation of the cytoplasmic functions of mammalian post-transcriptional regulatory proteins by methylation and acetylation: A key layer of regulation waiting to be uncovered. Biochem Soc Trans, 2015; 43(6): 1285–95
8. Kulasingam V, Diamandis EP: Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat Clin Pract Oncol, 2008; 5(10): 588–99
9. Chivu Economescu M, Necula LG, Dragu D et al: Identification of potential biomarkers for early and advanced gastric adenocarcinoma detection. Hepatogastroenterol, 2010; 57(104): 1453–64
10. He J, Jin Y, Chen Y et al: Downregulation of ALDOB is associated with poor prognosis of patients with gastric cancer. Oncotargets Ther, 2016; 9: 6099–109
11. Holbrook JD, Parker JS, Gallagher KT et al: Deep sequencing of gastric carcinoma reveals somatic mutations relevant to personalized medicine. J Transl Med, 2011; 9: 119
12. Barrett T, Wilhite SE, Ledoux P et al: NCBI GEO: archive for functional genomics data sets – update. Nucleic Acids Res, 2013; 41: 991–95
13. Ashburner M, Ball CA, Blake JA et al: Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet, 2000; 25: 25–29
14. Kanehisa M, Goto S: KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res, 2000; 28: 27–30
15. Dennis G Jr., Sherman BT, Hosack DA et al: DAVID: Database for annotation, visualization and integrated discovery. Genome Biol, 2003; 4: P3
16. Szklarczyk D, Franceschini A, Wyder S et al: STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res, 2015; 3: D447–52
17. Shannon P, Markiel A, Ozier O et al: Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res, 2003; 13: 2498–504
18. 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: W98–102
19. Chandrashekar DS, Bashel B, Balasubramanya SAH et al: UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia, 2017; 19: 649–58
20. Zhang H, Liu J, Fu X, Yang A: Identification of key genes and pathways in tongue squamous cell carcinoma using bioinformatics analysis. Med Sci Monit, 2017; 23: 5924–32
21. Gumuslu E, Cine N, Ertan Gökbayrak M et al: Exenatide alters gene expression of neural cell adhesion molecule (NCAM), intercellular cell adhesion molecule (ICAM), and vascular cell adhesion molecule (VCAM) in the hippocampus of type 2 diabetic model mice. Med Sci Monit, 2016; 22: 2664–69
22. Quan R, Ning Z, Wang Y et al: Prognostic value of upregulation of myristoylated alanine-rich C-kinase substrate in gastric cancer. Med Sci Monit, 2019; 25: 279–87
23. Chow CR, Ebine K, Knab LM et al: Cancer cell invasion in three-dimensional collagen is regulated differentially by Gα 13, protein and discoidin domain receptor 1-Par3 protein signaling. J Biol Chem, 2016; 291(4): 1605–18
24. Carbone L, Harris RA, Gnerre S: Gibbon genome and the fast karyotype evolution of small apes. Nature, 2014; 513(7517): 195–201
25. Quan Wang, JY: MiR-129-5p suppresses gastric cancer cell invasion and proliferation by inhibiting COL1A1. Biochem Cell Biol, 2018; 96(1): 19–25
26. Ma H-P, Chang H-L, Bamodu OA et al: Collagen 1A1 (COL1A1) is a reliable biomarker and putative therapeutic target for hepatocellular carcinogenesis and metastasis. Cancers (Basel), 2019; 11(6): 786
27. 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(Complete): 473–85
28. Liang Y, Diehn M, Bollen AW et al: Type I collagen is overexpressed in medulloblastoma as a component of tumor microenvironment. J Neurooncol, 2018; 86(2): 133–41
29. Wang XQ, Tang ZX, Yu D et al: Epithelial but not stromal expression of collagen alpha-1(iii) is a diagnostic and prognostic indicator of colorectal carcinoma. Oncotarget, 2016; 7(8): 8823–38
30. Zhang Q-N, Zhu H-L, Xia M-T et al: A panel of collagen genes are associated with prognosis of patients with gastric cancer and regulated by microRNA-29c-3p: An integrated bioinformatics analysis and experimental validation. Cancer Manag Res, 2019; 11: 4757–72
31. Miyake M, Hori S, Morizawa Y et al: Collagen type IV alpha 1 (COL4A1) and collagen type XIII alpha 1 (COL13A1) produced in cancer cells promote tumor budding at the invasion front in human urothelial carcinoma of the bladder. Oncotarget, 2017; 8(22): 36099–114
32. Zhao X, Cai H, Wang X, Ma L: Discovery of signature genes in gastric cancer associated with prognosis. Neoplasma, 2016; 63(2): 239–45
33. Bryce LF, Jordan M, Alison M et al: Beyond the matrix: The many non-ECM ligands for integrins. Int J Mol Sci, 2018; 19(2): pii: E449
34. Biddle A, Mackenzie IC: Cancer stem cells and EMT in carcinoma. Cancer Metastasis Rev, 2012; 31(1–2): 285–93
35. Rongkun L, Chun Z, Shuheng J et al: ITGBL1 predicts a poor prognosis and correlates EMT phenotype in gastric cancer. J Cancer, 2017; 8(18): 3764–73
36. Bornstein P, Armstrong LC, Hankenson KD et al: Thrombospondin 2, a matricellular protein with diverse functions. Matrix Biol, 2000; 19(7): 557–68
37. Sun R, Wu J, Chen Y et al: Downregulation of thrombospondin2 predicts poor prognosis in patients with gastric cancer. Mol Cancer, 2014; 13(1): 225–34
38. Song G, Xu S, Zhang H et al: TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway. J Exp Clin Cancer Res, 2016; 35(1): 148
39. Wang Y-Y, Li L, Zhao Z-S, et al. Clinical utility of measuring expression levels of KAP1, TIMP1 and STC2 in peripheral blood of patients with gastric cancer. World J Surg Oncol, 2013; 11: 81
40. Insua-Rodríguez J, Pein M, Hongu T et al: Stress signaling in breast cancer cells induces matrix components that promote chemoresistant metastasis. EMBO Mol Med, 2018; 10(10): e9003
41. Li T, Gao X, Han L et al: Identification of hub genes with prognostic values in gastric cancer by bioinformatics analysis. World J Surg Oncol, 2018; 16(1): 114–20
42. Seeruttun SR, Cheung WY, Wang W et al: Identification of molecular biomarkers for the diagnosis of gastric cancer and lymph-node metastasis. Gastroenterol Rep, 2019; 7(1): 57–66
43. Heldin CH, Westermark B: Mechanism of action and *in vivo* role of platelet-derived growth factor. Physiol Rev, 1999; 79: 1283–316
44. Chen J, Wang X, Hu B et al: Candidate genes in gastric cancer identified by constructing a weighted gene co-expression network. Peer J, 2018; 6: e4692
45. Wang JX, Zhou JF, Huang FK et al: Gli2 induces pdgfrb expression and modulates cancer stem cell properties of gastric cancer. Eur Rev Med Pharmacol Sci, 2017; 21(17): 3857–65
46. Fogh BS, Multhaupt HAB, Couchman JR: Protein kinase C, focal adhesions and the regulation of cell migration. J Histochem Cytochem, 2014; 62(3): 172–84
47. Bach CTT, Schevzov G, Bryce NS et al: Tropomyosin isoform modulation of focal adhesion structure and cell migration. Cell Adh Migr, 2010; 4(2): 226–34
48. Lu P, Weaver V M, Werb Z: The extracellular matrix: A dynamic niche in cancer progression. J Cell Biol, 2012; 196(4): 395–406

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