Am J Transl Res 2023;15(3):1640-1653
www.ajtr.org /ISSN:1943-8141/AJTR0145137

Original Article

A prognostic model based on the COL1A1-network in gastric cancer

Shiping Liu[1*], Long Chen[2*], Jing Zeng[3], Yuhua Chen[1]

[1]The Chongqing Seventh People's Hospital, Banan District, Chongqing, China; [2]Department of Immunology, Basic Medical College, Chengdu Medical College, Xindu District, Chengdu, Sichuan, China; [3]School of Life Advanced Agriculture Bioengineering, Yangtze Normal University, Chongqing, China. [*]Co-first authors.

Received July 4, 2022; Accepted December 13, 2022; Epub March 15, 2023; Published March 30, 2023

Abstract: Background: Gastric cancer (GC) is one of the most common malignancies worldwide with a poor prognosis due to the lack of early detection and effective treatments. As a biomarker, collagen type I alpha 1 (COL1A1) is often dysregulated in some cancer types. However, the expression profile of COL1A1 and functional mechanism in GC is still unclear. Methods: To screen for the different expression genes of GC vs. adjacent tissues, an RNA-seq dataset containing 30 clinical samples and multi-omics datasets of 478 samples were obtained from Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases, respectively. Then the functional enrichment analysis and survival analysis of dysregulated genes were performed. Furthermore, through constructing the protein-protein interactive network, the function mode of COL1A1 was studied. Finally, a prognostic model was built by least absolute shrinkage and selection operator (LASSO) Cox algorithm to assess the clinical value of COL1A1-network. Results: Firstly, a total of 89 different expression genes (58 down-regulated and 31 up-regulated) that appeared simultaneously in both GEO and TCGA datasets were detected and enriched in some functions regarding the extracellular matrix. However, only 12 genes were significantly correlative with overall survival of GC patients. Among them, ASPN, COL1A1, COL12A1, FNDC1, INHBA and MMP12 could form a network that might activate the epithelial-mesenchymal transition (EMT) pathway. Meanwhile, a prognostic model containing ASPN and INHBA was able to divide GC patients into 2 groups with different risks and predict 5-years survival accurately (AUC = 0.732, 95% CI (0.619, 0.845)). Conclusion: COL1A1 is up-regulated in GC and may result in a poor prognosis with a higher mRNA level. Moreover, the COL1A1-network may promote malignant metastasis via EMT pathway activation and act as a prognostic marker.

Keywords: Gastric cancer, COL1A1, EMT pathway, prognostic model

Introduction

In China, about 380,000 new GC patients are diagnosed annually. The incidence of GC ranks NO. 3 among all malignant tumors [1]. Although the incidence and mortality of GC have been declining recently, GC remains one of the most common and lethal malignancies [2, 3]. Although chemo/radiotherapy and targeted therapy have improved response rates, surgical resection is the most curative therapy currently. Due to the aggressive invasion, most patients are diagnosed in an advanced stage without specific early-stage symptoms. Advanced GC patients are generally incurable with a poor prognosis. The 5-year survival rate is only 20% [4-7]. The invasion and metastasis of tumor cells are the main causes of GC recurrence and death, which also seriously affect the therapeutic effect [8]. A better understanding of the mechanism of GC carcinogenesis and progression is significant for improving early diagnosis and treatment efficacies. In addition, the discovery of key molecules regulating the invasion and metastasis of GC cells is conducive to the design of targeted therapy to improve the prognosis.

Recent evidence has proven that the aberrant activation of epithelial-mesenchymal transition (EMT) plays a crucial role in the genesis, invasion and metastasis of various tumors [9, 10], like GC [11]. EMT is a process with a transformation from epithelial cells into cells with mesenchymal phenotypes, characterized by lost

Table 1. The clinical data of GC patients

Age	Median 43, Range 23-66
Gender	Male (n = 20), Female (n = 10)
Race	Ethnic Han
Stage	I 5 (~17%); II 8 (~27%); III 11 (~37%); IV 6 (~20%)

cellular polarity and adhesion while enhancing invasive and migratory properties. In GC aggressiveness, the EMT was also an important mode of responding to the tumor microenvironment. It might regulate malignant behaviors via EMT [12, 13]. A detailed investigation into the role of EMT in GC could further our understanding of GC initiation, invasion, and metastasis.

Type I collagen (COL-1), found in most connective and embryonic tissue, is an important member of the collagen family [14]. COL-1 has the function of creating matrix structural skeletons and forms the interstitial portion of most solid tumors [15]. Typically, type I collagen is composed of 2 COL1A1 chains and 1 COL1A2 chain [16]. Recently, similar to the other collagen family members, abnormal expression of COL1A1 has been reported in various cancer types [17-19], which was believed to be involved in carcinogenesis [20-22]. However, the regulation of COL1A1 in normal epithelium and tumor lesions of the stomach was rarely mentioned. Although via gene set enrichment analysis (GSEA) analysis, Zhao et al. [23] have demonstrated that in GC, COL1A1 may regulate the EMT pathway. The clinical significance and mechanism of COL1A1 in GC remains unclear. Meanwhile, the regulatory mode of COL1A1 in the EMT pathway also need further study.

In the present study, high-throughput sequencing data of 15 pair's of clinical samples (GC vs. adjacent tissues) were analyzed. The differential expression genes were screened followed by functional enrichment analysis. Through the construction of the PPI protein interaction network, the key pivotal proteins in GC were identified, which were also used to build a prognostic model of GC and preliminarily explore the clinical value and mechanism of COL1A1.

Methods

Sequencing datasets

The public GC sequencing dataset (GSE118916) was obtained from the GEO database. According to the corresponding study [24], tumor and matched normal tissues samples were obtained from the GC patients at the Affiliated Hospital of Xuzhou Medical University in 2014. These tissues were stored at 4°C until full penetration of RNAlater into the tissues and transferred to -80°C for storage. The selection criteria were as follows: (1) The subject presented was diagnosed with GC and had no history of other tumors; (2) Complete demographic and clinical data including age, gender, clinical manifestations, tumor size, the extent of resection, and date of relapse and/or death have been collected. The legal surrogates of those participants provided their written informed consent. This study was approved by the Medical Ethics Committee of the Affiliated Hospital of Xuzhou Medical University. The demographic and clinical features of the patient were summarized in **Table 1**.

Additionally, the RNA-seq data of STADs from the TCGA database were downloaded from the NIH National Cancer Institute GDC Data Portal (https://portal.gdc.cancer.gov/). Overall survival (OS) was identified from the diagnosis date until death or the end of follow up.

Microarray data and enrichment analysis

Total RNA from cells and tissues was isolated using Trizol extractions (Invitrogen). A total of 100 ng of total RNA was amplified using the Ambion® WT Expression Kit (4411973, Life Technologies). Then 5.5 μg of the cDNA was fragmented and labeled with the GeneChip® WT Terminal Labeling kit (901525, Affymetrix). Libraries were sequenced on Illumina HiSeq 2500 using v3 chemistry. Followed by background deletion, quantile normalization, and probe assembly. Differentially expression genes (DEGs) were detected [25]. The *p*-values were adjusted for multiple comparisons using the Benjamini-Hochberg procedure [26]. Genes with adjusted p-value < 0.05 and $|\log FC| \geq 2.0$ were considered as DEGs. Enrichment analysis of DEGs was performed with DAVID [27]. The enriched GO (BP: biological process; CC: cellular component; MF: molecular function) and pathway terms were listed with participant genes [28]. Some other databases used are listed in **Table 2**.

Table 2. List of database

Database ID	URL
GEO Dataset	https://www.ncbi.nlm.nih.gov/gds/?term=
TCGA	https://www.cancer.gov/
Pathview	https://pathview.uncc.edu/
cBioportal of cancer genomics	https://www.cbioportal.org/
TCPA	https://www.tcpaportal.org/tcpa/
DSA	http://cancer.digitalslidearchive.net/
The Human Protein Atlas	https://www.proteinatlas.org/
STRING	https://string-db.org/
GEPIA	http://gepia.cancer-pku.cn/index.html
GO	http://geneontology.org/
CAMOPI	https://www.camoip.net/

The protein-protein interaction network construction

The Retrieval of Interacting Genes (STRING v10) was used to analyze the interactive relationships among DEGs and construct a protein-protein interaction (PPI) network. Only experimentally validated interactions with a combined score (> 0.4) were selected as significant [29]. Cytoscape was used to construct the PPI network and Gephi was used for network visualization. The plug-in Molecular Complex Detection (MCODE) was used to select the prime module from the PPI network. The *p*-values < 0.05 were considered to be significant.

Immunohistochemistry (IHC)

Before IHC staining, GC specimens and normal tissues were fixed with 10% formalin and embedded with paraffin. IHC staining followed a standard protocol to evaluate the expression level of COL1A1 in human tissues. Staining was graded on a scale of 0-3 according to the intensity and the percentage of immune-positive cells.

Survival curves

Overall survival analyses were performed using the R package *survival* [30], and the patients were dichotomized based on the median expression. Kaplan-Meier estimator of survival was used to construct the survival curves. Log-rank tests (corresponding to a two-sided z test) were used to compare overall survival between patients in different groups. The hazard ratio (HR) (95% confidence interval) was provided for comparison of the two groups. The *p*-values were adjusted for multiple testing based on the false discovery rate (FDR) according to the Benjamini-Hochberg method [31]. Proportional hazard assumptions were tested.

Statistical analyses

Available samples from TCGA data were adequate because sufficient power using equivalent tests was observed in a previous study [32]. To test for differential expression across the two groups (tumor and normal), the R package DESeq2 was used on raw count data [33]. For comparison of two patient groups, the two-sided Student's t-test and Wilcoxon-rank sum test was used. Distributions of data are shown either as individual data points, as box-and-whisker plots, or as violin plots. The *p*-values were adjusted according to the Benjamini-Hochberg method.

Results

The DEGs of GC vs. adjacent tissue

Cell carcinogenesis is often accompanied by abnormal regulatory modes of abundant genes which can be reflected by transcriptome (mRNA level). The transcriptomic difference analysis between tumors and normal tissues is useful for studying pathogenesis and screening novel targets. To identify transcriptome differences between GC and adjacent tissues, one RNA-seq dataset containing 30 clinical samples (GSE118916) was used for DEG screening. A total of 204 down-regulated and 62 up-regulated genes were identified with the classical thresholds (|log FC| ≥ 2.0 and adjusted *p-value* < 0.05). The top 10 significant up- and down-regulated genes were marked with symbols (eg. KRT20 and GUCY1A3) (**Figure 1A**). In addition, the standard RNA sequencing reads of each DEG in normal and GC tissues were shown in a heatmap with the word. D2 cluster analysis (**Figure 1B**).

However, 15 pairs of clinical samples were too few to fully reflect the changes of gene expression profile in GC patients. In this study, the GC patient transcriptome sequencing data from the TCGA database were also analyzed for DEG screening. A total of 845 DEGs were identified

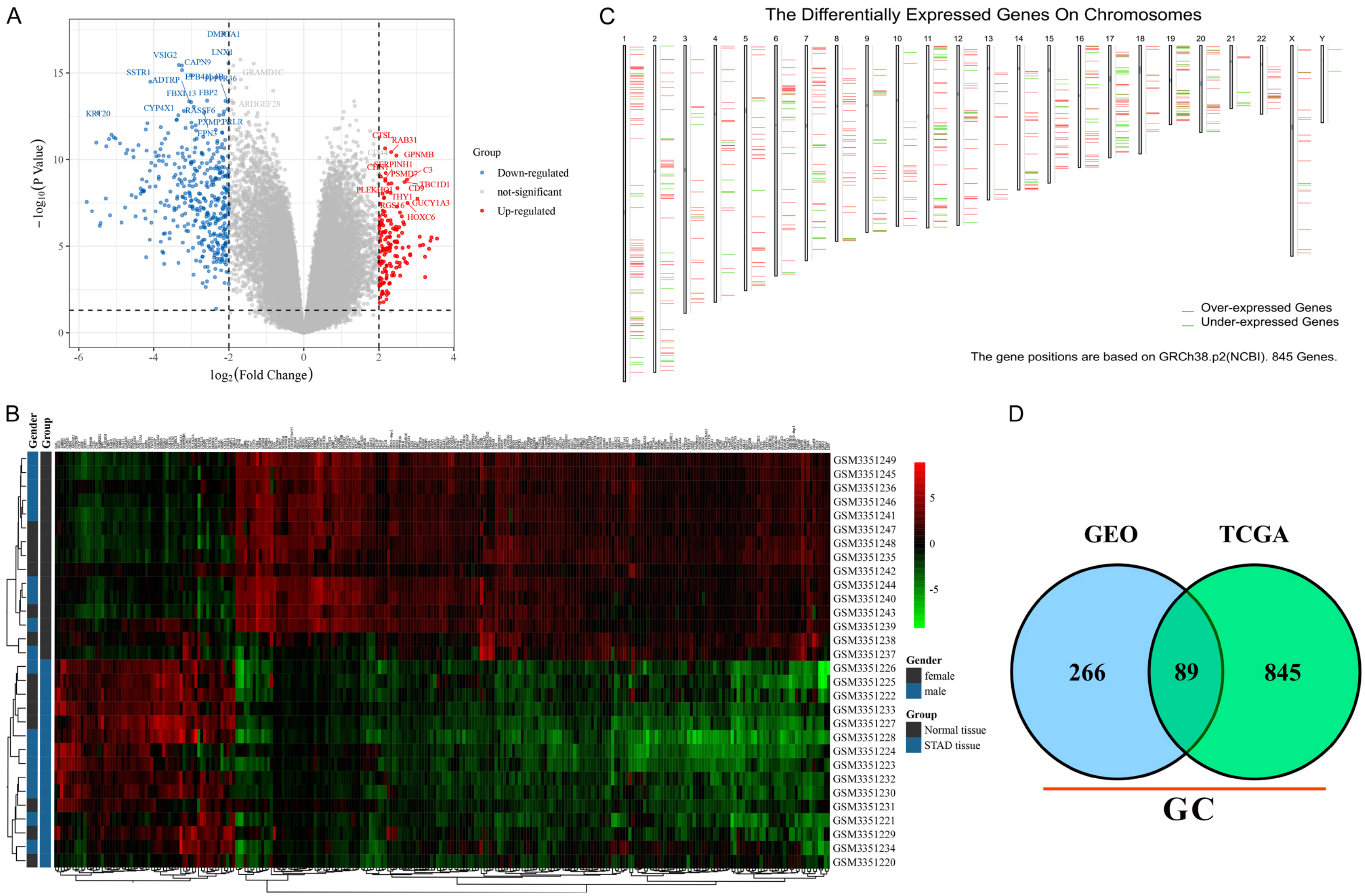


Figure 1. The DEGs of GC vs. adjacent tissue. A. The volcano plot of DEGs in GSE118916. The red points represent up-regulated genes and the blue ones represent down-regulated genes. The gray points represent genes without significant difference. B. The cluster heatmap of DEGs in GSE118916. The color shade represents the gene expression level. The legend was listed on the right. C. The DEG positions on a sketch map of chromosomes based on TCGA. The red segments represent up-regulated genes and the green ones represent down-regulated genes. D. The Venn diagram of GC DEGs in both GSE118916 and TCGA datasets.

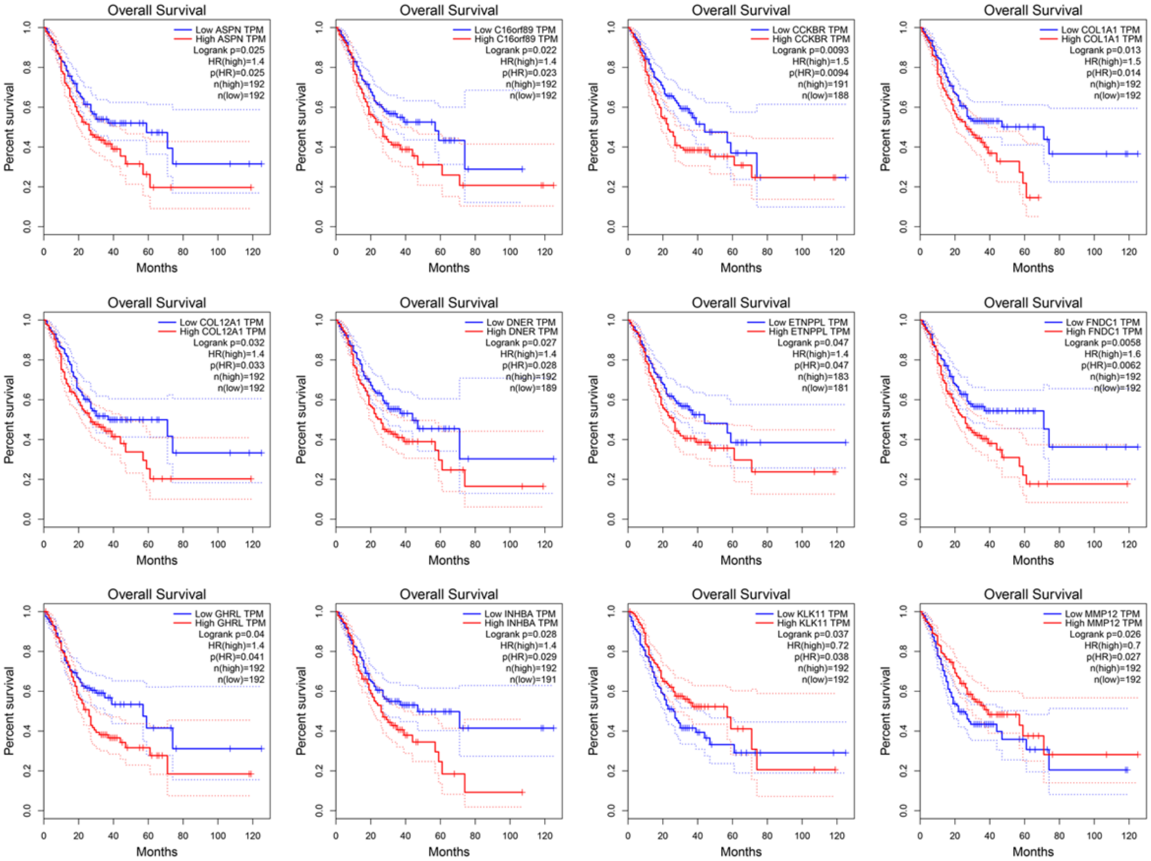


Figure 2. The overall survival curves of 12 DEGs. The red line represents the samples with a higher level of gene expression and the blue one represents the samples with a lower level, respectively. The legend was listed on the right.

with |log FC| ≥ 2.0 and *p*-value < 0.05. The DEGs positions in chromosomes are shown by a sketch map based on GRCh38.p2 (**Figure 1C**). After comparison, a total of 89 DEGs (58 down-regulated and 31 up-regulated) that appeared simultaneously in both GEO and TCGA datasets were regarded as the key genes involved in GC pathology (**Figure 1D**). Followed by function enrichment analysis, the down-regulated genes were enriched in 12 MF, 11 BP, 2 CC and 2 pathways like gastric acid secretion (Figure S1), while the up-regulated genes were enriched in 15 MF, 89 BP, 13 CC and 6 pathways like cell migration (Figure S2). The results suggested that some DEGs may participate in the GC process via these enriched functions.

The correlation between DEGs level and prognosis

The main reference index for evaluating the clinical application value of a gene is the correlation between expression level and survival time. Thus, the overall survival curves of 89 DEGs were drawn. According to the expression level of each gene, the samples were equally divided into high and low expression groups (n = 190). After analyzing, only 12 DEGs (ASPN, C16orf89, CCKBR, COL1A1, COL12A1, DNER, ETNPPL, FNDC1, GHRL, INHBA, KLK11 and MMP12) were found to significantly affect prognosis with different expression levels (*p*-value < 0.05) (**Figure 2**). Among them, only KLK11 (*p*-value = 0.037) and MMP12 (*p*-value = 0.026) may result in a worse prognosis with lower expression levels, while the other genes have the opposite trend (*p*-value < 0.05). In addition, the level of FNDC1 (*p*-value = 0.0058) was the most significantly correlative with prognosis.

In the carcinogenesis process, gene expression alterations are often beneficial for cancer progression and lead to poor prognosis. Based on both GEO and TCGA datasets, the ASPN, COL1A1, COL12A1, FNDC1, INHBA and MMP12 were up-regulated while the others were down-regulated when compared with normal tissue (*p*-value < 0.05) (**Figure 3**). Combined with sur-

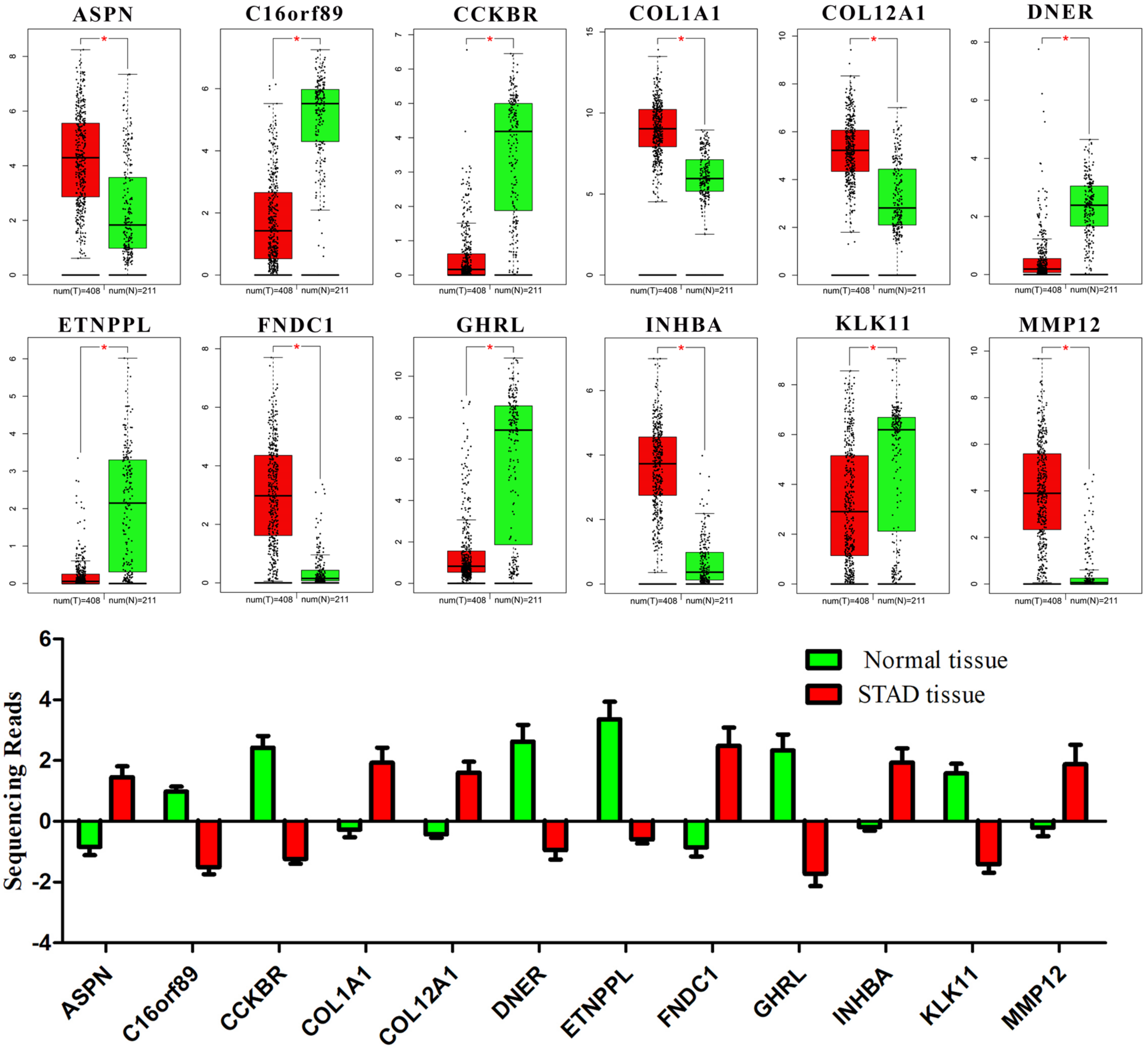

ASPN
C16orf89
CCKBR
COL1A1
COL12A1
DNER
ETNPPL
FNDC1
GHRL
INHBA
KLK11
MMP12
*
num(T)=408
num(N)=211
Sequencing Reads
6
4
2
0
-2
-4
🟩 Normal tissue
🟥 STAD tissue
ASPN
C16orf89
CCKBR
COL1A1
COL12A1
DNER
ETNPPL
FNDC1
GHRL
INHBA
KLK11
MMP12

Figure 3. The expression of 12 DEGs in normal and GC tissues. The sequencing data were obtained from GEO and TCGA databases. The red box represents GC tissue and the green one represents normal tissue. In the TCGA dataset, there were 408 GC samples and 211 normal samples. *, *p*-value < 0.05.

vival analysis results, it may suggest that only ASPN, COL1A1, COL12A1, FNDC1, INHBA and KLK11 could promote GC progression when their expression levels were altered. Notably, in this study, only the correlation between the expression of key genes and the prognosis of patients was calculated. The specific underlying molecular mechanism needs to be further studied.

The molecular aberrations of DEGs in GC samples

The correlation between expression level of DEGs and overall survival has been calculated. Moreover, in multi-type tumors, gene expression profile might often be regulated by molecular aberration like mutation, copy number and methylation. To further study the clinical function of DEGs, multi-omics data of 443 GC patients in the TCGA database were analyzed (**Figure 4**). A total of 148 (33%) patients have molecular aberration of the 12 DEGs. After analysis, it was shown that the gene aberration has no significant correlation with clinical parameters like gender, age or cancer stage. Among those genes, COL12A1 is the most prone to aberration (15%) like missense mutation. In addition, there were 68 mutation types have been detected in all cancer types containing 51 missenses, 15 truncating and 2 splice mutations (Figure S3). Interestingly, there was no mutation site in MMP12 DNA sequence which implied it is highly conserved. Certainly, the copy number alteration occurs in each gene containing both amplification and deep deletion. Furthermore, according to the methylation analysis, it was demonstrated that in GC samples, ASPN, C16orf89, CCKBR and COL1A1 have high methylation levels. Followed by drawing the overall survival curve of the altered (n = 146) and unaltered (n = 289) group, the median months overall (95% CI) was 30.88 and 28.55 without significant difference (Log rank Test *p*-value: 0.33).

The correlation between DEGs and classical cancer pathways

Genes often interact with each other to form a molecular network and participate in multi biological processes. To further study the function mode of the above 12 DEGs, a PPI analysis was performed according to STRING database [34]. Those genes were be clustered into 2 groups. Among them, COL1A1, COL12A1, ASPN, FNDC1, INHBA and MMP12 could construct a network while there was no proven correlation among the other genes (**Figure 5A** and **5B**). To study the function of this network, an enrichment analysis was performed. A total of 3 MF, 9 BP, 8 CC and 1 pathway were enriched which are most interrelated with the extracellular matrix (Figure S4). The GC has a strong invasion ability and abnormal regulation of extracellular matrix would promote metastasis of cancer cells. Among the 10 classic cancer signaling pathways, the EMT pathway is closely related to the extracellular matrix. By calculating the relationship between genes and pathway status in GC patients, it was found that all genes in the PPI network can activate the EMT pathway more frequently than their inhibitory effects, besides ETNPPL and CCKBR (**Figure 5C**). Of course, there may not be a change in EMT status in the majority of patients. Interestingly, COL1A1 was the gene with the highest frequency of activating the EMT pathway (> 50%). Therefore, it is believed that the molecular network composed of these 12 differential genes may promote the malignant transformation of GC cells by activating the EMT pathway.

Notably, in the PPI network, COL1A1 has the most interaction relationships (MMP12, CLO12A1, ASPN and INHBA) and activates the EMT pathway in most GC patients simultaneously (> 50%). It may be suggested that the COL1A1 was the central gene in this molecular network. Moreover, in this network, the ASPN, COL12A1, INHBA and MMP12 interact with COL1A1 directly. The correlation calculation results indicated that the ASPN (R = 0.42), COL12A1 (R = 0.74) and INHBA (R = 0.63) had positive correlative with COL1A1 (*p*-value < 0.001) while MMP12 was uncorrelated (R = 0.061, *p*-value = 0.2) (**Figure 6A-D**). The immunohistochemical pictures of normal and GC tissues showed that the protein level of COL1A1 was increased and the cell morphology was more irregular with increased intercellular

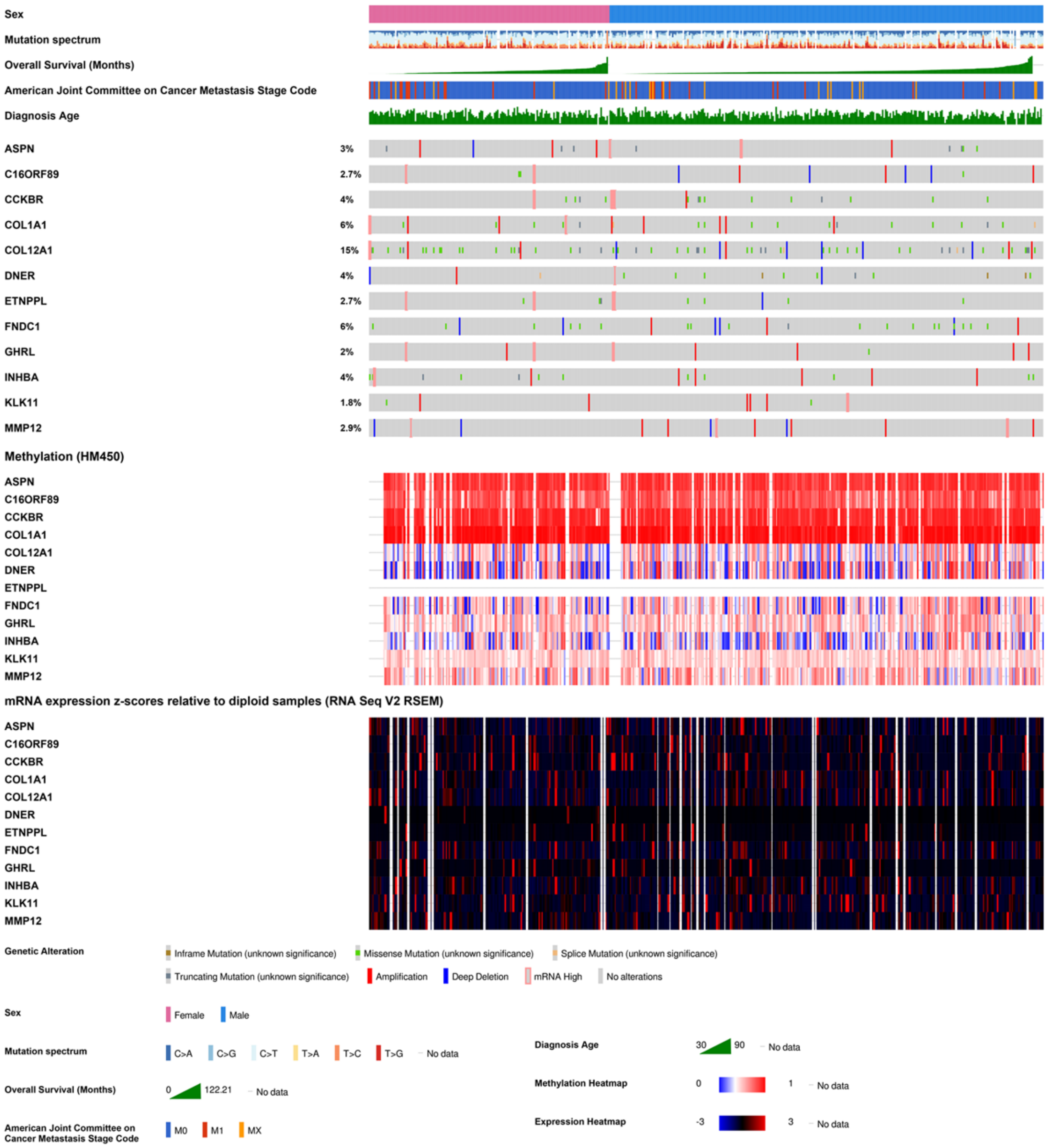


Figure 4. 12 DEGs in GC patients based on TCGA. The integrated plot of clinical data and molecular aberrations of DEGs in 443 GC samples. From top to bottom panels indicate: sex, mutation spectrum, American Joint Committee on Cancer tumor stage code, diagnosis age, overall survival (months), the heatmap of DEGs' aberrations, methylation and RNA expression. The key to the color-coding is at the bottom.

space and less compact arrangement (**Figure 6E**). This was probably because COL1A1 activating the EMT pathway results in cellular polarity/adhesion being lost and invasive/migratory properties being enhanced.

Construction of a prognostic model based on the COL1A1-network

The expression level of each gene in the COL1A1-network was significantly correlative with prognosis and the network has been confirmed to potentially active the EMT pathway, a key mechanism of GC metastasis. To determine the value of the network in the prognosis estimation of GC patients, a prognostic model based on COL1A1-network was constructed with the LASSO Cox regression model. In this model, all GC patients could be divided into 2 groups with different risk scores (high and low risk) based on the expression of ASPN and

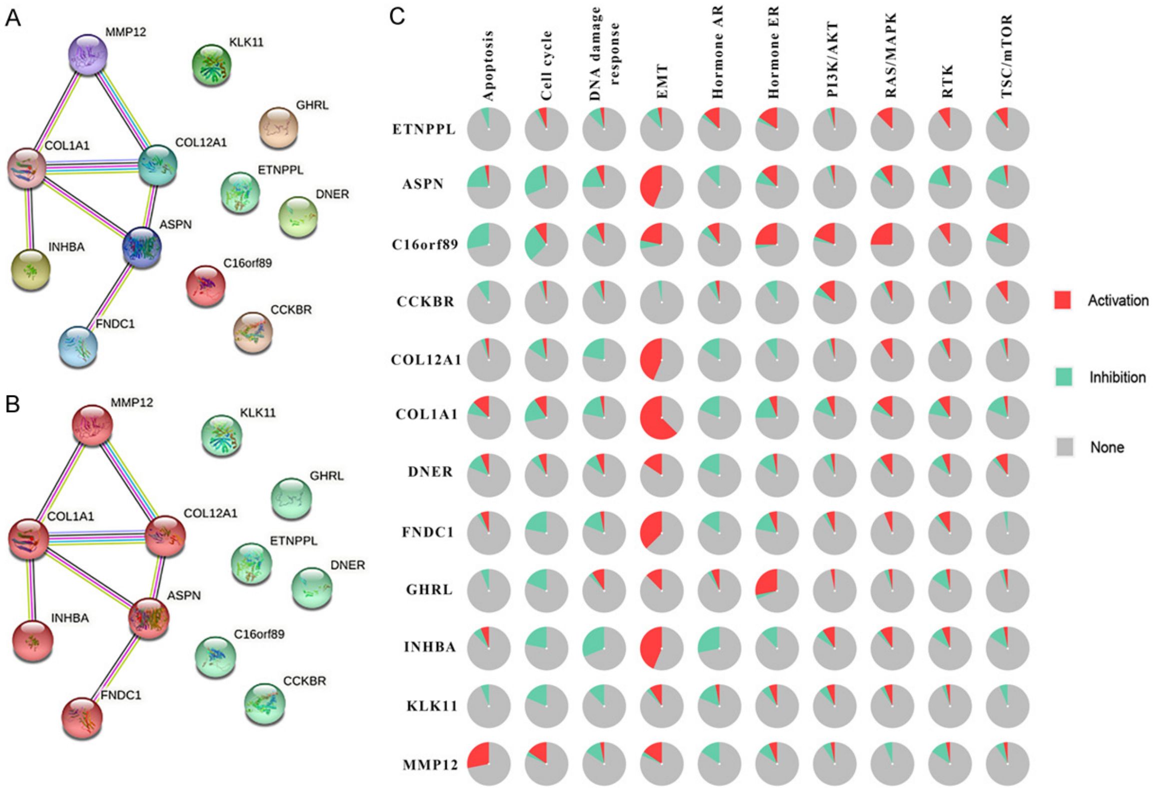


Figure 5. The function analysis of DEGs network. A, B. The PPI network of 12 DEGs. Each circle represents one single gene. The edge with different colors represents different interactions. C. The pie chart shows the activation and inhibition percentage of each pathway among GC patients. The legend was listed on the right.

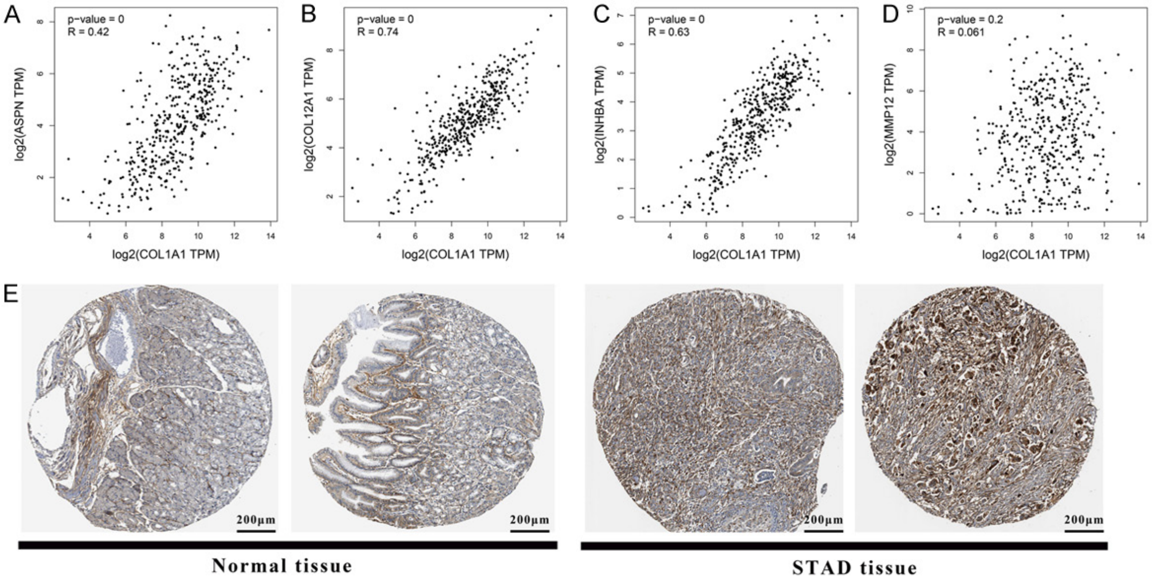


Figure 6. The correlation between COL1A1 and interactive genes. A-D. The correlation between COL1A1 and ASPN, COL12A1, INHBA and MMP12, respectively. Each point represents one single sample. The *p*-value and R-value were listed on the left. E. The immunohistochemical pictures of normal and tumor tissues which were stained by COL1A1 antibody.

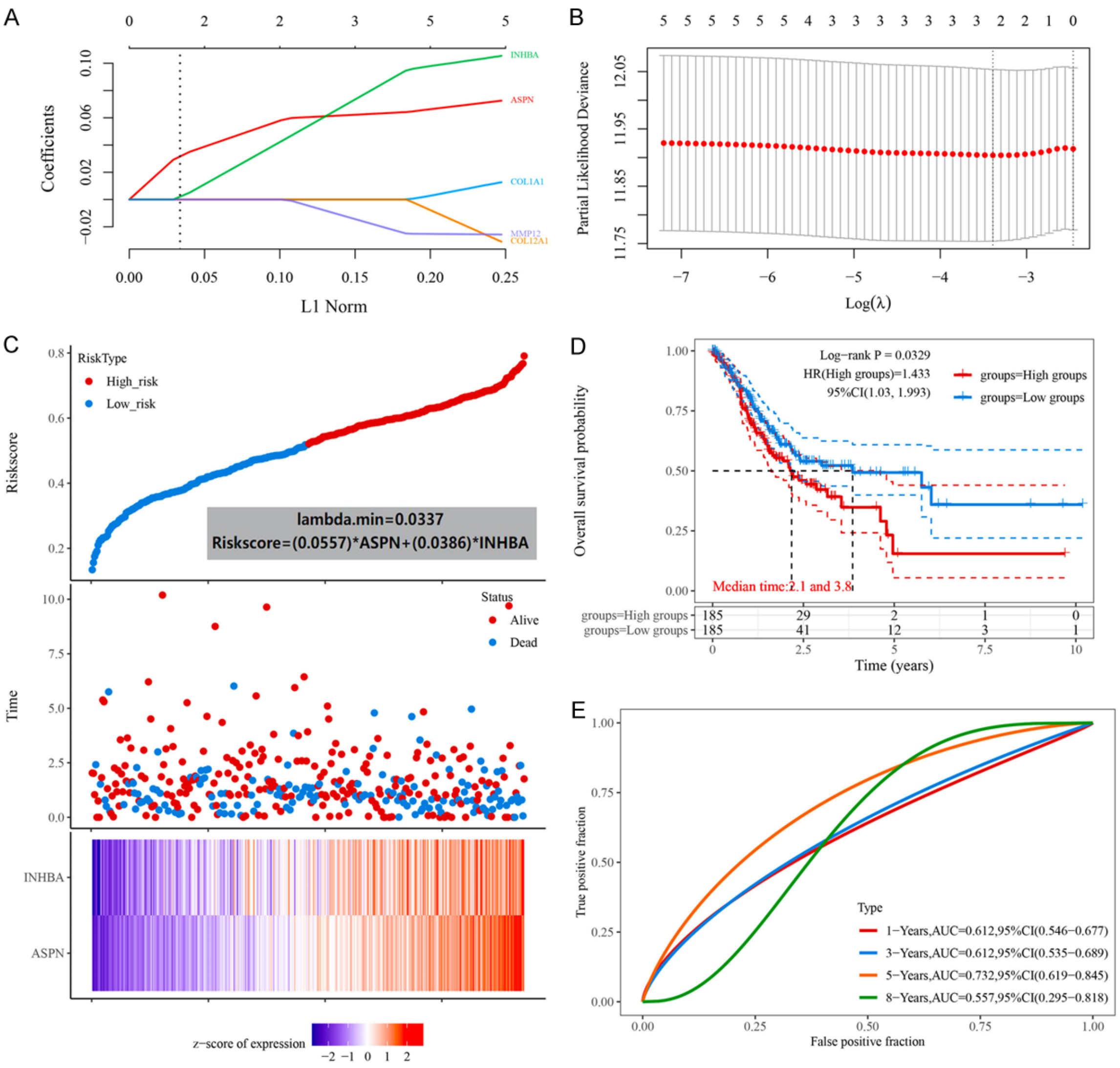


Figure 7. The prognostic mode based on COL1A1-network. A. The coefficients of selected features are shown by lambda parameter. The abscissa represents the value of lambda, and the ordinate represents the coefficients of the independent variable. B. The relationship between partial likelihood deviance and log (λ) is plotted using LASSO Cox regression model. C. The Riskscore, survival time and survival status of selected dataset. The top scatterplot represents the Riskscore from low to high. Different colors represent different groups. The scatter plot distribution represents the Riskscore of different samples correspond to the survival time and survival status. The buttom heatmap show the expression of ASPN and INHBA. D. Kaplan-Meier survival analysis of different groups was made by log-rank test. HR (High exp) represents the hazard ratio of the low-expression sample relatives to the high-expression sample. E. The ROC curve and AUC of the gene. The higher values of AUC corresponding to higher predictive power.

INHBA (**Figure 7A-C**). This indicated that the expression levels of ASPN and INHBA are the most important for patient prognosis in this molecular network prognostic model. The formula of risk score was:

Riskscore = (0.0557) * ASPN + (0.0386) * INHBA

According to the risk score, the patients were divided into high-risk and low-risk groups (n = 185). The median survival times were 2.1 and 3.8 years, respectively (**Figure 7D**). The Log-rank *p*-value = 0.0329, HR (high group) = 1.433 and 95% CI (1.03, 1.993). This indicated that the prognostic model can divide patients into 2 groups with significant differences in survival. To evaluate whether this model can objectively predict prognosis, the AUC values of 1-years (0.612, 95% CI (0.546, 0.677)), 3-years (0.612, 95% CI (0.535, 0.689)), 5-years (0.732, 95% CI

(0.619, 0.845)) and 8-years (0.557, 95% CI (0.295, 0.818)) were calculated (**Figure 7E**). Among them, the AUC value greater than 0.7 indicated that this model had a better prognostic evaluation effect and clinical reference value. In conclusion, this prognostic model based on the COL1A1-network could predict 5-years prognosis of GC patients. However, although ASPN and INHBA may increase the risk of GC in patients as main factors in this prognostic model based on the COL1A1 molecular network, this indirectly indicates the clinical value of COL1A1 in the diagnosis and treatment of GC.

Discussion

The GC has a high degree of malignancy. Due to the lack of early diagnostic markers, the pathological tissue is difficult to detect in time only by gastroscopy. The GC patients were often in the advanced stage when diagnosed, accompanied by multiple organ metastases of tumor cells. Currently, gastrectomy is the common clinical treatment, which seriously affects the life quality and has poor efficacy. Therefore, screening effective early diagnostic markers and potential therapeutic targets are important for the prevention and treatment of GC.

In the present study, the DEGs of GC vs. adjacent tissues were screened firstly. After enrichment analysis, the DEGs may be involved in GC carcinogenesis and regarded as potential markers or therapeutic targets. To improve the accuracy, the bulk RNA-seq datasets of GC and adjacent from both GEO and TCGA databases were obtained for transcriptome analysis. There were 89 overlapping DEGs which were detected followed by functional enrichment and survival analysis. Among them, only 12 genes were significantly correlative with overall survival in different expression levels. In addition, once C16orf89, CCKBR, DNER, ETNPPL, GHRL and MMP12 were dysregulated in GC, the prognosis may be better. For example, in a study by Li et al. [35], it was confirmed that stable downregulation of CCKBR could result in reduced proliferation, migration and invasion of GC cell lines (BGC-823 and SGC-7901). Downregulation of CCKBR also inhibited the growth of gastric tumors *in vivo*. This is consistent with the results of our study, suggesting that CCKBR can improve the prognosis of GC patients. However, the evidence verifying the correlation between other genes and GC is scant.

COL1A1 is a member of the collagen family, and the abnormal expression of COL1A1 has been confirmed to be associated with the occurrence of some cancers [18, 36]. In addition, it has also been proven to promote the progression of multiple cancers [22]. COL1A1 is a reliable biomarker and a potential therapeutic target for hepatocellular carcinogenesis and metastasis [37]. However, the role and molecular mechanism of COL1A1 in GC development remain unclear. According to the PPI network in our study, the ASPN was correlative with both COL1A1 and COL12A1, which belong to the collagen family. In Marco et al.'s study [38], an *Aspn*$^{-/-}$ mouse model and the investigation of the *Aspn*$^{-/-}$ skin phenotype were constructed successfully. Functionally, *Aspn*$^{-/-}$ mice had an increased skin mechanical toughness, although there were no structural changes present on histology or immunohistochemistry. Electron microscopy showed 7% thinner collagen fibrils in *Aspn*$^{-/-}$ mice. Several matrix genes were upregulated, including collagens COL1A1. It further confirmed that the ASPN may negatively regulate COL1A1. In conclusion, it suggested the ASPN might be the upstream regulator of the collagen family.

Among those 12 DEGs, COL1A1 was an activating gene of the EMT pathway with the most frequency. Nevertheless, the mechanism is still unknown. In Guo et al.'s study [39], it was detected that the COL1A1 was upregulated in GC tissue, while the miR-133b was significantly down-regulated. Overexpression of miR-133b could suppress the migration and invasion of GC cells. In addition, the EMT process was inhibited as well. COL1A1 has been verified as a target gene of miR-133b and its overexpression had a significant impact on the prognosis of GC patients. GSEA pathway enrichment results showed that COL1A1 was markedly enriched in the TGF-β signaling pathway. In addition, COL1A1 overexpression induced the activation of the TGF-β signaling pathway to promote proliferation and migration of GC cells. As well known, the correlation between the TGF-β signaling pathway and EMT pathway has been proved adequately [40]. In Wang et al.'s study [41], it has also been proved in colon cancer, the overexpression of

COL1A1 promoted cell viability, migration, invasion and EMT pathway. In conclusion, these findings may offer new insights into the development of new treatments against GC. Our results may also assist in identifying potential biomarkers for the timely diagnosis and prognosis of GC. In addition, more basic experiments may be needed to elucidate the mechanism.

Acknowledgements

The authors are grateful to the school health care staff who participated in this study. This study was approved by the Medical Ethics Committee of the Affiliated Hospital of Xuzhou Medical University.

Disclosure of conflict of interest

None.

Abbreviations

GC, Gastric Carcinoma; EMT, Epithelial-Mesenchymal Transition; CSCs, Cancer Stem Cells; COL-1, Type I Collagen; GSEA, Gene Set Enrichment Analysis; BP, Biological Process; CC, Cellular Component; MF, Molecular Function; PPI, Protein-Protein Interaction; MCODE, Molecular Complex Detection; HR, Hazard Ratio; FDR, False Discovery Rate; IHC, Immunohistochemistry; HSD, Honest Significant Difference.

Address correspondence to: Yuhua Chen, The Chongqing Seventh People's Hospital, Banan District, Chongqing 401320, China. E-mail: 569347808@qq.com

References

[1] Zhu X and Li J. Gastric carcinoma in China: current status and future perspectives (review). Oncol Lett 2010; 1: 407-412.

[2] Jemal A, Bray F, Center MM, Ferlay J, Ward E and Forman D. Global cancer statistics. CA Cancer J Clin 2011; 61: 69-90.

[3] Xu AM, Huang L, Liu W, Gao S, Han WX and Wei ZJ. Neoadjuvant chemotherapy followed by surgery versus surgery alone for gastric carcinoma: systematic review and meta-analysis of randomized controlled trials. PLoS One 2014; 9: e86941.

[4] Memon MA, Subramanya MS, Khan S, Hossain MB, Osland E and Memon B. Meta-analysis of D1 versus D2 gastrectomy for gastric adenocarcinoma. Ann Surg 2011; 253: 900-911.

[5] Xu AM, Huang L, Zhu L and Wei ZJ. Significance of peripheral neutrophil-lymphocyte ratio among gastric cancer patients and construction of a treatment-predictive model: a study based on 1131 cases. Am J Cancer Res 2014; 4: 189-195.

[6] Huang L, Xu A, Li T, Han W, Wu S and Wang Y. Detection of perioperative cancer antigen 72-4 in gastric juice pre- and post-distal gastrectomy and its significances. Med Oncol 2013; 30: 651.

[7] Xu AM, Huang L, Han WX and Wei ZJ. Monitoring of peri-distal gastrectomy carbohydrate antigen 19-9 level in gastric juice and its significance. Int J Clin Exp Med 2014; 7: 230-238.

[8] Chaffer CL and Weinberg RA. A perspective on cancer cell metastasis. Science 2011; 331: 1559-1564.

[9] Montemayor-Garcia C, Hardin H, Guo Z, Larrain C, Buehler D, Asioli S, Chen H and Lloyd RV. The role of epithelial mesenchymal transition markers in thyroid carcinoma progression. Endocr Pathol 2013; 24: 206-212.

[10] Liang Q, Li L, Zhang J, Lei Y, Wang L, Liu DX, Feng J, Hou P, Yao R, Zhang Y, Huang B and Lu J. CDK5 is essential for TGF-β1-induced epithelial-mesenchymal transition and breast cancer progression. Sci Rep 2013; 3: 2932.

[11] Zhao L, Li W, Zang W, Liu Z, Xu X, Yu H, Yang Q and Jia J. JMJD2B promotes epithelial-mesenchymal transition by cooperating with β-catenin and enhances gastric cancer metastasis. Clin Cancer Res 2013; 19: 6419-6429.

[12] Sleeman JP and Thiery JP. SnapShot: the epithelial-mesenchymal transition. Cell 2011; 145: 162, e1.

[13] Huang L, Xu AM, Liu S, Liu W and Li TJ. Cancer-associated fibroblasts in digestive tumors. World J Gastroenterol 2014; 20: 17804-17818.

[14] Chowdhury SR, Mh Busra MF, Lokanathan Y, Ng MH, Law JX, Cletus UC and Binti Haji Idrus R. Collagen type I: a versatile biomaterial. Adv Exp Med Biol 2018; 1077: 389-414.

[15] Zhu X, Luo X, Jiang S and Wang H. Bone morphogenetic protein 1 targeting COL1A1 and COL1A2 to regulate the epithelial-mesenchymal transition process of colon cancer SW620 cells. J Nanosci Nanotechnol 2020; 20: 1366-1374.

[16] Exposito JY, Valcourt U, Cluzel C and Lethias C. The fibrillar collagen family. Int J Mol Sci 2010; 11: 407-426.

[17] Ibanez de Caceres I, Dulaimi E, Hoffman AM, Al-Saleem T, Uzzo RG and Cairns P. Identification of novel target genes by an epigenetic reactivation screen of renal cancer. Cancer Res 2006; 66: 5021-5028.

[18] Hayashi M, Nomoto S, Hishida M, Inokawa Y, Kanda M, Okamura Y, Nishikawa Y, Tanaka C, Kobayashi D, Yamada S, Nakayama G, Fujii T, Sugimoto H, Koike M, Fujiwara M, Takeda S and Kodera Y. Identification of the collagen type 1 α 1 gene (COL1A1) as a candidate survival-related factor associated with hepatocellular carcinoma. BMC Cancer 2014; 14: 108.

[19] Bonazzi VF, Nancarrow DJ, Stark MS, Moser RJ, Boyle GM, Aoude LG, Schmidt C and Hayward NK. Cross-platform array screening identifies COL1A2, THBS1, TNFRSF10D and UCHL1 as genes frequently silenced by methylation in melanoma. PLoS One 2011; 6: e26121.

[20] Kita Y, Mimori K, Tanaka F, Matsumoto T, Haraguchi N, Ishikawa K, Matsuzaki S, Fukuyoshi Y, Inoue H, Natsugoe S, Aikou T and Mori M. Clinical significance of LAMB3 and COL7A1 mRNA in esophageal squamous cell carcinoma. Eur J Surg Oncol 2009; 35: 52-58.

[21] Wu YH, Chang TH, Huang YF, Huang HD and Chou CY. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. Oncogene 2014; 33: 3432-3440.

[22] Ramaswamy S, Ross KN, Lander ES and Golub TR. A molecular signature of metastasis in primary solid tumors. Nat Genet 2003; 33: 49-54.

[23] Zhao Q, Xie J, Xie J, Zhao R, Song C, Wang H, Rong J, Yan L, Song Y, Wang F and Xie Y. Weighted correlation network analysis identifies FN1, COL1A1 and SERPINE1 associated with the progression and prognosis of gastric cancer. Cancer Biomark 2021; 31: 59-75.

[24] Li L, Zhu Z, Zhao Y, Zhang Q, Wu X, Miao B, Cao J and Fei S. FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics. Sci Rep 2019; 9: 7827.

[25] Tian Y, Ma L, Cai X and Zhu J. Statistical method based on bayes-type empirical score test for assessing genetic association with multilocus genotype data. Int J Genomics 2020; 2020: 4708152.

[26] Chen X. False discovery rate control for multiple testing based on discrete *p*-values. Biom J 2020; 62: 1060-1079.

[27] Chen L, Lu D, Sun K, Xu Y, Hu P, Li X and Xu F. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. Gene 2019; 692: 119-125.

[28] Huang da W, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4: 44-57.

[29] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ and von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015; 43: D447-452.

[30] Li R, Yang YE, Yin YH, Zhang MY, Li H and Qu YQ. Methylation and transcriptome analysis reveal lung adenocarcinoma-specific diagnostic biomarkers. J Transl Med 2019; 17: 324.

[31] Ladas EJ, Blonquist TM, Puligandla M, Orjuela M, Stevenson K, Cole PD, Athale UH, Clavell LA, Leclerc JM, Laverdiere C, Michon B, Schorin MA, Greene Welch J, Asselin BL, Sallan SE, Silverman LB and Kelly KM. Protective effects of dietary intake of antioxidants and treatment-related toxicity in childhood leukemia: a report from the DALLT cohort. J Clin Oncol 2020; 38: 2151-2159.

[32] Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, Waldner MJ, Bindea G, Mlecnik B, Galon J and Trajanoski Z. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. Genome Biol 2015; 16: 64.

[33] Wang Z, Tang W, Yuan J, Qiang B, Han W and Peng X. Integrated analysis of RNA-binding proteins in glioma. Cancers (Basel) 2020; 12: 892.

[34] von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P and Snel B. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res 2003; 31: 258-261.

[35] Li M, Chang J, Ren H, Song D, Guo J, Peng L, Zhou X, Zhao K, Lu S, Liu Z and Hu P. Downregulation of CCKBR expression inhibits the proliferation of gastric cancer cells, revealing a potential target for immunotoxin therapy. Curr Cancer Drug Targets 2022; 22: 257-268.

[36] Lv J, Guo L, Wang JH, Yan YZ, Zhang J, Wang YY, Yu Y, Huang YF and Zhao HP. Biomarker identification and trans-regulatory network analyses in esophageal adenocarcinoma and Barrett's esophagus. World J Gastroenterol 2019; 25: 233-244.

[37] Ma HP, Chang HL, Bamodu OA, Yadav VK, Huang TY, Wu ATH, Yeh CT, Tsai SH and Lee WH. Collagen 1A1 (COL1A1) is a reliable biomarker and putative therapeutic target for hepatocellular carcinogenesis and metastasis. Cancers (Basel) 2019; 11: 786.

[38] Maccarana M, Svensson RB, Knutsson A, Giannopoulos A, Pelkonen M, Weis M, Eyre D, Warman M and Kalamajski S. Asporin-deficient mice have tougher skin and altered skin gly-

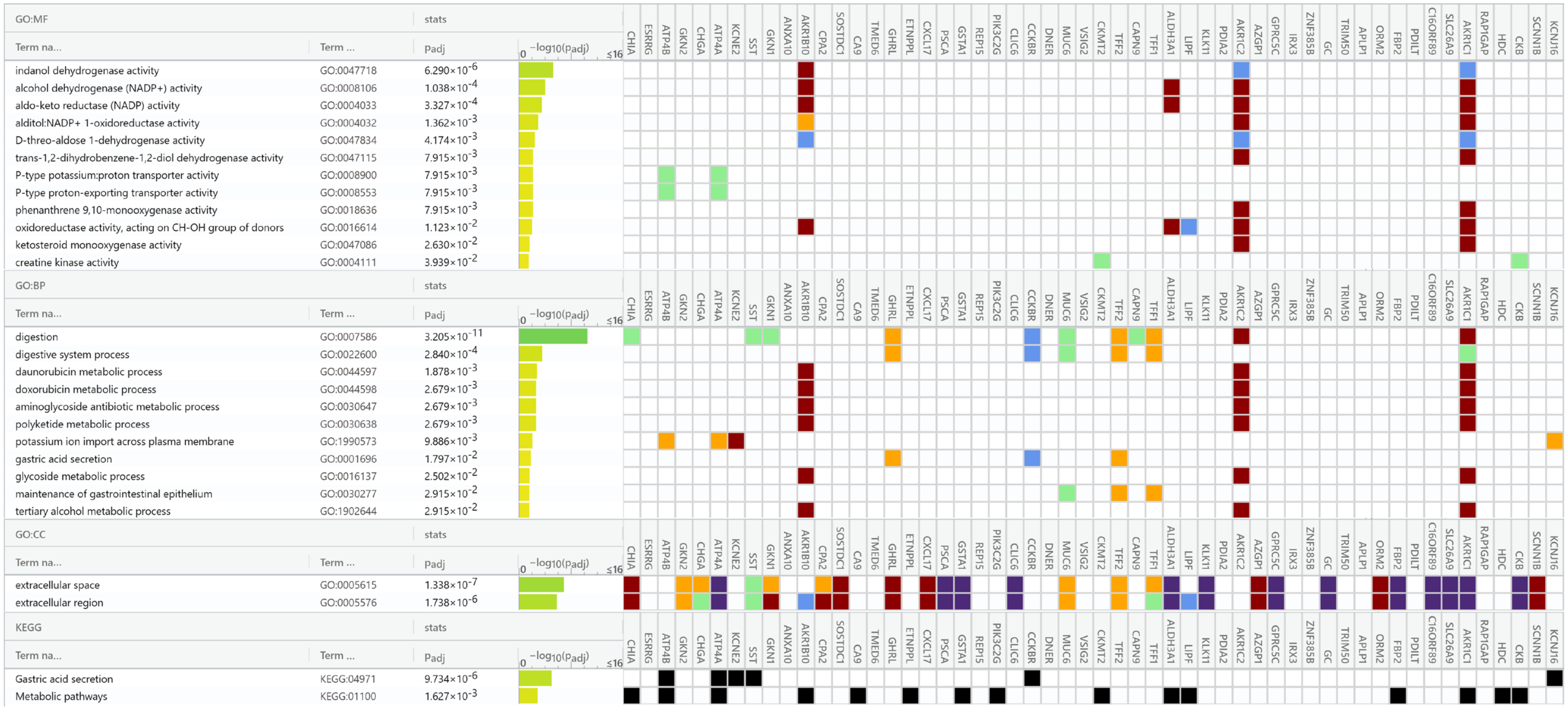
cosaminoglycan content and structure. PLoS One 2017; 12: e0184028.

[39] Guo Y, Lu G, Mao H, Zhou S, Tong X, Wu J, Sun Q, Xu H and Fang F. miR-133b suppresses invasion and migration of gastric cancer cells via the COL1A1/TGF-β axis. Onco Targets Ther 2020; 13: 7985-7995.

[40] Song J and Shi W. The concomitant apoptosis and EMT underlie the fundamental functions of TGF-β. Acta Biochim Biophys Sin (Shanghai) 2018; 50: 91-97.

[41] Wang X, Song Z, Hu B, Chen Z, Chen F and Cao C. MicroRNA-642a-5p inhibits colon cancer cell migration and invasion by targeting collagen type I α1. Oncol Rep 2021; 45: 933-944.

The function of COL1A1 in GC



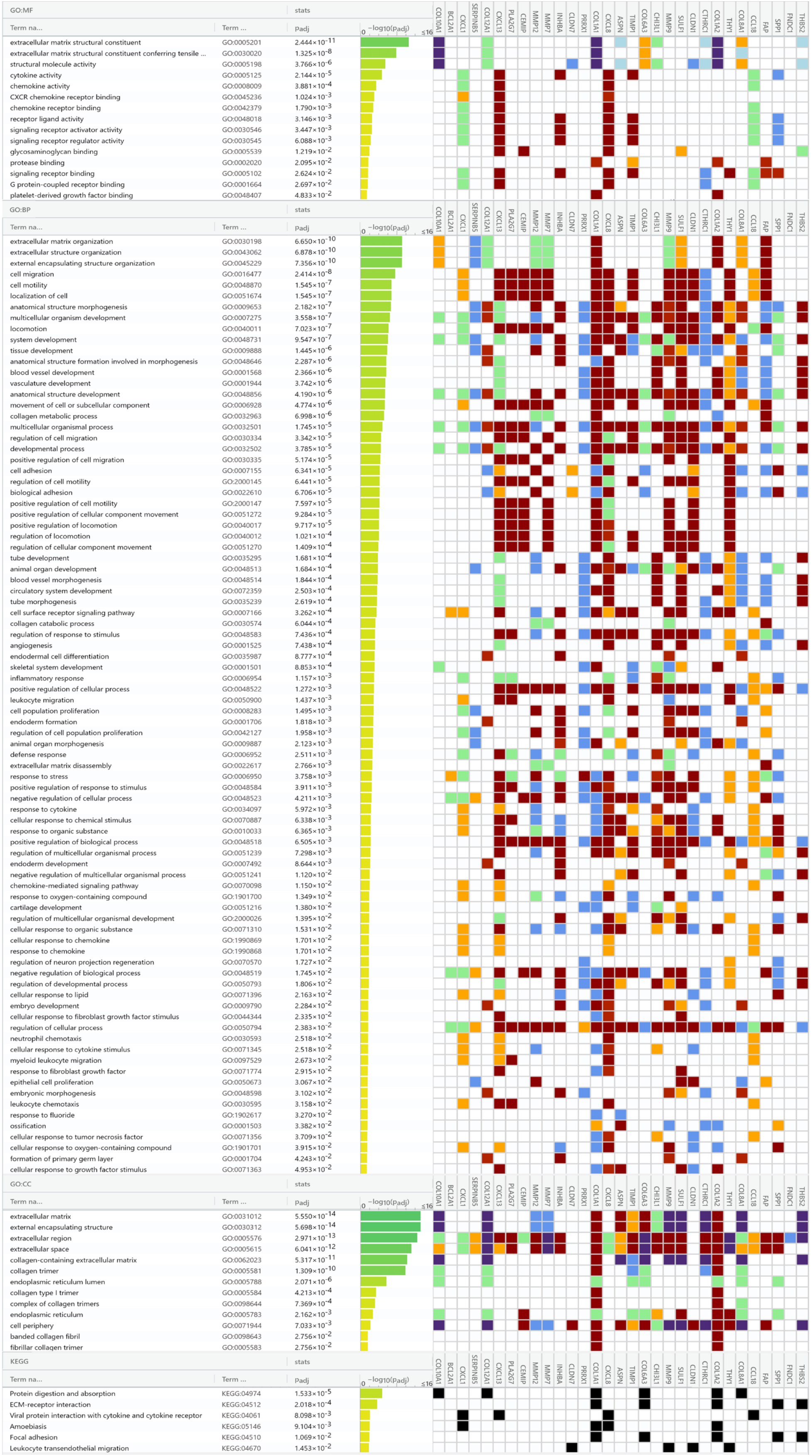
GO:MF		stats	
Term na...	Term ...	Padj	0 −log10(Padj) ≤16
indanol dehydrogenase activity	GO:0047718	6.290×10^{-6}	
alcohol dehydrogenase (NADP+) activity	GO:0008106	1.038×10^{-4}	
aldo-keto reductase (NADP) activity	GO:0004033	3.327×10^{-4}	
alditol:NADP+ 1-oxidoreductase activity	GO:0004032	1.362×10^{-3}	
D-threo-aldose 1-dehydrogenase activity	GO:0047834	4.174×10^{-3}	
trans-1,2-dihydrobenzene-1,2-diol dehydrogenase activity	GO:0047115	7.915×10^{-3}	
P-type potassium:proton transporter activity	GO:0008900	7.915×10^{-3}	
P-type proton-exporting transporter activity	GO:0008553	7.915×10^{-3}	
phenanthrene 9,10-monooxygenase activity	GO:0018636	7.915×10^{-3}	
oxidoreductase activity, acting on CH-OH group of donors	GO:0016614	1.123×10^{-2}	
ketosteroid monooxygenase activity	GO:0047086	2.630×10^{-2}	
creatine kinase activity	GO:0004111	3.939×10^{-2}	

GO:BP		stats	
Term na...	Term ...	Padj	0 −log10(Padj) ≤16
digestion	GO:0007586	3.205×10^{-11}	
digestive system process	GO:0022600	2.840×10^{-4}	
daunorubicin metabolic process	GO:0044597	1.878×10^{-3}	
doxorubicin metabolic process	GO:0044598	2.679×10^{-3}	
aminoglycoside antibiotic metabolic process	GO:0030647	2.679×10^{-3}	
polyketide metabolic process	GO:0030638	2.679×10^{-3}	
potassium ion import across plasma membrane	GO:1990573	9.886×10^{-3}	
gastric acid secretion	GO:0001696	1.797×10^{-2}	
glycoside metabolic process	GO:0016137	2.502×10^{-2}	
maintenance of gastrointestinal epithelium	GO:0030277	2.915×10^{-2}	
tertiary alcohol metabolic process	GO:1902644	2.915×10^{-2}	

GO:CC		stats	
Term na...	Term ...	Padj	0 −log10(Padj) ≤16
extracellular space	GO:0005615	1.338×10^{-7}	
extracellular region	GO:0005576	1.738×10^{-6}	

KEGG		stats	
Term na...	Term ...	Padj	0 −log10(Padj) ≤16
Gastric acid secretion	KEGG:04971	9.734×10^{-6}	
Metabolic pathways	KEGG:01100	1.627×10^{-3}	

Figure S1. The function enrichment analysis of down-regulated genes in GC. The MF, BP, CC and KEGG pathway were list with term name, ID, *p*-value and enriched genes. The significance threshold was g.SCS threshold.



GO:MF

Term na...	Term ...	padj
extracellular matrix structural constituent	GO:0005201	2.444×10^{-11}
extracellular matrix structural constituent conferring tensile ...	GO:0030020	1.325×10^{-8}
structural molecule activity	GO:0005198	3.766×10^{-6}
cytokine activity	GO:0005125	2.144×10^{-5}
chemokine activity	GO:0008009	3.881×10^{-4}
CXCR chemokine receptor binding	GO:0045236	1.024×10^{-3}
chemokine receptor binding	GO:0042379	1.790×10^{-3}
receptor ligand activity	GO:0048018	3.146×10^{-3}
signaling receptor activator activity	GO:0030546	3.447×10^{-3}
signaling receptor regulator activity	GO:0030545	6.088×10^{-3}
glycosaminoglycan binding	GO:0005539	1.219×10^{-2}
protease binding	GO:0002020	2.095×10^{-2}
signaling receptor binding	GO:0005102	2.624×10^{-2}
G protein-coupled receptor binding	GO:0001664	2.697×10^{-2}
platelet-derived growth factor binding	GO:0048407	4.833×10^{-2}

GO:BP

Term na...	Term ...	padj
extracellular matrix organization	GO:0030198	6.650×10^{-10}
extracellular structure organization	GO:0043062	6.878×10^{-10}
external encapsulating structure organization	GO:0045229	7.356×10^{-10}
cell migration	GO:0016477	2.414×10^{-8}
cell motility	GO:0048870	1.545×10^{-7}
localization of cell	GO:0051674	1.545×10^{-7}
anatomical structure morphogenesis	GO:0009653	2.182×10^{-7}
multicellular organism development	GO:0007275	3.558×10^{-7}
locomotion	GO:0040011	7.023×10^{-7}
system development	GO:0048731	9.547×10^{-7}
tissue development	GO:0009888	1.445×10^{-6}
anatomical structure formation involved in morphogenesis	GO:0048646	2.287×10^{-6}
blood vessel development	GO:0001568	2.366×10^{-6}
vasculature development	GO:0001944	3.742×10^{-6}
anatomical structure development	GO:0048856	4.190×10^{-6}
movement of cell or subcellular component	GO:0006928	4.774×10^{-6}
collagen metabolic process	GO:0032963	6.998×10^{-6}
multicellular organismal process	GO:0032501	1.745×10^{-5}
regulation of cell migration	GO:0030334	3.342×10^{-5}
developmental process	GO:0032502	3.785×10^{-5}
positive regulation of cell migration	GO:0030335	5.174×10^{-5}
cell adhesion	GO:0007155	6.341×10^{-5}
regulation of cell motility	GO:2000145	6.441×10^{-5}
biological adhesion	GO:0022610	6.706×10^{-5}
positive regulation of cell motility	GO:2000147	7.597×10^{-5}
positive regulation of cellular component movement	GO:0051272	9.284×10^{-5}
positive regulation of locomotion	GO:0040017	9.717×10^{-5}
regulation of locomotion	GO:0040012	1.021×10^{-4}
regulation of cellular component movement	GO:0051270	1.409×10^{-4}
tube development	GO:0035295	1.681×10^{-4}
animal organ development	GO:0048513	1.684×10^{-4}
blood vessel morphogenesis	GO:0048514	1.844×10^{-4}
circulatory system development	GO:0072359	2.503×10^{-4}
tube morphogenesis	GO:0035239	2.619×10^{-4}
cell surface receptor signaling pathway	GO:0007166	3.262×10^{-4}
collagen catabolic process	GO:0030574	6.044×10^{-4}
regulation of response to stimulus	GO:0048583	7.436×10^{-4}
angiogenesis	GO:0001525	7.438×10^{-4}
endodermal cell differentiation	GO:0035987	8.777×10^{-4}
skeletal system development	GO:0001501	8.853×10^{-4}
inflammatory response	GO:0006954	1.157×10^{-3}
positive regulation of cellular process	GO:0048522	1.272×10^{-3}
leukocyte migration	GO:0050900	1.437×10^{-3}
cell population proliferation	GO:0008283	1.495×10^{-3}
endoderm formation	GO:0001706	1.818×10^{-3}
regulation of cell population proliferation	GO:0042127	1.958×10^{-3}
animal organ morphogenesis	GO:0009887	2.123×10^{-3}
defense response	GO:0006952	2.511×10^{-3}
extracellular matrix disassembly	GO:0022617	2.766×10^{-3}
response to stress	GO:0006950	3.758×10^{-3}
positive regulation of response to stimulus	GO:0048584	3.911×10^{-3}
negative regulation of cellular process	GO:0048523	4.211×10^{-3}
response to cytokine	GO:0034097	5.972×10^{-3}
cellular response to chemical stimulus	GO:0070887	6.338×10^{-3}
response to organic substance	GO:0010033	6.365×10^{-3}
positive regulation of biological process	GO:0048518	6.505×10^{-3}
regulation of multicellular organismal process	GO:0051239	7.298×10^{-3}
endoderm development	GO:0007492	8.644×10^{-3}
negative regulation of multicellular organismal process	GO:0051241	1.120×10^{-2}
chemokine-mediated signaling pathway	GO:0070098	1.150×10^{-2}
response to oxygen-containing compound	GO:1901700	1.349×10^{-2}
cartilage development	GO:0051216	1.380×10^{-2}
regulation of multicellular organismal development	GO:2000026	1.395×10^{-2}
cellular response to organic substance	GO:0071310	1.531×10^{-2}
cellular response to chemokine	GO:1990869	1.701×10^{-2}
response to chemokine	GO:1990868	1.701×10^{-2}
regulation of neuron projection regeneration	GO:0070570	1.727×10^{-2}
negative regulation of biological process	GO:0048519	1.745×10^{-2}
regulation of developmental process	GO:0050793	1.806×10^{-2}
cellular response to lipid	GO:0071396	2.163×10^{-2}
embryo development	GO:0009790	2.284×10^{-2}
cellular response to fibroblast growth factor stimulus	GO:0044344	2.335×10^{-2}
regulation of cellular process	GO:0050794	2.383×10^{-2}
neutrophil chemotaxis	GO:0030593	2.518×10^{-2}
cellular response to cytokine stimulus	GO:0071345	2.518×10^{-2}
myeloid leukocyte migration	GO:0097529	2.673×10^{-2}
response to fibroblast growth factor	GO:0071774	2.915×10^{-2}
epithelial cell proliferation	GO:0050673	3.067×10^{-2}
embryonic morphogenesis	GO:0048598	3.102×10^{-2}
leukocyte chemotaxis	GO:0030595	3.158×10^{-2}
response to fluoride	GO:1902617	3.270×10^{-2}
ossification	GO:0001503	3.382×10^{-2}
cellular response to tumor necrosis factor	GO:0071356	3.709×10^{-2}
cellular response to oxygen-containing compound	GO:1901701	3.915×10^{-2}
formation of primary germ layer	GO:0001704	4.243×10^{-2}
cellular response to growth factor stimulus	GO:0071363	4.953×10^{-2}

GO:CC

Term na...	Term ...	padj
extracellular matrix	GO:0031012	5.550×10^{-14}
external encapsulating structure	GO:0030312	5.698×10^{-14}
extracellular region	GO:0005576	2.971×10^{-13}
extracellular space	GO:0005615	6.041×10^{-12}
collagen-containing extracellular matrix	GO:0062023	5.317×10^{-11}
collagen trimer	GO:0005581	1.309×10^{-10}
endoplasmic reticulum lumen	GO:0005788	2.071×10^{-6}
collagen type I trimer	GO:0005584	4.213×10^{-4}
complex of collagen trimers	GO:0098644	7.369×10^{-4}
endoplasmic reticulum	GO:0005783	2.162×10^{-3}
cell periphery	GO:0071944	7.033×10^{-3}
banded collagen fibril	GO:0098643	2.756×10^{-2}
fibrillar collagen trimer	GO:0005583	2.756×10^{-2}

KEGG

Term na...	Term ...	padj
Protein digestion and absorption	KEGG:04974	1.533×10^{-5}
ECM-receptor interaction	KEGG:04512	2.018×10^{-4}
Viral protein interaction with cytokine and cytokine receptor	KEGG:04061	8.098×10^{-3}
Amoebiasis	KEGG:05146	9.104×10^{-3}
Focal adhesion	KEGG:04510	1.069×10^{-2}
Leukocyte transendothelial migration	KEGG:04670	1.453×10^{-2}

Figure S2. The function enrichment analysis of up-regulated genes in GC. The MF, BP, CC and KEGG pathway were list with term name, ID, *p*-value and enriched genes. The significance threshold was g.SCS threshold.

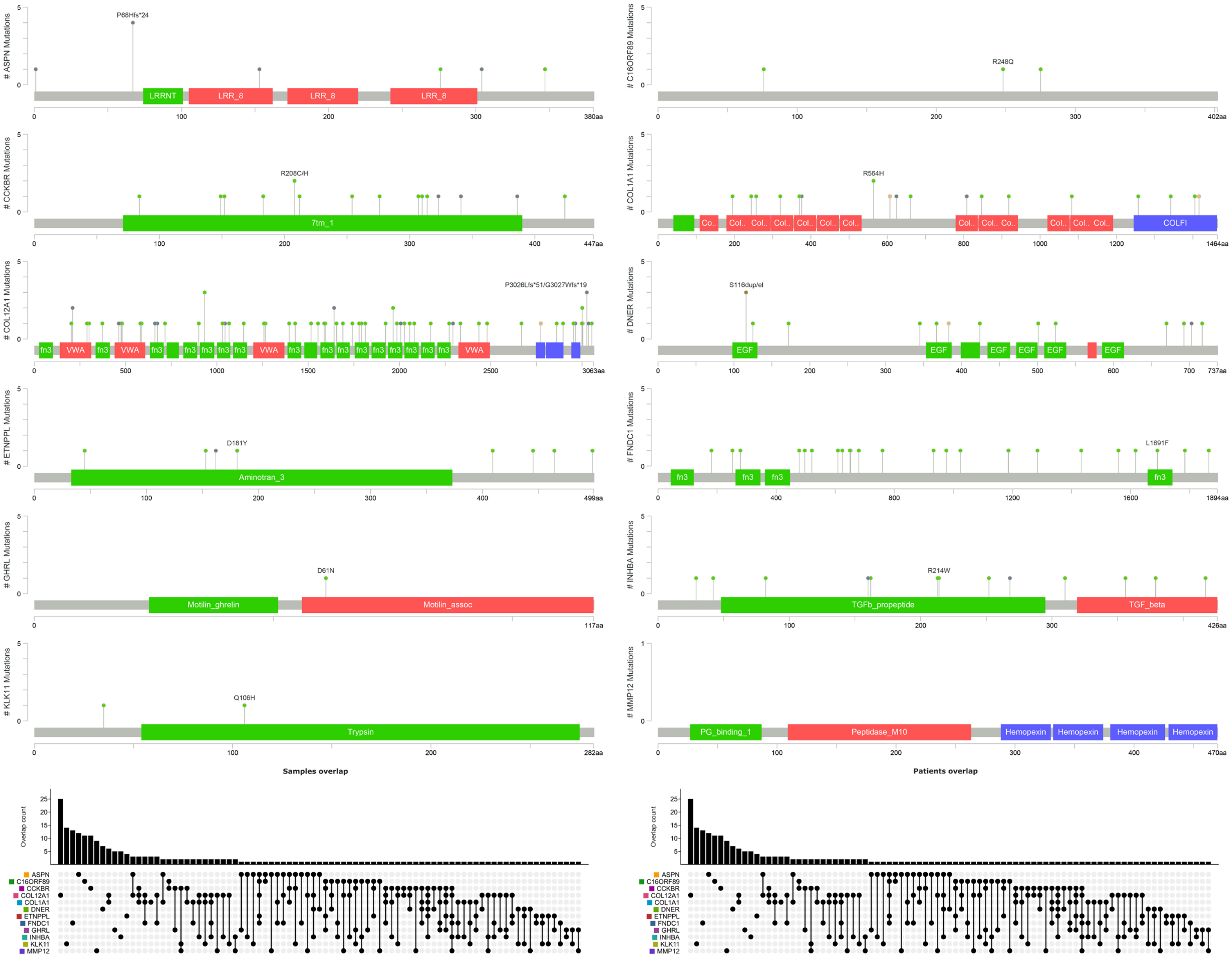
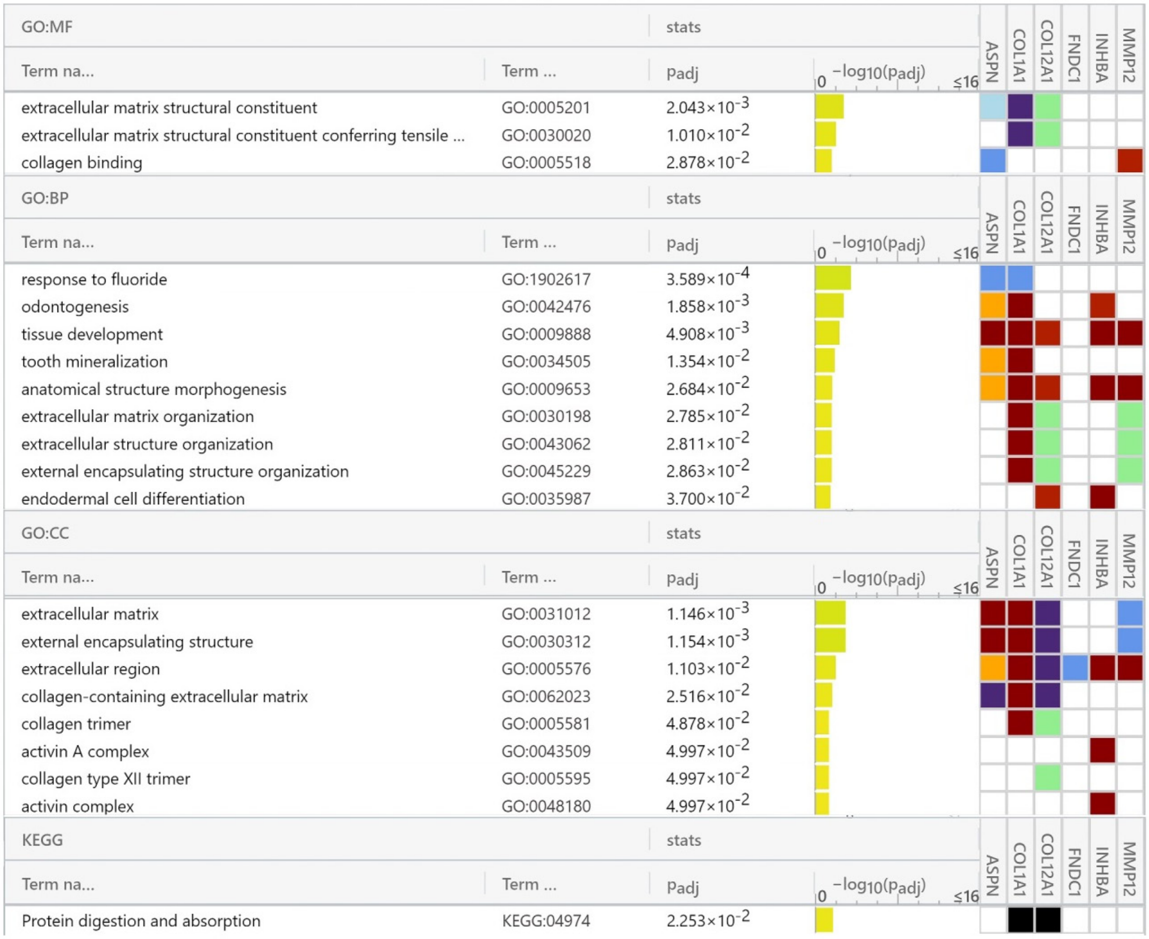


Figure S3. The mutation diagram of 12 DEGs (ASPN, C16orf89, CCKBR, COL1A1, COL12A1, DNER, ETNPPL, FNDC1, GHRL, INHBA, KLK11 and MMP12). The mutation data was obtained from TCGA. The overlap diagram of patients or samples were listed at the bottom.



GO:MF		stats							
Term na...	Term ...	padj	0 −log10(padj) ≤16	ASPN	COL1A1	COL12A1	FNDC1	INHBA	MMP12
extracellular matrix structural constituent	GO:0005201	2.043×10^{-3}		🟦	🟪	🟩			
extracellular matrix structural constituent conferring tensile ...	GO:0030020	1.010×10^{-2}			🟪	🟩			
collagen binding	GO:0005518	2.878×10^{-2}		🟦					🟥

GO:BP		stats							
Term na...	Term ...	padj	0 −log10(padj) ≤16	ASPN	COL1A1	COL12A1	FNDC1	INHBA	MMP12
response to fluoride	GO:1902617	3.589×10^{-4}		🟦	🟦				
odontogenesis	GO:0042476	1.858×10^{-3}		🟧	🟥			🟥	
tissue development	GO:0009888	4.908×10^{-3}		🟥	🟥	🟥		🟥	🟥
tooth mineralization	GO:0034505	1.354×10^{-2}		🟧	🟥				
anatomical structure morphogenesis	GO:0009653	2.684×10^{-2}		🟧	🟥	🟥		🟥	🟥
extracellular matrix organization	GO:0030198	2.785×10^{-2}			🟥	🟩			🟩
extracellular structure organization	GO:0043062	2.811×10^{-2}			🟥	🟩			🟩
external encapsulating structure organization	GO:0045229	2.863×10^{-2}			🟥	🟩			🟩
endodermal cell differentiation	GO:0035987	3.700×10^{-2}				🟥		🟥	

GO:CC		stats							
Term na...	Term ...	padj	0 −log10(padj) ≤16	ASPN	COL1A1	COL12A1	FNDC1	INHBA	MMP12
extracellular matrix	GO:0031012	1.146×10^{-3}		🟥	🟥	🟪			🟦
external encapsulating structure	GO:0030312	1.154×10^{-3}		🟥	🟥	🟪			🟦
extracellular region	GO:0005576	1.103×10^{-2}		🟧	🟥	🟪	🟦	🟥	🟥
collagen-containing extracellular matrix	GO:0062023	2.516×10^{-2}		🟪	🟥	🟪			
collagen trimer	GO:0005581	4.878×10^{-2}			🟥	🟩			
activin A complex	GO:0043509	4.997×10^{-2}						🟥	
collagen type XII trimer	GO:0005595	4.997×10^{-2}				🟩			
activin complex	GO:0048180	4.997×10^{-2}						🟥	

KEGG		stats							
Term na...	Term ...	padj	0 −log10(padj) ≤16	ASPN	COL1A1	COL12A1	FNDC1	INHBA	MMP12
Protein digestion and absorption	KEGG:04974	2.253×10^{-2}			⬛	⬛			

Figure S4. The function enrichment analysis of COL1A1, COL12A1, ASPN, FNDC1, INHBA and MMP12 in GC. The MF, BP, CC and KEGG pathway were list with term name, ID, *p*-value and enriched genes. The significance threshold was g.SCS threshold.