Bioinformatics analysis of aberrantly methylated-differentially expressed genes in gastric cancer

Song Siyuan

Affiliated Hospital of Nanjing University of Chinese Medicine

Shu Peng (✉ shupengnjucm@163.com)

Affiliated Hospital of Nanjing University of Chinese Medicine

Research Article

Bioinformatics analysis of aberrantly methylated-differentially expressed genes in gastric cancer

Siyuan Song[1,2,3], Peng Shu[1,2,3*]

[1] Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing 210029, Jiangsu Province, China

[2] Nanjing University of Chinese Medicine, Nanjing 210029, Jiangsu Province, China

[3] Jiangsu Provincial Hospital of Chinese Medicine, Nanjing 210029, Jiangsu Province, China

Abstract

Background: This study was carried out to identify the aberrantly methylated-differentially expressed genes in gastric cancer (GC). *Methods:* We downloaded data of gene expression microarrays GSE118916 and gene methylation microarrays GSE25869 from the Gene Expression Omnibus (GEO) database. The DEGs and DMGs were analyzed by the limma software package and Venn diagram. The PPI network was mapped and the enrichment analysis was conducted by the DAVID database. GEPIA online tool, Oncomine database, HPA, and cBioPortal tool were used to verify hub genes. *Result:* We obtained 110 Hypo-HGs, 9 high-regulation hypomethylation oncogenes, 23 Hyper-LGs, and 2 low-regulation hypermethylation tumor suppressor genes. Hypo-HGs biological process mainly involves cell adhesion and extracellular matrix organization, Hyper-LGs biological process mainly involves response to nicotine and xenobiotic metabolic process. KEGG analysis showed that Hypo-HGs significantly enriched in Focal adhesion, PI3K-Akt signaling pathway, and ECM-receptor interaction. Hyper-LGs significantly enriched in Drug metabolism-cytochrome P450, Chemical carcinogenesis, and Metabolism of xenobiotics by cytochrome P450. The database identified the hub genes were COL1A1, THBS1, COL5A2, COL12A1, and CXCR. *Conclusion:* COL1A1, THBS1, COL5A2, COL12A1, and CXCR4 can be used as a target for precise diagnosis and treatment of GC. Focal adhesion, PI3K-Akt signaling pathway, and ECM-receptor interaction are important mechanisms of GC.

Keywords: Bioinformatics; Gastric Cancer; Differentially expressed genes;

Methylation;

1. Introduction

Gastric cancer (GC) refers to an epithelial malignant tumor that originates in the stomach and is the third most common cause of cancer death, second only to lung cancer and liver cancer [1]. The occurrence of GC is affected by many factors, including dietary factors, environmental factors, and genetic factors [2-4]. The treatment methods of GC mainly include surgery, chemotherapy, radiotherapy, immunotherapy, and targeted therapy [5]. Although the GC treatment methods are constantly evolving, the survival rate is still low due to the prone to recurrence and metastasis of advanced GC [6]. According to statistics, the 5-year survival rate of GC patients is only 20% to 30% [7]. Therefore, searching for hub genes and biomarkers that affect the occurrence and development of GC is of great significance for the early diagnosis, treatment, prognosis, and drug discovery of GC.

Epigenetics was considered to be a heritable change in gene expression, not mediated by changes within the DNA sequence [8]. DNA methylation in the gene promoter region was related to the silencing of oncogenes and tumor suppressor, and was considered to be a hallmark of many tumors [9]. Although some studies have confirmed that certain genes have abnormal DNA hypermethylation or hypomethylation in GC [10, 11], but it is still difficult to determine the comprehensive profile and pathways of the interaction network.

In recent years, high-throughput platform microarrays can be used to screen for genetic or epigenetic changes in cancer [12]. Many gene expression profiling microarray analyses and abnormal methylation studies have been carried out via this tool, and various differentially expressed genes (DEGs) [13] and differentially methylated genes (DMGs) [14] in GC have been discovered. However, these studies did not conduct joint analysis, which may lead to the lack of some hub genes. In this study, we combined gene expression profiles and gene methylation microarray data to identify abnormally methylated and differentially expressed genes and pathways between GC tissue and normal tissue. Then used GEPIA database, Oncomine database, HPA, and cBioPortal

tool to identify the hub genes involved in the pathogenesis of GC. The protocol of our study procedures is shown in Fig. 1.
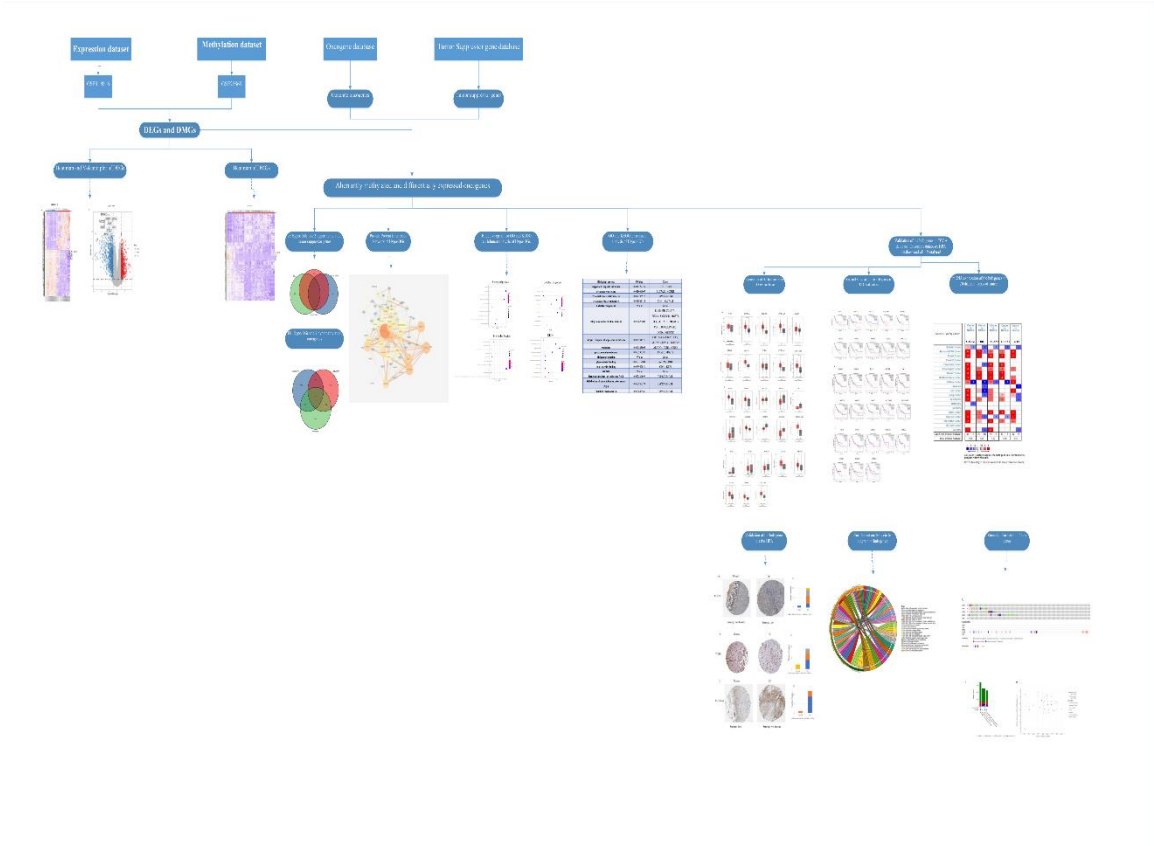

Fig.1: The protocol of our study procedures

2. Method

2.1 Microarray data

The GC gene expression dataset GSE118916 and the methylation dataset GSE25869 were received by searching the keyword "gastric cancer" in the GEO database (https://www.ncbi.nlm.nih.gov/geo/). GSE118916 (platform: GPL15207 Affymetrix Human Gene Expression Array) included 30 samples, including 15 in the GC group and 15 in the normal group. GSE25869 (platform GPL8490 Illumina HumanMethylation27 BeadChip) included 72 samples, including 24 in the GC group and 48 in the normal group.

2.2 Data processing

The R software [15] was used to process the data set, using FDR<0.05 and |logFC |>1 as the inclusion criteria for screening DEGs, and FDR<0.05 and |logFC |>0.1 as the

[在此处键入]

inclusion criteria for screening DMGs. Generate oncogenes and tumor suppressor genes from the Oncogene database (http://ongene.bioinfo-minzhao.org/) and Tumor Suppressor gene database (https://bioinfo.uth.edu/TSGene/index.html). The online Venn diagram [16] (http://bioinfogp.cnb.csic.es/tools/venny/index.html) was used to identify up-regulated hypomethylated oncogenes and down-regulated hypermethylated tumor suppressor genes.

2.3 Protein-Protein Interaction Network Construction

The obtained DEGs and DMGs were imported into a String database (https://string-db.org/cgi/input.pl) for PPI network, and visualized using Cytoscape3.7.2.

2.4 Gene ontology and pathway functional enrichment analysis

110 mutually inclusive hypomethylated/up-expression and 23 hypermethylated/down-expression genes were imported into the DAVID database[17] (https://david.ncifcrf.gov/) for GO function [18]and KEGG pathway enrichment analysis[19], the species was qualified as Homo sapiens. GO enrichment analysis was mainly composed of biological process (BP), cellular component (CC), and molecular function (MF).

2.5 Validation of the hub genes in the TCGA database

The Hypo-HGs, Hyper-LGs which Degree value of the top 10 genes, 9 upregulated-hypomethylated oncogenes, and 2 downregulated-hypermethylated were input into an online tool of GEPIA (http://gepia.cancer-pku.cn/index.html) to verify its expression in TCGA-STAD and draw Kaplan-Meier survival curve (OS). And the hub genes were identified.

2.6 Analysis of hub genes in Oncomine database

Oncomine (https://www.oncomine.org/)[20] is a microarray-based gene database and integrated data-mining online cancer microarray database. The Oncomine database was used to confirm the expression of 5 hub genes in 20 different types of cancer, and to explore the mRNA expression differences between GC and normal gastric tissue.

2.7 Validation in HPA and Genetic information of the hub genes

In this paper, the protein expression and distribution of hub genes were investigated in GC tissues and compared to normal tissues in HPA (https://www.proteinatlas.org/) [21].

[在此处键入]

And the hub genes enrichment analysis circle diagram was drawn. The cBioPortal tool (http://www.cbioportal.org/) was used to discover the genetic information of GC hub genes and the correlation between messenger RNA (mRNA) expression and DNA methylation.

3. Results

3.1 DEGs and DMGs in GC

The GSE118916 expression matrix has 1163 DEGs, including 528 up-regulated genes and 635 down-regulated genes, drawn into cluster heat maps (Fig. 2A) and volcano maps (Fig. 2B). A total of 2589 DMGs were obtained in GSE25869, including 680 hypermethylated genes and 1909 hypomethylated genes, which were drawn into a cluster heat map (Fig. 3);
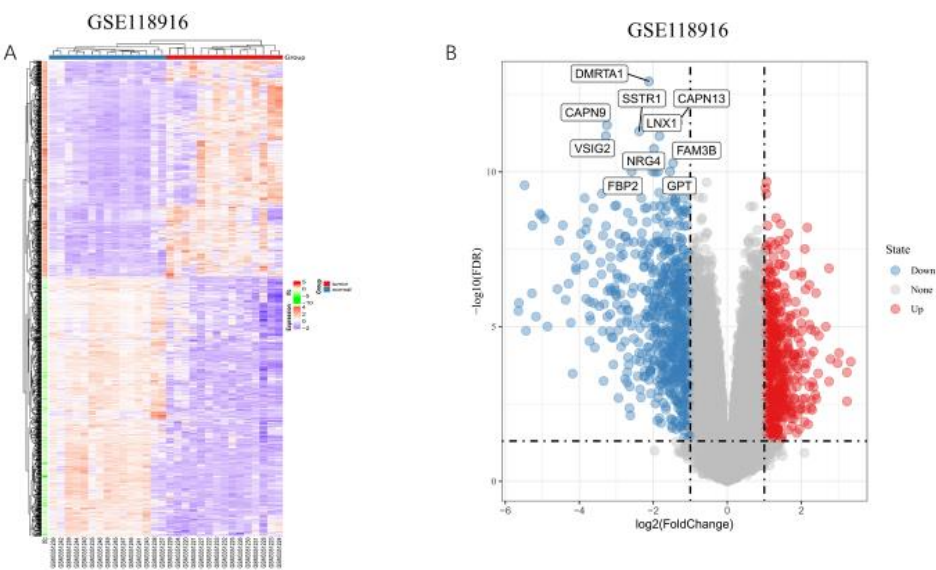

Fig. 2 Heat map and Volcano plot of DEGs

2A：The hierarchical clustering of the heat map reveals the DEGs in the GC. Orange and purple indicate higher and lower expression levels, respectively.

2B：The volcano map visualizes all DEGs. The red dots represent up-regulated genes, the blue dots represent down-regulated genes, and the gray dots represent genes that are not differentially expressed.
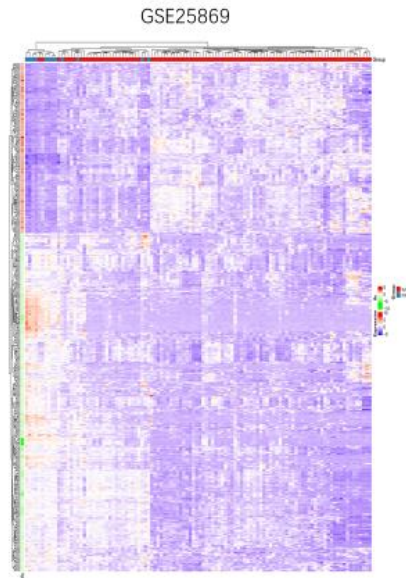
[在此处键入]

Fig. 3 Heat map of DMGs

The hierarchical clustering of the heat map reveals the DMGs in the GC. Orange and purple indicate

higher and lower expression levels, respectively.

3.2 Aberrantly methylated DEGs

By overlapping 1909 hypomethylation genes, 528 upregulated genes, and 803 oncogenes, we obtained 110 Hypo/HGs, and 9 up-regulation hypomethylation oncogenes, which may potentially trigger tumorigenesis by up-regulating gene expression after methylation (Fig. 4A). At the same time, 23 Hyper/LGs were obtained, and 2 hypermethylation tumor suppressor genes were downregulated. Hypermethylation genes may trigger tumorigenesis by downregulating the expression of these genes (Fig. 4B).
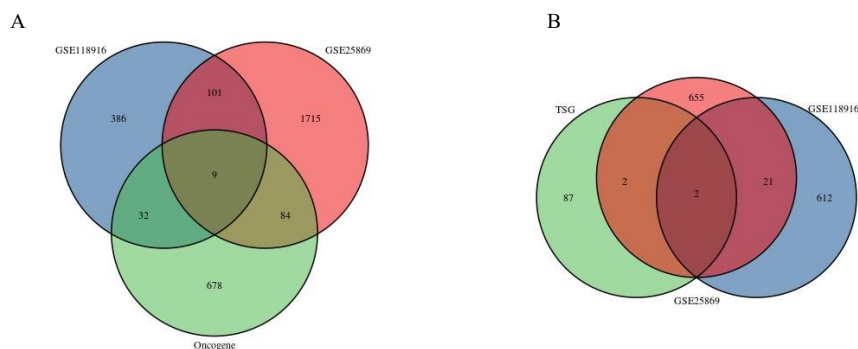

Fig. 4 Aberrantly methylated and differentially expressed oncogenes

[在此处键入]

4A: 110 hypomethylated and up-regulated genes were identified, including 9 up-regulation

hypomethylation oncogenes. 4B: 23 hypermethylated and down-regulated genes were identified,

including 2 hypermethylation tumor suppressor genes (TSGs).

3.3 Protein-Protein Interaction Network Construction

110 Hypo-HGs and 23 Hyper-LGs were imported into the string database. Cytoscape

3.7.2 was used for analysis, sorted by Degree value, FN1, COL3A1, COL1A1,

COL1A2, MMP2 in Hypo-HGs (Fig. 5), CDH1, FOXA1, and KLF4 in Hyper-LGs

are at the core.

Fig. 5 Protein-Protein Interaction Network of Hypo-HGs

A larger Degree or combined-score value indicates a larger node size and deeper node color.

3.4 Gene ontology and pathway functional enrichment analysis

The GO analysis of Hypo-HGs showed that BP was mainly concentrated in cell

adhesion and extracellular matrix organization, CC was mainly concentrated in the

extracellular region, extracellular space, and extracellular exosome, and MF was

[在此处键入]

mainly concentrated in protein-binding, calcium ion binding, and heparin binding. Enriched. KEGG analysis results showed that Hypo-HGs were significantly enriched in Focal adhesion, PI3K-Akt signaling pathway, and ECM-receptor interaction. R language was used to draw the GO and KEGG bubble charts according to the count value of the TOP10 (Fig. 6). The GO analysis results of Hyper-LGs showed that BP was mainly enriched in response to nicotine and xenobiotic metabolic process, CC was mainly enriched in integral component of membrane, an integral component of the plasma membrane, endosome, and MF were mainly enriched in glycoprotein binding, beta-Catenin binding. KEGG analysis results showed that Hyper-LGs were significantly enriched in Drug metabolism-cytochrome P450, Chemical carcinogenesis, and Metabolism of xenobiotics by cytochrome P450 (Table 1).
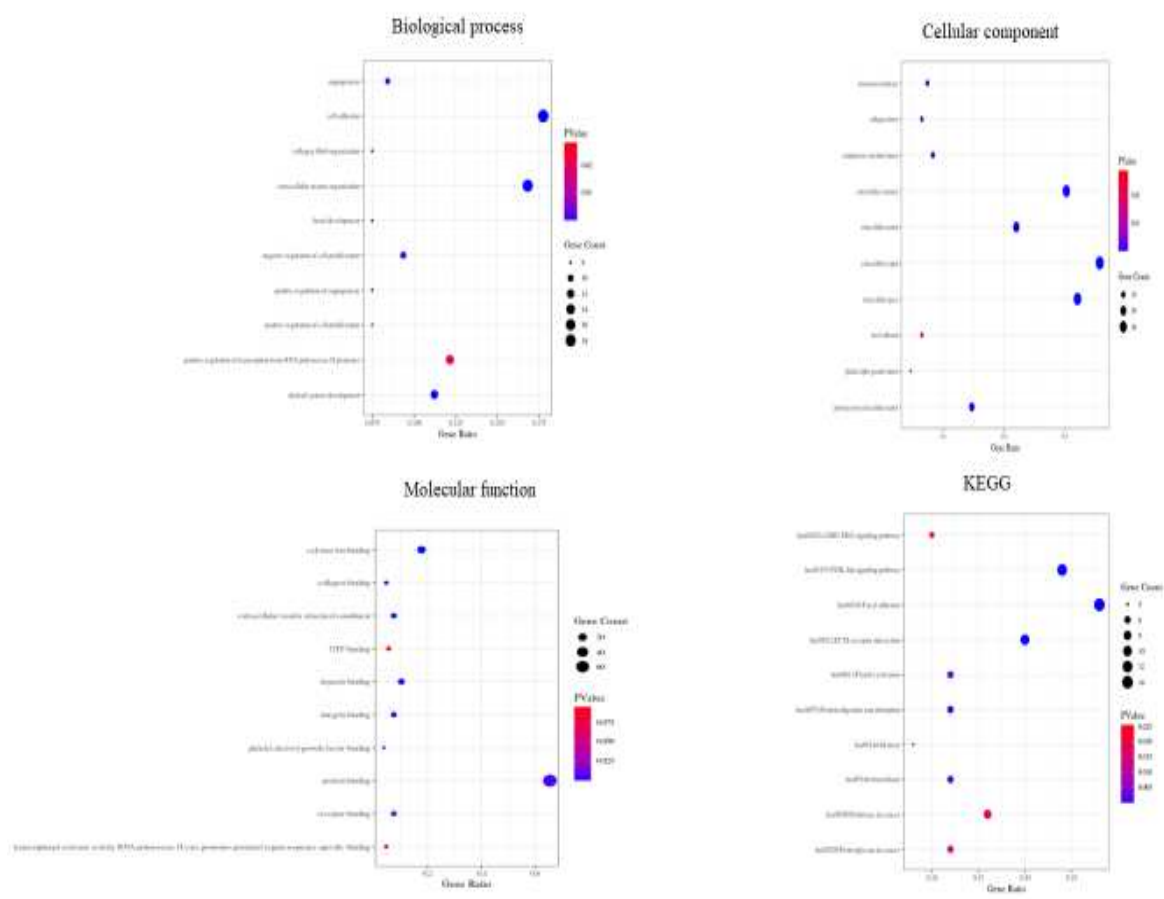

Fig. 6 Bubble diagram for GO and KEGG enrichment analysis of Hypo-HGs

The size of the bubble in the figure represents the number of enriched genes, and the difference in

[在此处键入]

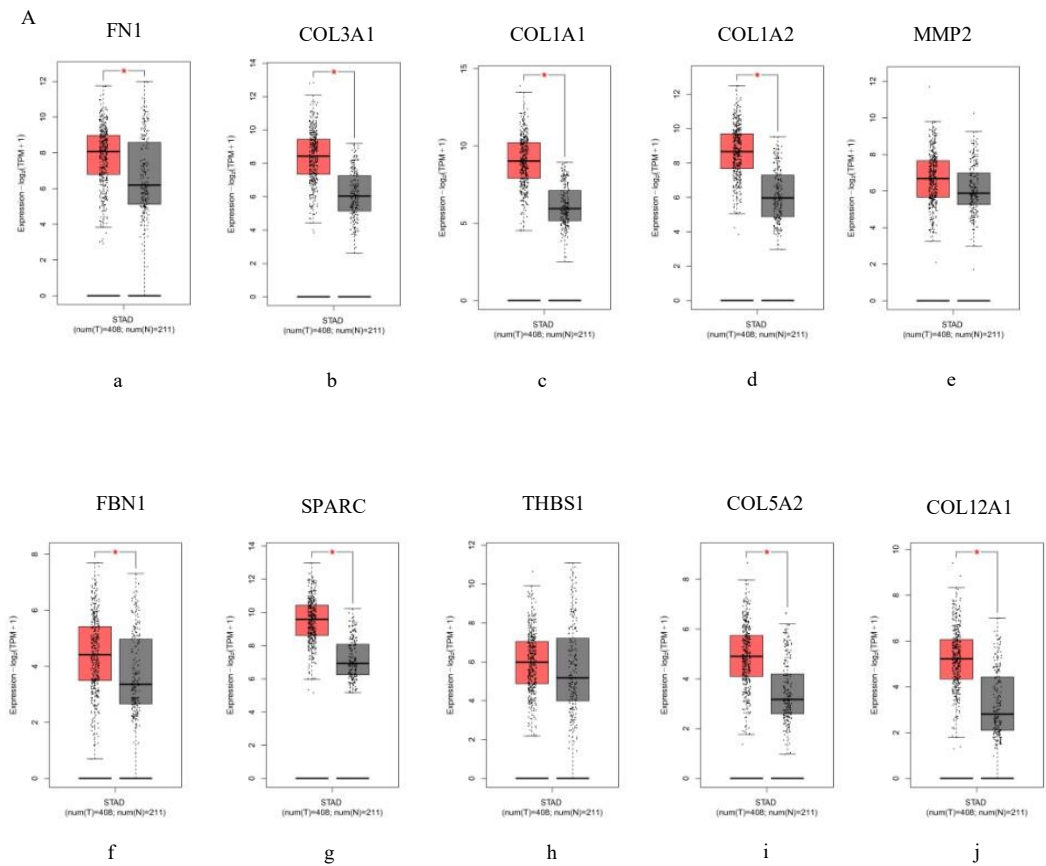bubble color represents the significance of gene enrichment.

Table 1 GO and KEGG enrichment analysis of Hyper-LGs

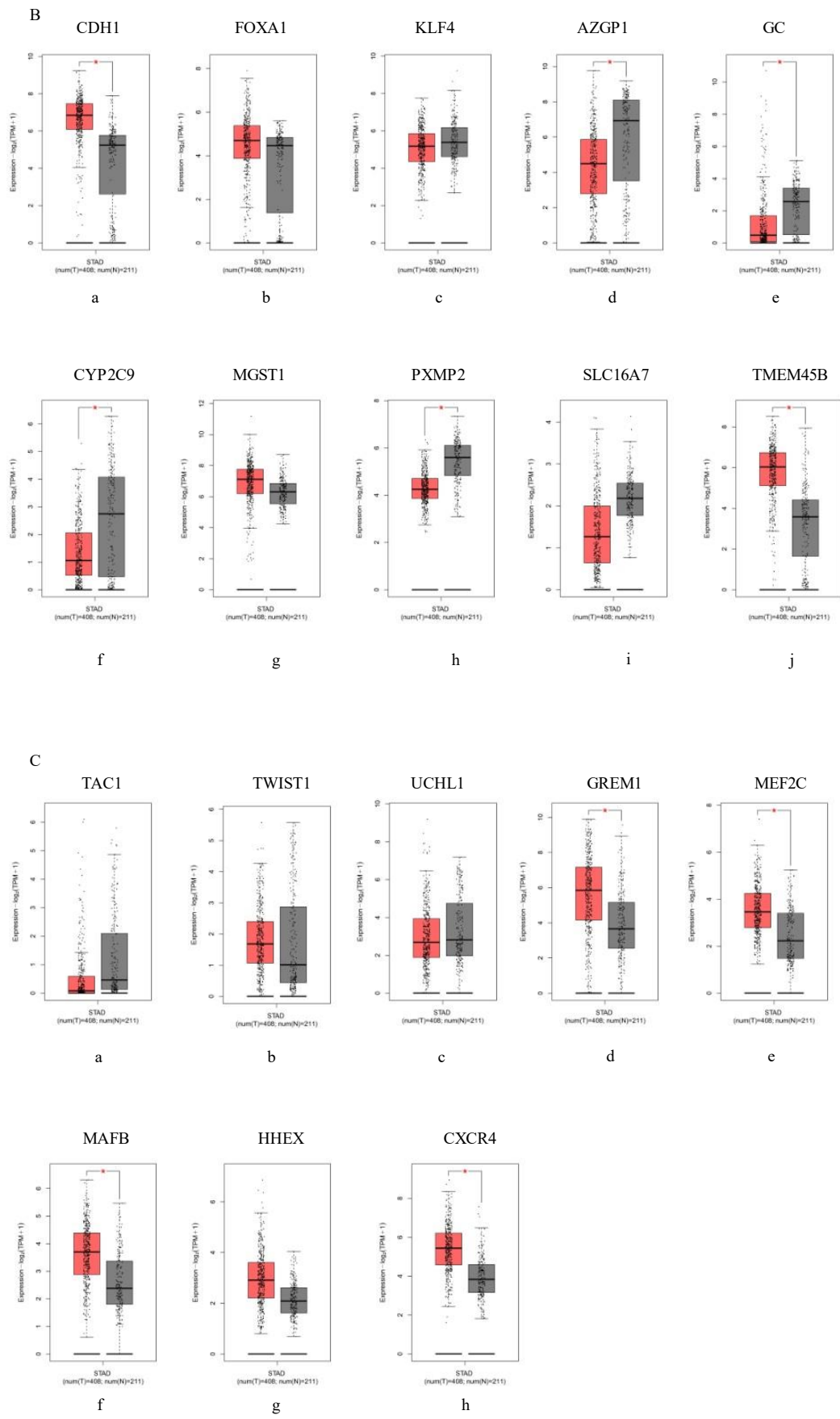| Biological process	PValue	Genes
response to organic substance	0.031227253	CDH1, S100P
response to nicotine	0.041066807	SLC7A11, KCNK1
xenobiotic metabolic process	0.084705572	CYP2C9, MGST1
response to toxic substance	0.091965444	CDH1, SLC7A11
Cellular component	*PValue*	*Genes*
integral component of membrane	0.008623408	TMEM45B, COL17A1, REG3A, SIGLEC11, MGST1, LIFR, SLC7A11, TMEM116, CDH1, DPCR1, PXMP2, KCNK1, GPRC5C
integral component of plasma membrane	0.024234976	COL17A1, SLC16A7, LIFR, SLC7A11, KCNK1, GPRC5C
endosome	0.029805805	ARRDC4, CDH1, KCNK1
peroxisomal membrane	0.063235381	PXMP2, MGST1
Molecular function	*PValue*	*Genes*
glycoprotein binding	0.067118903	AZGP1, CDH1
beta-catenin binding	0.083957851	CDH1, KLF4
KEGG	*PValue*	*Genes*
Drug metabolism - cytochrome P450	0.057884899	CYP2C9, MGST1
Metabolism of xenobiotics by cytochrome P450	0.062855379	CYP2C9, MGST1
Chemical carcinogenesis	0.067803986	CYP2C9, MGST1

3.5 Validation of the hub genes in TCGA database

[在此处键入]

GEPIA database was used to view Hypo-HGs (Fig. 7A), Hyper-LGs (Fig. 7B) Degree value of the top 10 genes, 9 up-regulated hypomethylation oncogenes TAC1, TWIST1, UCHL1, SPARC, GREM1, MEF2C, MAFB, HHEX, CXCR4 (Fig. 7C) and 2 down-regulated hypermethylated TSGs AZGP1, CDH1 (7A) in the STAD samples in the TCGA database. The results showed that COL3A1, COL1A2, COL1A2, SPARC, CDH1, and TMEM45B were highly expressed in GC tissues, and PXMP2 was lowly expressed in GC tissues. Survival analysis of the above-mentioned genes was performed to draw a Kaplan-Meier survival curve (Fig. 8). Patients with high expression of COL1A1, THBS1, COL5A2, COL12A1, CXCR4 were associated with shorter overall survival (P＜0.05). Although the p-value was not statistically significant, from the image trend point of view, Patients with high expression of FN1, COL1A2, MMP2, FBN1, SPARC, TAC1, and GREM1 were associated with shorter OS.
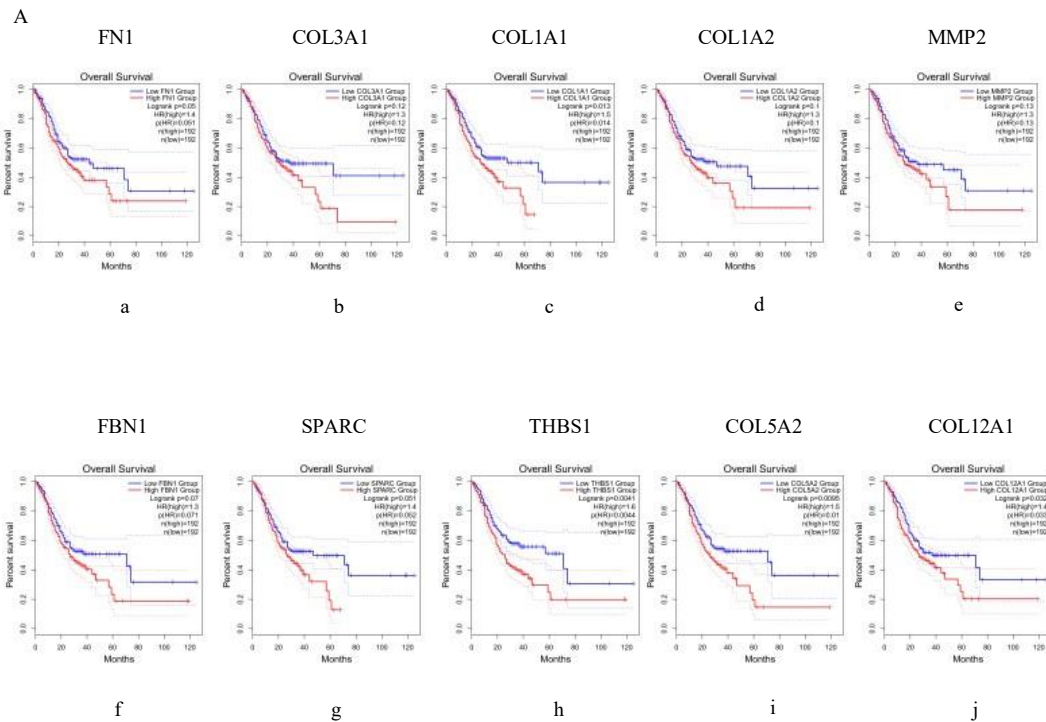

[在此处键入]

B

CDH1
a

FOXA1
b

KLF4
c

AZGP1
d

GC
e

CYP2C9
f

MGST1
g

PXMP2
h

SLC16A7
i

TMEM45B
j

C

TAC1
a

TWIST1
b

UCHL1
c

GREM1
d

MEF2C
e

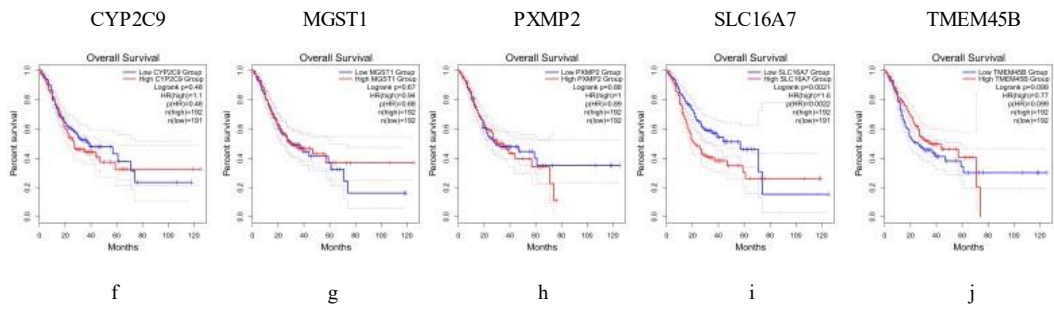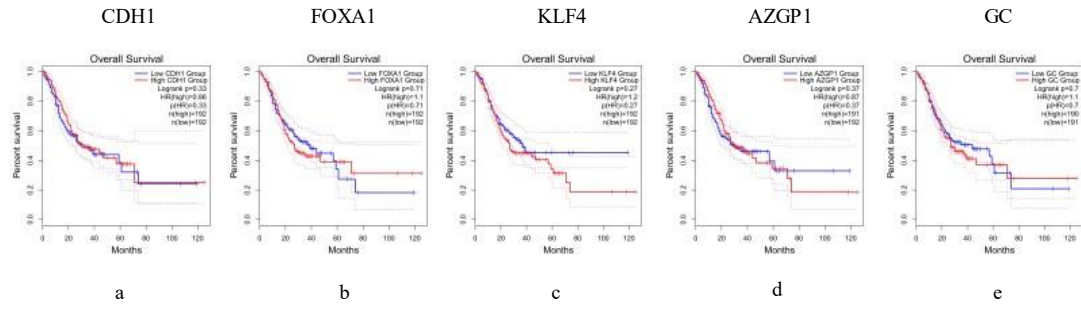MAFB
f

HHEX
g

CXCR4
h

[在此处键入]

Fig. 7 Expression of the hub genes in TCGA database

The box plots showed that the genes expression of (Aa)FN1, (Ab)COL3A1, (Ac)COL1A1, (Ad)COL1A2, (Ae)MMP2, (Af)FBN1, (Ag)SPARC, (Ah)THBS1, (Ai)COL5A2, (Aj)COL12A1, (Ba)CDH1, (Bb)FOXA1, (Bc)KLF4, (Bd)AZGP1, (Be)GC, (Bf)CYP2C9, (Bg)MGST1, (Bh)PXMP2, (Bi)SLC16A7, (Bj)TMEM45B, (Ca)TAC1, (Cb)TWIST1, (Cc)UCHL1, (Cd)GREM1, (Ce)MEF2C, (Cf)MAFB, (Cg)HHEX, and (Ch)CXCR4 in GEPIA. Red represents Tumor, Gray represents normal. (Ab)COL3A1, (Ad)COL1A2, (Ag)SPARC, (Ba)CDH1, (Aj)TMEM45B were highly expressed in GC tissues, and PXMP2 was lowly expressed in GC tissues.
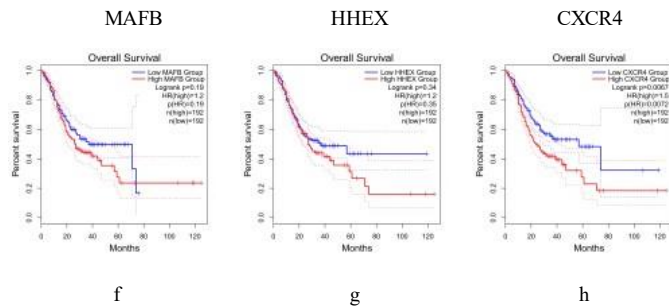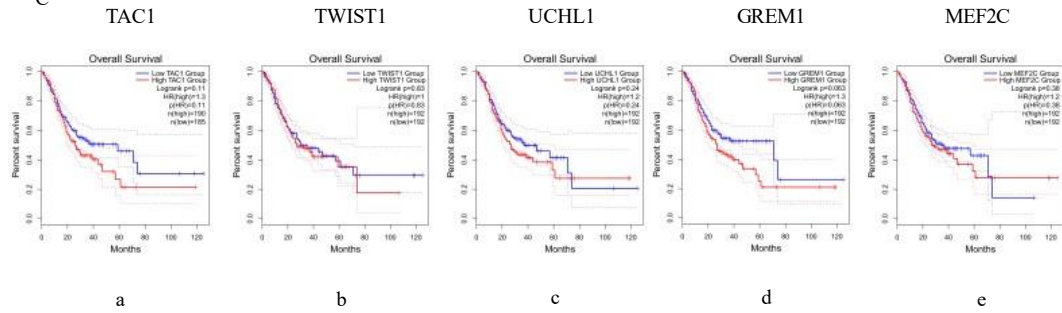

[在此处键入]

B

CDH1

a

FOXA1

b

KLF4

c

AZGP1

d

GC

e

CYP2C9

f

MGST1

g

PXMP2

h

SLC16A7

i

TMEM45B

j

C

TAC1

a

TWIST1

b

UCHL1

c

GREM1

d

MEF2C

e

MAFB

f

HHEX

g

CXCR4

h

[在此处键入]

Fig. 8 Overall Survival of the hub genes in TCGA database

The line charts showed that the OS of hub gene (Aa)FN1, (Ab)COL3A1, (Ac)COL1A1, (Ad)COL1A2, (Ae)MMP2, (Af)FBN1, (Ag)SPARC, (Ah)THBS1, (Ai)COL5A2, (Aj)COL12A1, (Ba)CDH1, (Bb)FOXA1, (Bc)KLF4, (Bd)AZGP1, (Be)GC, (Bf)CYP2C9, (Bg)MGST1, (Bh)PXMP2, (Bi)SLC16A7, (Bj)TMEM45B, (Ca)TAC1, (Cb)TWIST1, (Cc)UCHL1, (Cd)GREM1, (Ce)MEF2C, (Cf)MAFB, (Cg)HHEX, and (Ch)CXCR4 in GEPIA. The survival curve comparing the patients with high (red) and low (blue) expression in GC. Prognostic value of (Ac)COL1A1, (Ah)THBS1, (Ai)COL5A2, (Aj)COL12A1, (Ch)CXCR4 had a significant difference（P＜0.05）. Although the p value was not statistically significant, from the image trend point of view, patients with high expression of (Aa)FN1, (Ad)COL1A2, (Ae)MMP2, (Af)FBN1, (Ag)SPARC, (Ca)TAC1, and (Cd)GREM1 were associated with shorter OS.

3.6 Analysis of the hub genes in Oncomine database

The results of the Oncomine database showed that COL1A1, THBS1, COL5A2, COL12A1 were all expressed in GC, which was statistically significant. Although CXCR4 was not expressed in GC, 41 studies have shown its expression in other tumors. (Fig. 9).
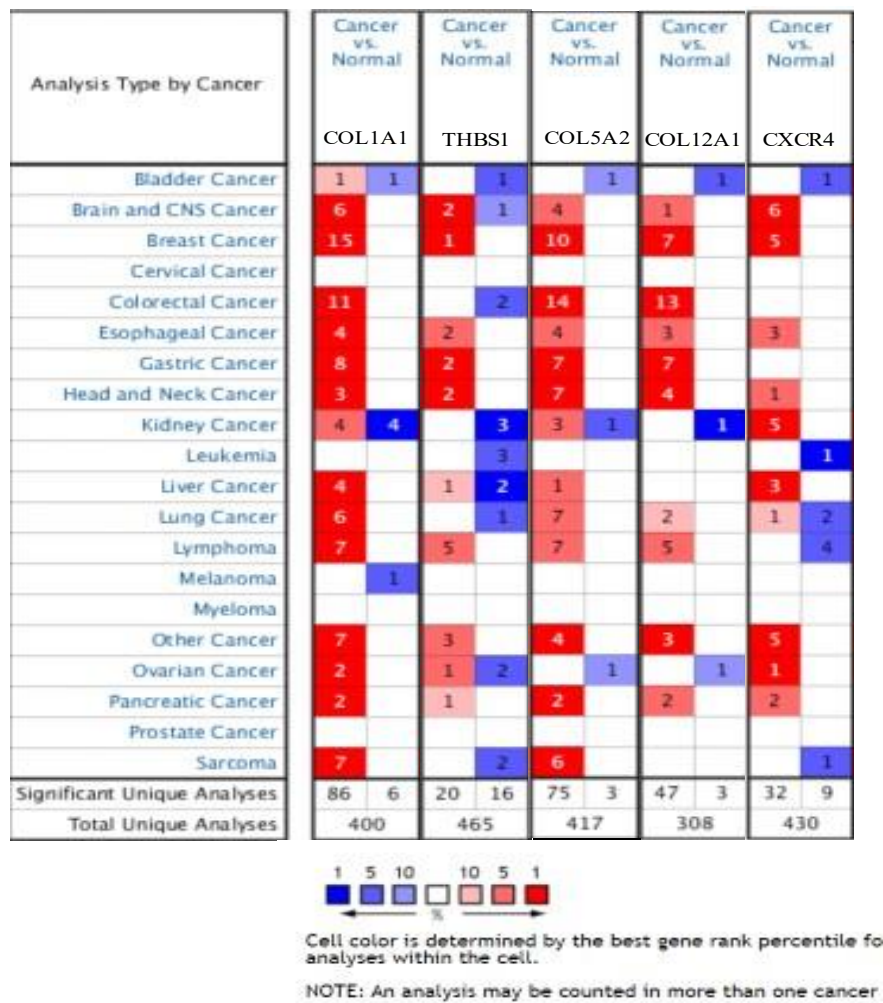
Analysis Type by Cancer	COL1A1 (Cancer vs. Normal)		THBS1 (Cancer vs. Normal)		COL5A2 (Cancer vs. Normal)		COL12A1 (Cancer vs. Normal)		CXCR4 (Cancer vs. Normal)	
Bladder Cancer	1	1		1		1		1		1
Brain and CNS Cancer	6		2	1	4		1		6	
Breast Cancer	15		1		10		7		5	
Cervical Cancer										
Colorectal Cancer	11			2	14		13			
Esophageal Cancer	4		2		4		3		3	
Gastric Cancer	8		2		7		7			
Head and Neck Cancer	3		2		7		4		1	
Kidney Cancer	4	4		3	3	1		1	5	
Leukemia				3						1
Liver Cancer	4		1	2	1				3	
Lung Cancer	6			1	7		2		1	2
Lymphoma	7		5		7		5			4
Melanoma		1								
Myeloma										
Other Cancer	7		3		4		3		5	
Ovarian Cancer	2		1	2		1		1	1	
Pancreatic Cancer	2		1		2		2		2	
Prostate Cancer										
Sarcoma	7			2	6					1
Significant Unique Analyses	86	6	20	16	75	3	47	3	32	9
Total Unique Analyses	400		465		417		308		430	

1 5 10   10 5 1

%

Cell color is determined by the best gene rank percentile fo analyses within the cell.

NOTE: An analysis may be counted in more than one cancer

Fig. 9 mRNA expression of the hub genes in 20 different types of cancer

3.7 Validation in HPA and Genetic information of the hub genes

The HPA online tool was used to analyze the protein expression of COL1A1, THBS1, COL5A2, COL12A1, and CXCR4 (Fig. 10). The results showed that the COL1A1 protein gene was not expressed in normal gastric tissues and was low expressed in GC tissues. THBS1 protein gene was highly expressed in normal gastric tissues, but not expressed in GC tissues. The COL12A1 protein gene was low expressed in normal gastric tissues, but not expressed in GC tissues. There is no pathological map of COL5A2, CXCR4 expression in the HPA database. Then we drew an enrichment analysis circle diagram of the hub genes (Fig. 11). The cBioPortal tool showed that 101 of 393 patients with gastric adenocarcinoma (26%) had genetic mutations in these five genes (Fig.12A). Overview of genetic variation of 5 hub genes was also analyzed

[在此处键入]

(Fig.12B). Fig.12C showed the correlation between COL12A1 mRNA and DNA methylation. There is no data plot between COL1A1, THBS1, COL5A2, CXCR4 mRNA, and DNA methylation in the database.
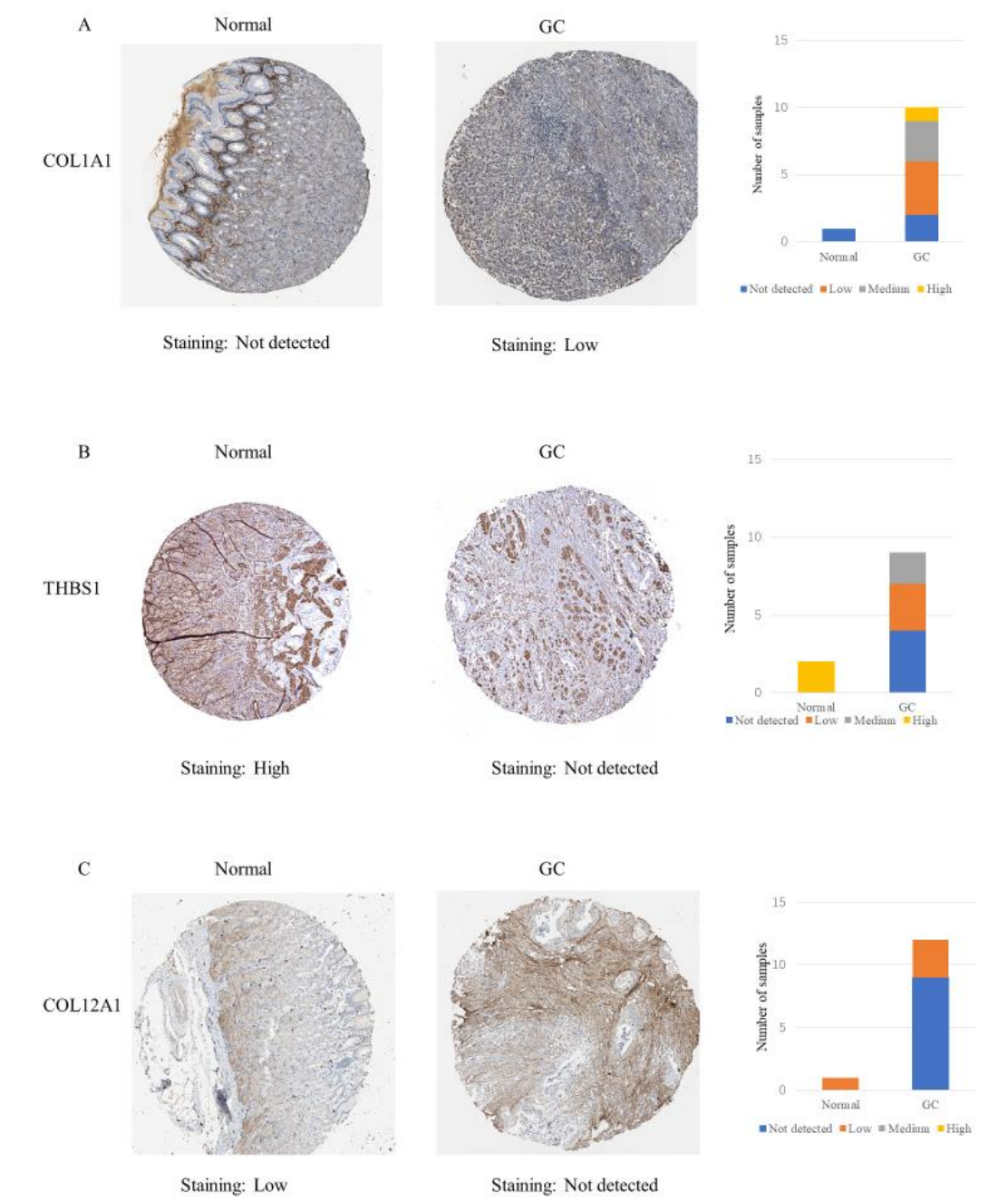

Fig. 10 Validation of the hub genes via the HPA

Representative immunohistochemistry images of (A) COL1A1, (B) THBS1, and (C) COL12A1 in GC and noncancerous stomach tissues derived from the HPA database. The staining strengths were annotated as Not detected, Low, Medium and High. The bar plots indicated the number of samples with different

staining strength.

Fig. 11 Enrichment analysis circle diagram of hub genes

In order to explore the functions of 5 hub genes in GC, we conducted enrichment analysis about them. The results showed that the 5 hub genes BP mainly include collagen fibril organization，collagen catabolic process.CC mainly include endoplasmic reticulum lumen，extracellular matrix. MF include extracellular matrix structural constituent. The enriched pathways mainly include ECM-receptor interaction，Protein digestion and absorption，Focal adhesion and PI3K-Akt signaling pathway.

Fig. 12 Genetic information of hub genes

Fig. 12A Data from TCGA of gastric adenocarcinoma showed that 101 of 393 patients (26%) had genetic mutations in these 5 genes. Fig.12B showed overview of genetic variation of 5 genes. Fig.12C showed the correlation between COL12A1 mRNA and DNA methylation. There was no data plot between COL1A1, THBS1, COL5A2, CXCR4 mRNA and DNA methylation in the database.

4. Discussion

In recent years, despite the continuous development of the treatment process of GC, the treatment and prognosis of GC are still poor due to the lack of difficulties in early diagnosis. Explaining the potential molecular mechanism of GC will help the early diagnosis and prognosis of GC. Bioinformatics analysis is increasingly used to screen possible target biomolecules that have a guiding role in tumor diagnosis and treatment [22].

In this study, bioinformatics tools were used to analyze gene expression data set GSE118916 and methylation data set GSE25869, and 110 Hypo-HGs were obtained, 9 up-regulated hypomethylated oncogenes. 23 Hyper-LGs and 2 down-regulated hypermethylated TSGs. The PPI network showed FN1, COL3A1, COL1A1, COL1A2, MMP2 in Hypo-HGs, CDH1, FOXA1, and KLF4 in Hyper-LGs were at the core. GO analysis results showed that the biological process of Hypo-HGs mainly involved cell

[在此处键入]

adhesion and extracellular matrix organization. Cell adhesion is involved in the pathological and physiological processes of a variety of tumor cells [23], and changes in cell-cell adhesion and cell-matrix adhesion can promote cancer cell metastasis [24]. Intercellular adhesion molecule-1 (ICAM-1) is a member of the immunoglobulin superfamily of adhesion molecules (IGSF). It is a key protein for intercellular signal communication which is related to a variety of pathological processes. When the body has inflammation, infection, and immunity under stress and other conditions, ICAM-1 can be over-activated and expressed, and participates in regulating the immune response of the body's cells [25]. High-level expression of ICAM-1 can be detected in GC cells with a high metastasis rate, which shows that the expression of ICAM-1 is significantly related to the invasion and metastasis of GC, and can be effectively used for clinical monitoring of gastric cancer's blood-borne lymphoid transfer [26]. The extracellular matrix is a key component that plays an active role in all cancer characteristics [27] and mediates cell-microenvironment interactions [28]. The biological process of Hyper-LGs mainly involved response to nicotine and xenobiotic metabolic process. Nicotine can significantly up-regulate the expression of matrix metalloproteinase 7 (MMP7), and high expression of MMP7 has been shown to play a key role in cancer invasion, and smoking addiction increases the risk of GC [29]. The xenobiotic metabolic process may regulate the sensitivity of GC [30]. KEGG analysis results showed that Hypo-HGs were significantly enriched in Focal adhesion, PI3K-Akt signaling pathway, and ECM-receptor interaction. Studies have found that focal adhesion was involved in the occurrence and metastasis of GC, and calcium release-activated calcium regulation 2 (ORAI2) promotes the tumorigenicity and metastasis of GC through PI3K/Akt signal transduction and MAPK-dependent focal adhesion decomposition [31]. The PI3K-Akt pathway is widely distributed in various cells and can regulate a variety of biological behaviors of cells [32]. The abnormal PI3K-Ak pathway may trigger the occurrence and development of cancer [33]. Studies have shown that it can promote the proliferation of GC cells and inhibit cell apoptosis, which is closely related to the invasion and metastasis of GC cells [34]. ECM is an important

[在此处键入]

part of the tumor microenvironment [35]. The ECM-receptor interaction plays a crucial role in many cancers [36, 37]. Hyper-LGs were significantly enriched in Drug metabolism-cytochrome P450, Chemical carcinogenesis, and Metabolism of xenobiotics by cytochrome P450. Cytochrome P450 family genes participate in the development of GC through the xenobiotic metabolism of cytochrome P450 [38]. Overexpression of cytochrome P450 family 2 subfamily E polypeptide 1 (CYP2E1) promotes the proliferation and invasion of GC, and inhibits their apoptosis [39]. Studies have proved the dose-dependent enhancement of salt in the gastric chemical carcinogenesis of Mongolian gerbils infected with Helicobacter pylori (HP). Reducing salt intake may be one of the most important chemoprevention methods for human GC [40]. Therefore, the occurrence and development of GC are related to the activation of these abnormal pathways, and detecting these abnormal signal pathways can accurately predict the occurrence and development of tumors.

GEPIA database verified FN1, COL3A1, COL1A1, COL1A2, MMP2, FBN1, SPARC, THBS1, COL5A2, CDH1, FOXA1, KLF4, AZGP1, GC, CYP2C9, MGST1, PXMP2, SLC16A7, TMEM45B, COL12A1, TAC1, TWIST1, UCHL1, GREM1, MEF2C, MAFB, HHEX, and CXCR4 28 genes, the results showed that COL3A1, COL1A2, SPARC, CDH1, TMEM45B were highly expressed in GC tissues, PXMP2 was low in GC tissues; Most of the genes have been studied in GC. It has been confirmed that COL3A1 was overexpressed in other bladder cancers and glioblastomas [41, 42], but the mechanism of influence in GC has not been fully understood. COL1A2 is related to the invasion and metastasis of GC [43]. The high expression of COL1A2 may indicate a poor clinical prognosis in patients with GC [44]. The high SPARC expression increases tumor cell activity and enhances epithelial-mesenchymal transition and angiogenesis [45]. Pathogenic mutations and germline deletions of CDH1 are important pathogenic factors for early-onset diffuse GC [46]. TMEM45B is abnormally expressed in many types of tumors. Knockdown of TMEM45B can inhibit the JAK2/STAT3 signaling pathway, thereby inhibiting the proliferation, migration, and invasion of GC cells [47]. PXMP2 plays a vital role in lipid and reactive oxygen metabolism, which is

a new target for studying depression [48], but its effect in tumors has not been reported. Survival analysis showed that patients with high expression of COL1A1, THBS1, COL5A2, COL12A1, CXCR4 were associated with shorter overall survival (P＜0.05). Oncomine database verified the expression of 5 hub genes in GC, and the results showed that COL1A1, THBS1, COL5A2, COL12A1 were expressed in GC and had significant statistical significance. Although CXCR4 was not expressed in GC, 41 studies have shown that CXCR4 was expressed in other tumors. The HPA online database analyzed the protein expression of 5 hub genes, and the verification results were similar to the above. Enrichment analysis of 5 hub genes found that BP mainly included fibril organization and collagen catabolic process. KEGG mainly included ECM-receptor interaction, Protein digestion and absorption, Focal adhesion, and PI3K-Akt signaling pathway. Most of them have been analyzed above. cBioPortal showed that 26% of patients in the gastric adenocarcinoma TCGA database have genetic mutations in these five genes. These findings supported the accuracy of our bioinformatics analysis to a certain extent.

COL1A1, COL5A2, COL12A1 belong to the collagen-forming gene family [49], and each collagen is composed of 3 polypeptide chains numbered with Arabic numerals [50]. The collagen-forming gene family is involved in the formation of collagen in extracellular matrix proteins [51] and overexpressed in a variety of cancers [52]. Collagen is the main component of the extracellular matrix of GC cells and the main component of the mesenchymal microenvironment. It can induce tumor cell migration [53]. When GC occurs, collagen synthesis increases and induces epithelial-mesenchymal transition, leading to tumor cell infiltration and metastasis [54]. The expression of COL1A1 is higher in GC tissues than normal tissues [55], which is related to the prognosis of GC patients [56]. COL5A2 is related to the pathological process of osteosarcoma [57], bladder cancer [58], and GC [59]. COL12A1 is known to be abnormally expressed in connective tissue diseases, and COL12A1 mutations are associated with poor prognosis [60]. COL12A1 is highly expressed in intestinal diffuse GC which is associated with poor OS and PFS [61]. Studies have found that THBS1

[在此处键入]

mutations are related to early GC. THBS1 may become a new prognostic target of GC by affecting tumor purity, TMB, TME score, and multiple oncogenic signaling pathways [62]. CXCR4 is related to the aggressiveness of GC [63] and lymph node metastasis [64]. Blocking CXCR4 can inhibit the growth and invasion of GC cells [65]. Consistent with this study, CXCR4 may potentially trigger the occurrence of GC by up-regulating gene expression after methylation.

5. Conclusion

This study combined the gene expression microarrays and gene methylation microarrays, conducted a comprehensive bioinformatics analysis, and identified abnormally methylated and differentially expressed tumor-promoting genes and TSGs in GC tissues, as well as related Function and pathways. The five hub genes verified by TCGA, Oncomine, HPA, and cBioPortal databases include COL1A1, THBS1, COL5A2, COL12A1, CXCR4, which can be used as targets for precise diagnosis and treatment of GC. However, due to the limitations of bioinformatics, further experimental studies are needed to verify its molecular mechanism in detail.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

Peng Shu designed the research. Siyuan Song analyzed the data and wrote the paper. All authors read and approved the submitted version.

Acknowledgments

[在此处键入]

Abbreviations

GC: gastric cancer

DEG: differentially expressed gene

DMG: differentially methylated gene

TSGs: tumor suppressor genes

GEO: gene expression omnibus

GEPIA: gene expression profiling interactive analysis

BP: biological process

CC: cellular component

MF: molecular function

Hypo-HGs: hypomethylated and highly expressed genes

Hyper-LGs: hypermethylated and lowly expressed genes

PPI: protein–protein interaction

STRING: search tool for the retrieval of interacting genes

OS: overall survival

PFS: progression-free survival

TMB: tumor mutation burden

TME: tumor microenvironment

References

[1] Smyth E, Nilsson M, Grabsch H, et al. Gastric cancer. Lancet 2020;10251(396):635-648.

[2] Crew K D, Neugut A I. Epidemiology of gastric cancer. World Journal of Gastroenterology 2006;12(3):354-362.

[在此处键入]

[3] De Martel C, Forman D, Plummer M. Gastric cancer: epidemiology and risk factors. Gastroenterology Clinics of North America 2013;42(2):219-240.

[4] Yusefi A R, Lankarani K B, Bastani P. Risk factors for gastric cancer: a systematic review. Asian Pacific Journal of Cancer Prevention 2018;19(3):591-603.

[5] Song X, Qi W, Guo J, et al. Immune checkpoint inhibitor combination therapy for gastric cancer: Research progress. Oncol Lett 2020;20(4):46.

[6] Thrift A, El-Serag H. Burden of Gastric Cancer. Clin Gastroenterol Hepatol 2019; 2:124.

[7] Li G, Hu Y, Liu H. Current status of randomized controlled trials for laparoscopic gastric surgery for gastric cancer in China. Asian J Endosc Surg 2015;8(3):263-267.

[8] Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 2002;16(1):6-21.

[9] Liang G, Weisenberger D. DNA methylation aberrancies as a guide for surveillance and treatment of human cancers. Epigenetics 2017; 12:416-432.

[10] Amini M, Hejazi M, Ghorban K, et al. Identification of functional methylated CpG loci in PD-L1 promoter as the novel epigenetic biomarkers for primary gastric cancer. Gene 2020; 23:772.

[11] Yan H, Chen W, Ge K. Value of Plasma Methylated SFRP2 in Prognosis of Gastric Cancer. Dig Dis Sci 2020; 10:14-18.

[12] Kulasingam V, Diamandis E. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat Clin Pract Oncol 2008;5(10):588-599.

[13] M L, X W, J L, et al. Identification of Core Prognosis-Related Candidate Genes in Chinese Gastric Cancer Population Based on Integrated Bioinformatics. Biomed Res Int 2020; 10:115.

[14] Li H, Liu J, Liu S. Bioinformatics-Based Identification of Methylated-Differentially Expressed Genes and Related Pathways in Gastric Cancer. Dig Dis Sci 2017; 62(11):3029-3039.

[15] Law C, Alhamdoosh M, Su S. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Res 2016; 5:1408.

[16] Chen H, Boutros P. (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 2011; 12:35.

[17] Dennis G J, Sherman B, Hosack D. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003; 4(5): P3.

[18] Consortium G O. The gene ontology (GO) project in 2006. Nucleic Acids Res 2006; 34: D322-

[在此处键入]

D326.

[19] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000; 28(1):27-30.

[20] Liu Y, Cui S, Li W. PAX3 is a biomarker and prognostic factor in melanoma: Database mining. Oncol Lett 2019; 17:4985-4993.

[21] Lin C, Zhang Y, Zhang K. Fever promotes T lymphocyte trafficking via a thermal sensory pathway involving heat shock protein 90 and α4 integrins. Immunity 2019; 50(1):137-151.

[22] Kulasingam V, Diamandis E. (2008) Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat Clin Pract Oncol; 5:588-599.

[23] Wu X, Liu GS, Luo Y. Expression and significance of cell adhesion factor-1 in lung tissue of neonatal rat model of hyperoxic lung injury. Chinese Journal of Neonatology 2012; 8(2):124.

[24] Liu X, Chu K. E-cadherin and gastric cancer: Cause, consequence, and applications. Biomed Res Int 2014; 2014:637308.

[25] Xin X, He HP, Liu Y. Detection of serum adhesion molecules, high-sensitivity C-reactive protein and homocysteine in patients with hyperlipidemia and its clinical significance. Modern Medicine 2014; 42(1):38-41.

[26] Ke H, Chen YH. The clinical significance of ICAM-1 expression in childhood tumors. The Practical Journal of Cancer 2013; 8(4):350-352.

[27] Andreuzzi E, Capuano A, Poletto E. Role of extracellular matrix in gastrointestinal cancer-associated angiogenesis. International Journal of Molecular Sciences 2020; 21(10):3686.

[28] Insua-Rodríguez J, Oskarsson T. The extracellular matrix in breast cancer. Adv. Drug. Deliv. Rev 2016; 97:41-55.

[29] Kesh K, Subramanian L, Ghosh N. Association of MMP7-181A→G Promoter Polymorphism with Gastric Cancer Risk. J Biol Chem 2015; 290(23):14391-14406.

[30] de Castro A, Fernandes M, de Carvalho D. Polymorphisms of xenobiotic-metabolizing and transporter genes, and the risk of gastric and colorectal cancer in an admixed population from the Brazilian Amazon. Am J Transl Res 2020; 12(10):6626-6636.

[31] S W, M C, J H. ORAI2 Promotes Gastric Cancer Tumorigenicity and Metastasis Through PI3K/Akt Signaling and MAPK-dependent Focal Adhesion Disassembly. Cancer Res 2020; 49.

[在此处键入]

[32] Osaki M, Oshimura M, Ito H. PI3K-Akt pathway: its functions and alterations in human cancer. Apoptosis 2004; 9(6):667-676.

[33] Juric D, Krop I, Ramanathan R. Phase I dose-escalation study of taselisib, an oral PI3K inhibitor, in patients with advanced solid tumors. Cancer Discovery; 2017,7(7):704-715.

[34] Li D, Gao SY. Research progress of PI3K/Akt and gastric cancer. Journal of Qiqihar Medical College 2020; 41(1):90-94.

[35] Reinhard J, Brösicke N, Theocharidis U. The extracellular matrix niche microenvironment of neural and cancer stem cells in the brain. Int. J. Biochem. Cell Biol 2016; 81:174-183.

[36] Rahbari N, Kedrin D, Incio J, et al. Anti-VEGF therapy induces ECM remodeling and mechanical barriers to therapy in colorectal cancer liver metastases. Sci Transl Med 2016; 8:135r-360r.

[37] Bao Y, Wang L, Shi L, et al. Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. Cell Mol Biol Lett 2019; 24:38.

[38] Zhang W, Liu S, Zhan H, et al. Transcriptome sequencing identifies key pathways and genes involved in gastric adenocarcinoma. Molecular Medicine Reports 2018; 18.

[39] Wang R, Chen X, Zhang W. CYP2E1 changes the biological function of gastric cancer cells via the PI3K/Akt/mTOR signaling pathway. Mol Med Rep 2020; 21(2):842-850.

[40] Kato S, Tsukamoto T, Mizoshita T. High salt diets dose-dependently promote gastric chemical carcinogenesis in Helicobacter pylori-infected Mongolian gerbils associated with a shift in mucin production from glandular to surface mucous cells. Int J Cancer 2006; 7(119):1558-1566.

[41] Gao Y, Mao X, Zhu T. COL3A1 and SNAP91: novel glioblastoma markers with diagnostic and prognostic value. Oncotarget 2016; 7:70494-70503.

[42] Vastrad B, Vastrad C, Godavarthi A. Molecular mechanisms underlying gliomas and glioblastoma pathogenesis revealed by bioinformatics analysis of microarray data. Med Oncol 2017; 34:182.

[43] Yasui W, Oue N, Ito R. Search for new biomarkers of gastric cancer through serial analysis of gene expression and its clinical implications. Cancer Science 2004; 95(5):385-392.

[44] Rong L, Huang W, Tian S. COL1A2 is a Novel Biomarker to Improve Clinical Prediction in

[在此处键入]

Human Gastric Cancer: Integrating Bioinformatics and Meta-Analysis. Pathol. Oncol. Res 2018; 24:129-134.

[45] Atorrasagasti C, Onorato A, Gimeno M. SPARC is required for the maintenance of glucose homeostasis and insulin secretion in mice. Clin Sci 2019; 133:351.

[46] García-Ruvalcaba A, Rizo De La Torre L, Magaña-Torres M. Predicting Pathogenicity of CDH1 Gene Variants in Patients with Early-onset Diffuse Gastric Cancer from Western Mexico. Rev Invest Clin; 2021.

[47] Shen K, Yu W, Yu Y. Knockdown of TMEM45B inhibits cell proliferation and invasion in gastric cancer. Biomed Pharmacother 2018; 104:576-581.

[48] Ciobanu L, Sachdev P, Trollor J. Differential gene expression in brain and peripheral tissues in depression across the life span: A review of replicated findings. Neurosci Biobehav Rev 2016; 71:281-293.

[49] Huang R, Gu W, Sun B. dentification of COL4A1 as a potential gene conferring trastuzumab resistance in gastric cancer based on bioinformatics analysis. Mol Med Rep 2018; 17(5):6387-6396.

[50] Ricard-Blum S. The collagen family. Cold Spring Harb Perspect Biol 2011; 3(1): a4978.

[51] Li Z, Liu Z, Shao Z. Identifying multiple collagen gene family members as potential gastric cancer biomarkers using integrated bioinformatics analysis. Peer J 2020; 8: e9123.

[52] Liu X, Xu Y, Zhou Q. PI3K in cancer: its structure, activation modes and role in shaping tumor microenvironment. Future Oncology 2018; 14(7):665-674.

[53] Zhou Z, Ji C, Xiao H. Reorganized Collagen in the Tumor Microenvironment of Gastric Cancer and Its Association with Prognosis. J Cancer 2017; 8(8):1466-1476.

[54] Jang M, Koh I, Lee S. Droplet-based microtumor model to assess cell-ECM interactions and drug resistance of gastric cancer cells. Sci Rep 2017; 27(7):41541.

[55] Wang Q, Yu J. MiR-129-5p suppresses gastric cancer cell invasion and proliferation by inhibiting COL1A1. Biochem Cell Biol 2018; 96(1):19-25.

[56] Li J, Ding Y, Li A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. World J Surg Oncol 2016; 14(1):297.

[57] Wu D, Chen K, Bai Y. Screening of diagnostic markers for osteosarcoma. Mol Med Rep 2015; 10(5):2415-2420.

[在此处键入]

[58] Zeng X, Liu X, Liu T. The clinical significance of COL5A2 in patients with bladder cancer: A retrospective analysis of bladder cancer gene expression data. Medicine (Baltimore) 2018; 97(10): e91.

[59] Wu K, Xu X, Zhan C. dentification of a nine-gene prognostic signature for gastric carcinoma using integrated bioinformatics analyses. World Journal of Gastrointestinal Oncology 2020; 12(9):975-991.

[60] Punetha J, Kesari A, Hoffman E. Novel COL12A1 variant expands the clinical picture of congenital myopathies with extracellular matrix defects. Muscle Nerve 2017; 55:277-281.

[61] Jiang X, Wu M, Xu X. COL12A1, a novel potential prognostic factor and therapeutic target in gastric cancer. Mol Med 2019; 20(4):3103-3112.

[62] Li Y, Wang J, Zhang T. Identification of New Therapeutic Targets for Gastric Cancer With Bioinformatics. Front Genet 2020; 11:865.

[63] Zheng S, Shi L, Zhang Y. Expression of SNCG, MAP2, SDF-1 and CXCR4 in gastric adenocarcinoma and their clinical significance. Int J Clin Exp Pathol 2014; 7(10):6606-6615.

[64] Ying J, Xu Q, Zhang G. The expression of CXCL12 and CXCR4 in gastric cancer and their correlation to lymph node metastasis. Medical oncology (Northwood, London, England) 2012; 29(3):1716-1722.

[65] Lin X, Xu Q, Tang L. Regorafenib inhibited gastric cancer cells growth and invasion via CXCR4 activated Wnt pathway. PloS one 2017; 12(5): e177335.
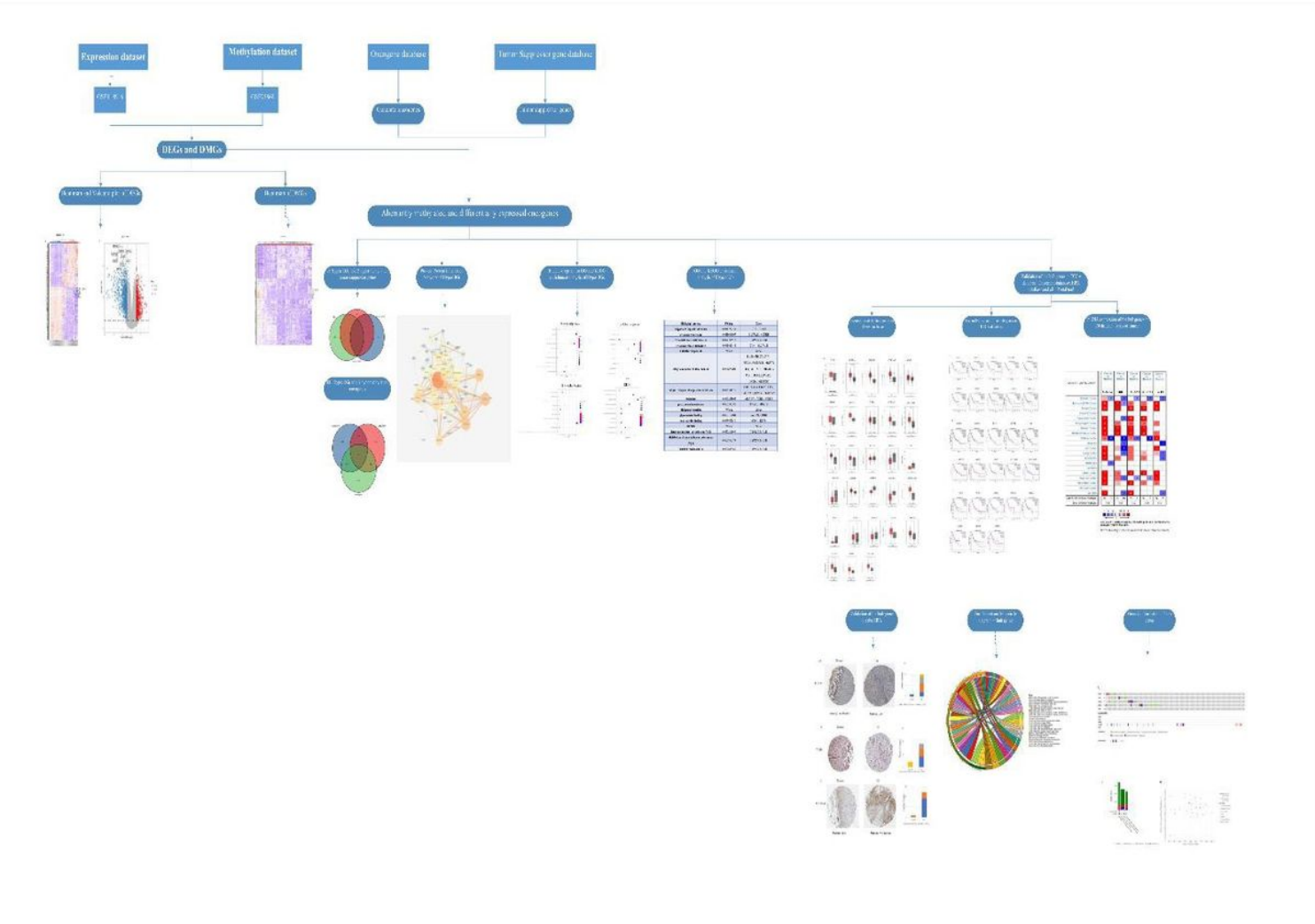
[在此处键入]

Figures

Figure 1
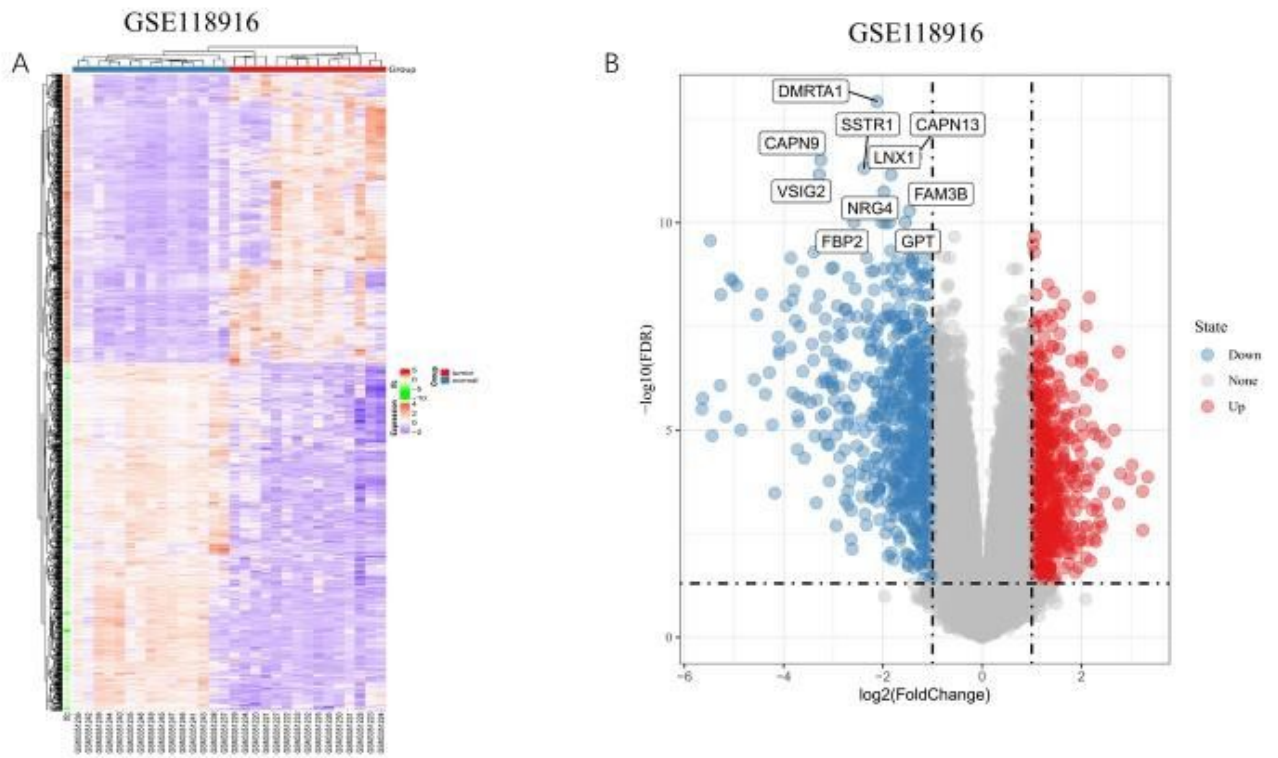
The protocol of our study procedures

Figure 2

Heat map and Volcano plot of DEGs. 2A The hierarchical clustering of the heat map reveals the DEGs in the GC. Orange and purple indicate higher and lower expression levels, respectively. 2B The volcano map visualizes all DEGs. The red dots represent up-regulated genes, the blue dots represent down-regulated genes, and the gray dots represent genes that are not differentially expressed.
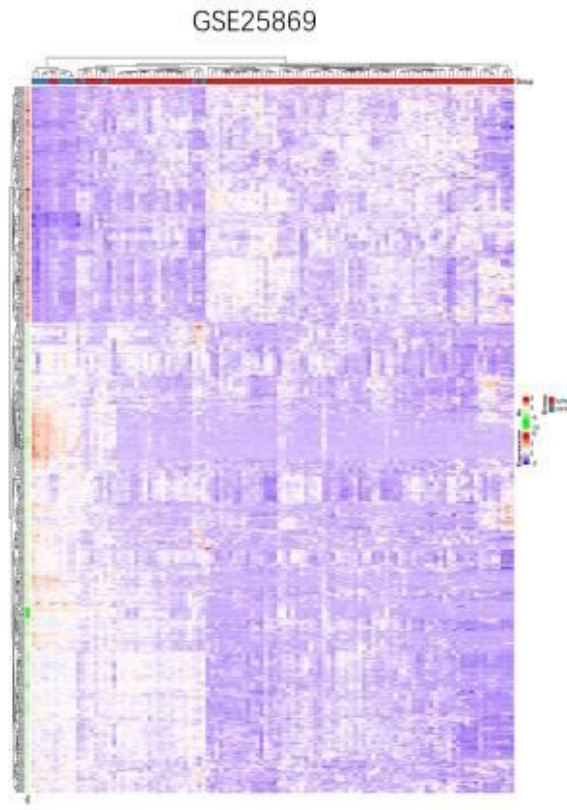
GSE25869

Figure 3

Heat map of DMGs. The hierarchical clustering of the heat map reveals the DMGs in the GC. Orange and purple indicate higher and lower expression levels, respectively.
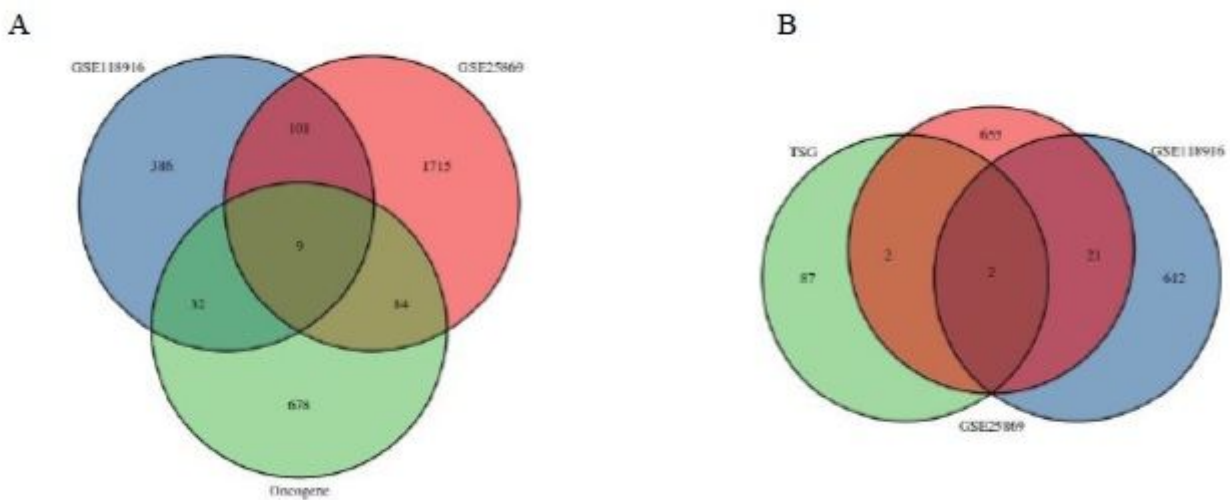

Figure 4

Aberrantly methylated and differentially expressed oncogenes AB. 4A: 110 hypomethylated and up-regulated genes were identified, including 9 up-regulation hypomethylation oncogenes. 4B: 23

hypermethylated and down-regulated genes were identified, including 2 hypermethylation tumor suppressor genes (TSGs).
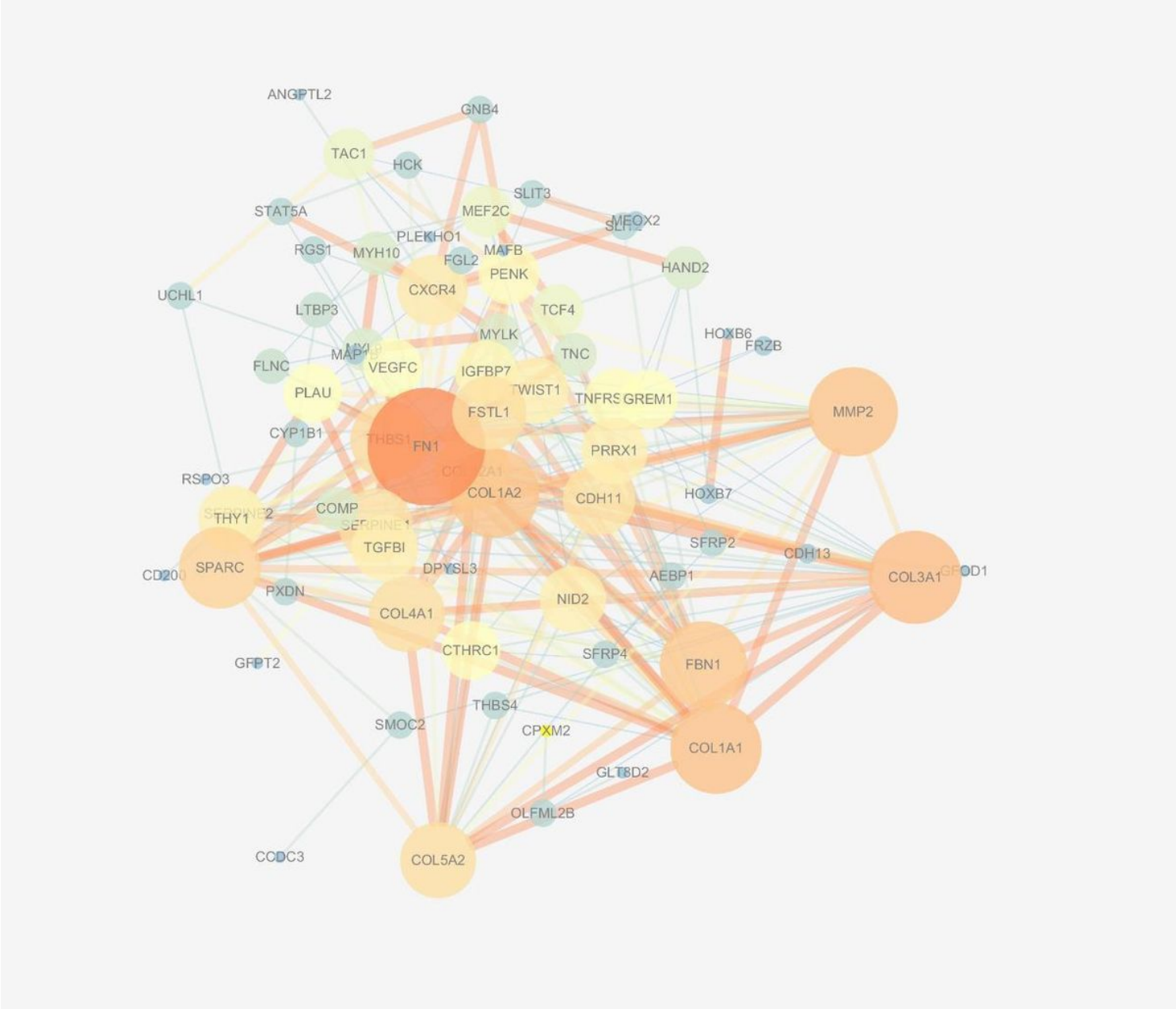

Figure 5

Protein-Protein Interaction Network of Hypo-HGs. A larger Degree or combined-score value indicates a larger node size and deeper node color.
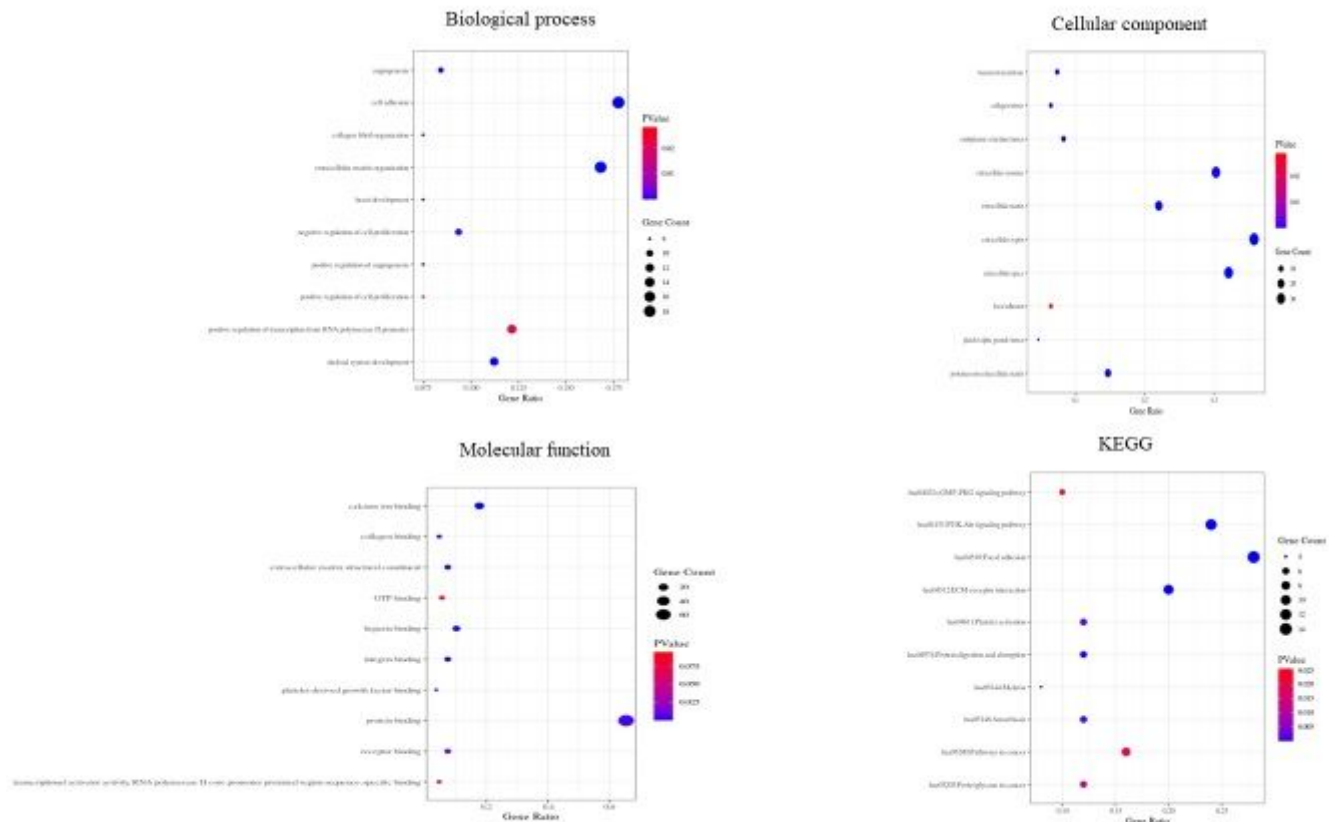
Figure 6

Bubble diagram for GO and KEGG enrichment analysis of Hypo-HGs. The size of the bubble in the figure represents the number of enriched genes, and the difference in bubble color represents the significance of gene enrichment.
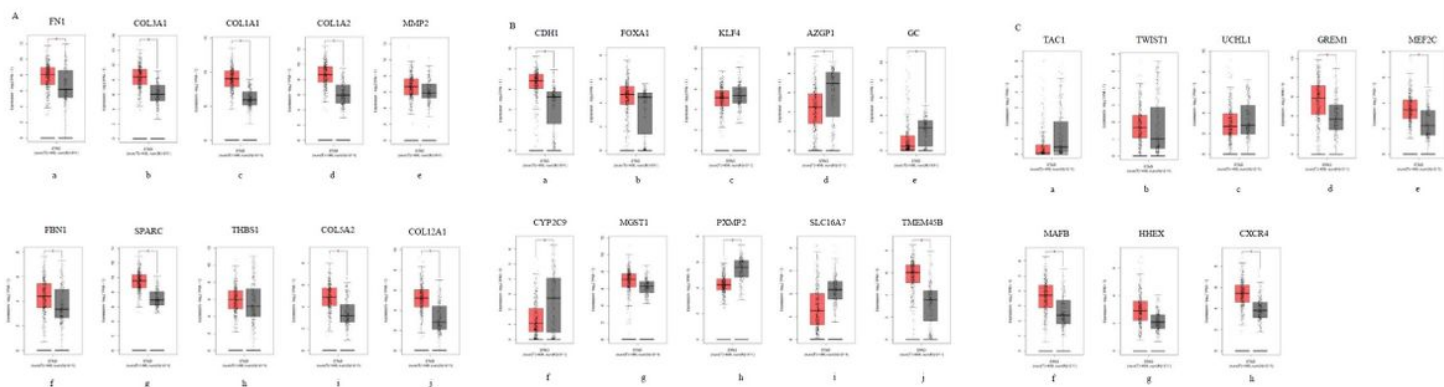

Figure 7

Expression of the hub genes in TCGA database The box plots showed that the genes expression of (Aa)FN1, (Ab)COL3A1, (Ac)COL1A1, (Ad)COL1A2, (Ae)MMP2, (Af)FBN1, (Ag)SPARC, (Ah)THBS1, (Ai)COL5A2, (Aj)COL12A1, (Ba)CDH1, (Bb)FOXA1, (Bc)KLF4, (Bd)AZGP1, (Be)GC, (Bf)C P2C9,

(Bg)MGST1, (Bh)PXMP2, (Bi)SLC16A7, (Bj)TMEM45B, (Ca)TAC1, (Cb)TWIST1, (Cc)UCHL1, (Cd)GREM1, (Ce)MEF2C, (Cf)MAFB, (Cg)HHEX, and (Ch)CXCR4 in GEPIA. Red represents Tumor, Gray represents normal. (Ab)COL3A1, (Ad)COL1A2, (Ag)SPARC, (Ba)CDH1, (Aj)TMEM45B were highly expressed in GC tissues, and PXMP2 was lowly expressed in GC tissues.
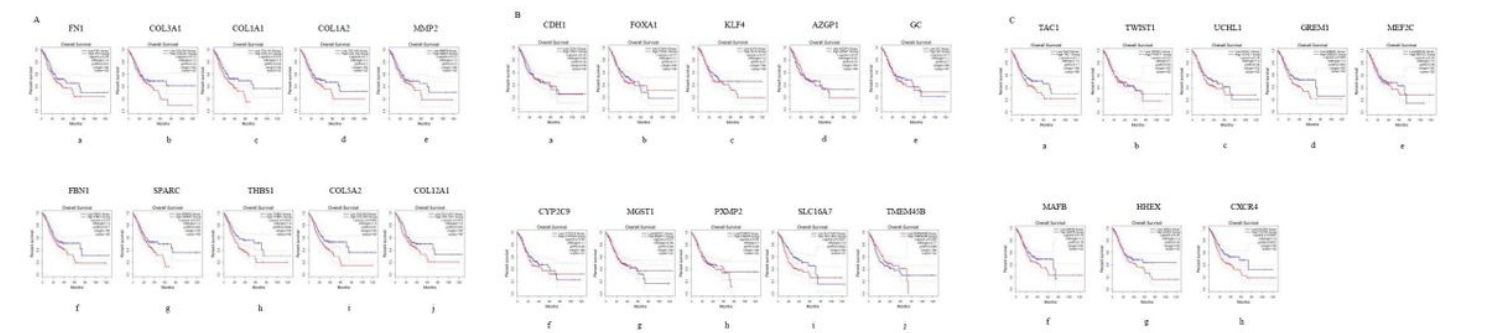

Figure 8

Overall Survival of the hub genes in TCGA database The line charts showed that the OS of hub gene (Aa)FN1, (Ab)COL3A1, (Ac)COL1A1, (Ad)COL1A2, (Ae)MMP2, (Af)FBN1, (Ag)SPARC, (Ah)THBS1, (Ai)COL5A2, (Aj)COL12A1, (Ba)CDH1, (Bb)FOXA1, (Bc)KLF4, (Bd)AZGP1, (Be)GC, (Bf)C P2C9, (Bg)MGST1, (Bh)PXMP2, (Bi)SLC16A7, (Bj)TMEM45B, (Ca)TAC1, (Cb)TWIST1, (Cc)UCHL1, (Cd)GREM1, (Ce)MEF2C, (Cf)MAFB, (Cg)HHEX, and (Ch)CXCR4 in GEPIA. The survival curve comparing the patients with high (red) and low (blue) expression in GC. Prognostic value of (Ac)COL1A1, (Ah)THBS1, (Ai)COL5A2, (Aj)COL12A1, (Ch)CXCR4 had a significant difference P 0.05⬚⬚. Although the p value was not statistically significant, from the image trend point of view, patients with high expression of (Aa)FN1, (Ad)COL1A2, (Ae)MMP2, (Af)FBN1, (Ag)SPARC, (Ca)TAC1, and (Cd)GREM1 were associated with shorter OS.

Analysis Type by Cancer	Cancer vs. Normal COL1A1		Cancer vs. Normal THBS1		Cancer vs. Normal COL5A2		Cancer vs. Normal COL12A1		Cancer vs. Normal CXCR4	
Bladder Cancer	1	1		1		1		1		1
Brain and CNS Cancer	6		2	1	4		1		6	
Breast Cancer	15		1		10		7		5	
Cervical Cancer										
Colorectal Cancer	11			2	14		13			
Esophageal Cancer	4		2		4		3		3	
Gastric Cancer	8		2		7		7			
Head and Neck Cancer	3		2		7		4		1	
Kidney Cancer	4	4		3	3	1		1	5	
Leukemia				3						1
Liver Cancer	4		1	2	1				3	
Lung Cancer	6			1	7		2		1	2
Lymphoma	7		5		7		5			4
Melanoma		1								
Myeloma										
Other Cancer	7		3		4		3		5	
Ovarian Cancer	2		1	2		1		1	1	
Pancreatic Cancer	2		1		2		2		2	
Prostate Cancer										
Sarcoma	7			2	6					1
Significant Unique Analyses	86	6	20	16	75	3	47	3	32	9
Total Unique Analyses	400		465		417		308		430	

1 5 10   10 5 1

Cell color is determined by the best gene rank percentile fo analyses within the cell.

NOTE: An analysis may be counted in more than one cancer

Figure 9

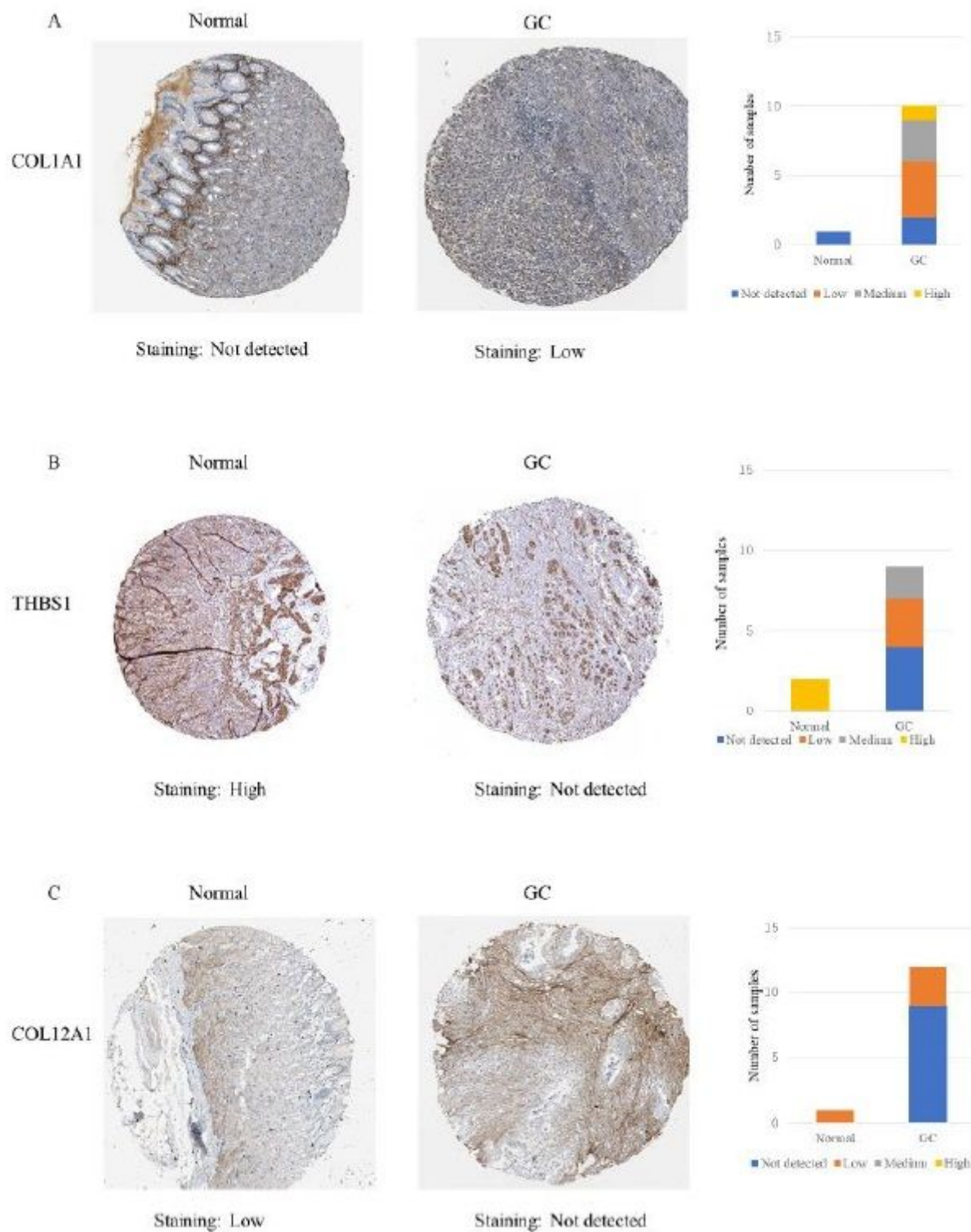mRNA expression of the hub genes in 20 different types of cancer

Figure 10

Validation of the hub genes via the HPA. Representative immunohistochemistry images of (A) COL1A1, (B) THBS1, and (C) COL12A1 in GC and noncancerous stomach tissues derived from the HPA database. The staining strengths were annotated as Not detected, Low, Medium and High. The bar plots indicated the number of samples with different staining strength.
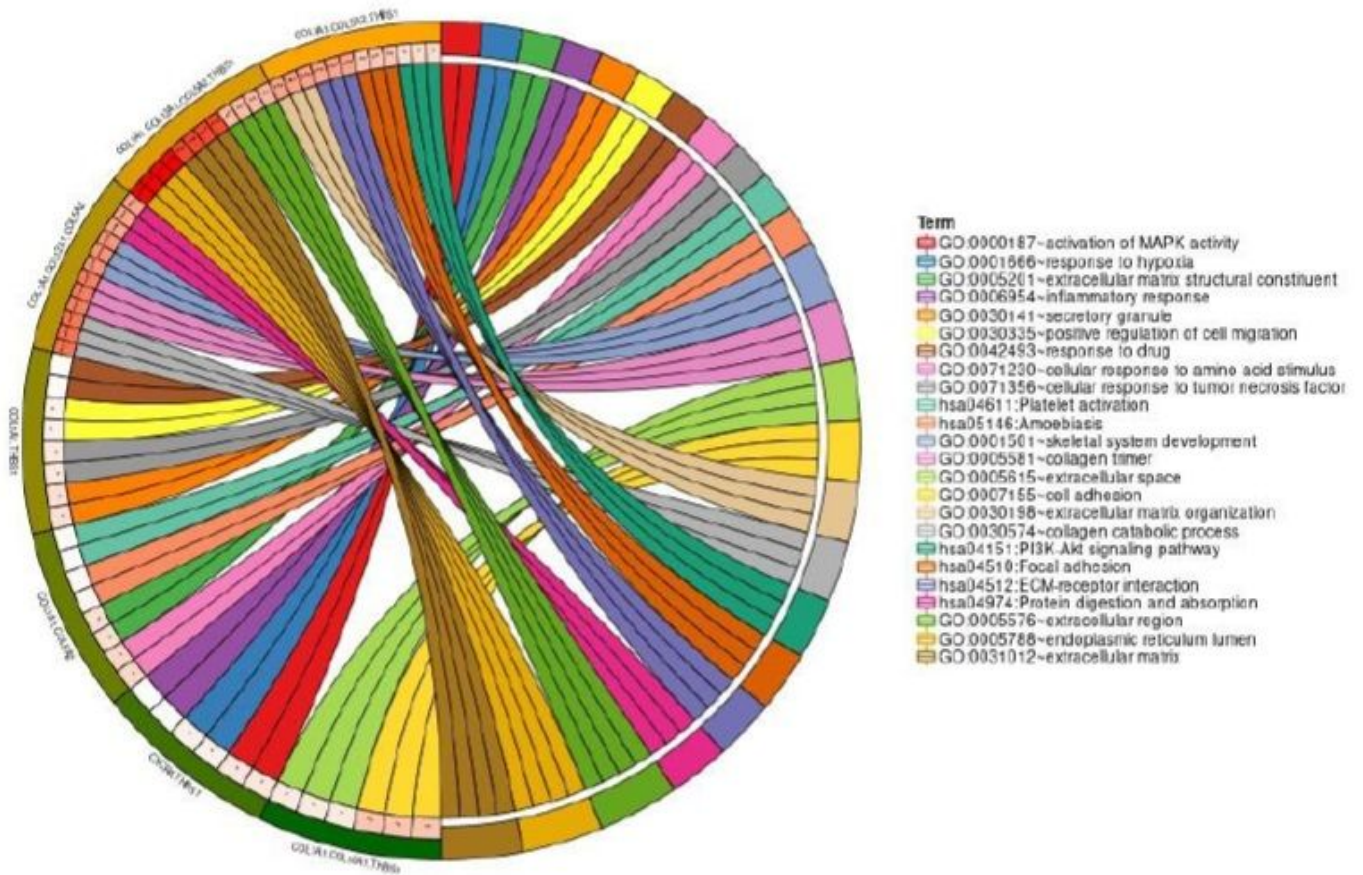
Figure 11

Enrichment analysis circle diagram of hub genes In order to explore the functions of 5 hub genes in GC, we conducted enrichment analysis about them. The results showed that the 5 hub genes BP mainly include collagen fibril organization、collagen catabolic process.CC mainly include endoplasmic reticulum lumen、extracellular matrix. MF include extracellular matrix structural constituent. The enriched pathways mainly include ECM-receptor interaction、Protein digestion and absorption、Focal adhesion and PI3K-Akt signaling pathway.
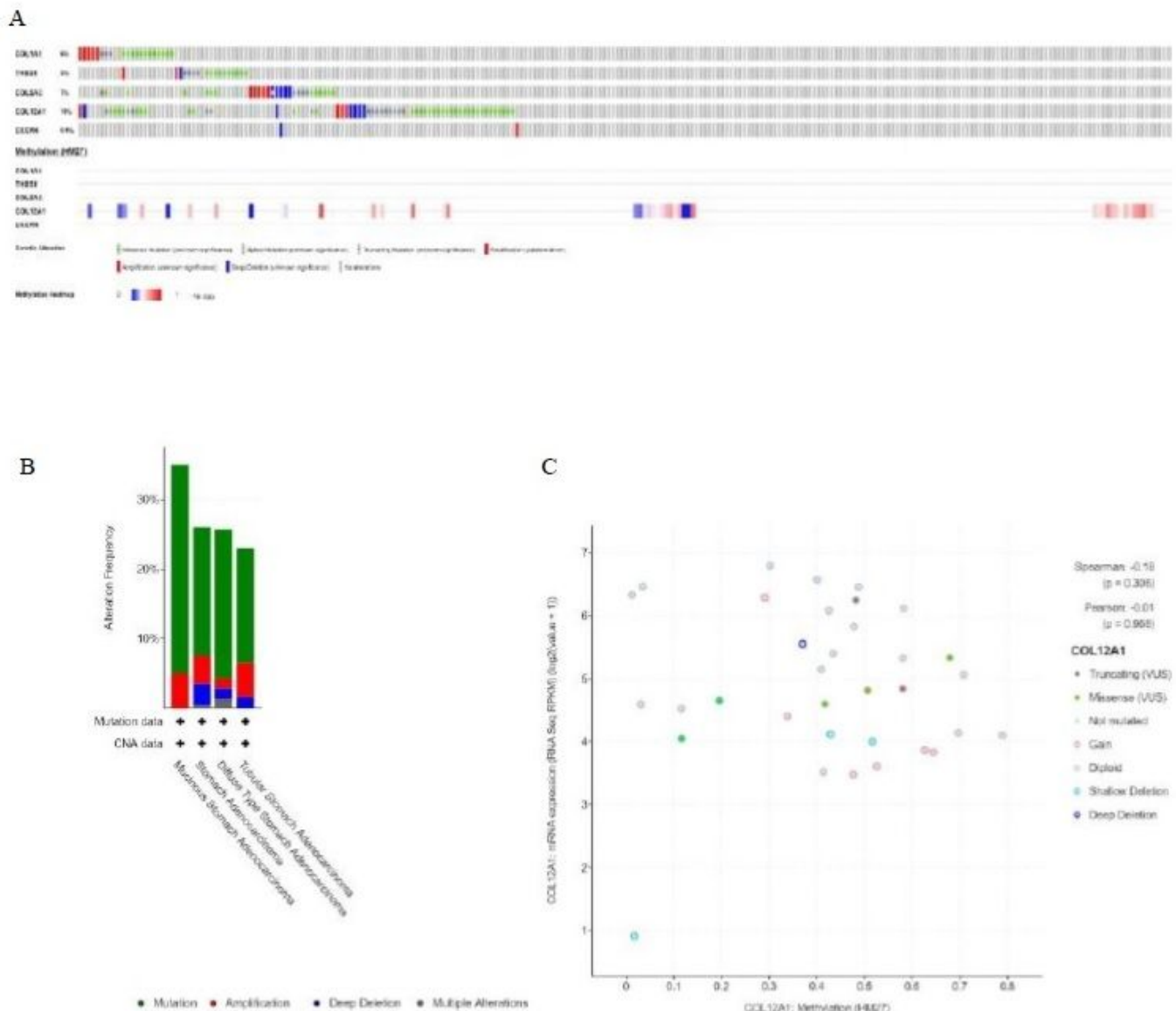
Figure 12

Genetic information of hub genes Fig. 12A Data from TCGA of gastric adenocarcinoma showed that 101 of 393 patients (26%) had genetic mutations in these 5 genes. Fig.12B showed overview of genetic variation of 5 genes. Fig.12C showed the correlation between COL12A1 mRNA and DNA methylation. There was no data plot between COL1A1, THBS1, COL5A2, CXCR4 mRNA and DNA methylation in the database.