The TCGA and GEO databases identify COL5A2 as a poorly prognostic gene in patients with advanced gastric cancer

Yu-en Tan

China medical university hospital

Qing-chuan Chen

China Medical University Hospital

Yao Xing

china medical university

Chao Zhang

China Medical University Hospital

Si-wei Pan

China Medical University Hospital

Wen An

China Medical University Hospital

Huimian Xu（✉ xuhuimian@126.com ）

China Medical University Hospital   https://orcid.org/0000-0002-1177-4445

Primary research

Abstract

Background

Gastric cancer (GC) metastasis determines the prognosis of patients, and exploring the molecular mechanism of GC metastasis is expected to provide a theoretical basis for clinical treatment. Recent studies have shown that extracellular matrix protein is closely related to GC metastasis. This study aimed to explore the expression profile and role of COL5A2 (Collagen V-type α2), as an extracellular matrix protein, in GC.

Methods

The expression, overall survival and progression-free survival data of COL5 family members were extracted from The Cancer Genome Atlas(TCGA)database, respectively. Paraffin immunohistochemistry and RT-qPCR in GC and matched normal tissues were used to analyze the expression of the target genes. Weighted gene co-expression network analysis of the GSE62229 database was performed out to identify modules and associated genes, and the functions were predicted by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses.

Results

COL5A2 was selected as our research target in the TCGA database, and was also verified in the GSE62229 and GSE15459 datasets. COL5A2 was upregulated in GC tissues by paraffin immunohistochemistry and RT-qPCR. The analysis of clinicopathological characteristics showed a significant difference in the Borrmann type (P = 0.036), histological type (P = 0.013) and T stage(P = 0.011). The prognosis of patients with low COL5A2 expression was better than that of patients with high COL5A2 expression. GSEA results showed that the TCGA and GSE62229 samples were significantly enriched in several well-known cancer-related pathways, such as the TGF-β, MAPK, and JAK2 signaling pathways.

Conclusion

COL5A2 was most closely related to advanced GC among COL5 family members. High COL5A2 expression is associated with a poor prognosis in GC, and may be a novel therapeutic target for GC.

Background

Gastric cancer (GC) is a common malignant tumor of the digestive tract, and the global incidence and mortality of GC are ranked fifth and second, respectively [1, 2]. Presently, the survival of GC patients has been significantly prolonged by the combination of radical surgery with radiotherapy and chemotherapy;

however, the prognosis of advanced or metastatic patients remains unsatisfactory [3, 4]. Because the symptoms of early GC lack specificity, most of the patients are diagnosed in the middle and late stages. Thus, identifying abnormally expressed genes in GC and intervening are important strategies to prolong the survival time of GC patients.

Collagen is the main component of the extracellular matrix (ECM), which can be divided into types I–V [5]. Type V collagen (COL5), an important component of the ECM, can regulate the diameter of fibers by interacting with type I collagen during fiber development [6]. The COL5 family comprises three main isomers, with three different polypeptide α chains, A1, A2 and A3. The abnormal expression of the COL5 family in tumors affects malignancy and progression, but the clinical role and molecular mechanism of the COL5 family in GC remain unclear [7–9].

Previously, high-throughput bioinformatics approaches, such as gene chip and gene sequencing, have been widely used to identify cancer biomarkers [10]. Some high-throughput storage databases are publicly available [11, 12], and investigators can reuse these databases for data mining according to their study design. gene co-expression network analysis (WGCNA) is a powerful biology method to analyze the correlation patterns among genes in RNA-seq or microarray samples [13, 14]. The method clusters highly correlated genes into the same module and connects them with clinical traits, which may be more conducive to the identification of clinical biomarkers for diagnosis and treatment. This method has been generally recognized in cancer research and has successfully identified targeting modules and hub genes [15, 16].

In the current study, we performed The Cancer Genome Atlas (TCGA) analysis on expression, overall survival (OS), and progression-free survival (PFS) microarray data to identify the COL5A family gene that is significantly associated with GC metastasis. Moreover, we explored the related genes and predicted the pathway through WGCNA analysis of GSE62229 database.

Materials And Methods

Data Sources and Data Preprocessing

The TCGA Stomach Adenocarcinoma (STAD) data set contains 408 cancer cases and 211 matched paracancerous tissues. We used GEPIA (http://gepia.cancer-pku.cn/) to compare the gene expression differences, OS and PFS of COL5A1, COL5A2 and COL5A3 in TCGA, so as to select the most significant different expression genes (DEGs) of COL5 family.

The pretreated expression profiles of the GSE62229 and GSE15459 datasets,with a high quality and quantity of GC cases, were downloaded from the GEO database. The OS and PFS of the two patient databases were detected using the K-M plotter [17]. GSE62229 is a microarray dataset containing 300 cancer tissue samples and 100 cases of paracancerous tissues, and its clinical characteristics are very complete. This dataset selected as the training data for further study.

Samples and patients

We used 48 pairs of fresh specimens and adjacent non-cancerous tissues from the First Affiliated Hospital of China Medical University in 2018. We also used 126 paraffin-embedded GC tissues and 60 adjacent normal tissues from patients treated between 2011 and 2012. All the patients were confirmed to have gastric adenocarcinoma pathoogically, no tumor was found in other regions, and no radiotherapy or chemotherapy was performed before the operation. The patients or their families sign informed consent. This study was approved by the research ethics committee of our institute.

Screening of DEGs

The R software based on the "Limma" R package was used to screen the DEGs between GC tissue and adjacent normal tissue for GSE62229. A false discovery rate (FDR) < 0.05 and |log 2 (FC)| ≥0.263 were regarded as the cut-off thresholds.

Construction of the Co-expression network

After determining the DEGs' expression data from the GSE62229 dataset, a co-expression network was conducted for downstream analysis using the "WGCNA" R package. WGCNA could effectively combine gene expression information with the clinicopathological features to identify potential modules. Next, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEEG) enrichment analyses were used to assess the functional role of the module genes based on R software [18, 19].

Gene Set Enrichment Analysis (GSEA)

To determine the possible pathway through which COL5A2 functions in the development of GC, the expression data from GSE62229 and TCGA were also used to perform Gene Set Enrichment Analysis (GSEA) [20, 21]. According to the differences in expression, the database cases were uniformly divided into low-expression and high-expression groups.

Real time quantitative PCR (RT-qPCR) analysis

The tissues were cut and homogenized. After extracting the total RNA according to the instructions, cDNA templates were generated by reverse transcription by PrimeScriptTM RT kit (TaKaRa, Japan). Real-time polymerase chain reaction was performed to calculate relative expressions of mRNA according to the reaction system. The number of cycles was set to 40. GAPDH was chosen as the reference gene. The primer sequences of COL5A2 were 5'-CAGGCTCCATAGGAATCAGAGG – 3' (sense) and 5'-CCAGCATTTCCTGCTTC TCCAG-3' (antisense).

Immunohistochemistry

Immunohistochemistry (IHC) staining was performed according to standard protocols. IHC staining was assessed by scores based on the percentage of positive cells (0: < 5%; 1: 5–25%; 2: 25–50%; 3: 50–75%; 4: > 75%) multiplied by scores based on the intensity of staining, (0: colorless; 1: light yellow; 2: brown; 3: dark brown), with 6–12 considered high expression and 0–4 considered low expression. The primary

antibody against COL5A2 used in IHC testing was purchased from LifeSpan BioSciences, Inc (Seattle, WA, USA).

Statistical analysis

Statistical analysis was performed using SPSS 22.0 statistical software and GraphPad Prism7.0 mapping software. Student's t- test was used to compare the two groups. The Kaplan–Meier method was used to calculate OS. $P < 0.05$ was considered statistically significant.

Results

COL5A2 is upregulated in GC tissues and correlates with poor survival in the TCGA and GEO databases.

First, TCGA-STAD was used to predict the mRNA expression levels of three major isomers of the COL5 family in GC and adjacent normal tissues. COL5A1 and COL5A2 were up-regulated in GC compared with COL5A3 ($P<0.05$) (Figure 1A). To evaluate the prognostic value of the COL5 family mRNA expression in GC, Kaplan-Meier analysis and the log-rank test were used to verify the relationship between mRNA expression and OS or PFS in GC patients. In patients with high COL5A2 expression, OS and PFS were significantly reduced ($P<0.05$), however, COL5A1 was only have a significant trend in OS ($P=0.12$) and PFS ($P=0.14$) (Figure 1B and C). Analysis of T stage showed that COL5A2 expression in advanced GC was significantly higher than that in early GC (Figure 1E). The above analysis showed that high COL5A2 expression indicated a poor prognosis of GC. Therefore, we chose COL5A2 for further exploration (Figure 1D).

To verify the findings in the TCGA database, the GSE62229 and GSE15459 datasets were selected to evaluate the expression and prognosis of COL5A2. COL5A2 expression in cancer tissue was significantly higher than that in adjacent normal tissues ($P<0.001$) (Figure 2B). Additionally, in the two GEO databases, patients with low COL5A2 expression showed longer OS and PFS (Figure 2A and C).

High COL5A2 expression indicates a poor prognosis in GC tissues

To validate the possible role of COL5A2 in GC progression, the expression pattern of COL5A2 was explored in paired clinical tissue samples in our patient samples. Thus, 126 paraffin-embedded GC tissues and 60 adjacent normal tissues with complete clinicopathological variable and follow-up information were collected. The COL5A2 protein level was significantly higher in GC tissues than in normal tissues ($P<0.001$; Figure 3A, 3B). Next, we used RT-qPCR to assess the expression pattern of COL5A2 in 48 pairs of fresh specimens and adjacent non-cancerous tissues (Figure 3C); the findings were consistent with the IHC results. Taken together, these results confirmed that COL5A2 is highly expressed in GC tissues.

Next, the prognostic role of COL5A2 was confirmed in our samples. Based on the COL5A2 expression levels, patients with complete follow-up information were divided into the COL5A2 low-expression group (negative or weakly positive expression, n=64) and COL5A2 high expression group (moderately or

strongly positive expression, n=64). Kaplan-Meier curves confirmed that patients with high COL5A2 expression had a significantly shorter OS than those with low COL5A2 expression (P=0.0085, Figure 3D). Additionally, we verified the significance of COL5A2 in the survival of advanced GC (P=0.018; Figure 3E).

The association between COL5A2 expression and clinicopathological parameters in patients with GC was further evaluated. As shown in Table 1, COL5A2 expression in GC was correlated with Borrmann type (P=0.036), histological type (P=0.013), and T stage (P<0.011). A significant correlation was not found between COL5A2 and age, sex, tumor location, tumor size, or N stage. These results confirmed that COL5A2 expression is associated with the malignant phenotype of GC.

Weighted co-expression network construction and module identification

After quality evaluation and data preprocessing, an expression matrix was formed from the 298 GC samples of the GSE62229 dataset. The clinical traits were shown in the heatmap of the clustering dendrogram (Figure 4A). With the variance in the top 25%, 5407 genes were screened out and used for subsequent co-expression analysis. When choosing the soft threshold, we calculated the network topology with power values from 1 to 20. As shown in Figure 4B, the power value of 3, which was the lowest power of the scale-free topological fit index of 0.9, was pitched on. Additionally, the mean connectivity met the scale-free network distribution at the power value of 3. After merging similar clusters, thirteen different modules were identified that contained groups of genes with similar connection strengths (Figure 4C).

Finally, we found that COL5A2 was enriched into the salmon module (Figure 5A). and was highly correlated with T stage and Lauren stage (Figure 5B, r = 0.32, P=3e-8 and r = 0.31, P=4e-8). Interestingly, the salmon module was also found to be related to pStage (r = 0.23, P=8e-5) and survival status (r = 0.23, P=9e-5). Additionally, we selected the top 100 genes related to COL5A2 and constructed a visualized network using Cytoscope software (Figure 5C).

Functional Annotation and GSEA in the GSE62229 dataset and TCGA database

To understand the biological correlation of COL5A2, GO enrichment and KEGG pathway analyses were carried out. The top GO terms are shown in Figure 6A. The most enriched GO terms were as follows: BP (biological process), such as the extracellular matrix and structure organization, epithelial cell proliferation, and cell- substrate adhesion, CC (cellular component) such as the extracellular matrix, endoplasmic reticulum lumen, collagen trimer, and basement membrane, and MF (molecular function) such as cell adhesion molecular binding, glycosaminoglycan binding, and growth factor binding. Additionally, these genes were mainly enriched in the PI3K-Akt signaling pathway and focal adhesion, suggesting that the tumor microenvironment plays an important role in metastasis development (Figure 6B).

We performed GSEA of the GSE62229 dataset and TCGA database which revealed that COL5A2 was enriched in focal adhesion, ECM receptor interaction and regulation of actin cytoskeleton (Supplementary

Figure S1). The GSEA results also showed that metastasis samples were significantly enriched in several well-known cancer-related pathways, such as the TGF-β, MAPK and JAK2 signaling pathways (Figure 7A, B). The results provide clues into the in-depth mechanism of metastasis development.

Discussion

GC is a biologically and pathologically heterogeneous disease [22]. The prognosis of advanced GC has shown little improvement, and it is necessary to identify efficient prognostic biomarkers and therapeutic targets. In the current study, we first focused on the COL5 family, and chose COL5A2 as our target, according to the expression, OS and PFS data of the TCGA database. The analyses showed that COL5A2 was associated with T stage and Lauren stage and is involved in cancer-related pathways.

The expression level of COL5A2 is increased in various types of cancers, such as pancreatic cancer and colon cancer [9, 23]. The upregulation of COL5A2 is correlated with a poor prognosis in tongue cancer [8], a finding that was consistent with ours. Moreover, higher COL5A2 expression was associated with the Borrmann type, histological type and T stage in the GC samples of our department, suggesting that COL5A2 might be a potential biomarker for GC tumorigenesis and progression.

WGCNA is a method that can highlight functional co-expression gene modules, and plays an important role in determining the potential mechanisms of malignancies, including breast cancer and colon cancer [16, 24]. One main advantage of our study was that the WGCNA model of GSE62229 was constructed identify the module of COL5A2, and further explore the role of COL5A2 in GC. Eventually, we found that COL5A2 was enriched in the salmon module and was associated with T stage and Lauren stage, findings that are consistent with our IHC data. However, our study possessed the limitation of a small sample size and more databases need to be incorporated into future research.

Disorders of functions and cancer-related pathways are common in cancers [25, 26]. Regarding GO and KEGG enrichment analyses, COL5A2 was involved in the extracellular matrix, focal adhesion, and PI3K-Akt signaling pathway. During cancer cell migration, Paluch et al [27] proposed that adhesion to the matrix through a specific site is an essential step. Additionally, the PI3K-Akt signaling pathway plays an important role in cell migration, angiogenesis, and survival in GC [28, 29]. In GSEA enrichment, cancer-related pathways, such as the TGF-β, MAPK and JAK2 signaling pathways, were significantly identified. Notably, our previous study showed that TGF-β was an independent factor of the peritoneal metastasis of GC [30]. These results reveal the deeper mechanism of COL5A2 in the metastasis development of GC.

Conclusions

In conclusion, we aimed to select a COL5 family member with expression and survival significance and identified its potential molecular mechanism in advanced GC using bioinformatics analyses and clinical samples. Eventually, we used the TCGA database to select COL5A2 as our research target. WGCNA showed that COL5A2 was enriched in the salmon module, which was connected with the T stage and

Lauren stage. Functional annotation demonstrated COL5A2 might be involved in the formation of the extracellular matrix, focal adhesion, and some cancer-related pathways. However, because this study is mainly based on the analysis of open available datasets and clinical samples, further detailed experimental studies are needed to confirm the results in the future.

Declarations

Authors' contributions Tan and Xing work at paper writing, and Chen and Pan work at data analysis. Zhang and An is responsible for data download. Xu provides ideas of research.

Consent of publication: Not applicable.

Competing interests: The authors declare that they have no competing interest.

Ethics approval and consent to participate: The experiment was approved by the Medical Ethics Research Association of the first affiliated Hospital of China Medical University, and each GC patient signed a written informed consent form.

Availability of data and materials: All data generated or analyzed during this study are included in this article.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018; 68: 394-424.

2. Zheng R, Zeng H, Zhang S, Chen W. Estimates of cancer incidence and mortality in China, 2013. Chin J Cancer. 2017; 36: 66.

3. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, Scarffe JH, Lofts FJ, Falk SJ, Iveson TJ, Smith DB, Langley RE, Verma M, et al. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. N Engl J Med. 2006; 355: 11-20.

4. Bang YJ, Kim YW, Yang HK, Chung HC, Park YK, Lee KH, Lee KW, Kim YH, Noh SI, Cho JY, Mok YJ, Kim YH, Ji J, et al. Adjuvant capecitabine and oxaliplatin for gastric cancer after D2 gastrectomy (CLASSIC): a phase 3 open-label, randomised controlled trial. Lancet. 2012; 379: 315-21.

5. Walma DAC, Yamada KM. The extracellular matrix in development. 2020; 147.

6. Sun M, Chen S, Adams SM, Florer JB, Liu H, Kao WW, Wenstrup RJ, Birk DE. Collagen V is a dominant regulator of collagen fibrillogenesis: dysfunctional regulation of structure and function in a corneal-
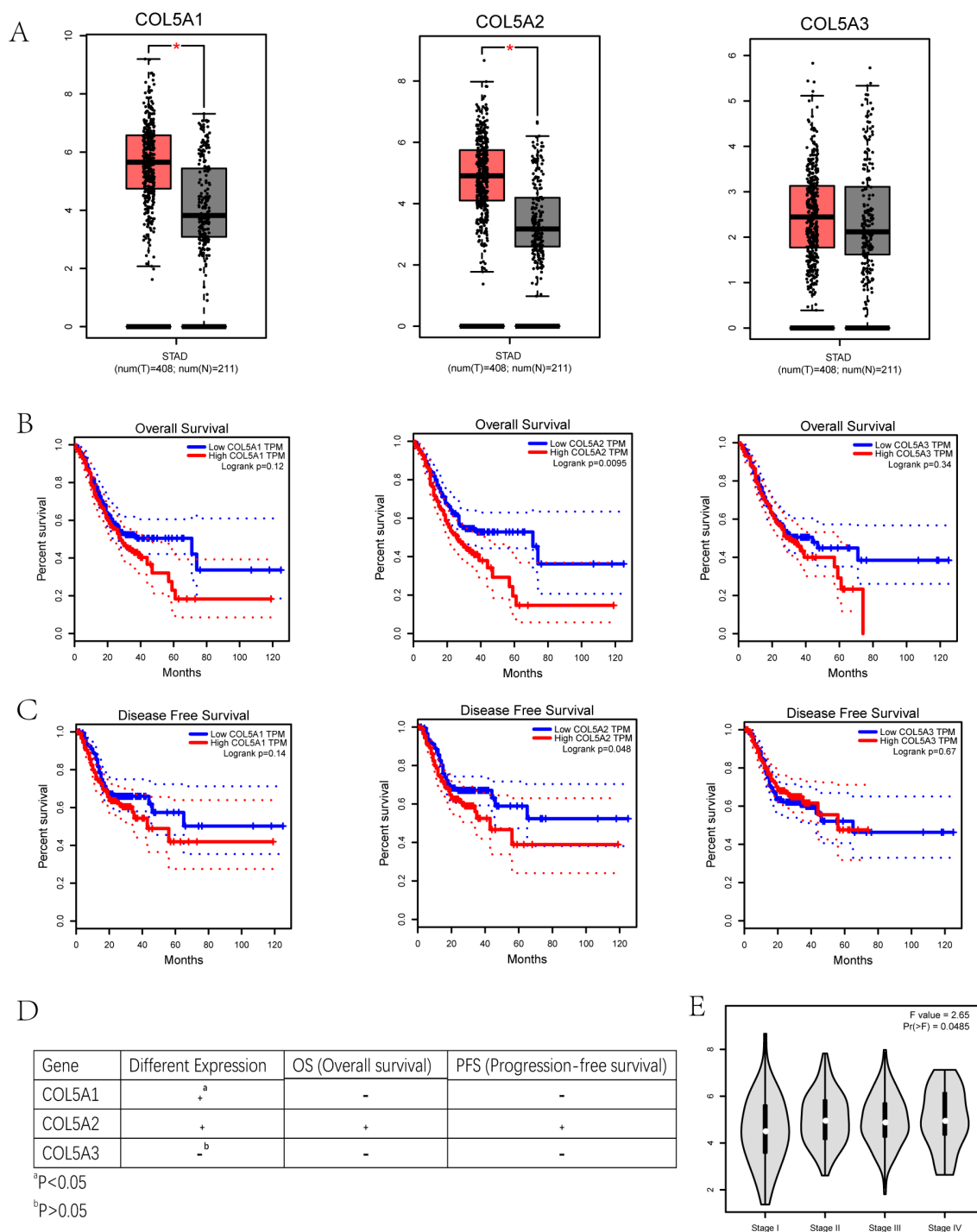
stroma-specific Col5a1-null mouse model. J Cell Sci. 2011; 124: 4096-105.

7. Meng XY, Shi MJ, Zeng ZH, Chen C, Liu TZ, Wu QJ, Li S, Li S. The Role of COL5A2 in Patients With Muscle-Invasive Bladder Cancer: A Bioinformatics Analysis of Public Datasets Involving 787 Subjects and 29 Cell Lines. Front Oncol. 2018; 8: 659.

8. Chen HC, Tseng YK, Shu CW, Weng TJ, Liou HH, Yen LM, Hsieh IC, Wang CC, Wu PC, Shiue YL, Fu TY, Tsai KW, Ger LP, et al. Differential clinical significance of COL5A1 and COL5A2 in tongue squamous cell carcinoma. 2019; 48: 468-76.

9. Berchtold S, Grünwald B, Krüger A, Reithmeier A, Hähl T, Cheng T, Feuchtinger A, Born D, Erkan M, Kleeff J, Esposito I. Collagen type V promotes the malignant phenotype of pancreatic ductal adenocarcinoma. Cancer Lett. 2015; 356: 721-32.

10. Dupain C, Gracia C, Harttrampf AC, Rivière J, Geoerger B, Massaad-Massade L. Newly identified LMO3-BORCS5 fusion oncogene in Ewing sarcoma at relapse is a driver of tumor progression. Oncogene. 2019; 38: 7200-15.

11. Blum A, Wang P, Zenklusen JC. SnapShot: TCGA-Analyzed Tumors. Cell. 2018; 173: 530.

12. Jia D, Li S, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. Aging (Albany NY). 2018; 10: 592-605.

13. Zhang H, Guo L, Zhang Z, Sun Y, Kang H, Song C, Liu H, Lei Z, Wang J, Mi B, Xu Q, Guan H, Li F. Co-Expression Network Analysis Identified Gene Signatures in Osteosarcoma as a Predictive Tool for Lung Metastasis and Survival. J Cancer. 2019; 10: 3706-16.

14. Chen L, Peng T, Luo Y, Zhou F, Wang G, Qian K, Xiao Y, Wang X. ACAT1 and Metabolism-Related Pathways Are Essential for the Progression of Clear Cell Renal Cell Carcinoma (ccRCC), as Determined by Co-expression Network Analysis. Front Oncol. 2019; 9: 957.

15. Xiao H, Chen P, Zeng G, Xu D, Wang X, Zhang X. Three novel hub genes and their clinical significance in clear cell renal cell carcinoma. J Cancer. 2019; 10: 6779-91.

16. Lv Y, Xie B, Bai B, Shan L, Zheng W, Huang X, Zhu H. Weighted gene coexpression analysis indicates that PLAGL2 and POFUT1 are related to the differential features of proximal and distal colorectal cancer. Oncol Rep. 2019; 42: 2473-85.

17. Nagy Á, Lánczky A, Menyhárt O, Győrffy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. 2018; 8: 9227.

18. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25: 25-9.

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

20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102: 15545-50.

21. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003; 34: 267-73.

22. Serra O, Galán M, Ginesta MM, Calvo M, Sala N, Salazar R. Comparison and applicability of molecular classifications for gastric cancer. Cancer Treat Rev. 2019; 77: 29-34.

23. Fischer H, Stenling R, Rubio C, Lindblom A. Colorectal carcinogenesis is associated with stromal expression of COL11A1 and COL5A2. Carcinogenesis. 2001; 22: 875-8.

24. Lan L, Xu B, Chen Q, Jiang J, Shen Y. Weighted correlation network analysis of triple-negative breast cancer progression: Identifying specific modules and hub genes based on the GEO and TCGA database. Oncol Lett. 2019; 18: 1207-17.

25. Bu L, Baba H, Yoshida N, Miyake K, Yasuda T, Uchihara T, Tan P, Ishimoto T. Biological heterogeneity and versatility of cancer-associated fibroblasts in the tumor microenvironment. 2019; 38: 4887-901.

26. Lin Y, Liang R, Qiu Y, Lv Y, Zhang J, Qin G, Yuan C, Liu Z, Li Y, Zou D, Mao Y. Expression and gene regulation network of RBM8A in hepatocellular carcinoma based on data mining. Aging (Albany NY). 2019; 11: 423-47.

27. Paluch EK, Aspalter IM, Sixt M. Focal Adhesion-Independent Cell Migration. Annu Rev Cell Dev Biol. 2016; 32: 469-90.

28. Hu M, Zhu S, Xiong S, Xue X, Zhou X. MicroRNAs and the PTEN/PI3K/Akt pathway in gastric cancer (Review). Oncol Rep. 2019; 41: 1439-54.

29. Liu JY, Jiang L, He T, Liu JJ, Fan JY, Xu XH, Tang B, Shi Y, Zhao YL, Qian F, Wang Y, Cui YH, Yu PW. NETO2 promotes invasion and metastasis of gastric cancer cells via activation of PI3K/Akt/NF-κB/Snail axis and predicts outcome of the patients. Cell Death Dis. 2019; 10: 162.

30. Yin S, Miao Z, Tan Y, Wang P, Xu X, Zhang C, Hou W, Huang J, Xu H. SPHK1-induced autophagy in peritoneal mesothelial cell enhances gastric cancer peritoneal dissemination. 2019; 8: 1731-43.

Table

Please see the supplementary files section to view the table.

Figures

Figure 1

Expression and survival analysis of the COL5 family in the TCGA-STAD cohort. (A) Box plot of the expression levels of the three genes in GC from the TCGA database. The x-axis shows the number of GC samples and normal samples, and the y-axis shows the gene expression levels. The P value was determined using Student's t-test, and error bars were represented as means ± s.d. (B) and (C) Kaplan-Meier survival curves of the OS and PFS for three genes were plotted. The P value was determined using

the log-rank test. (D) COL5A2 was found to be statistically significant in both expression and survival in patient samples, compared with COL5A1 and COL5A3. (E) Relationship between COL5A2 and T stage based on the TCGA-STAD data. P<0.05 represents statistical significance.
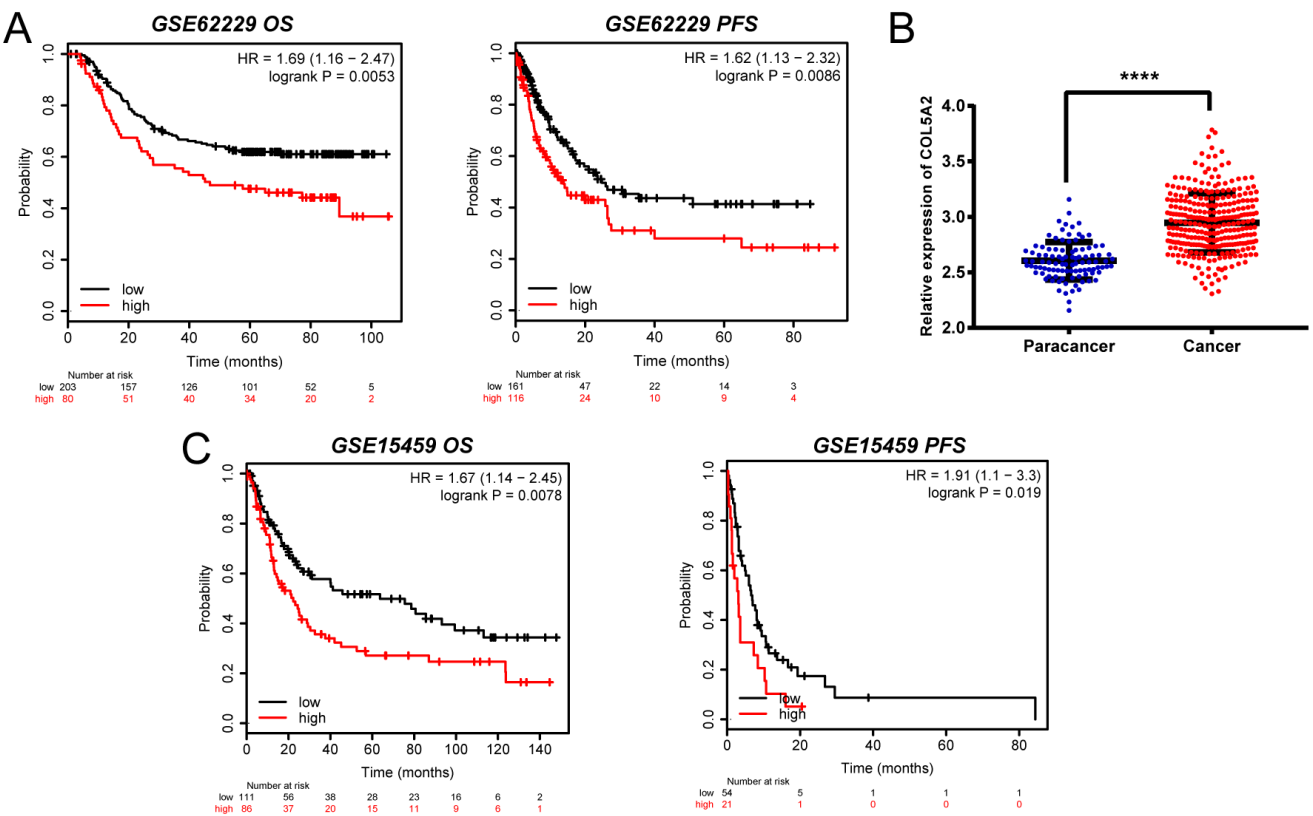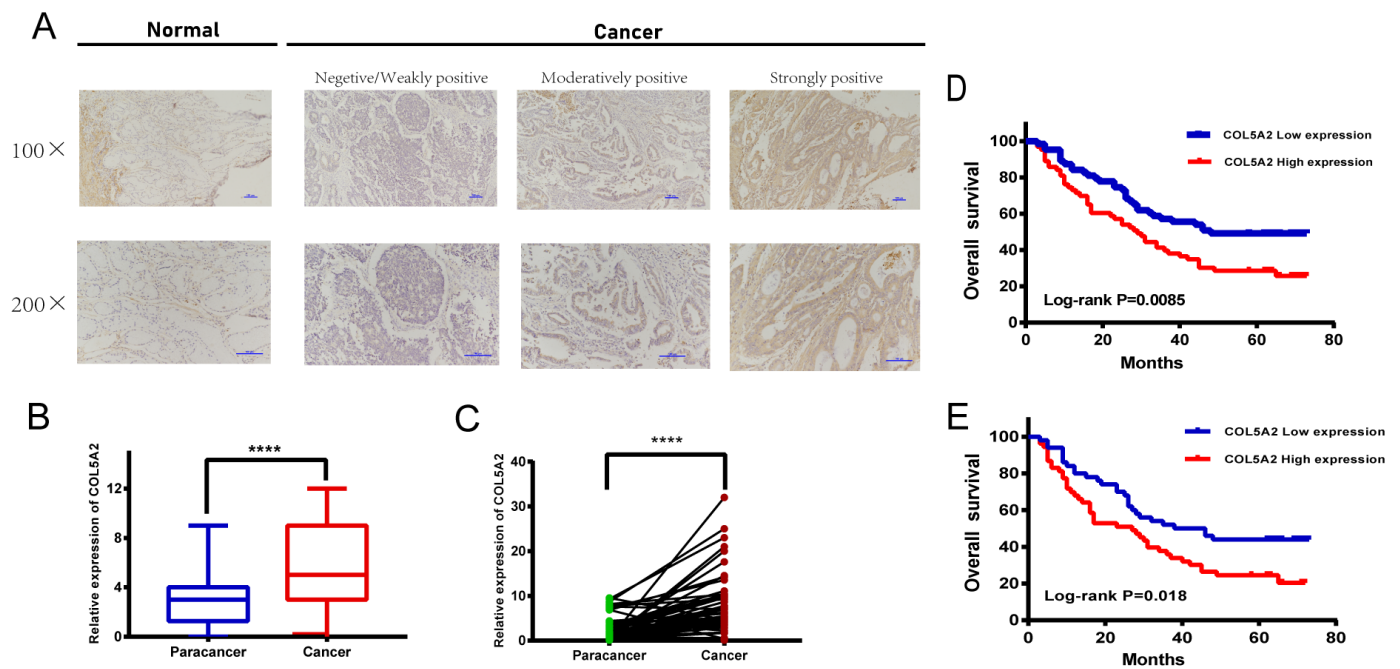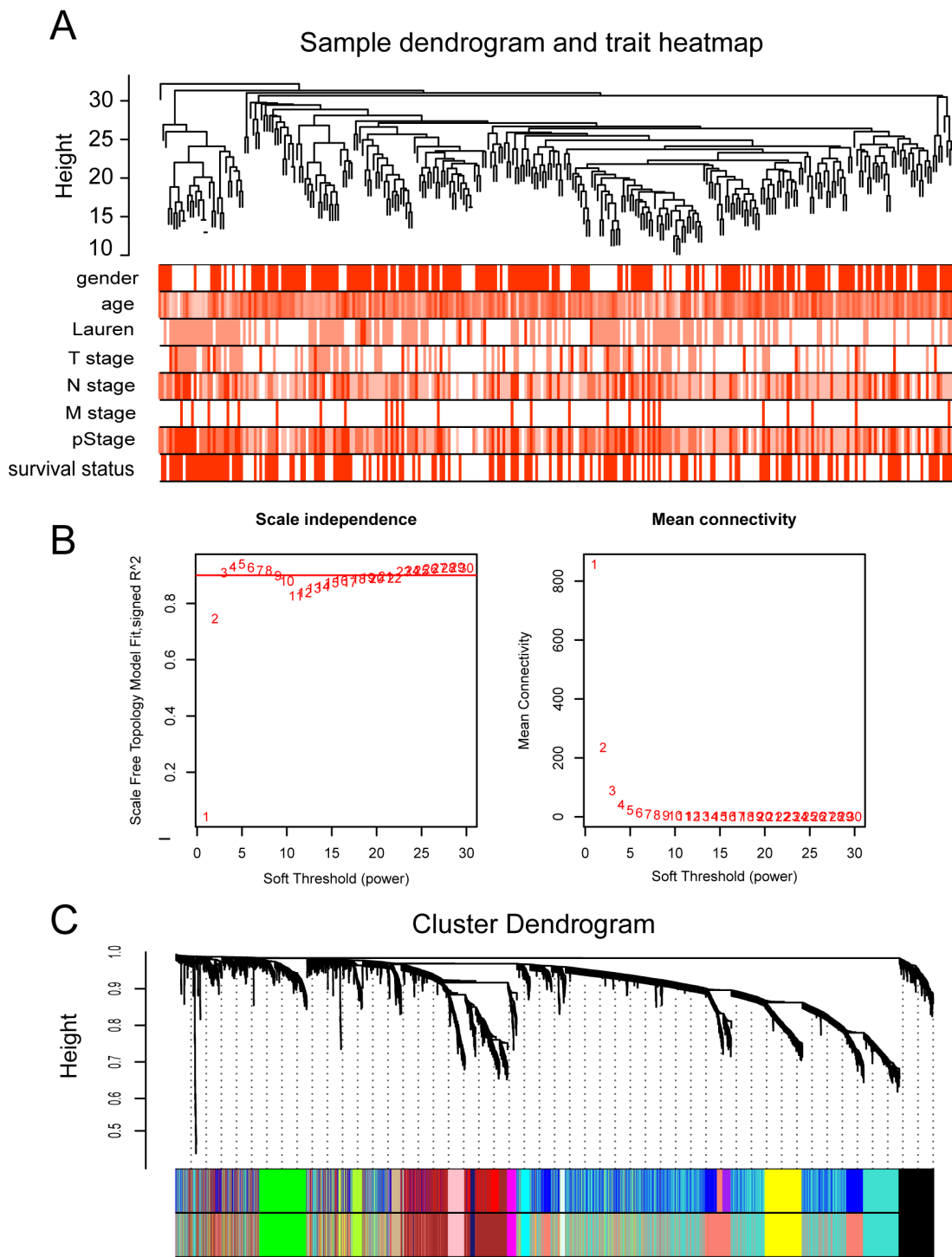

Figure 2

COL5A2 is upregulated in GC tissues and correlates with poor survival. (A) and (C) The K-M plotter shows the survival curves of COL5A2 in GSE62229 and GSE15459. Overall survival=OS, and progression-free survival=PFS. P<0.05 represents statistical significance. (B) The point map shows the COL5A2 expression level in GC tissue (n=200) and paracancerous tissue (n=100) from GSE62229. ****, P<0.001. P<0.05 represents statistical significance.

Figure 3

Detection of COL5A2 protein and mRNA expression in patient tissues. (A) Representative images of COL5A2 protein expression in GC and normal tissues. IHC scoring based on the staining intensity was performed. The scale bar is displayed. (B) COL5A2 protein expression was significantly increased in primary tumor tissues (n = 126), compared with that in adjacent normal tissues (n = 60) by IHC. (C) RT-qPCR detected the COL5A2 mRNA level of the 48 paired fresh specimens and adjacent non-tumor tissues. (D) Kaplan-Meier analysis of COL5A2 in GC patients. Patients with low COL5A2 expression had longer OS than those with high COL5A2 expression (P =0.0085). (E) Kaplan-Meier analysis of COL5A2 in advanced GC patients (P =0.018). ****, P<0.001. P<0.05 represents statistical significance.
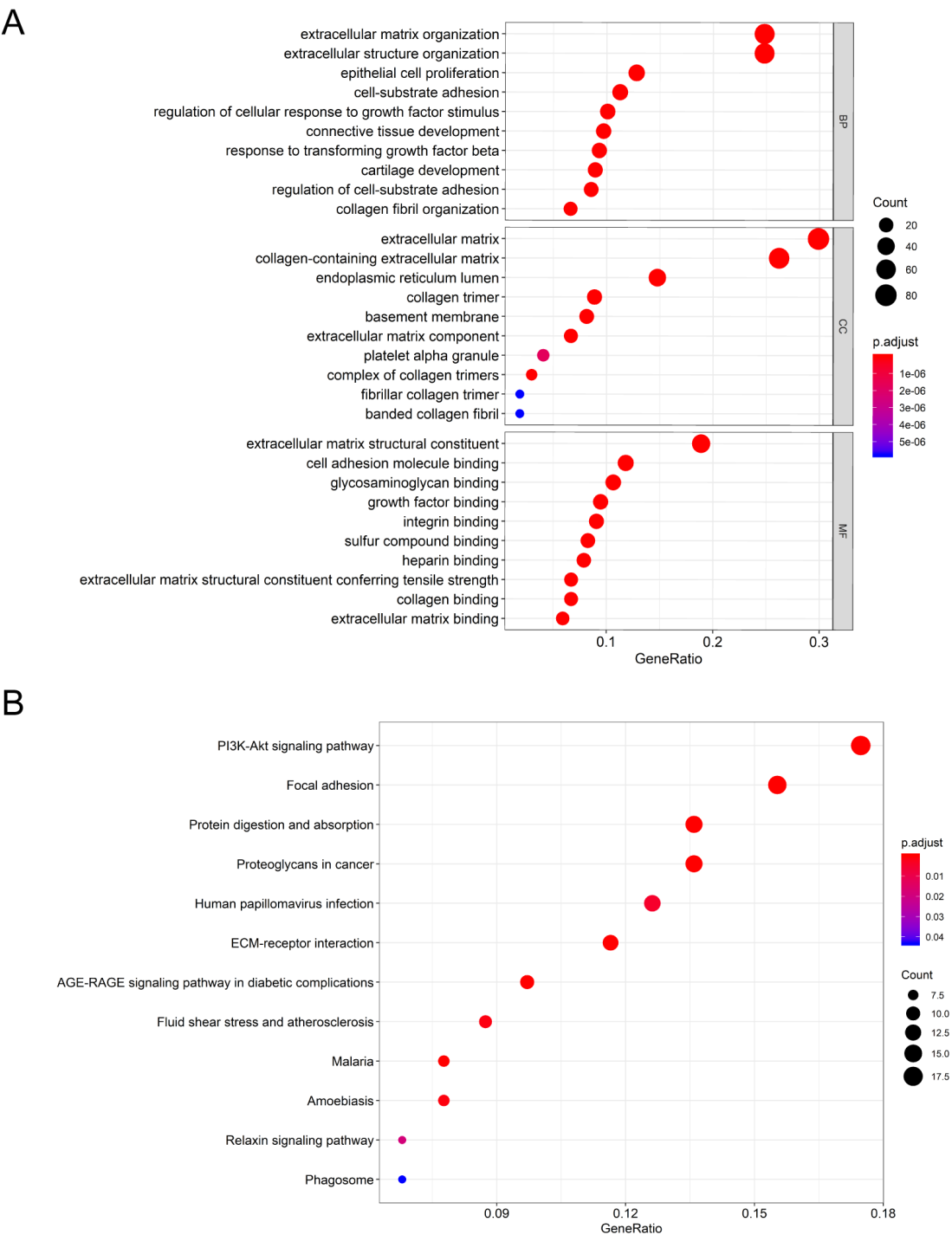
Figure 4

Construction of the weighted gene co-expression network. (A) Clustering dendrogram of the clinical traits showed no obvious exception value. (B) Network topology analysis showed that when β=3, the network topology met the threshold of 0.9 scale-free topology. The scale-free fit index in the left panel represents the soft-threshold power, and the mean connectivity in the right panels represents the soft-threshold

power. (C) Clustering dendrogram of genes based on topological overlap. Each module served as a cluster of co-associated genes and was assigned a unique color.

Figure 5

Analysis of the weighted gene co-expression network. (A) Heatmap of the correlation and significant differences between the gene modules and clinical traits of GC. The correlation is displayed in the rectangle, while significant differences are shown in parentheses. (B) Two scatterplots of Gene

Significance (GS) for the Lauren type (upper) and T stage (lower) vs. Module Membership (MM) in the salmon module. (C) Protein-protein interactions (PPI) network of COL5A2 in the salmon module.
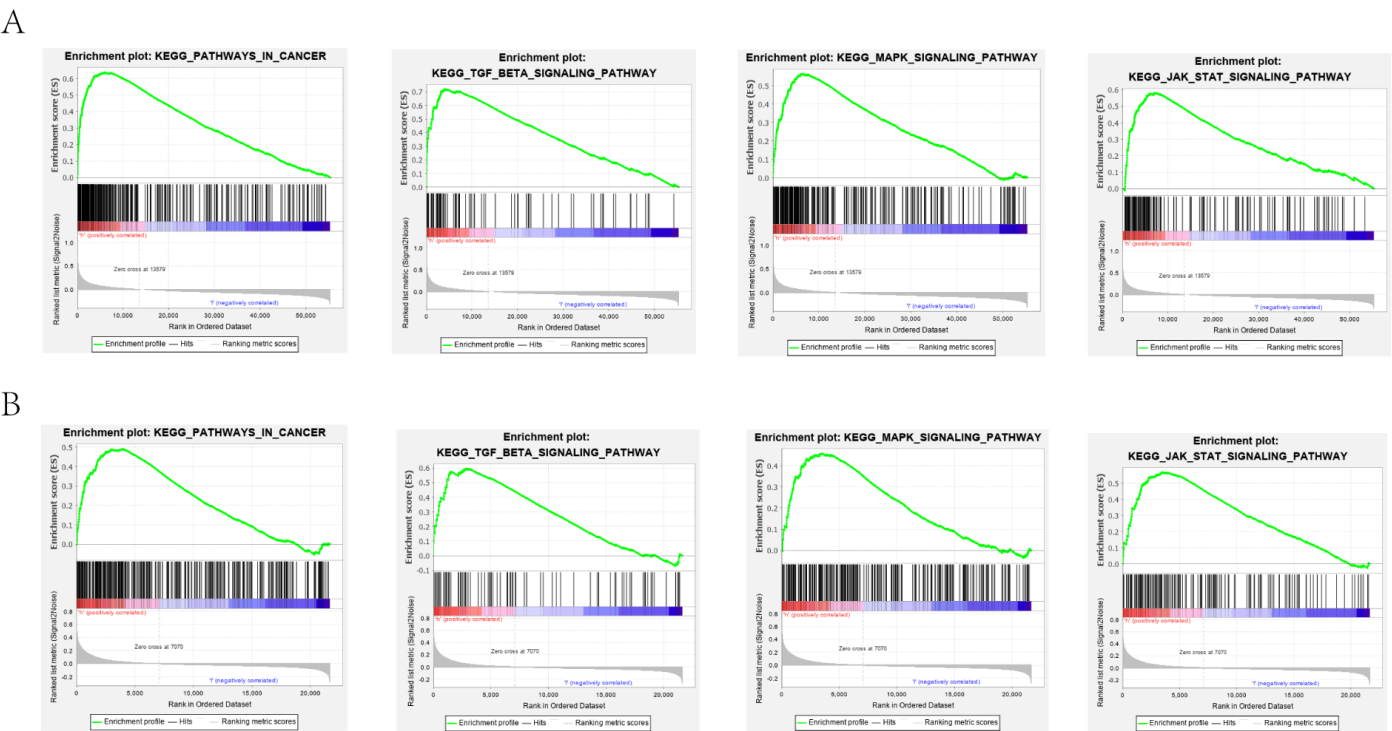

Figure 6

Functional annotation for COL5A2 in the salmon module. (A) Enriched biological processes (BP), cellular components (CC) and molecular functions (MF) of the salmon module. The x-axis displays the ratio of

the total genes, and the y-axis displays the GO terms. (B) Enriched KEGG pathways of the salmon module.

A

B

Figure 7

Gene set enrichment analysis (GSEA) results of cancer-related pathways. (A) GSEA of COL5A2 using the TCGA database. (B) GSEA of COL5A2 using the GSE62229 dataset.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- table1.pdf
- supplement.docx