Original Article

# Could CTSK and COL4A2 be specific biomarkers of poor prognosis for patients with gastric cancer in Asia?—a microarray analysis based on regional population

**Zhijun Feng[1#], Ruili Qiao[2#], Zhijian Ren[1], Xiaofeng Hou[1], Jie Feng[1], Xiaodong He[1], Dongdong Chen[3]**

[1]Department of General Surgery, The Second Clinical Medical College, Lanzhou University, Lanzhou 730030, China; [2]Department of VIP Internal Medicine, Lanzhou University First Hospital, Lanzhou 730000, China; [3]Department of The First General Surgery, Gansu Provincial Hospital, Lanzhou 730000, China

*Contributions:* (I) Conception and design: X He, D Chen; (II) Administrative support: X He, D Chen; (III) Provision of study materials or patients: Z Feng, R Qiao; (IV) Collection and assembly of data: Z Ren, X Hou, J Feng; (V) Data analysis and interpretation: Z Feng, R Qiao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

*Correspondence to:* Xiaodong He. Department of General Surgery, The Second Clinical Medical College, Lanzhou University, No. 82, Cuiyingmen, Chengguan District, Lanzhou 730030, China. Email: hxd@lzu.edu.cn; Dongdong Chen. Department of The First General Surgery, Gansu Provincial Hospital, No. 204, Donggang West Road, Chengguan District, Lanzhou 730000, China. Email: chendd18@lzu.edu.cn.

**Background:** In the purpose of identifying reliable biomarkers for evaluating prognosis, monitoring recurrence and exploring new therapeutic targets, it is quite necessary to screen for the genetic changes and potential molecular mechanisms of the occurrence and development of gastric cancer (GC) from the aspects of race and region.

**Methods:** Target datasets were retrieved from Gene Expression Omnibus (GEO) database with "gastric cancer" as the key word, and corresponding data was downloaded. The differentially expressed genes (DEGs) were obtained by using limma R package, and the Gene Ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway for DEGs were analyzed in Enirchr database. Protein-protein interaction (PPI) network and molecular module were also constructed through STRING database and Cytoscape software. Survival analyses were completed for DEGs in GEO and Kaplan-Meier plotter database via cross validation. Finally, the correlation between gene expression and the infiltration cell levels in tumor microenvironment (TME) was explored based on the tumor immune estimation resource (TIMER) database.

**Results:** Five GC-related microarray datasets were selected and used for differential analysis, and 222 DEGs were identified. GO analyses of DEGs were mainly involved in cell metabolism and the formation of extracellular matrix (ECM). The top enriched pathways of DEGs were protein digestion and absorption, ECM-receptor interaction, focal adhesion (FA), PI3K-Akt signaling pathway. Survival analyses of DEGs revealed that the expression levels of *CTSK* and *COL4A2* were significantly associated with poor prognosis of GC patients in Asian. Specifically, the high expression of *CTSK* had a closely related to the infiltration level of inflammatory cell in TME.

**Conclusions:** *CTSK* and *COL4A2* could play a critical role in the pathogenesis of GC and act as the promising prognostic biomarkers. *CTSK* could induce the formation of immunosuppressive TME and promote the immune escape of GC cells.

**Keywords:** Stomach neoplasms; biomarkers; integrated bioinformatics; microarray analysis

Submitted Dec 12, 2019. Accepted for publication Feb 14, 2020.
doi: 10.21037/jgo.2020.03.01
**View this article at:** http://dx.doi.org/10.21037/jgo.2020.03.01

 

## Introduction

Nowadays, it is widely considered that gastric cancer (GC) is an inflammation-driven disease, Helicobacter pylori and Epstein-Barr (EB) virus infections are thought to be the important risk factors of GC. Chronic inflammation in gastric mucosa leading to changes in microenvironment followed by molecular alterations, causes neoplasia gradually, but the specific mechanisms are still uncertain (1). The proposition of molecular subtypes of GC not only analyzes the molecular changes of GC and its corresponding biological behavior characteristics from genetic level, but also provides a favorable guidance for the selection of anti-tumor drugs (2-4). Other genes used for grouping GC have been focused by several studies in order to guide GC treatment and evaluate prognosis as well, such as human epidermal growth factor receptor 2 (*HER2*) and tumor protein 53 (*TP53*) (5,6). But beyond that, ethnic differences in response to anti-tumor therapies in GC patients are always concerned. For instance, patients from Asia have a better prognosis and response to treatments than Caucasian (7,8); survival differences are independent of clinical and pathological factors among different races and ethnicities patients (9); and significant differences found in the frequencies of somatic mutation from diverse geographic populations (10). Accordingly, GC is highly heterogeneous, and its biological behavior has wide genetic and epigenetic differences between different individuals or between different lesions of the same individual, which might further result in different prognosis and treatment outcome. Therefore, it is quite necessary to explore the genetic changes and potential molecular mechanisms of the occurrence and development of GC from the aspects of race and region. On one hand, it can help us to find more specific biomarkers for diagnosis and assessing prognosis. On the other hand, it has a very significant guiding role for us to better design individualized regimens, especially targeting therapies (11).

To find reliable biomarkers, it is a feasible method to obtain gene expression profiles of GC from public functional genomics databases like GEO (12) and TCGA, and then perform bioinformatics analysis. In this study, we selected GC expression profiles of Asian population in GEO database, screened out hub genes for GC patients through microarray analysis. For one, the work helps us to understand the genetic changes clearly in GC groups in Asia. For another, during the analysis, we mainly discuss whether *COL4A2* and *CTSK* can be used as effective biomarkers for early diagnosis, prognosis assessment, recurrence monitoring, and therapeutic target in patients with GC in Asia. And the work offers a new direction for the diagnosis and treatment of GC.

## Methods

### *Microarray data*

Datasets were retrieved from GEO database (https://www.ncbi.nlm.nih.gov/geo/) with "gastric cancer" as the key word. The filters of organism and study type were limited as "Homo sapiens" and "Expression profiling by array", respectively. Then the gene expression matrix and the corresponding platform TXT files of target datasets were downloaded. R software (R 3.6.1, https://cran.r-project.org/) and related packages (http://www.bioconductor.org/) were used for data processing. The datasets utilized to differential analysis must be satisfied the following conditions: (I) all of the sequencing samples were from the patients with GC in China; (II) the datasets for differential gene analysis contained controls for cancer and cancer-adjacent tissues; (III) the sample size of each dataset was at least 20; (IV) the information of the platform annotation was available. The datasets used for prognosis analysis must be contained detailed survival data.

### *Screening for DEGs*

A volcano map was plotted to assess the differential expression of all genes by the ggplot2 package (https://cran.r-project.org/web/packages/ggplot2/index.html). The limma package was used for screening the DEGs in candidate dataset (13). As a DEG, it is necessary to satisfy both statistical P value <0.05 and |log fold change (FC)| >1. The co-expressed genes of DEGs were visualized by UpSetR package (14).

### *Construction of PPI network and molecular modules analysis*

The PPI network of DEGs was constructed in the STRING database (https://string-db.org/) through the following setting: meaning of network edges was set as "confidence", minimum required interaction score was selected as "medium confidence (0.400)" and finally display simplification was to hide disconnected nodes in the network (15). The PPI-data was downloaded and then identified hub genes and molecular modules by using the cytoHubba and MCODE plug-in Cytoscape software (version 3.7.1, https://cytoscape.org/), respectively. In

 

**Table 1** The information of ten microarray datasets from GEO

| Series | Country | Sample (tumor) | Sample (normal) | Platform | Version |
|---|---|---|---|---|---|
| GSE118916 | China | 15 | 15 | GPL15207 | Affymetrix Human Gene Expression Array |
| GSE19826 | China | 12 | 12 | GPL570 | Affymetrix Human Genome U133 Plus 2.0 Array |
| GSE65801 | China | 32 | 32 | GPL14550 | Agilent-028004 SurePrint G3 Human GE 8x60K Microarray |
| GSE79973 | China | 10 | 10 | GPL570 | Affymetrix Human Genome U133 Plus 2.0 Array |
| GSE54129 | China | 111 | 21 | GPL570 | Affymetrix Human Genome U133 Plus 2.0 Array |
| GSE57302 | China | 131 | 0 | GPL4091 | Agilent-014693 Human Genome CGH Microarray 244A |
| GSE26253 | South Korea | 432 | 0 | GPL8432 | Illumina HumanRef-8 WG-DASL v3.0 |
| GSE28541 | USA (MDACC Cohort) | 40 | 0 | GPL13376 | Illumina HumanWG-6 v2.0 expression BeadChip |
| GSE62254 | USA (ACRG Cohort) | 300 | 0 | GPL570 | Affymetrix Human Genome U133 Plus 2.0 Array |

MDACC, MD Anderson Cancer Center; ACRG, Asian Cancer Research Group; GEO, Gene Expression Omnibus.

MCODE, filters were based on the default parameters as "Degree Cutoff =2," "Node Score Cutoff =0.2," "K-Core =2" and "Max.Depth =100" (16).

### *GO and KEGG pathway analyses*

Functional annotation and pathway analyses for DEGs were done in the Enrichr database (http://amp.pharm.mssm.edu/Enrichr/), which is a comprehensive database for GO and KEGG pathway enrichment analysis (17). And the results of which were visualized by using GOplot package and Cytoscape software, respectively. Adjusted P value <0.05 was as a selecting criterion.

### *Prognosis analyses*

Firstly, survival analysis was conducted for DEGs using the Kaplan Meier analysis in the one of the datasets from GEO database so as to identify hub genes associated with prognosis of GC. Secondly, the racial survival differences of hub genes were evaluated again in the Kaplan-Meier plotter database (http://kmplot.com/analysis/), which included gene chip and RNA-seq data-sources from GEO database and TCGA (18). Thirdly, the hub genes were re-analyzed the prognostic value in multiple datasets from different cohort in GEO database. The results were visualized by forestplot (https://cran.r-project.org/web/packages/forestplot) and survival package (https://cran.r-project.org/web/packages/survival/) in R.

### *TIMER database analysis*

The correlations between the hub genes expression and the infiltration levels of inflammatory cells, including B cells, *CD4*$^+$ T cells, *CD8*$^+$ T cells, neutrophils, macrophages, and dendritic cells, were analyzed based on TIMER database (19). Moreover, we also explored the relationship of hub genes expression and molecular markers that had been reported in published studies (20,21), including markers of tumor-associated macrophages (TAMs), M2 macrophages, myeloid-derived suppressor cells (MDSCs), natural killer (NK) cells, dendritic cells (DCs), regulatory T cells (Tregs), and exhausted T cells. Correlation strength was classified according to the absolute value of partial correlation coefficient as follow: 0.00–0.19 "very weak", 0.20–0.39 "weak", 0.40–0.59 "moderate", 0.60–0.79 "strong", 0.80–1.0 "very strong" (22). The gene expression level was displayed with log2 RSEM.

## Results

### *Microarray data*

The GSE118916 (23), GSE19826 (24), GSE65801 (25), GSE79973 (26,27), and GSE54129 dataset were used for differential analysis, including 90 normal gastric tissue samples and 180 GC samples of China. The GSE57302 (28), GSE26253 (29,30), GSE28541 (29) and GSE62254 (6,29) dataset were used for prognosis analysis. The information of ten datasets was shown in *Table 1*. The five datasets for differential gene analysis were normalized, which of the

 

results were revealed in *Figure 1*.

### *Identification of DEGs*

The DEGs from five dataset was shown in *Figure 2A,B,C,D,E*. Two hundred twenty-two co-expressed genes were obtained by integrating bioinformatics analysis of all DEGs, covering 82 up-regulated expression genes and 140 down-regulated expression genes. The number of all and co-expressed DEGs were shown in *Figure 2F*. The co-expressed genes were provided in *Table S1*. The cluster heat map of all genes from five datasets was shown in *Figure S1*.

### *PPI network and molecular module*

The PPI network of 222 integrated DEGs was built through STRING database and the result was shown in *Figures 3A,S2*. The top 20 hub genes were *FN1*, *COL3A1*, *COL1A2*, *COL5A2*, *BGN*, *FBN1*, *THBS2*, *TIMP1*, *SPARC*, *COL6A3*, *COL5A1*, *SPP1*, *CDH11*, *COL12A1*, *VCAN*, *PDGFRB*, *COL4A2*, *ASPN*, *SERPINH1*, *COL10A1* and the matching information from GSE118916 was listed in *Table 2*. Six molecular modules were identified by using MCODE, the most important of which contained 32 genes, as visualized in *Figure 3B*.

### *Functional enrichment analyses of DEGs*

The total results of GO functional analyses for DEGs were shown in *Table S2*. It was evidence that the top 20 hub genes were mainly involved in the biological process (BP) of cell metabolism and the formation of extracellular matrix (ECM). Simultaneously, the molecular functions of hub genes were enriched in various binding of BP, as shown in *Figure 4*. Beyond that, the KEGG pathway enrichment of DEGs were mainly focus on protein digestion and absorption, ECM-receptor interaction, focal adhesion (FA) and phosphoinositide 3-kinase (PI3K)-protein kinase B (AKT) signaling pathway, as shown in *Figure 5* and *Table S3*.

### *Prognosis analysis of DEGs*

It turned out that six genes were associated with prognosis of GC patients from GSE57302, and of which *FBN1* (P=0.009), RARRES1 (P=0.001), *GPT2* (P=0.041) were related to good prognosis in GC patients, while *COL4A2* (P<0.001), *CTSK* (P=0.018), *GCNT2* (P=0.002) were related to poor prognosis (*Figure 6*). To further examine the prognostic potential of *CTSK*, *COL4A2* and *GCNT2* in different races of GC, we assessed the correlation of the OS and these genes expression between Asian and White patients in the Kaplan-Meier plotter databases, and found that over-expressions of *CTSK* and *COL4A2* reflected a worse OS in total patients (P<0.05). Specifically, it was worth mentioning that the *CTSK* expression level was correlated with worse OS in Asian GC patients (HR =6.53, P=0.01), but was not associated with OS of White patients (HR =1.60, P=0.069) (*Figure 7*). Then we continued to re-analyze the survival differences of *CTSK* and *COL4A2* in the GSE26253, GSE28541 and GSE62254 dataset and the eventual outcomes also showed that the high-expression of *CTSK* and *COL4A2* had a significant association with poor prognosis in the cohort of Asian GC patients (*Figure 8*).

### *The relationship between CTSK, COL4A2 and GCNT2 expression level and inflammatory cell infiltration*

We analyzed whether the expression of *CTSK*, *COL4A2* and *GCNT2* were correlated with inflammatory cell infiltration levels in GC. The results showed that the level of *CTSK* expression had significant correlations with tumor purity, macrophage, neutrophil and dendritic cell (*Figure 9*). We also investigated the relationships between *CTSK* expression and biomarkers of different immune cells, included TAMs, M2 macrophages, MDSCs, NK cells, DCs, Tregs and exhausted T cells. We found that *CCL2*, *IL10* of TAMs, *CD163*, VSIG4, *MS4A4A* of M2 phenotype, *CD33*, *ITGAM*, *CD14*, *CSF1R* of MDSCs, *NRP1*, *IL3RA*, *ITGAX* of DCs, *TGFB1* of Tregs, and *TIM3* of T cell exhaustion were significantly correlated with *CTSK* expression in GC (P<0.0001; *Table 3*).

## Discussion

The occurrence of tumors shows significant racial differences (31). Changes in genetic level and ethnic heterogeneities have always been the focus of tumor researches. To improve early diagnostic rate, and screen for therapeutic targets on ethnic characteristics for GC, it is urgently necessary to identify sensitive and specific prognostic biomarkers.

In our study, with the differential analysis of five datasets comprising the GSE118916, GSE19826, GSE65801 GSE79973 and GSE54129 by using limma package in R software, and 222 co-expressed DEGs were found. Then, a functional analysis, a PPI network and the most important

 

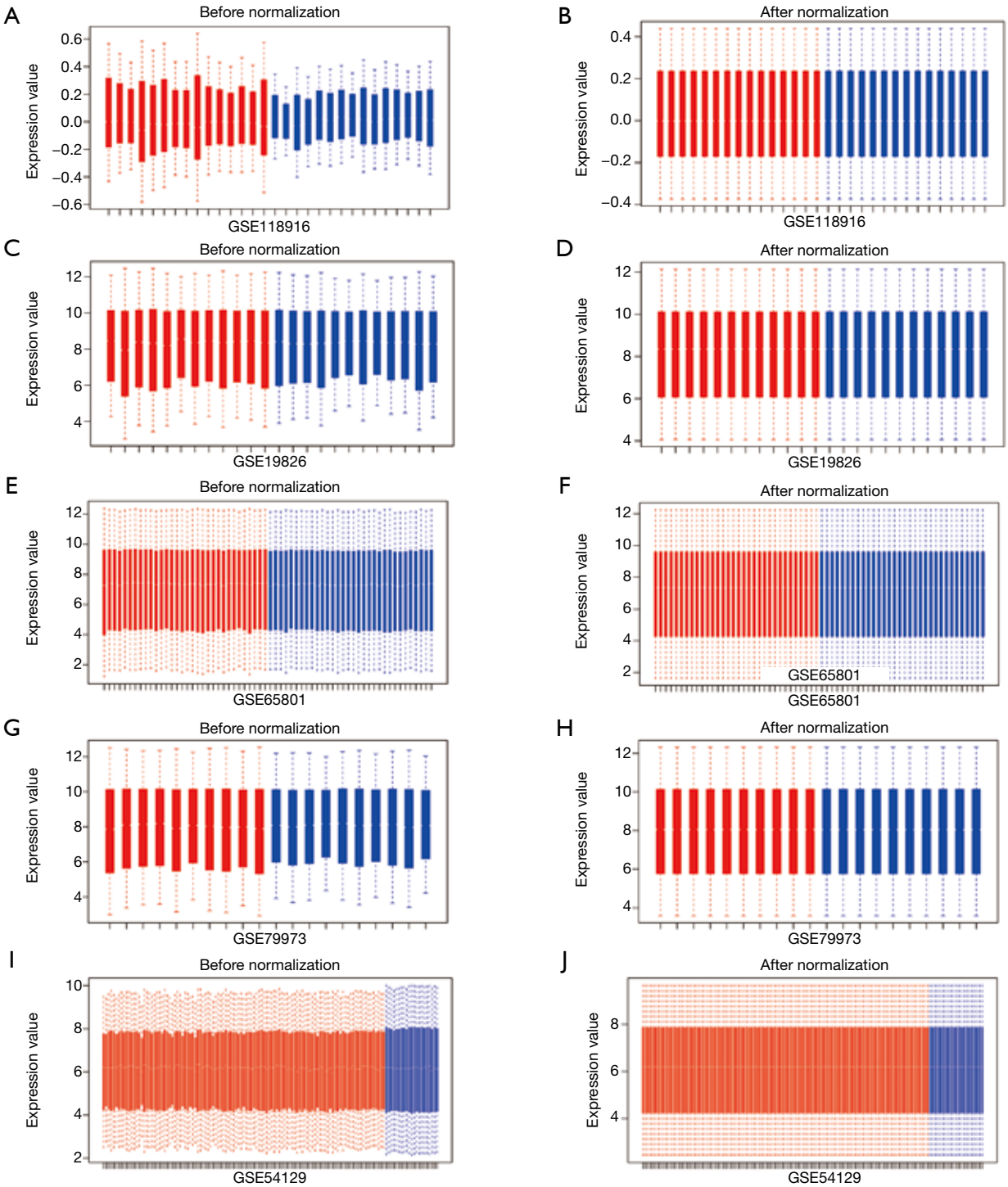


**Figure 1** Normalization of gene expression. (A,B) Normalization of the GSE118916 data set; (C,D) normalization of the GSE19826 data set; (E,F) normalization of the GSE65801 data set; (G,H) normalization of the GSE79973 data set; (I,J) normalization of the GSE54129 data set. Red represents gastric cancer tissues and blue represents normal tissues.

 

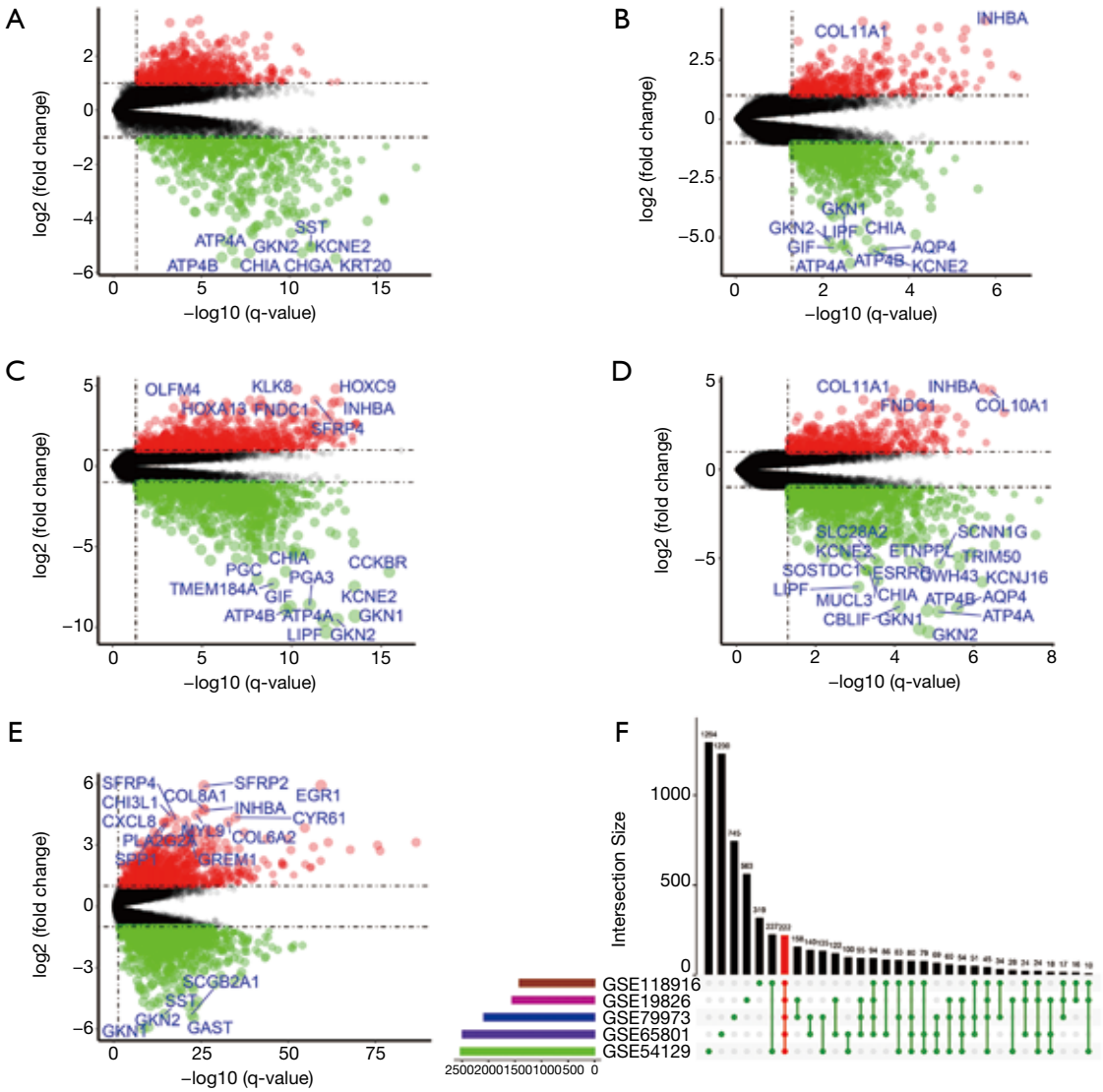


**Figure 2** Differentially expressed genes between the two groups of samples in each data set. (A) GSE118916, (B) GSE19826, (C) GSE65801, (D) GSE79973, (E) GSE54129. The red dots represent the up-regulated genes based on an adjusted P<0.05 and logFC >1 (if logFC >4 then the gene symbol is labeled); the green dots represent the down-regulated genes based on an adjusted P<0.05 and |logFC| >1 (if |logFC| >6 then the gene symbol is labeled); the black spots represent genes with no significant difference in expression.

molecular module of DEGs were performed through online databases and Cytoscape software. A prognostic analysis of DEGs was conducted by using survival package in R software, which of the results showed that *CTSK* (Cathepsin K, *CatK*), *GCNT2* [Glucosaminyl (N-acetyl) transferase2] and *COL4A2* (Collagen type IV alpha 2 chain) were related to poor prognosis, while *FBN1* (Fibrillin 1), *RARRES1* (Retinoic acid receptor responder 1), *GTP2* (Glutamic pyruvate transaminase 2) were associated with desirable prognosis in GC patients in Chinese population. Furthermore, we found that the high-expression levels of *CTSK* and *COL4A2* are significantly correlated with poor prognosis of GC in Asian through cross-validation between databases.

*CTSK*, belonging to the cathepsin L-like cluster of *C1A* family, has been confirmed to play a key role in ECM-remodeling, regulation of cytokine level and tumor growth factor, process of lymph node and bone metastasis in a variety of cancers such as breast and prostate cancer (32-35). In gastric and oral squamous cell carcinoma, researchers

 

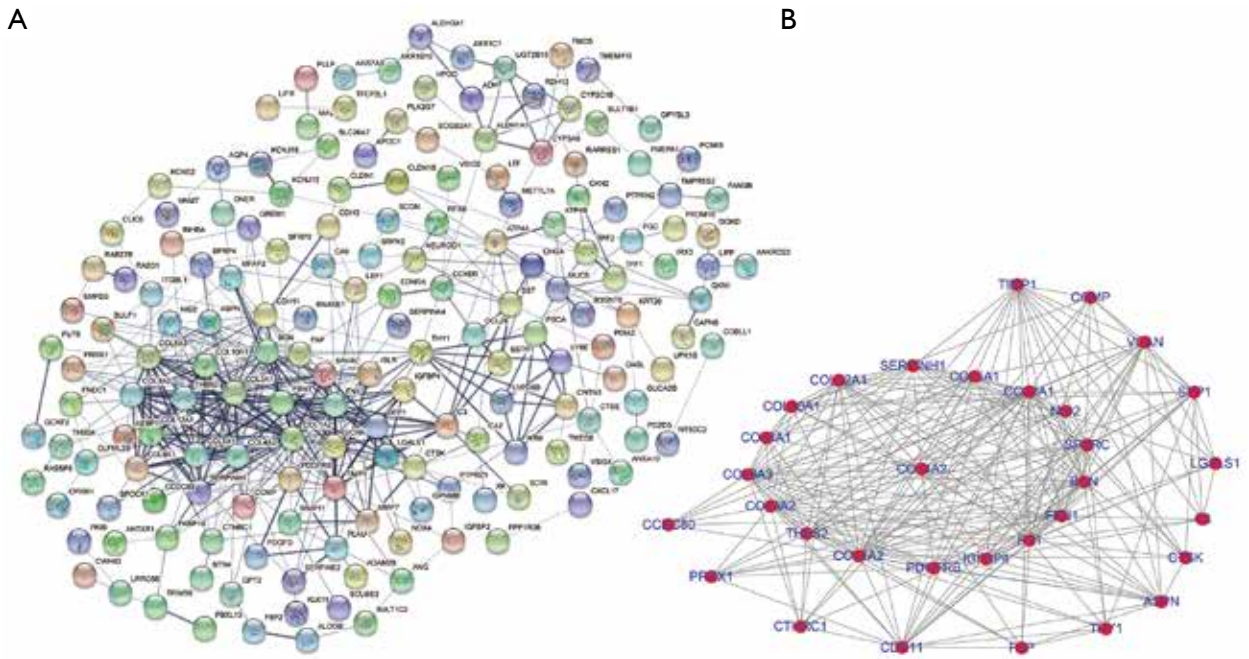


**Figure 3** The PPI network and the most important molecular module of DEGs. (A) PPI network of DEGs constructed in STRING database; (B) the most important molecular module of DEGs. Circles represent genes, lines represent interactions between gene-encoded proteins and line width represents evidence of interactions between proteins.

**Table 2** The top 20 hub genes of integrated DEGs

| Gene symbol | LogFC (GSE118916) | P value (GSE118916) | Gene name |
|---|---|---|---|
| *ASPN* | 1.89 | ** | Asporin |
| *BGN* | 1.44 | ** | Biglycan |
| *CDH11* | 1.56 | ** | Cadherin 11 |
| *COL10A1* | 2.01 | ** | Collagen type X alpha 1 chainX |
| *COL12A1* | 2.03 | ** | Collagen type XII alpha 1 chain |
| *COL1A2* | 2.48 | ** | Collagen type I alpha 2 chain |
| *COL3A1* | 1.81 | ** | Collagen type III alpha 1 chain |
| *COL4A2* | 1.54 | ** | Collagen type IV alpha 2 chain |
| *COL5A1* | 1.61 | ** | Collagen type V alpha 1 chain |
| *COL5A2* | 1.61 | ** | Collagen type V alpha 2 chain |
| *COL6A3* | 2.46 | ** | Collagen type VI alpha 3 chain |
| *FBN1* | 1.16 | ** | Fibrillin 1 |
| *FN1* | 2.12 | ** | Fibronectin 1 |
| *PDGFRB* | 1.69 | ** | Platelet derived growth factor receptor beta |
| *SERPINH1* | 1.99 | ** | Serpin family H member 1 |
| *SPARC* | 2.16 | ** | Secreted protein acidic and cysteine rich |
| *SPP1* | 2.48 | ** | Secreted phosphoprotein 1 |
| *THBS2* | 2.56 | ** | Thrombospondin 2 |
| *TIMP1* | 2.31 | ** | TIMP metallopeptidase inhibitor 1 |
| *VCAN* | 1.98 | ** | Versican |

**, P value <0.01. DEGs, differentially expressed genes.

 

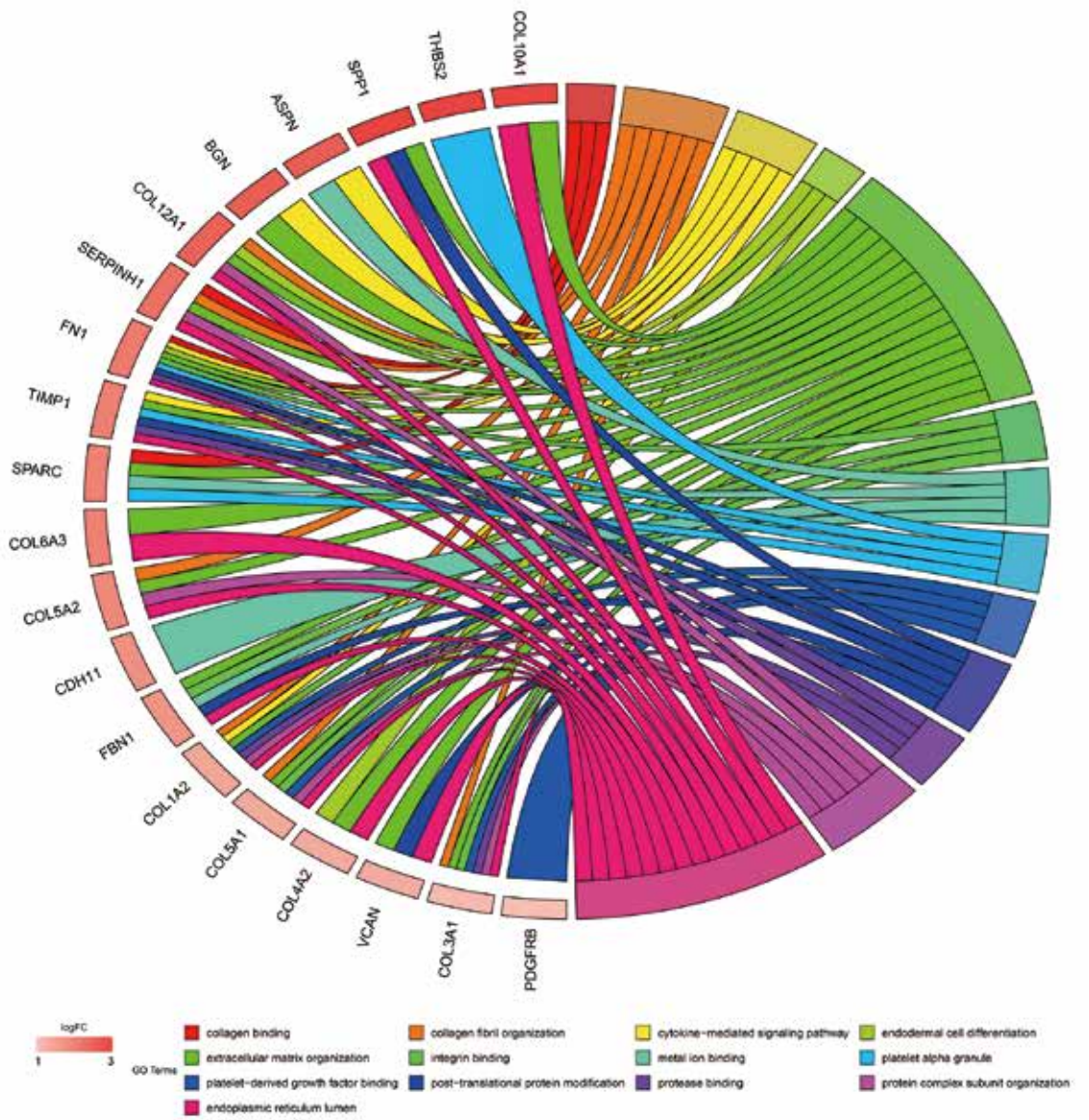


**Figure 4** Distribution of the top 20 hub genes in gastric cancer for different GO-enriched functions.

also report that up-regulation of *CTSK* associates with lymph node metastasis and poor prognosis (36,37). In terms of therapeutic significance, *CTSK* may be a therapeutic target for patients with prostate cancer at risk of bone metastases (33). Alongside these function, we also found that *CTSK* has significant correlations not only with macrophages, neutrophils and DCs in TME, but also with numerous gene marker sets of various inflammatory cells in GC, for instance *CCL2*, *IL10*, *VSIG4*, *MS4A4A*, *CD33*, *ITGAM*, *CD14*, *CSF1R*, *NRP1*, *IL3RA*, *ITGAX*, *TGFB1*, *CD4*, and *TIM3*. Although T cells recruited to TME have potential to kill tumor cells, more often, they are powerless, fading and exhausting due to high expression of genes like *TIM3* (38). TAMs, particularly M2 phenotype, play a vital role in promoting tumorigenesis through inducing neovascularization, regulating inflammatory responses and the reconstruction of ECM. Overexpression of *VSIG4* and *MS4A4A* can promote the M1-phenotype macrophages to transform into M2, and negatively regulate macrophage activation (39,40). High levels of *CD33*, *ITGAM*, *CD14*,

 

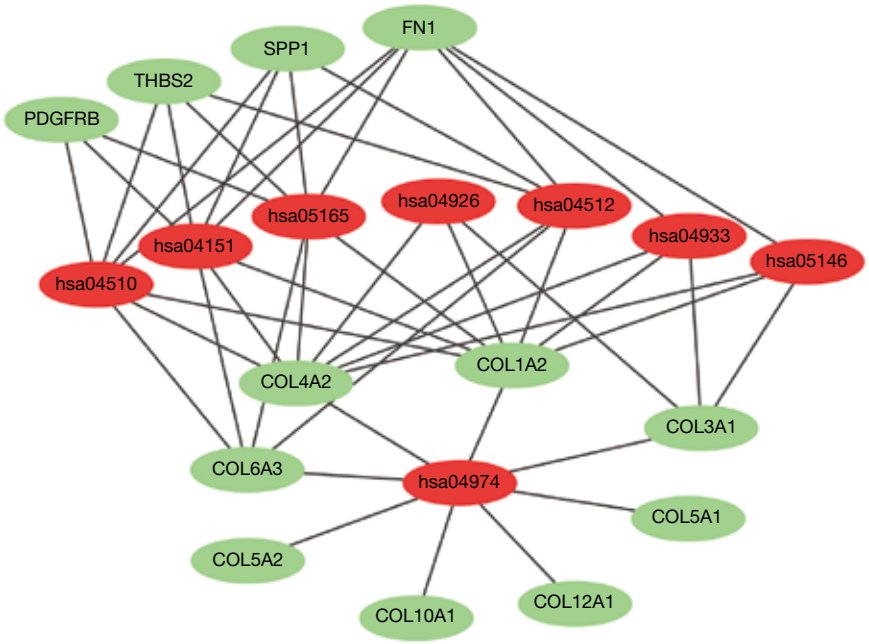


**Figure 5** Network map of enriched KEGG pathways. Red represents the pathways; green represents the hub genes.

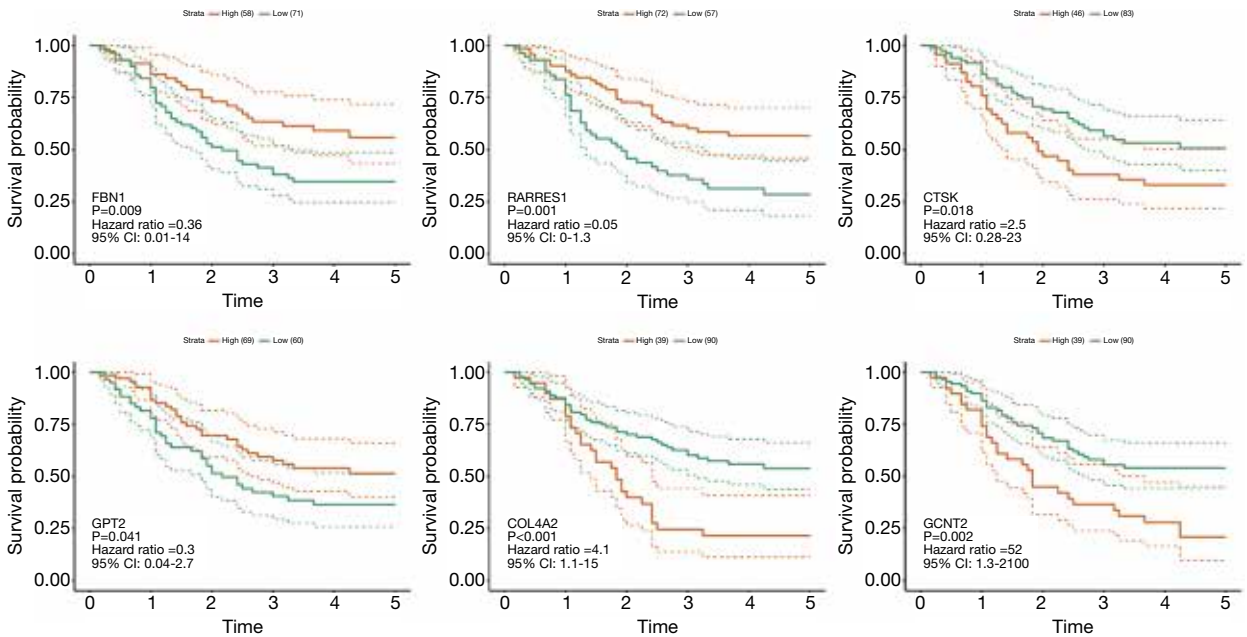


**Figure 6** Kaplan-Meier analysis results of hub genes.

*CSF1R*, *NRP1*, *IL3RA*, *ITGAX* and *TGFB1* have been confirmed to be associated with the aggregation of MDSCs and DC cells in TME. MDSCs family and DC cells also accelerate the formation of an immunosuppressive TME via coordinating the response of immune inflammatory cell (41-43). Several studies revealed that *CTSK* has a higher stromal expression in tumour-associated fibroblasts and macrophages of invasive tumors compared to non-invasive

 

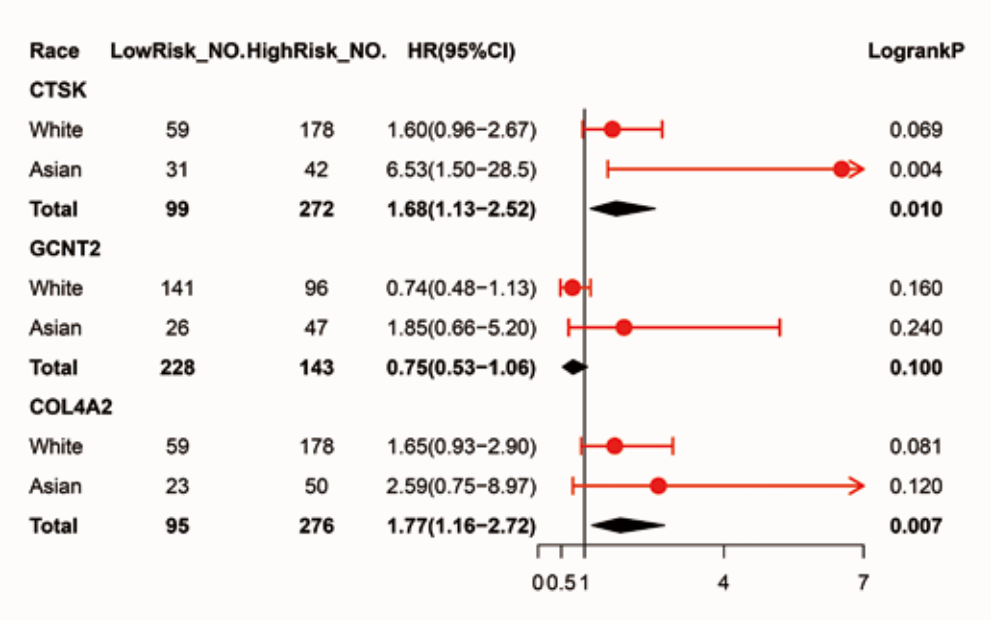


**Figure 7** Forest map showing the results of survival analyses for *CTSK GCNT2* and *COL4A2* among White and Asian gastric cancer patients.

carcinomas (44,45). In addition, according to a recent study, *CTSK* can accelerate metastasis in colorectal cancer via mediating *TLR4*-dependent M2 macrophage polarization with the help of gut microbiota (46).

GO analyses reflected that the BPs of DEGs are mainly enriched in ECM organization, protein complex subunit organization, endodermal cell differentiation, cellular response to cytokine stimulus, cellular protein modification process and platelet degranulation in our study. Additionally, the MFs are found to be in various biological binding like platelet-derived growth factor binding, integrin binding, collagen binding *et al.* The KEGG pathway enrichment revealed that the DEGs mainly act on the protein bio-utilization, ECM-receptor interaction, FA and PI3K-AKT signaling pathway. The ECM is an important part of cell micro-environment, whose structural function is essential to keep the normal activity of cell (47). However, recent studies have identified that the composition and mechanical properties of ECM take a significant part in the BPs associated with tumor development, such as escaping from apoptosis and regulating of cell growth, promoting tumor angiogenesis, gaining invasion and metastatic ability (48-51). Type IV collagen, one of the six subunits of which is encoded by *COL4A2*, is the most important component of the network structure of ECM and provides a tensile strength for underlying tissues, bounding diverse macromolecules and binding multiple cellular receptors. Similarly, evidences that *COL4A2* might promote cell-adhesion, activate migration, and stimulate proliferation of different cell types has been reported, which are related to tumorigenesis (52-54). Actually, mutations in *COL4A2* lead to its ectopic expression and eventually result in the occurrence of disease in the body. Even so, current researches on *COL4A2* mutations mainly focus on the field of cerebrovascular diseases, but the details are still not quite clear as well as involving in the development of tumors (55). These, together with our results, suggest that *CTSK* and *COL4A2* are indeed involved in matrix remodeling and enhance the invasion of tumor cells to a certain extent.

Based on the available evidences, we can attempt to explain the underlying molecular mechanisms of *CTSK* and *COL4A2* involved in tumorigenesis. To become cancerous, cell has to break down the original connection between cell and cell, remodel cell-matrix adhesion site, develop along a pathway with the participation of enzymes secreted by ECM and finally undergo epithelial-mesenchymal transformation (EMT) (56,57). Aberrant EMT activation reduces the epithelial features of gastric epithelial cells and makes them obtain more characteristics of mesenchymal cells that tend to be dedifferentiated and more cancerous,

 

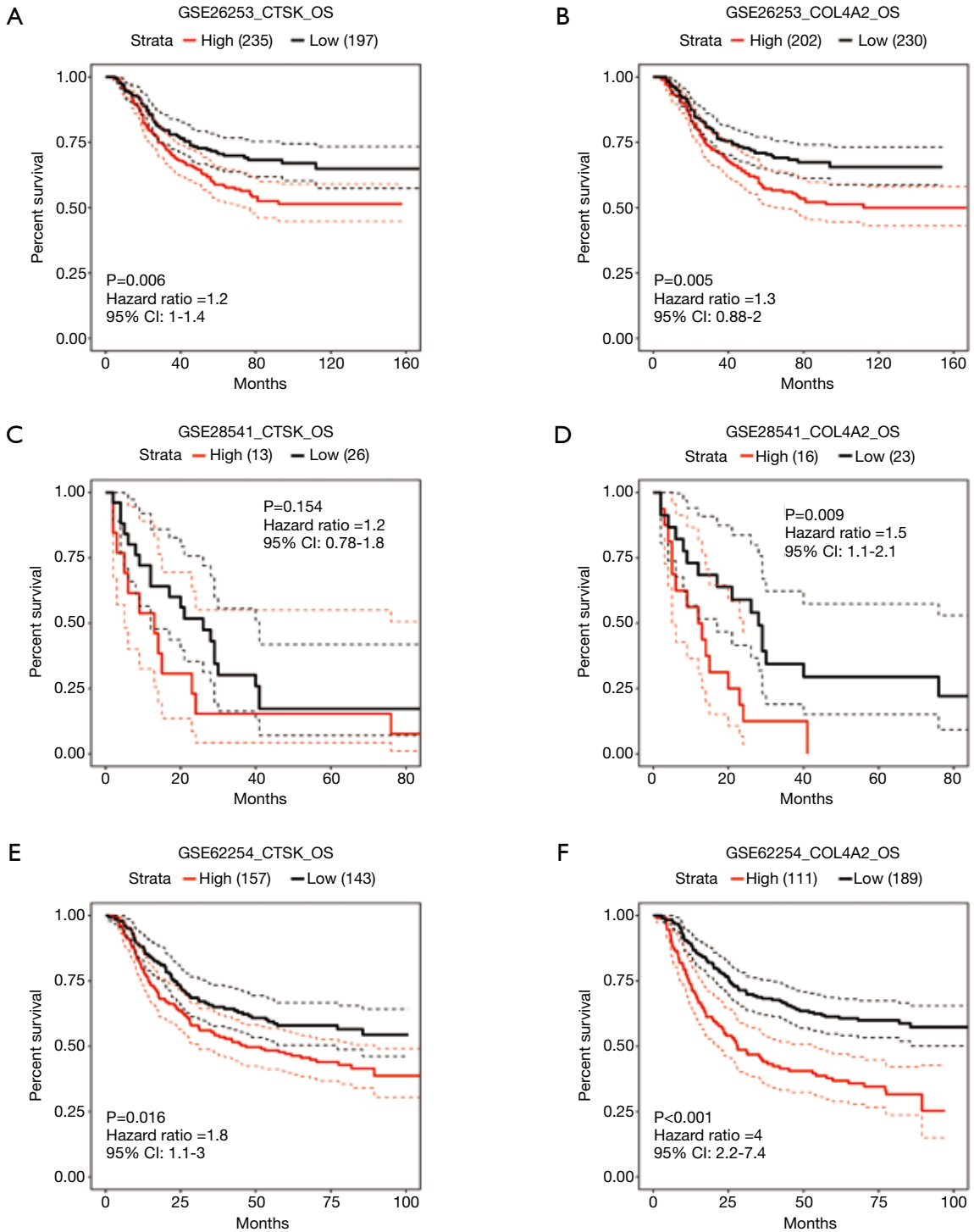


**Figure 8** The survival analysis of *CTSK* and *COL4A2* in the GSE26253, GSE28541 and GSE62254 dataset. OS, overall survival.

and obtain capability to invade (58,59). Pathways like PI3K/AKT and *TGF-β* signaling pathway are also activated during the process of EMT, which further contributes to induce angiogenesis and inflammatory cell recruitment in TME (60,61). More simply, we can divide these complex processes into two steps: the first is remodeling of ECM

 

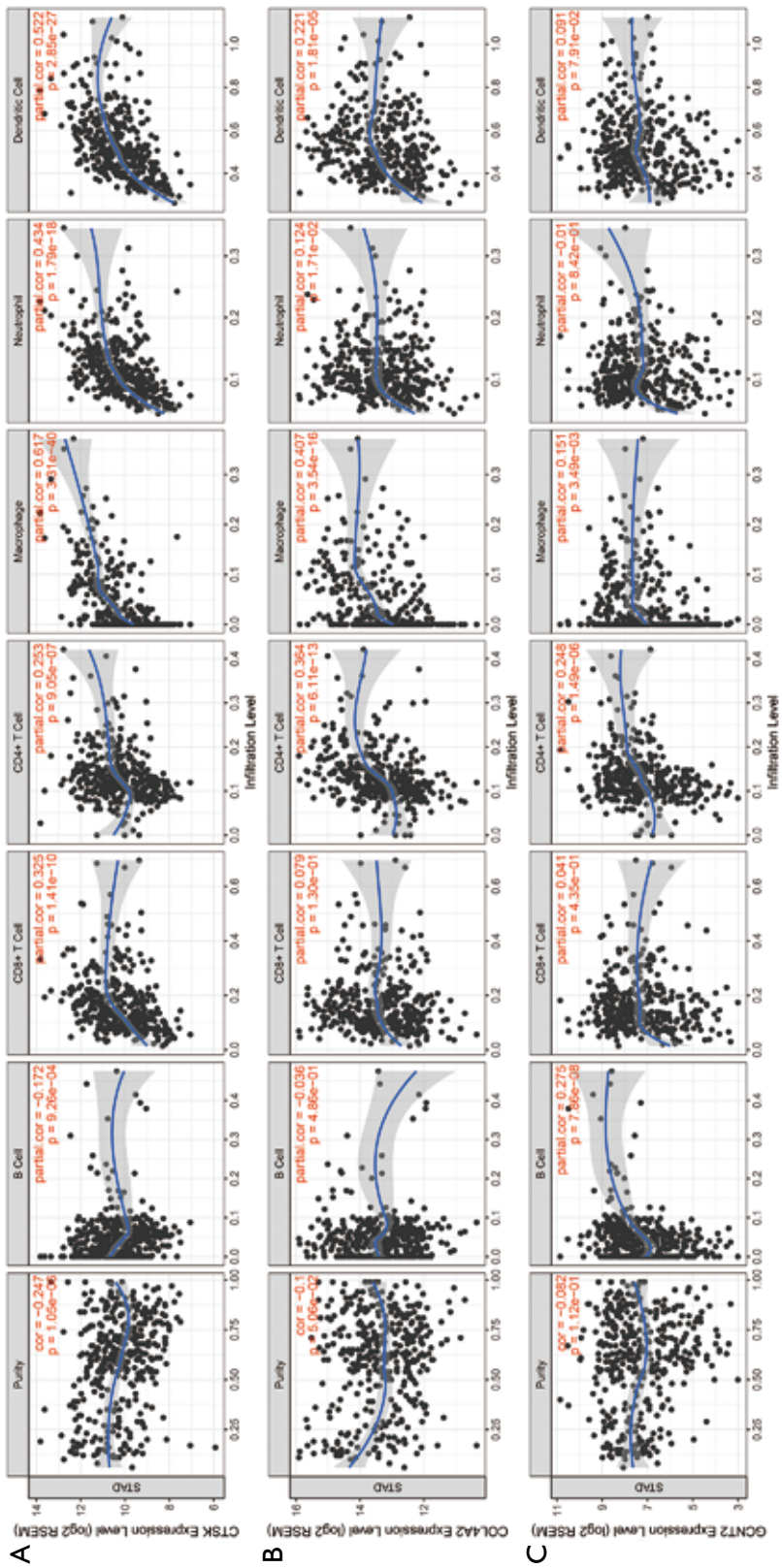


**Figure 9** Correlation of *CTSK COL4A2 GCNT2* expression with immune infiltration level in STAD. (A) *CTSK* expression is significantly negatively related to tumor purity and has significant positive correlations with infiltrating levels of macrophages, neutrophils, and dendritic cells in STAD; (B) COL4A2 expression has no significant correlations with tumor purity and infiltrating levels of B cells, *CD8+* T cells, *CD4+* T cells, neutrophils, and dendritic cells except macrophages in STAD; (C) *GCNT2* expression has no significant correlations with tumor purity and infiltrating levels of B cells, *CD8+* T cells, *CD4+* T cells, macrophages, neutrophils, and dendritic in STAD. STAD, Stomach adenocarcinoma.

 

**Table 3** *CTSK* expression correlated with the gene markers of tumor-infiltrating immune cells in STAD

| Description | Gene markers | None | | Purity | |
|---|---|---|---|---|---|
| | | Cor | P | Cor | P |
| TAMs | *CCL2* | 0.540 | *** | 0.502 | *** |
| | *CD68* | 0.371 | *** | 0.326 | *** |
| | *IL10* | 0.506 | *** | 0.473 | *** |
| M2 | *CD163* | 0.488 | *** | 0.455 | *** |
| | *VSIG4* | 0.560 | *** | 0.534 | *** |
| | *MS4A4A* | 0.567 | *** | 0.540 | *** |
| MDSCs | *FCGR3A* | 0.516 | *** | 0.497 | *** |
| | *CD33* | 0.549 | *** | 0.512 | *** |
| | *CD11B(ITGAM)* | 0.535 | *** | 0.517 | *** |
| | *CD80* | 0.370 | *** | 0.338 | *** |
| | *CD14* | 0.582 | *** | 0.543 | *** |
| | *CD115(CSF1R)* | 0.555 | *** | 0.526 | *** |
| | *CD64(FCGR1A)* | 0.544 | *** | 0.512 | *** |
| Natural killer cell | *CD16(FCGR3A)* | 0.516 | *** | 0.497 | *** |
| | *CD56(NCAM1)* | 0.355 | *** | 0.350 | *** |
| Dendritic cells | *CD11C(ITGAX)* | 0.526 | *** | 0.485 | *** |
| | *CD141(THBD)* | 0.412 | *** | 0.390 | *** |
| | *CD123(IL3RA)* | 0.554 | *** | 0.519 | *** |
| | *CD304(NRP1)* | 0.562 | *** | 0.546 | *** |
| | *CD74* | 0.302 | *** | 0.244 | *** |
| | *HLA-DQA1* | 0.413 | *** | 0.362 | *** |
| Tregs | *CKIT* | 0.347 | *** | 0.302 | *** |
| | *CCR8* | 0.413 | *** | 0.396 | *** |
| | *STAT5B* | 0.362 | *** | 0.365 | *** |
| | *TGFB1* | 0.559 | *** | 0.535 | *** |
| | *CD4* | 0.501 | *** | 0.459 | *** |
| | *CD25(IL2RA)* | 0.401 | *** | 0.358 | *** |
| T cell exhaustion | *PD1(PDCD1)* | 0.199 | *** | 0.153 | * |
| | *CTLA4* | 0.196 | *** | 0.140 | * |
| | *LAG3* | 0.202 | *** | 0.159 | * |
| | *TIM3(HAVCR2)* | 0.549 | *** | 0.519 | *** |
| | *GZMB* | 0.224 | *** | 0.158 | * |
| | *CD274* | 0.179 | ** | 0.144 | * |

Cor, Correlation coefficient; *, P<0.05; **, P<0.01; ***, P<0.001.

 

that caused by the abnormal expression of hub genes such as *COL4A2*, *CTSK et al.* and the second is activation of EMT-cascade reactions that promote gastric epithelial cells towards mesenchymal cells, and acquire the capability of carcinogenesis. Meanwhile, in our findings, over-expression of *CTSK* can induce more immunocytes with inhibitive function such as M2 macrophages into TME and provide a suitable environment for cancerous cells growth, and allow cancer cells to escape the immune response in the end. Unfortunately, the formation of immunosuppressive microenvironment can further precipitate EMT process and tumor progression partly, and eventually lead to a poor prognosis of tumor patients. In summary, the most important role of *CTSK* and *COL4A2* in cancers seems to be their contributions to the degradation of ECM-related proteins and the regulation of cytokines in TME.

However, many questions remain unanswered. Is *CTSK* involved in the regulation of tumor immunity as we have described? Is *CTSK* involved in regulating the tumor immune microenvironment continuous or intermittent? And what point in time does *CTSK* play a role in immunoregulation of normal gastric cells become cancerous? And what is the true relationship between *CTSK*, *COL4A2* and EMT? Thus, further researches on these problems are needed by means of more experiments.

## Conclusions

In conclusion, we identified two key genes (*COL4A2* and *CTSK*) that could play a vital role in the pathogenesis of GC in Asian and act as the promising diagnostic and prognostic biomarkers in patients with GC probably via integrated bioinformatics analysis. *CTSK* could induce the formation of immunosuppressive TME and promote the immune escape of GC cells.

## Acknowledgments

*Funding:* None.

## Footnote

*Conflicts of Interests*: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jgo.2020.03.01). The authors have no conflicts of interest to declare.

*Ethical Statement*: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.



## References

1. Nagtegaal ID, Odze RD, Klimstra D, et al. The 2019 WHO classification of tumours of the digestive system. Histopathology 2020;76:182-8.
2. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202-9.
3. Van Cutsem E, Sagaert X, Topal B, et al. Gastric cancer. Lancet 2016;388:2654-64.
4. Ott K, Vogelsang H, Mueller J, et al. Chromosomal instability rather than p53 mutation is associated with response to neoadjuvant cisplatin-based chemotherapy in gastric carcinoma. Clin Cancer Res 2003;9:2307-15.
5. Sawaki A, Ohashi Y, Omuro Y, et al. Efficacy of trastuzumab in Japanese patients with HER2-positive advanced gastric or gastroesophageal junction cancer: a subgroup analysis of the Trastuzumab for Gastric Cancer (ToGA) study. Gastric Cancer 2012;15:313-22.
6. Cristescu R, Lee J, Nebozhyn M, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med 2015;21:449-56.
7. Wang J, Sun Y, Bertagnolli MM. Comparison of gastric cancer survival between Caucasian and Asian patients treated in the United States: results from the Surveillance Epidemiology and End Results (SEER) database. Ann Surg Oncol 2015;22:2965-71.
8. Fujita K, Yamamoto W, Endo S, et al. CYP2A6 and the plasma level of 5-chloro-2, 4-dihydroxypyridine are determinants of the pharmacokinetic variability of tegafur and 5-fluorouracil, respectively, in Japanese patients with cancer given S-1. Cancer Sci 2008;99:1049-54.
9. Kim J, Sun CL, Mailey B, et al. Race and ethnicity correlate with survival in patients with gastric

 

adenocarcinoma. Ann Oncol 2010;21:152-60.
10. Jia F, Teer JK, Knepper TC, et al. Discordance of Somatic Mutations Between Asian and Caucasian Patient Populations with Gastric Cancer. Mol Diagn Ther 2017;21:179-85.
11. Biagioni A, Skalamera I, Peri S, et al. Update on gastric cancer treatments and gene therapies. Cancer Metastasis Rev 2019;38:537-48.
12. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 2013;41:D991-5.
13. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
14. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 2017;33:2938-40.
15. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019;47:D607-D613.
16. Otasek D, Morris JH, Boucas J, et al. Cytoscape Automation: empowering workflow-based network analysis. Genome Biol 2019;20:185.
17. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 2013;14:128.
18. Nagy Á, Lanczky A, Menyhart O, et al. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. Sci Rep 2018;8:9227.
19. Li T, Fan J, Wang B, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res 2017;77:e108-10.
20. Siemers NO, Holloway JL, Chang H, et al. Genome-wide association analysis identifies genetic correlates of immune infiltrates in solid tumors. PLoS One 2017;12:e0179726.
21. Danaher P, Warren S, Dennis L, et al. Gene expression markers of Tumor Infiltrating Leukocytes. J Immunother Cancer 2017;5:18.
22. Pan JH, Zhou H, Cooper L, et al. LAYN Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Gastric and Colon Cancers. Front Immunol 2019;10:6.
23. Li L, Zhu Z, Zhao Y, et al. FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics. Sci Rep 2019;9:7827.
24. Wang Q, Wen YG, Li DP, et al. Upregulated INHBA expression is associated with poor survival in gastric cancer. Med Oncol 2012;29:77-83.
25. Li H, Yu B, Li J, et al. Characterization of differentially expressed genes involved in pathways associated with gastric cancer. PLoS One 2015;10:e0125013.
26. Jin Y, He J, Du J, et al. Overexpression of HS6ST2 is associated with poor prognosis in patients with gastric cancer. Oncol Lett 2017;14:6191-7.
27. He J, Jin Y, Chen Y, et al. Downregulation of ALDOB is associated with poor prognosis of patients with gastric cancer. Onco Targets Ther 2016;9:6099-109.
28. Qian Z, Zhu G, Tang L, et al. Whole genome gene copy number profiling of gastric cancer identifies PAK1 and KRAS gene amplification as therapy targets. Genes Chromosomes Cancer 2014;53:883-94.
29. Oh SC, Sohn BH, Cheong JH, et al. Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. Nat Commun 2018;9:1777.
30. Lee J, Sohn I, Do IG, et al. Nanostring-based multigene assay to predict recurrence for gastric cancer patients after surgery. PLoS One 2014;9:e90133.
31. Ashktorab H, Kupfer SS, Brim H, et al. Racial Disparity in Gastrointestinal Cancer Risk. Gastroenterology 2017;153:910-23.
32. Le Gall C, Bellahcene A, Bonnelye E, et al. A cathepsin K inhibitor reduces breast cancer induced osteolysis and skeletal tumor burden. Cancer Res 2007;67:9894-902.
33. Munari E, Cima L, Massari F, et al. Cathepsin K expression in castration-resistant prostate carcinoma: a therapeutical target for patients at risk for bone metastases. Int J Biol Markers 2017;32:e243-e247.
34. Verbovšek U, Van Noorden CJ, Lah TT. Complexity of cancer protease biology: Cathepsin K expression and function in cancer progression. Semin Cancer Biol 2015;35:71-84.
35. Duong LT, Wesolowski GA, Leung P, et al. Efficacy of a cathepsin K inhibitor in a preclinical model for prevention and treatment of breast cancer bone metastasis. Mol Cancer Ther 2014;13:2898-909.
36. Ren G, Tian Q, An Y, et al. Coronin 3 promotes gastric cancer metastasis via the up-regulation of MMP-9 and cathepsin K. Mol Cancer 2012;11:67.
37. Leusink FK, Koudounarakis E, Frank MH, et al. Cathepsin K associates with lymph node metastasis and poor prognosis in oral squamous cell carcinoma. BMC Cancer 2018;18:385.
38. Wolf Y, Anderson AC, Kuchroo VK. TIM3 comes

 

of age as an inhibitory receptor. Nat Rev Immunol 2020;20:173-85.
39. Li J, Diao B, Guo S, et al. VSIG4 inhibits proinflammatory macrophage activation by reprogramming mitochondrial pyruvate metabolism. Nat Commun 2017;8:1322.
40. Sanyal R, Polyak MJ, Zuccolo J, et al. MS4A4A: a novel cell surface marker for M2 macrophages and plasma cells. Immunol Cell Biol 2017;95:611-9.
41. Kitano Y, Okabe H, Yamashita YI, et al. Tumour-infiltrating inflammatory and immune cells in patients with extrahepatic cholangiocarcinoma. Br J Cancer 2018;118:171-80.
42. Diakos CI, Charles KA, McMillan DC, et al. Cancer-related inflammation and treatment effectiveness. Lancet Oncol 2014;15:e493-503.
43. Ino Y, Yamazaki-Itoh R, Shimada K, et al. Immune cell infiltration as an indicator of the immune microenvironment of pancreatic cancer. Br J Cancer 2013;108:914-23.
44. Kleer CG, Bloushtain-Qimron N, Chen YH, et al. Epithelial and stromal cathepsin K and CXCL14 expression in breast tumor progression. Clin Cancer Res 2008;14:5357-67.
45. Rapa I, Volante M, Cappia S, et al. Cathepsin K is selectively expressed in the stroma of lung adenocarcinoma but not in bronchioloalveolar carcinoma. A useful marker of invasive growth. Am J Clin Pathol 2006;125:847-54.
46. Li R, Zhou R, Wang H, et al. Gut microbiota-stimulated cathepsin K secretion mediates TLR4-dependent M2 macrophage polarization and promotes tumor metastasis in colorectal cancer. Cell Death Differ 2019;26:2447-63.
47. Giussani M, Triulzi T, Sozzi G, et al. Tumor Extracellular Matrix Remodeling: New Perspectives as a Circulating Tool in the Diagnosis and Prognosis of Solid Tumors. Cells 2019;8:81.
48. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. EMBO Rep 2014;15:1243-53.
49. Eble JA, Niland S. The extracellular matrix in tumor progression and metastasis. Clin Exp Metastasis 2019;36:171-98.
50. Poltavets V, Kochetkova M, Pitson SM, et al. The Role of the Extracellular Matrix and Its Molecular and Cellular Regulators in Cancer Cell Plasticity. Front Oncol 2018;8:431.
51. Malik R, Lelkes PI, Cukierman E. Biomechanical and biochemical remodeling of stromal extracellular matrix in cancer. Trends Biotechnol 2015;33:230-6.
52. Boudko SP, Danylevych N, Hudson BG, et al. Basement membrane collagen IV: Isolation of functional domains. Methods Cell Biol 2018;143:171-85.
53. Ohlund D, Lundin C, Ardnor B, et al. Type IV collagen is a tumour stroma-derived biomarker for pancreas cancer. Br J Cancer 2009;101:91-7.
54. Brown CW, Brodsky AS, Freiman RN. Notch3 overexpression promotes anoikis resistance in epithelial ovarian cancer via upregulation of COL4A2. Mol Cancer Res 2015;13:78-85.
55. Gunda B, Mine M, Kovacs T, et al. COL4A2 mutation causing adult onset recurrent intracerebral hemorrhage and leukoencephalopathy. J Neurol 2014;261:500-3.
56. Thiery JP, Sleeman JP. Complex networks orchestrate epithelial-mesenchymal transitions. Nat Rev Mol Cell Biol 2006;7:131-42.
57. Yilmaz M, Christofori G. EMT, the cytoskeleton, and cancer cell invasion. Cancer Metastasis Rev 2009;28:15-33.
58. Peng Z, Wang CX, Fang EH, et al. Role of epithelial-mesenchymal transition in gastric cancer initiation and progression. World J Gastroenterol 2014;20:5403-10.
59. Mittal V. Epithelial Mesenchymal Transition in Tumor Metastasis. Annu Rev Pathol 2018;13:395-412.
60. Thiery JP, Acloque H, Huang RYJ, et al. Epithelial-Mesenchymal Transitions in Development and Disease. Cell 2009;139:871-90.
61. Polivka J Jr, Janku F. Molecular targets for cancer therapy in the PI3K/AKT/mTOR pathway. Pharmacol Ther 2014;142:164-75.

**Cite this article as:** Feng Z, Qiao R, Ren Z, Hou X, Feng J, He X, Chen D. Could CTSK and COL4A2 be specific biomarkers of poor prognosis for patients with gastric cancer in Asia?—a microarray analysis based on regional population. J Gastrointest Oncol 2020;11(2):386-401. doi: 10.21037/jgo.2020.03.01

 

**Table S1** The co-expressed DEGs

| Gene name | GSE118916 | | GSE19826 | | GSE54129 | | GSE65801 | | GSE79973 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | logFC | P value | logFC | P value | logFC | P value | logFC | P value | logFC | P value |
| *FNDC1* | 3.34 | 0.00 | 3.04 | 0.00 | 2.87 | 0.00 | 4.11 | 0.00 | 4.22 | 0.00 |
| *THBS4* | 3.23 | 0.00 | 2.18 | 0.00 | 3.22 | 0.00 | 3.18 | 0.00 | 2.32 | 0.01 |
| *FAP* | 2.79 | 0.00 | 3.77 | 0.00 | 3.28 | 0.00 | 3.40 | 0.00 | 3.99 | 0.00 |
| *C3* | 2.75 | 0.00 | 1.55 | 0.00 | 3.28 | 0.00 | 1.76 | 0.00 | 1.08 | 0.05 |
| *RARRES1* | 2.65 | 0.00 | 1.50 | 0.00 | 2.73 | 0.00 | 3.24 | 0.00 | 2.07 | 0.00 |
| *COL8A1* | 2.56 | 0.00 | 2.93 | 0.00 | 4.77 | 0.00 | 2.80 | 0.00 | 3.52 | 0.00 |
| *THBS2* | 2.56 | 0.00 | 2.73 | 0.00 | 3.79 | 0.00 | 3.25 | 0.00 | 3.61 | 0.00 |
| *CTHRC1* | 2.55 | 0.00 | 2.04 | 0.00 | 3.52 | 0.00 | 2.73 | 0.00 | 2.87 | 0.00 |
| *IGF2BP3* | 2.54 | 0.00 | 2.30 | 0.00 | 2.07 | 0.00 | 2.29 | 0.00 | 3.29 | 0.00 |
| *SFRP4* | 2.49 | 0.00 | 2.47 | 0.00 | 4.25 | 0.00 | 4.10 | 0.00 | 3.49 | 0.00 |
| *THY1* | 2.49 | 0.00 | 1.89 | 0.00 | 2.90 | 0.00 | 2.60 | 0.00 | 2.56 | 0.00 |
| *SPP1* | 2.48 | 0.00 | 2.39 | 0.00 | 4.06 | 0.00 | 3.56 | 0.00 | 3.39 | 0.00 |
| *COL1A2* | 2.48 | 0.00 | 1.28 | 0.00 | 2.20 | 0.00 | 2.14 | 0.00 | 1.48 | 0.00 |
| *COL6A3* | 2.46 | 0.00 | 1.87 | 0.00 | 2.60 | 0.00 | 2.32 | 0.00 | 2.06 | 0.00 |
| *SULF1* | 2.44 | 0.00 | 2.50 | 0.00 | 3.50 | 0.00 | 3.10 | 0.00 | 3.43 | 0.00 |
| *RAB31* | 2.33 | 0.00 | 1.32 | 0.00 | 1.81 | 0.00 | 1.18 | 0.00 | 1.51 | 0.00 |
| *TIMP1* | 2.31 | 0.00 | 1.82 | 0.00 | 2.06 | 0.00 | 2.46 | 0.00 | 2.09 | 0.00 |
| *SPOCK1* | 2.30 | 0.00 | 1.55 | 0.00 | 3.11 | 0.00 | 1.96 | 0.00 | 2.45 | 0.00 |
| *DPYSL3* | 2.29 | 0.00 | 1.21 | 0.01 | 2.60 | 0.00 | 1.26 | 0.01 | 1.96 | 0.00 |
| *SFRP2* | 2.26 | 0.01 | 1.85 | 0.00 | 5.88 | 0.00 | 2.96 | 0.00 | 1.69 | 0.02 |
| *CLDN1* | 2.20 | 0.00 | 2.19 | 0.00 | 1.77 | 0.00 | 3.87 | 0.00 | 2.66 | 0.00 |
| *SPARC* | 2.16 | 0.00 | 1.65 | 0.00 | 2.03 | 0.00 | 1.46 | 0.00 | 2.06 | 0.00 |
| *INHBA* | 2.13 | 0.00 | 4.16 | 0.00 | 4.68 | 0.00 | 4.00 | 0.00 | 4.56 | 0.00 |
| *FN1* | 2.12 | 0.00 | 1.72 | 0.00 | 3.07 | 0.00 | 2.29 | 0.00 | 2.18 | 0.00 |
| *CRISPLD1* | 2.11 | 0.00 | 1.43 | 0.01 | 3.27 | 0.00 | 1.96 | 0.00 | 2.57 | 0.00 |
| *AHNAK2* | 2.07 | 0.00 | 1.30 | 0.02 | 2.88 | 0.00 | 1.24 | 0.02 | 1.58 | 0.01 |
| *PLA2G7* | 2.07 | 0.00 | 1.34 | 0.01 | 2.11 | 0.00 | 2.19 | 0.00 | 1.20 | 0.00 |
| *COL12A1* | 2.03 | 0.00 | 1.30 | 0.00 | 1.86 | 0.00 | 1.95 | 0.00 | 2.56 | 0.00 |
| *NNMT* | 2.01 | 0.00 | 1.07 | 0.00 | 3.28 | 0.00 | 1.75 | 0.00 | 1.72 | 0.00 |
| *COL10A1* | 2.01 | 0.00 | 3.80 | 0.00 | 2.15 | 0.00 | 2.73 | 0.00 | 4.46 | 0.00 |
| *SERPINH1* | 1.99 | 0.00 | 2.79 | 0.00 | 1.77 | 0.00 | 1.91 | 0.00 | 2.34 | 0.00 |
| *VCAN* | 1.98 | 0.00 | 1.27 | 0.00 | 1.40 | 0.00 | 1.62 | 0.00 | 1.38 | 0.00 |
| *OLFML2B* | 1.98 | 0.00 | 1.38 | 0.00 | 2.50 | 0.00 | 1.64 | 0.00 | 1.89 | 0.00 |
| *SERPINE2* | 1.94 | 0.00 | 1.39 | 0.00 | 1.14 | 0.00 | 2.03 | 0.00 | 1.38 | 0.00 |
| *ITGBL1* | 1.94 | 0.00 | 1.13 | 0.01 | 3.44 | 0.00 | 1.48 | 0.00 | 1.37 | 0.02 |
| *ASPN* | 1.89 | 0.00 | 1.85 | 0.00 | 1.98 | 0.00 | 2.19 | 0.00 | 2.63 | 0.00 |
| *GPNMB* | 1.89 | 0.00 | 1.31 | 0.00 | 2.84 | 0.00 | 1.36 | 0.00 | 1.20 | 0.01 |
| *NRK* | 1.85 | 0.00 | 2.27 | 0.00 | 1.93 | 0.00 | 2.31 | 0.00 | 2.21 | 0.00 |
| *APOC1* | 1.85 | 0.00 | 2.50 | 0.00 | 1.15 | 0.00 | 2.31 | 0.00 | 2.51 | 0.00 |
| *NOX4* | 1.83 | 0.00 | 1.61 | 0.03 | 1.94 | 0.00 | 2.80 | 0.00 | 3.52 | 0.00 |
| *PMEPA1* | 1.81 | 0.00 | 1.64 | 0.00 | 1.62 | 0.00 | 1.82 | 0.00 | 1.94 | 0.00 |
| *COL3A1* | 1.81 | 0.00 | 1.16 | 0.00 | 1.41 | 0.00 | 2.03 | 0.00 | 1.15 | 0.00 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *CCDC80* | 1.81 | 0.00 | 1.06 | 0.01 | 3.32 | 0.00 | 1.39 | 0.00 | 1.11 | 0.05 |
| *PLAU* | 1.80 | 0.00 | 1.19 | 0.00 | 1.19 | 0.00 | 1.76 | 0.00 | 1.46 | 0.00 |
| *COMP* | 1.80 | 0.00 | 1.74 | 0.00 | 1.56 | 0.00 | 3.41 | 0.00 | 3.10 | 0.00 |
| *HEYL* | 1.75 | 0.00 | 1.54 | 0.00 | 2.11 | 0.00 | 1.90 | 0.00 | 1.11 | 0.00 |
| *IGFBP4* | 1.75 | 0.00 | 1.58 | 0.00 | 3.39 | 0.00 | 1.48 | 0.00 | 1.00 | 0.00 |
| *LGALS1* | 1.72 | 0.00 | 1.09 | 0.00 | 1.55 | 0.00 | 1.10 | 0.00 | 1.26 | 0.00 |
| *MMP7* | 1.70 | 0.02 | 1.72 | 0.02 | 1.65 | 0.00 | 2.83 | 0.00 | 2.04 | 0.01 |
| *PDGFRB* | 1.69 | 0.00 | 1.09 | 0.00 | 1.98 | 0.00 | 1.77 | 0.00 | 1.08 | 0.00 |
| *ISM1* | 1.67 | 0.00 | 1.59 | 0.00 | 2.17 | 0.00 | 2.21 | 0.00 | 1.39 | 0.05 |
| *LY6E* | 1.64 | 0.00 | 1.05 | 0.02 | 3.09 | 0.00 | 1.36 | 0.00 | 2.16 | 0.00 |
| *EDNRA* | 1.64 | 0.00 | 1.27 | 0.00 | 1.84 | 0.00 | 1.54 | 0.00 | 1.21 | 0.00 |
| *SRPX2* | 1.61 | 0.00 | 1.95 | 0.00 | 1.49 | 0.00 | 2.28 | 0.00 | 1.99 | 0.00 |
| *COL5A1* | 1.61 | 0.00 | 1.77 | 0.00 | 1.56 | 0.00 | 1.51 | 0.00 | 1.44 | 0.00 |
| *COL5A2* | 1.61 | 0.00 | 1.50 | 0.00 | 1.55 | 0.00 | 1.67 | 0.00 | 1.95 | 0.00 |
| *ANTXR1* | 1.59 | 0.00 | 1.01 | 0.00 | 1.04 | 0.00 | 1.42 | 0.00 | 1.57 | 0.00 |
| *ISLR* | 1.57 | 0.00 | 1.24 | 0.00 | 2.87 | 0.00 | 1.27 | 0.00 | 1.31 | 0.00 |
| *CDH11* | 1.56 | 0.00 | 1.44 | 0.00 | 2.32 | 0.00 | 1.73 | 0.00 | 1.81 | 0.00 |
| *BICC1* | 1.55 | 0.00 | 1.36 | 0.00 | 1.31 | 0.00 | 1.09 | 0.00 | 1.41 | 0.00 |
| *COL4A2* | 1.54 | 0.00 | 1.01 | 0.00 | 1.85 | 0.00 | 1.19 | 0.00 | 1.39 | 0.00 |
| *NTM* | 1.51 | 0.00 | 1.40 | 0.00 | 1.19 | 0.00 | 1.80 | 0.00 | 2.56 | 0.00 |
| *AEBP1* | 1.47 | 0.00 | 1.47 | 0.00 | 1.69 | 0.00 | 1.21 | 0.00 | 1.89 | 0.00 |
| *BGN* | 1.44 | 0.00 | 2.03 | 0.00 | 3.65 | 0.00 | 2.70 | 0.00 | 2.62 | 0.00 |
| *PRRX1* | 1.39 | 0.00 | 1.87 | 0.00 | 3.31 | 0.00 | 2.66 | 0.00 | 2.11 | 0.00 |
| *SULF2* | 1.38 | 0.00 | 1.36 | 0.00 | 1.28 | 0.00 | 1.72 | 0.00 | 1.41 | 0.00 |
| *FJX1* | 1.33 | 0.00 | 1.31 | 0.01 | 1.50 | 0.00 | 1.75 | 0.00 | 1.10 | 0.00 |
| *TNFSF4* | 1.30 | 0.00 | 1.14 | 0.00 | 1.52 | 0.00 | 1.47 | 0.00 | 1.68 | 0.00 |
| *CDH3* | 1.29 | 0.00 | 2.72 | 0.00 | 1.17 | 0.00 | 1.97 | 0.00 | 3.23 | 0.00 |
| *CPXM1* | 1.26 | 0.00 | 2.51 | 0.00 | 1.79 | 0.00 | 2.20 | 0.00 | 2.29 | 0.00 |
| *CTSK* | 1.26 | 0.00 | 1.08 | 0.00 | 2.27 | 0.00 | 1.23 | 0.00 | 1.09 | 0.00 |
| *MMP11* | 1.22 | 0.00 | 1.56 | 0.00 | 1.18 | 0.00 | 2.32 | 0.00 | 2.15 | 0.00 |
| *FKBP10* | 1.21 | 0.00 | 1.90 | 0.01 | 1.05 | 0.00 | 1.83 | 0.00 | 2.60 | 0.00 |
| *MFAP2* | 1.18 | 0.00 | 2.28 | 0.00 | 2.60 | 0.00 | 2.54 | 0.00 | 2.18 | 0.00 |
| *TMEM158* | 1.18 | 0.00 | 1.20 | 0.00 | 1.90 | 0.00 | 2.18 | 0.00 | 1.46 | 0.00 |
| *PCSK5* | 1.16 | 0.00 | 1.26 | 0.00 | 1.07 | 0.00 | 1.29 | 0.00 | 1.14 | 0.01 |
| *FBN1* | 1.16 | 0.00 | 1.58 | 0.00 | 1.94 | 0.00 | 1.23 | 0.00 | 1.77 | 0.00 |
| *NID2* | 1.10 | 0.00 | 1.69 | 0.00 | 1.85 | 0.00 | 1.95 | 0.00 | 2.28 | 0.00 |
| *CST1* | 1.10 | 0.01 | 3.93 | 0.00 | 3.04 | 0.00 | 3.67 | 0.00 | 3.56 | 0.00 |
| *GREM1* | 1.08 | 0.00 | 1.06 | 0.02 | 4.17 | 0.00 | 1.67 | 0.00 | 1.51 | 0.00 |
| *LEF1* | 1.05 | 0.00 | 1.21 | 0.00 | 1.58 | 0.00 | 1.88 | 0.00 | 1.19 | 0.00 |
| *NT5DC2** | 1.02 | 0.00 | 1.14 | 0.01 | 1.88 | 0.00 | 1.30 | 0.00 | 1.91 | 0.00 |
| *SCUBE2* | 1.02 | 0.02 | 1.31 | 0.04 | 1.04 | 0.00 | 1.45 | 0.00 | 1.69 | 0.01 |
| *FUT9* | 1.12 | 0.00 | 4.13 | 0.01 | 4.48 | 0.00 | 4.36 | 0.00 | 4.96 | 0.01 |
| *SCGN* | 1.12 | 0.00 | 1.83 | 0.01 | 1.90 | 0.00 | 2.65 | 0.00 | 1.88 | 0.01 |
| *OASL* | 1.14 | 0.00 | 1.20 | 0.00 | 2.07 | 0.00 | 1.14 | 0.01 | 2.12 | 0.00 |
| *PRDM16* | 1.17 | 0.00 | 1.30 | 0.02 | 2.92 | 0.00 | 1.93 | 0.00 | 1.32 | 0.00 |
| *TNFRSF17* | 1.17 | 0.02 | 1.62 | 0.03 | 2.53 | 0.00 | 1.78 | 0.01 | 2.93 | 0.00 |
| *PSAPL1* | 1.20 | 0.00 | 2.81 | 0.00 | 3.12 | 0.00 | 5.83 | 0.00 | 4.93 | 0.00 |
| *RNASE4* | 1.20 | 0.00 | 1.53 | 0.00 | 2.14 | 0.00 | 1.06 | 0.00 | 1.50 | 0.00 |
| *TFCP2L1* | 1.21 | 0.00 | 1.09 | 0.03 | 1.79 | 0.00 | 1.69 | 0.02 | 1.44 | 0.01 |
| *ITPKA* | 1.22 | 0.00 | 1.23 | 0.03 | 2.74 | 0.00 | 1.62 | 0.00 | 1.87 | 0.00 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *ADAM28* | 1.26 | 0.00 | 1.76 | 0.00 | 1.87 | 0.00 | 1.27 | 0.00 | 2.31 | 0.00 |
| *DHRS7* | 1.28 | 0.00 | 1.05 | 0.00 | 1.62 | 0.00 | 1.04 | 0.00 | 1.35 | 0.00 |
| *SLC41A2* | 1.30 | 0.00 | 1.29 | 0.00 | 2.26 | 0.00 | 1.20 | 0.00 | 1.55 | 0.00 |
| *MRAP2* | 1.31 | 0.00 | 1.72 | 0.02 | 2.37 | 0.00 | 1.35 | 0.00 | 1.58 | 0.00 |
| *MAGI3* | 1.32 | 0.00 | 1.04 | 0.00 | 2.07 | 0.00 | 1.14 | 0.00 | 1.31 | 0.00 |
| *CAPN13* | 1.34 | 0.00 | 3.08 | 0.01 | 1.57 | 0.00 | 3.28 | 0.00 | 3.80 | 0.00 |
| *FAM3B* | 1.34 | 0.00 | 2.64 | 0.01 | 3.80 | 0.00 | 2.70 | 0.00 | 3.08 | 0.00 |
| *PAIP2B* | 1.35 | 0.00 | 2.14 | 0.00 | 1.08 | 0.00 | 2.43 | 0.00 | 1.95 | 0.00 |
| *CYP3A5* | 1.37 | 0.00 | 1.50 | 0.00 | 2.98 | 0.00 | 1.69 | 0.02 | 2.09 | 0.00 |
| *MAL* | 1.41 | 0.00 | 2.09 | 0.00 | 3.22 | 0.00 | 3.01 | 0.00 | 4.10 | 0.00 |
| *NEUROD1* | 1.42 | 0.00 | 1.34 | 0.03 | 2.70 | 0.00 | 2.57 | 0.00 | 1.47 | 0.01 |
| *XK* | 1.42 | 0.00 | 1.21 | 0.02 | 2.54 | 0.00 | 2.09 | 0.00 | 2.11 | 0.00 |
| *SLC26A9* | 1.44 | 0.00 | 2.82 | 0.00 | 2.90 | 0.00 | 3.33 | 0.00 | 4.08 | 0.00 |
| *PTPRZ1* | 1.47 | 0.00 | 2.09 | 0.00 | 2.17 | 0.00 | 1.72 | 0.00 | 2.91 | 0.01 |
| *CCL28* | 1.47 | 0.00 | 1.08 | 0.01 | 2.51 | 0.00 | 1.23 | 0.01 | 2.00 | 0.01 |
| *MAP7D2* | 1.47 | 0.00 | 2.27 | 0.02 | 3.52 | 0.00 | 1.62 | 0.00 | 3.53 | 0.00 |
| *NTN4* | 1.48 | 0.00 | 1.13 | 0.00 | 1.51 | 0.00 | 1.30 | 0.00 | 1.64 | 0.00 |
| *SMPD3* | 1.51 | 0.00 | 1.23 | 0.00 | 2.53 | 0.00 | 1.41 | 0.01 | 1.48 | 0.00 |
| *GRAMD1C* | 1.51 | 0.00 | 1.21 | 0.00 | 2.55 | 0.00 | 1.15 | 0.00 | 1.17 | 0.00 |
| *PDZD3* | 1.52 | 0.00 | 1.78 | 0.01 | 1.15 | 0.00 | 1.20 | 0.05 | 2.09 | 0.01 |
| *RNASE1* | 1.55 | 0.00 | 1.61 | 0.00 | 2.15 | 0.00 | 1.21 | 0.00 | 2.51 | 0.00 |
| *AMPD1* | 1.55 | 0.00 | 1.72 | 0.00 | 2.87 | 0.00 | 2.44 | 0.00 | 2.68 | 0.00 |
| *LIFR* | 1.57 | 0.00 | 2.27 | 0.00 | 1.16 | 0.00 | 1.73 | 0.00 | 2.43 | 0.00 |
| *KCNJ15* | 1.59 | 0.00 | 2.77 | 0.00 | 1.33 | 0.00 | 2.65 | 0.00 | 3.46 | 0.00 |
| *ANG* | 1.59 | 0.00 | 1.29 | 0.00 | 2.50 | 0.00 | 1.27 | 0.00 | 2.06 | 0.00 |
| *SELENBP1* | 1.60 | 0.00 | 1.09 | 0.01 | 1.80 | 0.00 | 1.40 | 0.00 | 1.75 | 0.00 |
| *SCIN* | 1.62 | 0.00 | 1.28 | 0.01 | 4.08 | 0.00 | 1.64 | 0.00 | 2.26 | 0.00 |
| *GAS2* | 1.63 | 0.00 | 1.27 | 0.04 | 1.95 | 0.00 | 1.28 | 0.00 | 1.03 | 0.04 |
| *GPT2* | 1.66 | 0.00 | 1.42 | 0.00 | 1.12 | 0.00 | 2.21 | 0.00 | 1.01 | 0.02 |
| *UPK1B* | 1.66 | 0.00 | 2.65 | 0.01 | 4.87 | 0.00 | 1.35 | 0.01 | 3.50 | 0.00 |
| *UBL3* | 1.69 | 0.00 | 1.28 | 0.00 | 1.47 | 0.00 | 1.03 | 0.00 | 1.76 | 0.00 |
| *GCNT2* | 1.69 | 0.00 | 1.49 | 0.01 | 1.86 | 0.00 | 1.16 | 0.00 | 1.89 | 0.00 |
| *TMEM116* | 1.69 | 0.00 | 1.03 | 0.00 | 1.01 | 0.00 | 1.29 | 0.00 | 1.36 | 0.00 |
| *PDGFD* | 1.72 | 0.00 | 1.70 | 0.00 | 1.03 | 0.00 | 1.36 | 0.00 | 1.69 | 0.00 |
| *METTL7A* | 1.72 | 0.00 | 1.64 | 0.00 | 1.37 | 0.00 | 1.40 | 0.00 | 1.90 | 0.00 |
| *SERPINA4* | 1.74 | 0.00 | 1.48 | 0.04 | 1.21 | 0.00 | 2.69 | 0.00 | 1.20 | 0.02 |
| *RFX6* | 1.75 | 0.00 | 2.24 | 0.02 | 3.15 | 0.00 | 2.54 | 0.00 | 3.03 | 0.00 |
| *C1orf210* | 1.75 | 0.00 | 1.01 | 0.00 | 1.93 | 0.00 | 1.29 | 0.01 | 1.25 | 0.00 |
| *B3GNT6* | 1.77 | 0.00 | 2.47 | 0.00 | 1.80 | 0.00 | 3.16 | 0.00 | 2.99 | 0.00 |
| *BEX5* | 1.77 | 0.00 | 1.97 | 0.00 | 1.94 | 0.00 | 1.60 | 0.00 | 1.12 | 0.04 |
| *DGKD* | 1.78 | 0.00 | 1.34 | 0.00 | 1.71 | 0.00 | 1.64 | 0.00 | 1.56 | 0.00 |
| *LDHD* | 1.81 | 0.00 | 1.87 | 0.00 | 1.85 | 0.00 | 2.20 | 0.00 | 2.16 | 0.00 |
| *ZNF57* | 1.81 | 0.00 | 1.27 | 0.00 | 1.56 | 0.00 | 1.40 | 0.00 | 1.40 | 0.00 |
| *COBLL1* | 1.83 | 0.00 | 1.31 | 0.00 | 1.24 | 0.00 | 1.00 | 0.00 | 1.02 | 0.00 |
| *CNTN3* | 1.87 | 0.00 | 2.71 | 0.00 | 2.07 | 0.00 | 1.78 | 0.00 | 2.78 | 0.00 |
| *LRRC66* | 1.87 | 0.00 | 1.28 | 0.01 | 3.69 | 0.00 | 1.09 | 0.01 | 2.38 | 0.00 |
| *PTPRN2* | 1.94 | 0.00 | 1.10 | 0.00 | 1.80 | 0.00 | 1.34 | 0.00 | 1.37 | 0.01 |
| *FER1L4* | 1.94 | 0.00 | 1.11 | 0.01 | 2.61 | 0.00 | 1.61 | 0.00 | 1.87 | 0.00 |
| *IGFBP2* | 1.98 | 0.00 | 1.66 | 0.00 | 1.11 | 0.00 | 1.51 | 0.00 | 1.98 | 0.01 |
| *NRG4* | 2.00 | 0.00 | 2.13 | 0.01 | 1.05 | 0.00 | 2.20 | 0.00 | 2.41 | 0.00 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *SCNN1B* | 2.00 | 0.00 | 1.87 | 0.01 | 2.21 | 0.00 | 3.04 | 0.00 | 3.16 | 0.00 |
| *ANKRD22* | 2.02 | 0.00 | 1.26 | 0.00 | 2.20 | 0.00 | 1.92 | 0.01 | 1.45 | 0.00 |
| *PPP1R36* | 2.03 | 0.00 | 1.02 | 0.00 | 1.93 | 0.00 | 1.56 | 0.00 | 1.69 | 0.00 |
| *RAB27B* | 2.04 | 0.00 | 1.20 | 0.00 | 2.86 | 0.00 | 1.68 | 0.00 | 2.05 | 0.00 |
| *CYP2C18* | 2.08 | 0.00 | 2.10 | 0.00 | 3.64 | 0.00 | 2.56 | 0.00 | 2.82 | 0.00 |
| *LRRC31* | 2.12 | 0.00 | 1.29 | 0.04 | 1.45 | 0.00 | 2.20 | 0.00 | 2.60 | 0.00 |
| *RDH12* | 2.13 | 0.00 | 2.69 | 0.00 | 2.52 | 0.00 | 3.87 | 0.00 | 3.62 | 0.00 |
| *ALDH1A1* | 2.16 | 0.00 | 2.09 | 0.00 | 1.64 | 0.00 | 1.75 | 0.00 | 2.00 | 0.00 |
| *AKR1C1* | 2.17 | 0.00 | 1.28 | 0.01 | 2.45 | 0.00 | 1.59 | 0.00 | 2.73 | 0.00 |
| *PLLP* | 2.22 | 0.00 | 1.28 | 0.00 | 2.59 | 0.00 | 1.48 | 0.00 | 2.47 | 0.00 |
| *PXMP2* | 2.26 | 0.00 | 1.40 | 0.00 | 1.72 | 0.00 | 1.88 | 0.00 | 1.51 | 0.00 |
| *CWH43* | 2.30 | 0.00 | 2.67 | 0.00 | 3.65 | 0.00 | 4.37 | 0.00 | 5.28 | 0.00 |
| *SULT1B1* | 2.31 | 0.00 | 1.59 | 0.00 | 2.13 | 0.00 | 2.23 | 0.00 | 2.75 | 0.00 |
| *TMPRSS2* | 2.34 | 0.00 | 1.99 | 0.00 | 1.96 | 0.00 | 1.98 | 0.00 | 2.03 | 0.00 |
| *CLDN18* | 2.38 | 0.00 | 2.38 | 0.01 | 2.56 | 0.00 | 3.62 | 0.00 | 3.24 | 0.02 |
| *GUCA2B* | 2.40 | 0.00 | 3.33 | 0.00 | 1.11 | 0.00 | 1.63 | 0.00 | 2.95 | 0.00 |
| *FMO5* | 2.45 | 0.00 | 1.33 | 0.00 | 2.56 | 0.00 | 2.26 | 0.00 | 1.54 | 0.00 |
| *UGT2B15* | 2.46 | 0.00 | 2.00 | 0.02 | 3.52 | 0.00 | 3.42 | 0.00 | 2.69 | 0.00 |
| *C16orf89* | 2.49 | 0.00 | 2.82 | 0.00 | 2.17 | 0.00 | 2.80 | 0.00 | 3.35 | 0.01 |
| *ODAM* | 2.52 | 0.00 | 1.93 | 0.01 | 1.11 | 0.00 | 2.08 | 0.00 | 1.50 | 0.03 |
| *PKIB* | 2.58 | 0.00 | 1.40 | 0.01 | 3.24 | 0.00 | 2.01 | 0.00 | 2.93 | 0.00 |
| *FBP2* | 2.58 | 0.00 | 3.55 | 0.00 | 2.68 | 0.00 | 5.21 | 0.00 | 4.01 | 0.00 |
| *LTF* | 2.59 | 0.00 | 3.03 | 0.05 | 2.01 | 0.00 | 2.84 | 0.00 | 4.78 | 0.00 |
| *PBLD* | 2.64 | 0.00 | 1.44 | 0.00 | 2.95 | 0.00 | 1.27 | 0.00 | 2.66 | 0.00 |
| *EPN3* | 2.66 | 0.00 | 2.11 | 0.00 | 1.23 | 0.00 | 2.88 | 0.00 | 1.82 | 0.00 |
| *CTSE* | 2.67 | 0.00 | 2.23 | 0.01 | 3.02 | 0.00 | 2.87 | 0.00 | 2.80 | 0.02 |
| *TRIM50* | 2.72 | 0.00 | 3.68 | 0.00 | 1.46 | 0.00 | 5.35 | 0.00 | 5.44 | 0.00 |
| *IRX3* | 2.84 | 0.00 | 2.61 | 0.02 | 1.80 | 0.00 | 2.85 | 0.00 | 2.34 | 0.01 |
| *CAPN8* | 2.88 | 0.00 | 1.59 | 0.01 | 3.69 | 0.00 | 2.11 | 0.01 | 2.71 | 0.00 |
| *GPRC5C* | 2.88 | 0.00 | 1.43 | 0.00 | 1.21 | 0.00 | 1.86 | 0.00 | 1.67 | 0.00 |
| *RASSF6* | 2.94 | 0.00 | 1.85 | 0.00 | 3.26 | 0.00 | 1.46 | 0.00 | 2.02 | 0.00 |
| *HPGD* | 2.94 | 0.00 | 2.17 | 0.00 | 3.53 | 0.00 | 2.75 | 0.00 | 3.48 | 0.00 |
| *CA2* | 2.96 | 0.00 | 2.10 | 0.00 | 3.21 | 0.00 | 3.64 | 0.00 | 3.04 | 0.00 |
| *PDIA2* | 2.98 | 0.00 | 3.17 | 0.00 | 1.71 | 0.00 | 4.45 | 0.00 | 2.70 | 0.01 |
| *KLK11* | 2.99 | 0.00 | 2.78 | 0.00 | 1.96 | 0.00 | 2.70 | 0.00 | 1.84 | 0.01 |
| *FBXL13* | 3.01 | 0.00 | 2.32 | 0.00 | 1.87 | 0.00 | 2.04 | 0.00 | 2.38 | 0.00 |
| *LYPD6B* | 3.03 | 0.00 | 1.71 | 0.01 | 3.07 | 0.00 | 3.22 | 0.00 | 3.15 | 0.00 |
| *LIPF* | 3.08 | 0.00 | 5.33 | 0.00 | 2.94 | 0.00 | 10.30 | 0.00 | 6.61 | 0.00 |
| *ARL14* | 3.08 | 0.00 | 1.68 | 0.01 | 2.98 | 0.00 | 2.43 | 0.01 | 2.65 | 0.02 |
| *ALDH3A1* | 3.16 | 0.00 | 2.76 | 0.00 | 2.62 | 0.00 | 2.81 | 0.00 | 3.92 | 0.00 |
| *DUOX2* | 3.16 | 0.00 | 2.33 | 0.01 | 2.05 | 0.00 | 2.80 | 0.00 | 2.31 | 0.01 |
| *REG1A* | 3.18 | 0.00 | 2.99 | 0.01 | 2.66 | 0.00 | 3.50 | 0.00 | 2.42 | 0.04 |
| *TFF1* | 3.24 | 0.00 | 2.01 | 0.00 | 3.00 | 0.00 | 4.18 | 0.00 | 4.23 | 0.01 |
| *CAPN9* | 3.25 | 0.00 | 2.93 | 0.00 | 4.14 | 0.00 | 4.09 | 0.00 | 3.50 | 0.00 |
| *TFF2* | 3.33 | 0.00 | 3.37 | 0.00 | 3.88 | 0.00 | 5.09 | 0.00 | 4.31 | 0.01 |
| *VSIG2* | 3.34 | 0.00 | 2.72 | 0.00 | 3.93 | 0.00 | 3.64 | 0.00 | 3.58 | 0.00 |
| *TMEM171* | 3.36 | 0.00 | 1.73 | 0.00 | 1.71 | 0.00 | 2.15 | 0.00 | 1.59 | 0.00 |
| *SULT1C2* | 3.37 | 0.00 | 1.51 | 0.02 | 3.40 | 0.00 | 2.77 | 0.00 | 2.75 | 0.00 |
| *AKR7A3* | 3.39 | 0.00 | 2.34 | 0.00 | 1.90 | 0.00 | 2.33 | 0.00 | 2.12 | 0.00 |
| *MUC6* | 3.42 | 0.00 | 2.44 | 0.01 | 1.99 | 0.00 | 5.84 | 0.00 | 1.91 | 0.05 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *ESRRG* | 3.45 | 0.00 | 4.03 | 0.01 | 1.84 | 0.00 | 2.29 | 0.00 | 5.28 | 0.00 |
| *AQP4* | 3.46 | 0.00 | 5.49 | 0.00 | 1.23 | 0.00 | 5.55 | 0.00 | 7.84 | 0.00 |
| *AADAC* | 3.46 | 0.00 | 1.03 | 0.01 | 3.27 | 0.00 | 2.70 | 0.00 | 2.86 | 0.00 |
| *DNER* | 3.58 | 0.00 | 2.63 | 0.00 | 1.03 | 0.00 | 4.09 | 0.00 | 3.37 | 0.00 |
| *PGC* | 3.60 | 0.00 | 4.72 | 0.00 | 3.18 | 0.00 | 7.01 | 0.00 | 4.35 | 0.00 |
| *SCGB2A1* | 3.62 | 0.00 | 2.70 | 0.00 | 5.28 | 0.00 | 3.99 | 0.00 | 3.89 | 0.01 |
| *CCKBR* | 3.67 | 0.00 | 3.30 | 0.00 | 1.24 | 0.00 | 6.56 | 0.00 | 4.20 | 0.00 |
| *CLIC6* | 3.71 | 0.00 | 3.15 | 0.00 | 1.06 | 0.01 | 3.23 | 0.00 | 4.22 | 0.00 |
| *ALDOB* | 3.74 | 0.00 | 2.50 | 0.02 | 1.53 | 0.00 | 2.80 | 0.00 | 3.69 | 0.00 |
| *ADH7* | 3.74 | 0.00 | 2.00 | 0.00 | 3.21 | 0.00 | 3.62 | 0.00 | 4.27 | 0.00 |
| *PIK3C2G* | 3.76 | 0.00 | 2.11 | 0.03 | 2.61 | 0.00 | 3.81 | 0.00 | 3.71 | 0.00 |
| *PSCA* | 3.82 | 0.00 | 3.01 | 0.00 | 3.09 | 0.00 | 5.71 | 0.00 | 3.64 | 0.01 |
| *SLC26A7* | 3.86 | 0.00 | 2.52 | 0.02 | 2.22 | 0.00 | 3.02 | 0.00 | 3.61 | 0.00 |
| *CXCL17* | 3.87 | 0.00 | 3.38 | 0.00 | 2.91 | 0.00 | 5.19 | 0.00 | 4.58 | 0.00 |
| *SSTR1* | 4.10 | 0.00 | 2.48 | 0.00 | 4.02 | 0.00 | 3.57 | 0.00 | 3.59 | 0.00 |
| *TMED6* | 4.11 | 0.00 | 3.11 | 0.00 | 2.73 | 0.00 | 4.01 | 0.00 | 3.23 | 0.00 |
| *CA9* | 4.17 | 0.00 | 2.90 | 0.00 | 2.92 | 0.00 | 4.96 | 0.00 | 3.76 | 0.00 |
| *AKR1B10* | 4.26 | 0.00 | 2.35 | 0.00 | 4.09 | 0.00 | 4.12 | 0.00 | 4.55 | 0.00 |
| *KCNJ16* | 4.35 | 0.00 | 3.51 | 0.01 | 3.38 | 0.00 | 5.66 | 0.00 | 6.35 | 0.00 |
| *VSIG1* | 4.41 | 0.00 | 3.61 | 0.00 | 4.69 | 0.00 | 4.61 | 0.00 | 4.88 | 0.00 |
| *CPA2* | 4.48 | 0.00 | 3.55 | 0.00 | 1.92 | 0.00 | 4.96 | 0.00 | 3.94 | 0.00 |
| *ANXA10* | 4.54 | 0.00 | 2.94 | 0.01 | 3.63 | 0.00 | 4.57 | 0.00 | 3.44 | 0.04 |
| *GKN1* | 4.86 | 0.00 | 5.29 | 0.00 | 5.97 | 0.00 | 9.31 | 0.00 | 8.99 | 0.00 |
| *SST* | 5.02 | 0.00 | 4.04 | 0.00 | 5.27 | 0.00 | 5.17 | 0.00 | 3.19 | 0.02 |
| *KCNE2* | 5.07 | 0.00 | 5.54 | 0.00 | 3.03 | 0.00 | 7.47 | 0.00 | 5.63 | 0.00 |
| *ATP4A* | 5.17 | 0.00 | 6.09 | 0.00 | 3.31 | 0.00 | 9.58 | 0.00 | 8.02 | 0.00 |
| *CHGA* | 5.27 | 0.00 | 3.47 | 0.00 | 1.19 | 0.00 | 5.77 | 0.00 | 4.00 | 0.00 |
| *GKN2* | 5.28 | 0.00 | 5.15 | 0.01 | 5.88 | 0.00 | 9.52 | 0.00 | 9.18 | 0.00 |
| *ATP4B* | 5.44 | 0.00 | 5.54 | 0.00 | 4.08 | 0.00 | 8.92 | 0.00 | 7.96 | 0.00 |
| *KRT20* | 5.48 | 0.00 | 2.32 | 0.03 | 4.87 | 0.00 | 3.97 | 0.00 | 4.90 | 0.00 |

*, The genes above were up-regulated and the belows were down-regulated. The logFC values of down-regulated genes were absolute.

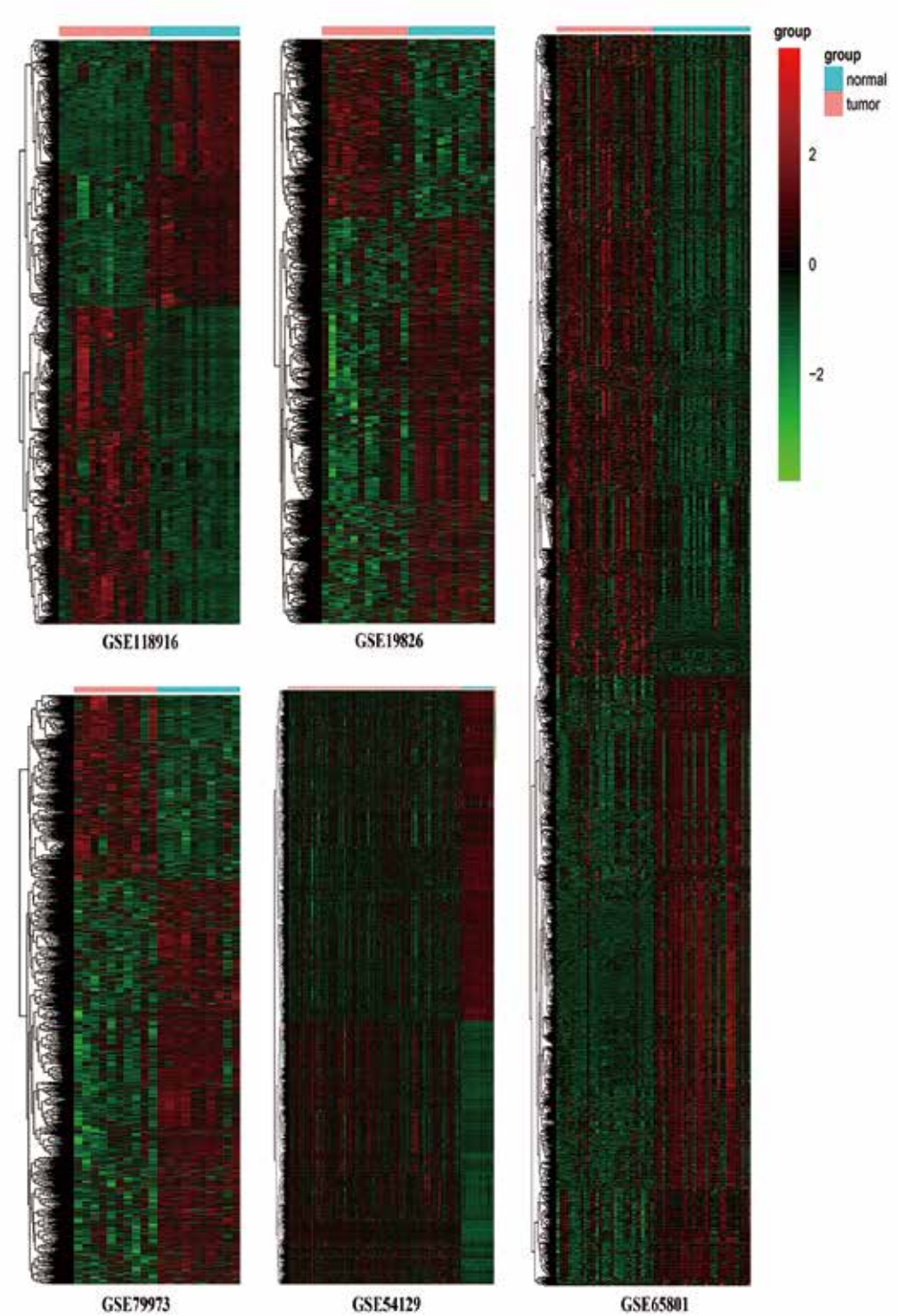


**Figure S1** The cluster heat map of five datasets.

**Table S2** The results of GO and KEGG of DEGs

| Category | ID | Term | Genes | Adj-P |
|---|---|---|---|---|
| BP | GO:0030198 | Extracellular matrix organization | *SPARC; MMP7; COL12A1; FN1; BGN; NID2; COMP; GREM1; MMP11; COL3A1; VCAN; COL1A2; COL4A2; COL5A1; CTSK; COL5A2; MFAP2; SERPINH1; SPP1; COL8A1; COL10A1; COL6A3; TIMP1; FBN1* | 0.000 |
| BP | GO:0030199 | Collagen fibril organization | *GREM1; COL3A1; COL1A2; COL5A1; COL12A1; COL5A2; SERPINH1* | 0.000 |
| BP | GO:0001501 | Skeletal system development | *COMP; VCAN; COL1A2; IGFBP4; COL12A1; CDH11; COL10A1; AEBP1; SULF1; PCSK5; FBN1; SULF2* | 0.000 |
| BP | GO:0071822 | Protein complex subunit organization | *GREM1; COL3A1; COL1A2; COL5A1; COL12A1; COL5A2; SERPINH1* | 0.001 |
| BP | GO:0006508 | Proteolysis | *CPA2; MMP7; TMPRSS2; AEBP1; PCSK5; PGC; CAPN13; CAPN8; CAPN9; FAP; PLAU; CTSK; CPXM1; CTSE* | 0.005 |
| BP | GO:0022617 | Extracellular matrix disassembly | *MMP11; MMP7; CTSK; SPP1; FN1; TIMP1; FBN1* | 0.024 |
| BP | GO:0035987 | Endodermal cell differentiation | *COL4A2; COL12A1; FN1; COL8A1; INHBA* | 0.022 |
| BP | GO:0010951 | Negative regulation of endopeptidase activity | *CST1; SERPINE2; SERPINH1; SPOCK1; TIMP1; SERPINA4; LTF* | 0.026 |
| BP | GO:0001706 | Endoderm formation | *COL4A2; COL12A1; FN1; COL8A1; INHBA* | 0.030 |
| BP | GO:0048592 | Eye morphogenesis | *COL5A1; COL5A2; MFAP2; FBN1* | 0.032 |
| BP | GO:0010107 | Potassium ion import | *ATP4B; ATP4A; KCNE2; KCNJ15; KCNJ16* | 0.032 |
| BP | GO:0042127 | Regulation of cell proliferation | *PDGFRB; GKN1; REG1A; LEF1; FN1; GKN2; LIFR; INHBA; SSTR1; GREM1; SFRP4; SFRP2; SCIN; CCKBR; GPNMB; SST; PDGFD; TNFSF4; GCNT2; NOX4; TIMP1* | 0.036 |
| CC | GO:0005788 | Endoplasmic reticulum lumen | *IGFBP4; COL12A1; FN1; C3; COL3A1; VCAN; LGALS1; COL1A2; COL4A2; COL5A1; PDGFD; COL5A2; SERPINH1; SPP1; COL8A1; COL10A1; COL6A3; TIMP1; FBN1* | 0.000 |
| MF | GO:0005178 | Integrin binding | *COL3A1; SFRP2; GPNMB; COL5A1; FAP; FN1; THY1; THBS4; FBN1* | 0.001 |
| MF | GO:0005518 | Collagen binding | *COMP; SPARC; CTSK; SERPINH1; FN1; ANTXR1; NID2* | 0.001 |
| MF | GO:0048407 | Platelet-derived growth factor binding | *PDGFRB; COL3A1; COL1A2; COL5A1* | 0.002 |
| MF | GO:0002020 | Protease binding | *COMP; PDZD3; CST1; COL3A1; COL1A2; FAP; FN1; TIMP1* | 0.016 |
| MF | GO:0004033 | Aldo-keto reductase activity | *ALDH3A1; AKR7A3; AKR1B10; AKR1C1* | 0.014 |
| MF | GO:0008106 | Alcohol dehydrogenase activity | *ALDH3A1; AKR1B10; RDH12; AKR1C1* | 0.012 |
| MF | GO:0005509 | Calcium ion binding | *COMP; SCGN; SPARC; CDH3; SCIN; DNER; CDH11; SPOCK1; DUOX2; ASPN; THBS4; FBN1* | 0.014 |
| MF | GO:0004866 | Endopeptidase inhibitor activity | *CST1; SERPINE2; SERPINH1; SPOCK1; TIMP1; SERPINA4; LTF* | 0.037 |
| KEGG | hsa04971 | Gastric acid secretion | *ATP4B; ATP4A; KCNE2; CCKBR; CA2; SST; KCNJ15; KCNJ16; SLC26A7* | 0.000 |
| KEGG | hsa04974 | Protein digestion and absorption | *CPA2; COL3A1; COL1A2; COL4A2; COL5A1; COL12A1; COL5A2; COL10A1; COL6A3* | 0.000 |
| KEGG | hsa04512 | ECM-receptor interaction | *COMP; COL1A2; COL4A2; SPP1; FN1; COL6A3; THBS2; THBS4* | 0.000 |
| KEGG | hsa04510 | Focal adhesion | *COMP; PDGFRB; COL1A2; COL4A2; PDGFD; FN1; SPP1; COL6A3; THBS2; THBS4* | 0.006 |
| KEGG | hsa00830 | Retinol metabolism | *RDH12; UGT2B15; ALDH1A1; ADH7; CYP3A5; CYP2C18* | 0.006 |
| KEGG | hsa00980 | Metabolism of xenobiotics by cytochrome P450 | *ALDH3A1; AKR7A3; AKR1C1; UGT2B15; ADH7; CYP3A5* | 0.009 |

**Table S3** The results of KEGG pathway of the hub genes

| ID | Term | Adj_P | Genes |
|---|---|---|---|
| hsa04974 | Protein digestion and absorption | 0.000 | *COL3A1, COL1A2, COL5A1, COL4A2, COL12A1, COL5A2, COL6A3, COL10A1* |
| hsa04512 | ECM-receptor interaction | 0.000 | *COL1A2, COL4A2, FN1, SPP1, COL6A3, THBS2* |
| hsa04510 | Focal adhesion | 0.000 | *PDGFRB, COL1A2, COL4A2, FN1, SPP1, COL6A3, THBS2* |
| hsa05165 | Human papillomavirus infection | 0.000 | *PDGFRB, COL1A2, COL4A2, FN1, SPP1, COL6A3, THBS2* |
| hsa04151 | PI3K-Akt signaling pathway | 0.000 | *PDGFRB, COL1A2, COL4A2, FN1, SPP1, COL6A3, THBS2* |
| hsa05146 | Amoebiasis | 0.000 | *COL3A1, COL1A2, COL4A2, FN1* |
| hsa04933 | AGE-RAGE signaling pathway in diabetic complications | 0.000 | *COL3A1, COL1A2, COL4A2, FN1* |
| hsa04926 | Relaxin signaling pathway | 0.011 | *COL3A1, COL1A2, COL4A2* |

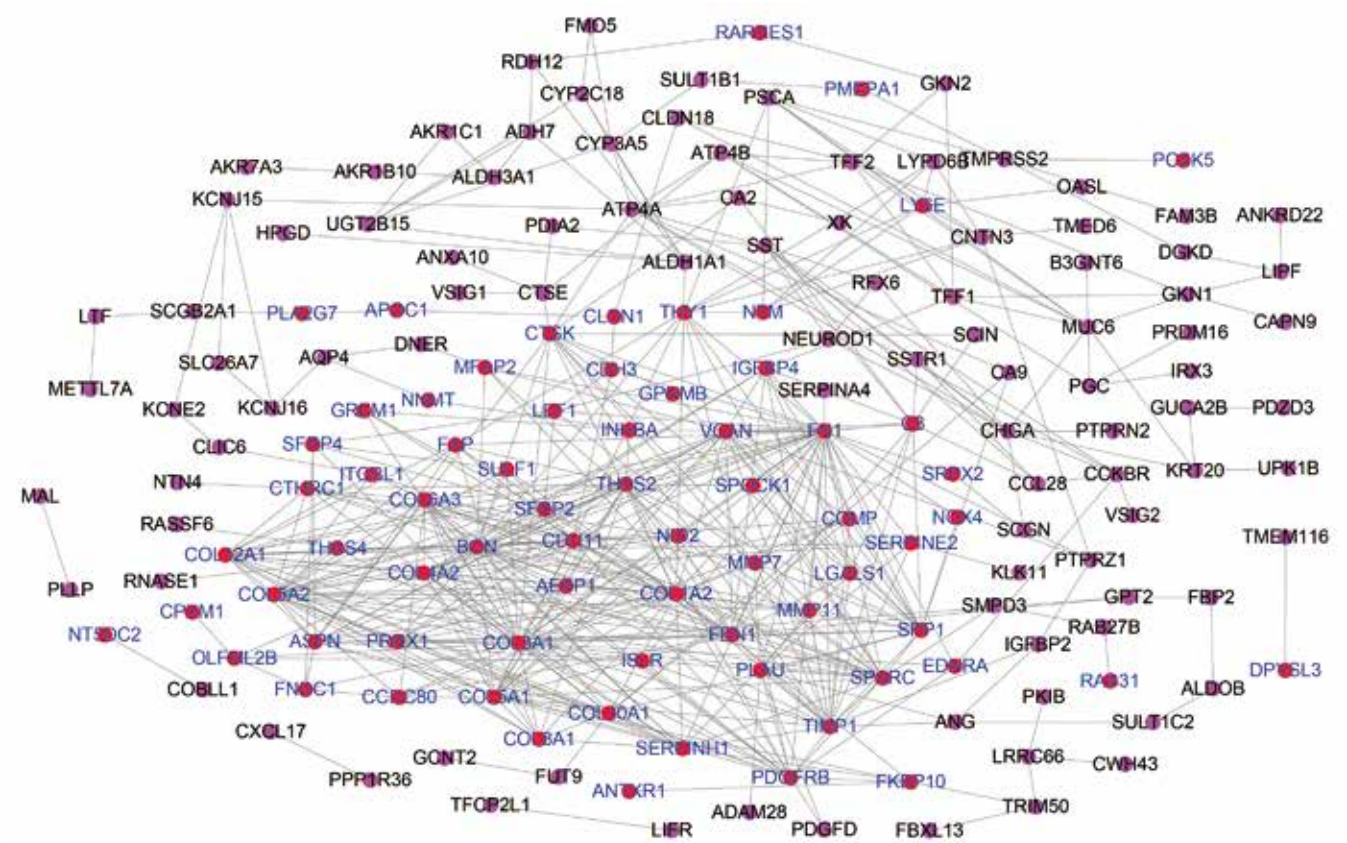


**Figure S2** The PPI network of DEGs. Red solid circles represent up-regulated DEGs, and purples represent down-regulated DEGs. DEGs, differentially expressed genes.