

Construction of Prognostic Prediction Model for Stomach Adenocarcinoma Based on the TCGA Database.

Zhengzhong Gu

Second Affiliated Hospital of Jilin University

Xiaohan Cui

Changzhou No.2 People's Hospital

Xudong Wang (✉ wangxudong1971@126.com)

Second Affiliated Hospital of Jilin University

Research Article

Keywords: Stomach adenocarcinoma, Bioinformatics analysis, COL3A1, TCGA, Tumorigenesis

Posted Date: December 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-114928/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Prognostic prediction models have been developed to detect new biomarkers of gastric cancer (GC). The identification of new biomarkers could provide theoretical foundations for the application of molecular targeted therapy in advanced GC. The aim of this study was to construct a prognostic prediction model for stomach adenocarcinoma (STAD) based on The Cancer Genome Atlas (TCGA) database.

Methods: First, we used the "limma" package to screen differentially expressed genes (DEGs) based on TCGA database. Gene ontology (GO) analysis was performed using the "ClusterProfiler" package. The interactions between proteins and the relationships between differentially expressed genes and clinical features were analyzed by protein-protein interaction (PPI) network analysis and weighted gene coexpression network analysis (WGCNA), respectively. Then, gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) were used to identify differentially enriched pathways. The GenVisR package and CIBERSORT were used to identify mutations and assess immune infiltration. Finally, the expression of COL3A1 in STAD tissues was verified by reverse transcription quantitative polymerase chain reaction (RT-qPCR) and western blotting.

Results: Six differentially expressed genes were screened out, namely, COL3A1, ADAMTS12, BGN, FNDC1, AEBP1 and HTRA3. The enrichment results showed that differentially expressed genes were involved in multiple pathways in STAD, such as those related to the extracellular matrix, extracellular structure organization, and extracellular matrix organization. The differentially expressed genes were related to immune infiltration via the mitogen-activated protein kinase (MAPK) and phosphatidylinositol 3-kinase/protein kinase B (PI3K/AKT) pathways. The western blotting and RT-qPCR results suggested that COL3A1 was overexpressed in STAD tissues compared with normal tissues.

Conclusion: COL3A1, ADAMTS12, BGN, FNDC1, AEBP1 and HTRA3 could play important roles in the tumorigenesis and progression of STAD via various pathways, including those involving the extracellular matrix, extracellular structure organization, and extracellular matrix organization. COL3A1, ADAMTS12, BGN, FNDC1, AEBP1, and HTRA3 act as oncogenes in most cancers and may be biomarkers. Additionally, the identification of COL3A1 as a candidate biomarker provides a direction for further research on the role of tumor immunity in gastric cancer.

Background

Gastric cancer (GC) is one of the most common tumors and one of the most common causes of tumor-related death [1]. Stomach adenocarcinoma (STAD) is the main pathological type of GC and can be further divided into the intestinal type and other pathological types [2]. At present, surgery is a significant method for the treatment of STAD. Other treatment methods, such as chemotherapy, radiotherapy and immunotherapy, are also very important. Most studies on STAD have focused on the unrevealed molecular mechanisms and the discovery of new STAD-related biomarkers [3, 4]. On the one hand, these

discoveries of effective biomarkers have contributed to the early diagnosis of stomach adenocarcinoma. On the other hand, these discoveries provide theoretical foundations for the application of molecular targeted therapy in advanced GC.

Traditional studies have revealed that oncogenes play important roles in the proliferation, metastasis and drug resistance of STAD. Recently, an increasing number of studies have found that oncogene expression is correlated with the tumor microenvironment and immune infiltration. For example, Pan and colleagues found that layilin (LAYN) expression was related to the infiltration of immune cells, especially in STAD patients [5]. Previous research showed that in STAD, C-X-C motif chemokine ligand 8, which is secreted by macrophages, could inhibit the function of CD8⁺ T cells by facilitating the expression of programmed death ligand-1 (PD-L1) on macrophages [6]. Therefore, further exploration of molecules related to the tumorigenesis, metastasis, and drug resistance of STAD would improve the understanding of the diagnosis, treatment and prevention of STAD. The discovery of new oncogenes that affect the infiltration of immune cells in STAD would provide a new direction for future research on STAD.

Bioinformatics analysis based on The Cancer Genome Atlas (TCGA) database revealed that the expressions of type III collagen α 1 chain (COL3A1), ADAM metallopeptidase with thrombospondin type 1 motif 12 (ADAMTS12), biglycan (BGN), fibronectin domain containing 1 (FNDC1), AE binding protein 1 (AEBP1) and HtrA serine peptidase 3 (HTRA3) were all upregulated in STAD compared to normal controls. The above upregulated mRNAs were associated with poor overall survival (OS). The biological characteristics of these six genes were analyzed using gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA). Mutation analysis and immune infiltration analysis revealed potential correlations between the six genes and immune infiltration in STAD. Therefore, the aim of this study was to construct a prognostic prediction model for STAD based on the TCGA database and reveal the potential regulatory mechanisms of COL3A1, ADAMTS12, BGN, FNDC1, AEBP1 and HTRA3 in STAD.

Methods

Data acquisition and preprocessing

The gene expression profiles and mutation data of STAD patients were downloaded from the TCGA database, which is the largest public cancer database, integrating data from 33 cancer genomes and clinical prognostic information. The clinical and prognostic information was downloaded in the form of bcr-xml files. Data preprocessing included background correction, data standardization, supplementation of missing values, and distinction between normal and tumor groups. The "Homo_sapiens.GRCh38.99.chr.gtf" file was used for ID conversion of the probe.

Differential gene analysis and gene ontology (GO) enrichment analysis

The "limma" package on the Bioconductor website, with thresholds of $|\log \text{fold-change (FC)}| > 1$ and $P\text{-value} < 0.05$, was used to identify differentially expressed genes (DEGS; T4 and T1 patients). The "clusterProfiler" package was utilized to perform GO enrichment analysis of DEGs. Interactions with a

composite score > 0.4 were considered statistically significant. The bioinformatics software Cytoscape (version 3.4) was subsequently utilized to visualize the protein-protein interaction (PPI) network. Finally, top nodes were calculated based on the maximal clique centrality (MCC) value of each node through the cytoHubba plug-in.

PPI network construction

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) website (<http://string-db.org>) was used to analyze the interactions between all DEGs to explore possible biological mechanisms.

Weighted gene coexpression network analysis (WGCNA)

To identify genes significantly related to the T clinical stage, a coexpression network was constructed using WGCNA. The program “Good Samples Genes” was used to check whether the differentially expressed mRNAs (DEmRNAs) of the data matrix met the standard and eliminate unqualified data. The software threshold function was used to calculate the value of the soft threshold power parameter β ($\beta=6$) to ensure the scale-lessness of the network. The hierarchical clustering method was used to construct a dendrogram to calculate the correlation between the modular characteristic genes and clinical traits, and these data were also analysed to screen the genes correlated with the T staging of STAD samples.

Prognostic model and nomogram construction

The Cox regression method was used to screen prognostic genes and construct a prognostic model, and each patient received a risk score. The formula for constructing the prognostic model was as follows: risk score = COL3A1 expression * 0.208 + ADAMTS12 expression * 0.172 + BGN expression * 0.005 + FNDC1 expression * 0.223 + AEBP1 expression * 0.053 + HTRA3 expression * 0.179. The risk score was combined with the clinical data of STAD patients to establish a nomogram to analyze the 1-, 3-, and 5-year survival rates of patients. The predictive performance of the nomogram was evaluated with a decision curve.

GSVA and GSEA

GSVA was performed by the GSVA software package in R software, an unsupervised gene enrichment method that calculates the changes in pathway activity among different samples. The patients were divided into groups according to their genes, and then GSEA was performed to evaluate their biological characteristics. The threshold was set to FDR<0.25 and P-value<0.05. The standard for gene sequencing was Signal2Noise. The high expression group was regarded as the experimental group, while the low expression group was regarded as the control group. The gene set used was c2.cp.kegg.v7.2.symbols.gmt.

Analysis of mutation and immune infiltration in high-risk and low-risk patients

After the data were sorted, the GenVisR package was used to analyze single nucleotide polymorphism (SNP) mutations in high-risk and low-risk patients. The CIBERSORT method was used to quantify 21 types of immune cells in the samples. R software was used to link the TIMER website online to analyze the correlation between immune cells and 6 model genes (COL3A1, ADAMTS12, BGN, FNDC1, AEBP1, and HTRA3).

Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

TRIzol reagent (Beyotime Biotechnology, Shanghai, China) was used to extract total RNA from the tissues following the manufacturer's instructions. Then, a ReverTra Ace- α - kit (ToYoBo) was used to reverse transcribe RNA into cDNA. Reverse transcription was performed at 37°C for 15 min and 98°C for 5 min. SYBR Green I Master Mix (ToYoBo) was utilized to conduct qPCR with the following steps: 95°C for 30 s, 98°C for 5 s, 58°C for 5 s, 72°C for 15 s with 40 cycles, 72°C for 10 s, followed by holding at 4°C. The $2^{-\Delta\Delta C_t}$ method was used to calculate the expression of COL3A1 relative to that of β -actin. The primer sequences were shown in Supplementary Table 1.

Western blotting (WB)

The protein expression of COL3A1 was analyzed by WB. The tissues were lysed with RAPI buffer (Beyotime Biotechnology, Shanghai, China) and centrifuged for 15 min at 12000 rpm (OHAUS, New Jersey, USA). The protein concentration was measured by a BCA protein assay kit. Then, tissue proteins were separated by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Protein electrophoresis was transferred to polyvinylidene fluoride membranes. The membrane was sealed with 5% skim milk for 60 min. After incubation overnight with the primary antibody, the fluorescent goat anti-rabbit secondary antibody was incubated with the membrane for 60 min. A Chemidoc-it2 imager (UVP) was used to detect the band intensity.

Results

Screening of DEGs between STAD and normal stomach tissues

The DEGs were visualized with a volcano map. Finally, a total of 1867 DEGs were identified, including 642 downregulated and 1225 upregulated genes (Figure 1A). The top ten pathways in terms of cellular component (CC), biological process (BP), and molecular function (MF) from the GO analysis were visualized. The results showed that the most enriched CC, BP and MF terms were concentrated in the extracellular matrix, extracellular structure organization, extracellular matrix organization, axon development, and receptor ligand activity-related pathways. The extracellular matrix, extracellular structure organization, extracellular matrix organization, axon development, and receptor ligand activity have been widely reported to be correlated with the tumorigenesis and progression of cancer (Figure 1B) [7].

PPI network construction

The top 74 and top 5 nodes were visualized by PPI analysis (Figure 1C). The top 5 nodes were COL3A1, FN1, COL1A2, POSTN and IL-6. Thus, these five genes might play key roles in the DEG protein interaction network (Figure 1D). In fact, the top 5 node genes are oncogenes that are known to play important roles in breast cancer, colorectal cancer, non-small-cell lung cancer, pancreatic cancer and GC [8-12].

WGCNA

Advanced GC refers to GC in which the tumor cells have infiltrated the submucosa and the muscle layer or have passed through the muscle layer to the serosa, regardless of the presence of lymph node metastasis. Due to the poor prognosis of advanced GC, it is of great necessity to identify the protein coding genes (PCGs) related to the pathological stage of GC. To determine the PCGs associated with the pathological stage of GC, WGCNA was performed. Average linkage clustering and Pearson correlation analysis were used to cluster the 375 STAD samples (Figure 2A). Next, network topology analysis was conducted with various soft-threshold powers. The WGCNA was set to have relatively balanced scale independence and average connectivity. In this research, the power of $\beta=4$ (R^2 without scale) was selected as the soft threshold parameter to ensure that the network was scale-free (Figure 2A). After merging modules with a difference of less than 25%, nine different PCG modules were identified (Figure 2B). Then, correlation analysis of the modular characteristic genes and T stage of each PCG module was assessed. Subsequently, black module ($Cor=0.89$, $P<0.001$) was identified as a PCG module that was highly correlated with T stage (Figure 2C). The ClueGO plug-in of Cytoscape was used to perform enrichment analysis of the black module. The results showed that it was enriched in DEGs related to the extracellular matrix, the complex of collagen trimers, the extracellular space, etc (Figure 2D).

Prognostic model and nomogram construction

Genes from the black module identified by WGCNA were used to construct the prognostic model. Then, Cox regression method was used to establish a prognostic model based on six genes to predict the OS of STAD patients (Figure 3A). The receiver operating characteristic (ROC) curve indicated that our model had great predictive efficiency (1-year survival $AUC=0.76$; 2-year survival $AUC=0.7$; 3-year survival $AUC=0.66$) (Figure 3B). Kaplan-Meier survival curves showed that the prognosis of patients with high expression of COL3A1, ADAMTS12, BGN, FNDC1, AEBP1 and HTRA3 was poorer than that of patients with low expression of these genes (Figure 3C). A nomogram was established by combining the prognostic model with clinical parameters such as age, sex, and T stage to better predict the 1-, 2-, and 3-year survival rates of STAD patients (Figure 3D). The decision curve indicated that the survival rates predicted by the nomogram were highly consistent with the actual OS values (Figure 3E).

GSEA and GSVA

Correlation analysis showed that the expression levels of the 6 model genes were highly correlated with a high T stage (Figure 4). GSEA showed that in samples with high expression of ADAMTS12, AEBP1 and FNDC1, the mitogen-activated protein kinase (MAPK) signaling was overactivated. In samples with high expression of BGN, COL3A1 and HTRA3, the MAPK and PI3K/AKT pathways were overactivated. GSVA

showed that collagen fiber organization and extracellular matrix structural constituents conferring tensile strength were activated with the increase in ADAMTS12, AEBP1, BGN, FNDC1, COL3A1 and HTRA3 expression.

Mutation analysis and immune infiltration analysis

The mutation analysis showed that in high-risk group, the mutation frequencies of SYNE1, AHNAK2, PCDH15, RYR2 and DNAH5 were significantly increased and the non-missense mutation rates of MUC16, TP53, and CSMD3 were lower than those in low-risk group (Figure 5A-B). The mutation analysis also showed that in the low-risk group, the MUC16, TP53, OBSCN, CSMD3, SYNE1, MDN1, PLEC, and NEB mutation rates were significantly increased. CIBERSORT was used to quantify the expression of markers of 21 immune cells in STAD (Figure 6A). The abundance of M0 macrophages, M1 macrophages, activated dendritic cells, activated mast cells, and neutrophils was different between the high- and low-risk groups (Figure 6B). The heat map showed the distribution of 21 types of immune cells in STAD tissues grouped according to T stage (Figure 6C). Under the background of STAD, the expression correlation of various immune cells was shown in the heat map (Figure 6D). The correlation between the 6 model genes and the infiltration of specific immune cell types was shown in Figure 6E-I. The tumor microenvironment (TME) has important clinicopathological significance in predicting outcome and treatment efficacy. For example, according to the "macrophage balance hypothesis", tumor-associated macrophages (TAMs) have dual effects: killing tumors and promoting tumor growth. Low purity and B cell infiltration were positively correlated with the occurrence of tumors. A decreased ratio of CD4+/CD8+ T cells and an increase in CD8+ T cells, neutrophils, macrophages and dendritic cells were both positively correlated with malignant tumors.

Expression of the 6 model genes in tumor and paired nontumor tissues

The box diagram showed that the 6 model genes were significantly expressed higher in cancer tissues than in normal tissues (Figure 7A-F). According to the Venn diagram, COL3A1 was identified among both the top 5 nodes of the PPI network and the 6 model genes (Figure 7G).

The expression level of COL3A1 in STAD

Since COL3A1 was the only screened gene that was common to the top 5 nodes of the PPI network and the 6 model genes, it might have a greater potential to play a crucial role in STAD than other model genes. To verify the expression of COL3A1 in STAD tissues, RT-qPCR and WB were performed. The results showed that COL3A1 expression in STAD tissues was notably upregulated compared to that in normal tissues (Figure 8A-B).

Discussion

Significant attention has been given to oncogenes in recent years, and the important role of oncogenes in STAD progression has been demonstrated by an increasing number of studies. For example, previous

research results showed that integrin 1 (ITGB1) expression was promoted by human telomerase reverse transcriptase (hTERT) [13]. The ubiquitin-mediated degradation of forkhead box O3 (FOXO3) promoted by hTERT was shown to lead to the metastasis of STAD [13]. Wang and his team revealed that P300 mediated the acetylation of methyltransferase like 3 (METTL3) promoter H3K27 and stimulated the m6A modification of heparin binding growth factor (HDGF) [14]. However, insulin-like growth factor 2 mRNA binding protein 3 (IGF2BP3) was found to recognize and combine with HDGF, enhancing the stability of HDGF [14]. This progression promotes glycolysis in STAD and stimulates the growth and liver metastasis of STAD [14]. Therefore, DEGs were initially screened, and the most valuable top 5 node genes were further mined through the PPI network; these genes were COL3A1, FN1, COL1A2, POSTN and IL-6. The upregulation of COL3A1 is positively related to the methylation of METTL3 in breast cancer [15]. FN1 is related to TNM staging, lymph node metastasis and OS and promotes the metastasis of GC cell lines [12]. IL-6-mediated signal transducer and activator of transcription 3 (STAT3) activation can promote the tumorigenesis of non-small-cell lung cancer through the activating NF- κ B pathway [16]. The overexpression of POSTN in tumors is an indicator of poor prognosis in colorectal cancer [8]. Furthermore, COL1A2 could promote the proliferation and migration of pancreatic cancer cells [9].

The OS of STAD patients was significantly related to pathological stage. Therefore, WGCNA and prognostic models were used to screen and verify prognostic genes related to STAD pathological staging, including ADAMTS12, AEBP1, BGN, COL3A1, FNDC1 and HTRA3.

To explore the potential pathways involving the six genes, GSEA and GSVA were performed. The GSEA results showed that these six genes were mainly enriched in the MAPK and PI3K/AKT pathways. Of course, many pathways could promote the tumorigenesis, progression and metastasis of STAD. For example, secreting-type ECM Spondin 2 (SPON2) is highly expressed in metastatic STAD, and by promoting epithelial-mesenchymal transition (EMT) in STAD cells, it promotes the activation of the extracellular signal-regulated protein kinase1/2 (ERK1/2) pathway and other signaling pathways, such as the MAPK cascade, which contributes to the recurrence of STAD [17]. MiR-135b can regulate ste20-like kinase 1 (MST1), thereby reducing the expression of p-p38 MAPK, p-ERK1/2, P-glycoprotein, p38 MAPK, ERK1/2, MDR1, MRP1, LRP and Bcl-2 and inhibiting the resistance of STAD cells to cisplatin [18]. In general, some lncRNAs involved in the STAD process can act as miRNA sponges. Some lncRNAs related to STAD are also responsible for the excessive proliferation and metastasis of STAD mediated by the MAPK/ERK pathway. The first study that revealed the role of lncRNAs in STAD focused on antisense H19 [19]. As a noncoding transcript, H19 can promote miR-675 to inhibit the apoptotic process of STAD cells [20]. In addition, miR-675 plays a role in colon cancer progression and regulates migration and invasion through the MAPK signaling pathway [21]. The sex-determining region Y-box 9 (SOX9)/miR-203a axis could promote the tumorigenesis of esophageal cancer by activating the PI3K/AKT signaling pathway [22]. According to the above evidence, the effects of the upregulation of ADAMTS12, AEBP1, BGN, FNDC1, COL3A1 and HTRA3 through the MAPK or PI3K/AKT signaling pathway on the progression of STAD should be explored in the future. The GSVA results showed that collagen fiber organization and extracellular matrix structural constituents conferring tensile strength were activated with the upregulation of ADAMTS12, AEBP1, BGN, FNDC1, COL3A1 and HTRA3. Collagen fiber organization and

extracellular matrix structural constituents conferring tensile strength could promote cancer metastasis and implantation [7, 23]. Therefore, ADAMTS12, AEBP1, BGN, FNDC1, COL3A1 and HTRA3 have the potential to affect the OS of STAD patients by promoting vascular infiltration, lymph node metastasis and distant metastasis of STAD cells.

There is evidence that oncogenes might promote tumor progression by affecting the tumor microenvironment. According to our results, six pathological stage-related genes were considered to be related to purity and the infiltration of B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages and dendritic cells. Immunotherapy plays a crucial role in the treatment of malignant tumors. Inhibitors against CTLA4, PD-1 and PD-L1 have achieved encouraging results in malignant melanoma and non-small-cell lung cancer [24, 25]. However, CTLA4 inhibitors have poor clinical effects in STAD, while PD-1 and PD-L1 m6A inhibitors achieve positive responses in some patients [26-28]. DCLK1 is a marker of clustered cells in the gastrointestinal tract that drives the type II immune response to inflammatory injury and has been reported to play an important role in the tumorigenesis and progression of various cancers [29, 30]. Interestingly, MAPK and PI3K signalings are thought to be related to tumor immunity. According to the results of previous research, the epidermal growth factor (EGF)-induced MAPK signaling is significantly involved in the expression of PD-L1, and these findings provide basic guidance for exploring the combination of MAPK pathway inhibitors and immunotherapy for hepatocellular carcinoma [31]. The PI3K and AKT pathways also play crucial roles in tumor immunity. T cell receptor signaling regulates forkhead box P3 (Foxp3) expression via PI3K, Akt, and mammalian Target of Rapamycin (mTOR) pathways [32]. In addition, immune active cells interact with humoral immune factors, such as Foxp3+ regulatory T cells and IDO, to form a complex tumor immune microenvironment, which ultimately affects tumor progression [33].

To identify the gene most significantly related to STAD, the top 5 node genes were overlapped with pathology-related genes. COL3A1 was found to be closely associated with the progression of STAD. COL3A1, also known as EDS4A, EDSVASC, and PMGEDSV, is one of the main types of fibrous collagen. Maintaining the structural integrity of the arteries, uterus, and intestine is the main function of COL3A1. COL3A1 is negatively regulated by miR-29a, and a regulatory relationship is thought to be associated with the resistance of ovarian cancer cell lines to topotekan [34]. As a new translocation partner gene of PLAG1, COL3A1 could be used to distinguish lipomas by affecting PLAG1 rearrangement and overexpression [11]. MiR-29a/b can promote cell migration and invasion by regulating COL3A1 expression and nasopharyngeal carcinoma progression [35]. Previous research has demonstrated that let-7d inhibits growth, metastasis and tumor macrophage infiltration by targeting COL3A1 and CCL7 in renal cell carcinoma [36]. Therefore, COL3A1 may affect tumor metastasis, invasion and the tumor microenvironment to promote the progression of STAD.

This study has several limitations that should be noted. First, the function of COL3A1 in STAD was not verified. Second, the study of COL3A1 in STAD was not thoroughly examined at the pathway level. Third, although we provided abundant evidence to prove the correlation between COL3A1 and immunity, more

evidence is needed to determine the specific effects of COL3A1 on the tumor immune microenvironment and immune cells.

Conclusion

In summary, COL3A1, ADAMTS12, BGN, FNDC1, AEBP1 and HTRA3 could play important roles in STAD and could affect the tumorigenesis and progression of STAD via many pathways. COL3A1, ADAMTS12, BGN, FNDC1, AEBP1 and HTRA3 act as oncogenes in most cancers and may be promising biomarkers for diagnosis and treatment of cancers, especially STAD. In addition, the identification of COL3A1 as a candidate biomarker provides a direction for further research on the role of tumor immunity in STAD.

Abbreviations

ADAMTS12, ADAM metalloproteinase with thrombospondin type 1 motif 12; AEBP1, AE binding protein 1; AKT, protein kinase B; BP, biological process; CC, cellular component; COL3A1, type III collagen α 1 chain; DEGs, differentially expressed genes; EGF, epidermal growth factor; FNDC1, fibronectin domain containing 1; Foxp3, forkhead box P3; GC, gastric cancer; GO, gene ontology; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; HTRA3, HtrA serine peptidase 3; MAPK, mitogen-activated protein kinase; MCC, maximal clique centrality; MF, molecular function; MST1, ste20-like kinase 1; mTOR, mammalian Target of Rapamycin; PD-L1, programmed death ligand-1; PI3K, phosphatidylinositol 3-kinase; PPI, protein-protein interaction; RT-qPCR, reverse transcription quantitative polymerase chain reaction; SNP, single nucleotide polymorphism; SOX9, sex-determining region Y-box 9; STAD, stomach adenocarcinoma; TCGA, The Cancer Genome Atlas; WB, western blotting; WGCNA, weighted gene coexpression network analysis.

Declarations

Ethics approval and consent to participate

This study was approved by the ethics committee of The Second Hospital of Jilin University and was conducted in accordance with the principles of Helsinki Declaration. All datasets were obtained from published literature and informed consent was obtained from the patients.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and analysed during the current study are available in TCGA database (<https://www.cancer.gov/tcga>) and can be acquired from the corresponding author on reasonable request.

Authors' contributions

ZZG and XDW designed the study. ZZG and XHC collected the data and performed the data analyses. ZZG obtained and validated the results and produced the initial draft of the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgments

None

Competing interests

All authors declare there is no competing interests in this work.

Funding

None.

Authors' Information

Affiliations

Zhengzhong Gu, Xudong Wang. Department of Gastrointestinal Nutrition and Hernia Surgery, The Second Hospital of Jilin University, Changchun, 130041, Jilin, P.R. China;

XiaoHan Cui. Department of General Surgery, the Affiliated Changzhou No. 2 People's Hospital of Nanjing Medical University, Changzhou, Jiangsu 213000, P.R. China.

Corresponding author:

Correspondence to Xudong Wang

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. **Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.** *Int J Cancer* 2015, **136**(5):E359-386.
2. Cancer Genome Atlas Research N. **Comprehensive molecular characterization of gastric adenocarcinoma.** *Nature* 2014, **513**(7517):202-209.
3. Du C, Li DQ, Li N, Chen L, Li SS, Yang Y, Hou MX, Xie MJ, Zheng ZD. **DDX5 promotes gastric cancer cell proliferation in vitro and in vivo through mTOR signaling pathway.** *Sci Rep* 2017, **7**:42876.
4. Kato H, Ishikawa S. **Genomic pathobiology of gastric carcinoma.** *Pathol Int* 2017, **67**(2):63-71.
5. Pan JH, Zhou H, Cooper L, Huang JL, Zhu SB, Zhao XX, Ding H, Pan YL, Rong L. **LAYN Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Gastric and Colon Cancers.** *Front Immunol* 2019,

10:6.

6. Lin C, He H, Liu H, Li R, Chen Y, Qi Y, Jiang Q, Chen L, Zhang P, Zhang H *et al.* **Tumour-associated macrophages-derived CXCL8 determines immune evasion through autonomous PD-L1 expression in gastric cancer.** *Gut* 2019, **68**(10):1764-1773.
7. Theocharis AD, Skandalis SS, Gialeli C, Karamanos NK. **Extracellular matrix structure.** *Advanced drug delivery reviews* 2016, **97**:4-27.
8. Oh HJ, Bae JM, Wen XY, Cho NY, Kim JH, Kang GH. **Overexpression of POSTN in Tumor Stroma Is a Poor Prognostic Indicator of Colorectal Cancer.** *Journal of pathology and translational medicine* 2017, **51**(3):306-313.
9. Wu J, Liu J, Wei X, Yu Q, Niu X, Tang S, Song L. **A feature-based analysis identifies COL1A2 as a regulator in pancreatic cancer.** *Journal of enzyme inhibition and medicinal chemistry* 2019, **34**(1):420-428.
10. Xiang Z, Li J, Song S, Wang J, Cai W, Hu W, Ji J, Zhu Z, Zang L, Yan R *et al.* **A positive feedback between IDO1 metabolite and COL12A1 via MAPK pathway to promote gastric cancer metastasis.** *J Exp Clin Cancer Res* 2019, **38**(1):314.
11. Yoshida H, Miyachi M, Ouchi K, Kuwahara Y, Tsuchiya K, Iehara T, Konishi E, Yanagisawa A, Hosoi H. **Identification of COL3A1 and RAB2A as novel translocation partner genes of PLAG1 in lipoblastoma.** *Genes Chromosomes Cancer* 2014, **53**(7):606-611.
12. Zhou Y, Cao G, Cai H, Huang H, Zhu X. **The effect and clinical significance of FN1 expression on biological functions of gastric cancer cells.** *Cellular and molecular biology (Noisy-le-Grand, France)* 2020, **66**(5):191-198.
13. Hu C, Ni Z, Li BS, Yong X, Yang X, Zhang JW, Zhang D, Qin Y, Jie MM, Dong H *et al.* **hTERT promotes the invasion of gastric cancer cells by enhancing FOXO3a ubiquitination and subsequent ITGB1 upregulation.** *Gut* 2017, **66**(1):31-42.
14. Wang Q, Chen C, Ding Q, Zhao Y, Wang Z, Chen J, Jiang Z, Zhang Y, Xu G, Zhang J *et al.* **METTL3-mediated m(6)A modification of HDGF mRNA promotes gastric cancer progression and has prognostic significance.** *Gut* 2020, **69**(7):1193-1205.
15. Shi Y, Zheng C, Jin Y, Bao B, Wang D, Hou K, Feng J, Tang S, Qu X, Liu Y *et al.* **Reduced Expression of METTL3 Promotes Metastasis of Triple-Negative Breast Cancer by m6A Methylation-Mediated COL3A1 Up-Regulation.** *Frontiers in oncology* 2020, **10**:1126.
16. Huang X, Xiao S, Zhu X, Yu Y, Cao M, Zhang X, Li S, Zhu W, Wu F, Zheng X *et al.* **miR-196b-5p-mediated downregulation of FAS promotes NSCLC progression by activating IL6-STAT3 signaling.** *Cell death & disease* 2020, **11**(9):785.
17. Lu H, Feng Y, Hu Y, Guo Y, Liu Y, Mao Q, Xue W. **Spondin 2 promotes the proliferation, migration and invasion of gastric cancer cells.** *J Cell Mol Med* 2020, **24**(1):98-113.
18. Zhou J, Chen Q. **Poor expression of microRNA-135b results in the inhibition of cisplatin resistance and proliferation and induces the apoptosis of gastric cancer cells through MST1-mediated MAPK signaling pathway.** *FASEB J* 2019, **33**(3):3420-3436.

19. Yang F, Bi J, Xue X, Zheng L, Zhi K, Hua J, Fang G. **Up-regulated long non-coding RNA H19 contributes to proliferation of gastric cancer cells.** *FEBS J* 2012, **279**(17):3159-3165.
20. Yan J, Zhang Y, She Q, Li X, Peng L, Wang X, Liu S, Shen X, Zhang W, Dong Y *et al.* **Long Noncoding RNA H19/miR-675 Axis Promotes Gastric Cancer via FADD/Caspase 8/Caspase 3 Signaling Pathway.** *Cell Physiol Biochem* 2017, **42**(6):2364-2376.
21. Yang Q, Wang X, Tang C, Chen X, He J. **H19 promotes the migration and invasion of colon cancer by sponging miR-138 to upregulate the expression of HMGA1.** *Int J Oncol* 2017, **50**(5):1801-1809.
22. Wang L, Zhang Z, Yu X, Li Q, Wang Q, Chang A, Huang X, Han X, Song Y, Hu J *et al.* **SOX9/miR-203a axis drives PI3K/AKT signaling to promote esophageal cancer progression.** *Cancer Lett* 2020, **468**:14-26.
23. Peng DH, Ungewiss C, Tong P, Byers LA, Wang J, Canales JR, Villalobos PA, Uraoka N, Mino B, Behrens C *et al.* **ZEB1 induces LOXL2-mediated collagen stabilization and deposition in the extracellular matrix to drive lung cancer invasion and metastasis.** *Oncogene* 2017, **36**(14):1925-1938.
24. Barbee MS, Ogunniyi A, Horvat TZ, Dang TO. **Current status and future directions of the immune checkpoint inhibitors ipilimumab, pembrolizumab, and nivolumab in oncology.** *Ann Pharmacother* 2015, **49**(8):907-937.
25. Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, Patnaik A, Aggarwal C, Gubens M, Horn L *et al.* **Pembrolizumab for the treatment of non-small-cell lung cancer.** *N Engl J Med* 2015, **372**(21):2018-2028.
26. Muro K, Chung HC, Shankaran V, Geva R, Catenacci D, Gupta S, Eder JP, Golan T, Le DT, Burtness B *et al.* **Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1b trial.** *Lancet Oncol* 2016, **17**(6):717-726.
27. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D *et al.* **PD-1 Blockade in Tumors with Mismatch-Repair Deficiency.** *N Engl J Med* 2015, **372**(26):2509-2520.
28. Overman MJ, McDermott R, Leach JL, Lonardi S, Lenz HJ, Morse MA, Desai J, Hill A, Axelson M, Moss RA *et al.* **Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study.** *Lancet Oncol* 2017, **18**(9):1182-1191.
29. Westphalen CB, Asfaha S, Hayakawa Y, Takemoto Y, Lukin DJ, Nuber AH, Brandtner A, Setlik W, Remotti H, Muley A *et al.* **Long-lived intestinal tuft cells serve as colon cancer-initiating cells.** *J Clin Invest* 2014, **124**(3):1283-1295.
30. Westphalen CB, Takemoto Y, Tanaka T, Macchini M, Jiang Z, Renz BW, Chen X, Ormanns S, Nagar K, Tailor Y *et al.* **Dcl1 Defines Quiescent Pancreatic Progenitors that Promote Injury-Induced Regeneration and Tumorigenesis.** *Cell Stem Cell* 2016, **18**(4):441-455.
31. Xing S, Chen S, Yang X, Huang W. **Role of MAPK activity in PD-L1 expression in hepatocellular carcinoma cells.** *J BUON* 2020, **25**(4):1875-1882.

32. Sauer S, Bruno L, Hertweck A, Finlay D, Leleu M, Spivakov M, Knight ZA, Cobb BS, Cantrell D, O'Connor E *et al.* **T cell receptor signaling controls Foxp3 expression via PI3K, Akt, and mTOR.** *Proc Natl Acad Sci U S A* 2008, **105**(22):7797-7802.
33. Li F, Sun Y, Huang J, Xu W, Liu J, Yuan Z. **CD4/CD8 + T cells, DC subsets, Foxp3, and IDO expression are predictive indicators of gastric cancer prognosis.** *Cancer Med* 2019, **8**(17):7330-7344.
34. Kazmierczak D, Jopek K, Sterzynska K, Ginter-Matuszewska B, Nowicki M, Rucinski M, Januchowski R. **The Significance of MicroRNAs Expression in Regulation of Extracellular Matrix and Other Drug Resistant Genes in Drug Resistant Ovarian Cancer Cell Lines.** *Int J Mol Sci* 2020, **21**(7).
35. Qiu F, Sun R, Deng N, Guo T, Cao Y, Yu Y, Wang X, Zou B, Zhang S, Jing T *et al.* **miR-29a/b enhances cell migration and invasion in nasopharyngeal carcinoma progression by regulating SPARC and COL3A1 gene expression.** *PLoS One* 2015, **10**(3):e0120969.
36. Su B, Zhao W, Shi B, Zhang Z, Yu X, Xie F, Guo Z, Zhang X, Liu J, Shen Q *et al.* **Let-7d suppresses growth, metastasis, and tumor macrophage infiltration in renal cell carcinoma by targeting COL3A1 and CCL7.** *Mol Cancer* 2014, **13**:206.

Figures

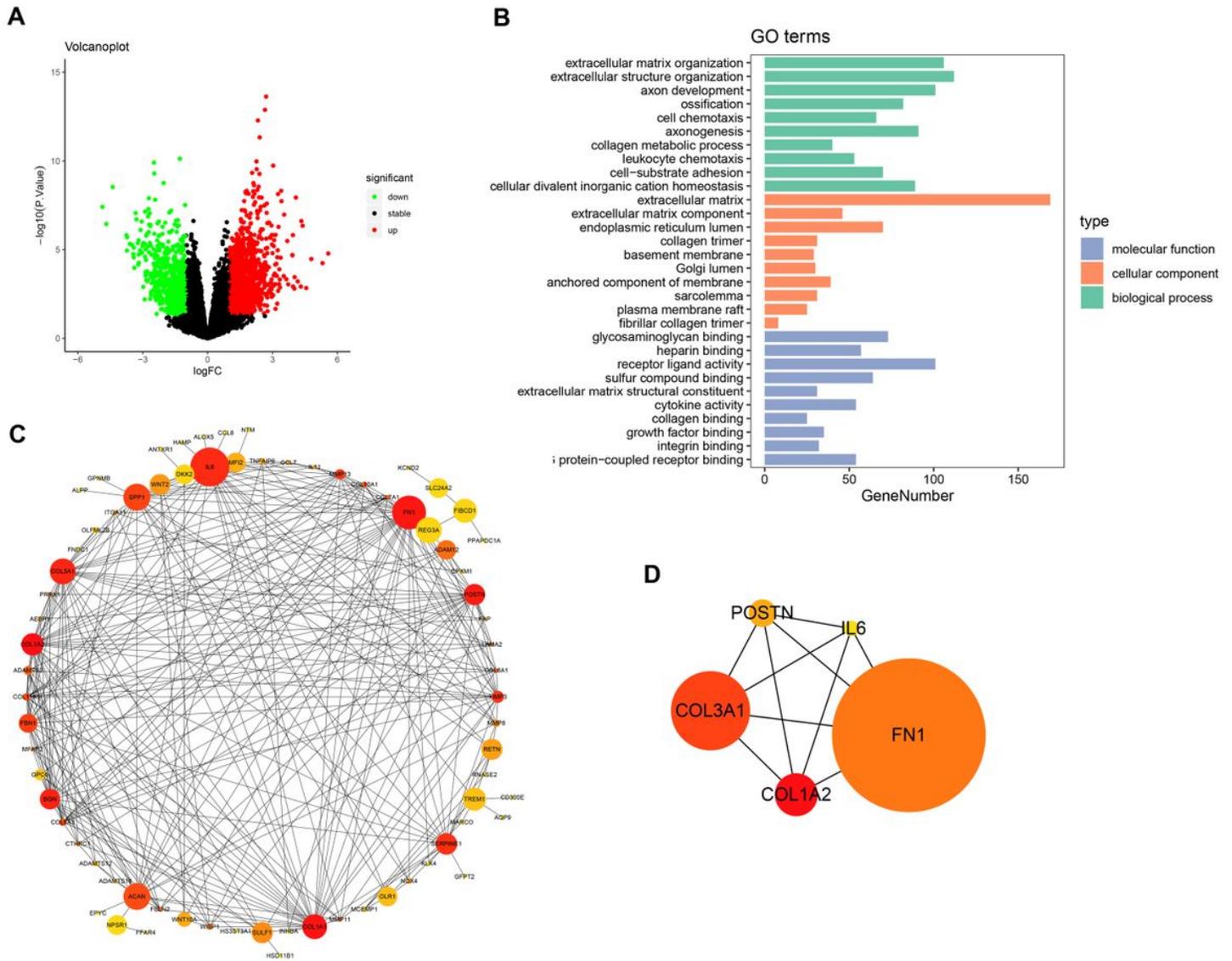


Figure 1

Screening of DEGs and enriched pathways. A. The volcano map showed the DEGs. B. The GO analysis results of DEGs, including MF, CC and BP categories. C. Top 74 nodes in the PPI network analysis. D. Top 5 nodes in the PPI network analysis.

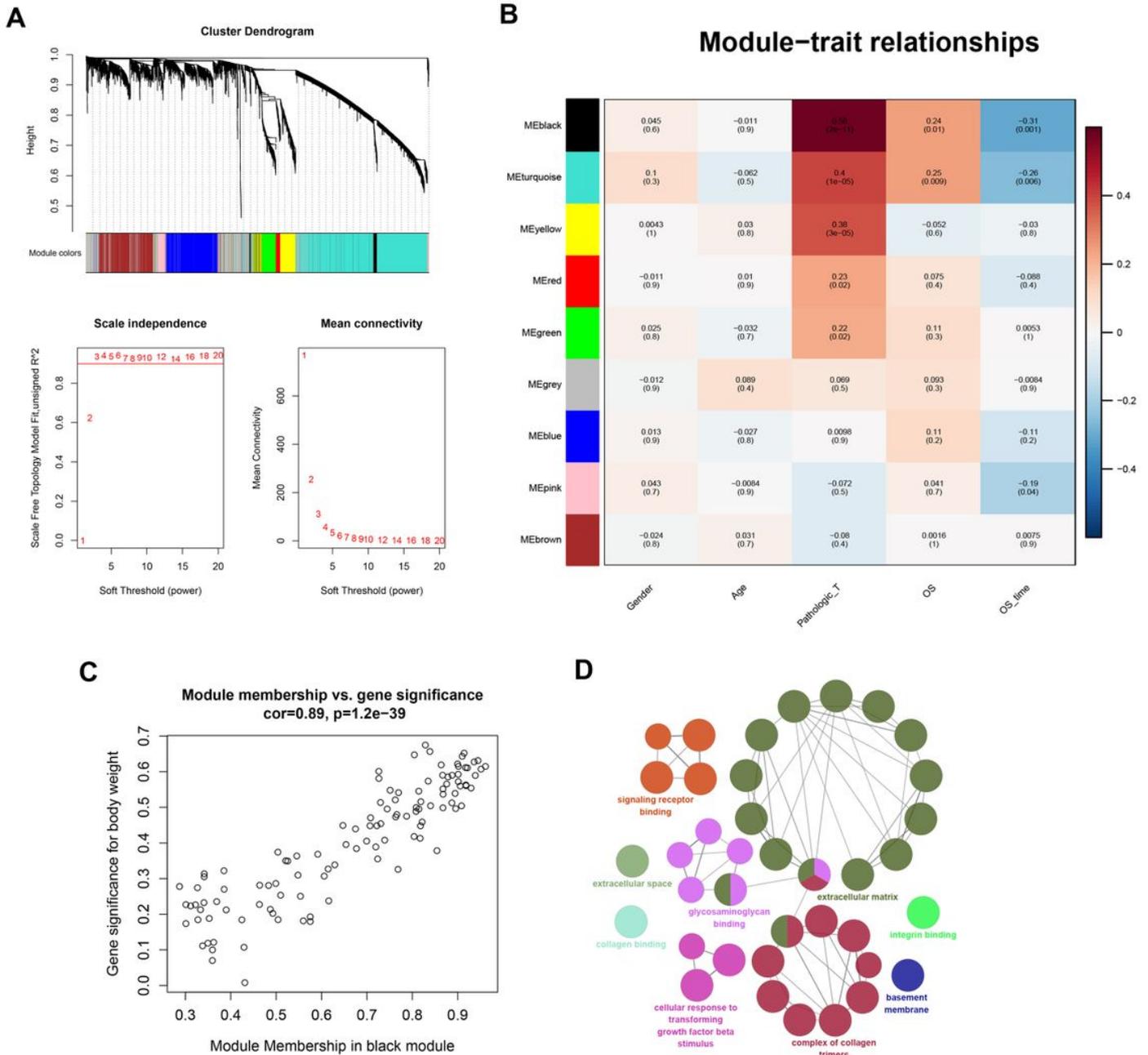


Figure 2

PCGs related to the T stage of STAD. A. Cluster analysis and network topology analysis of tumor samples from 375 patients with STAD. B. Nine different PCG modules. C. The black module was highly correlated with T stage (Cor=0.89, P<0.001). D. The enrichment results of the black modules were concentrated on the extracellular matrix, complex of collagen trimers, and extracellular space.

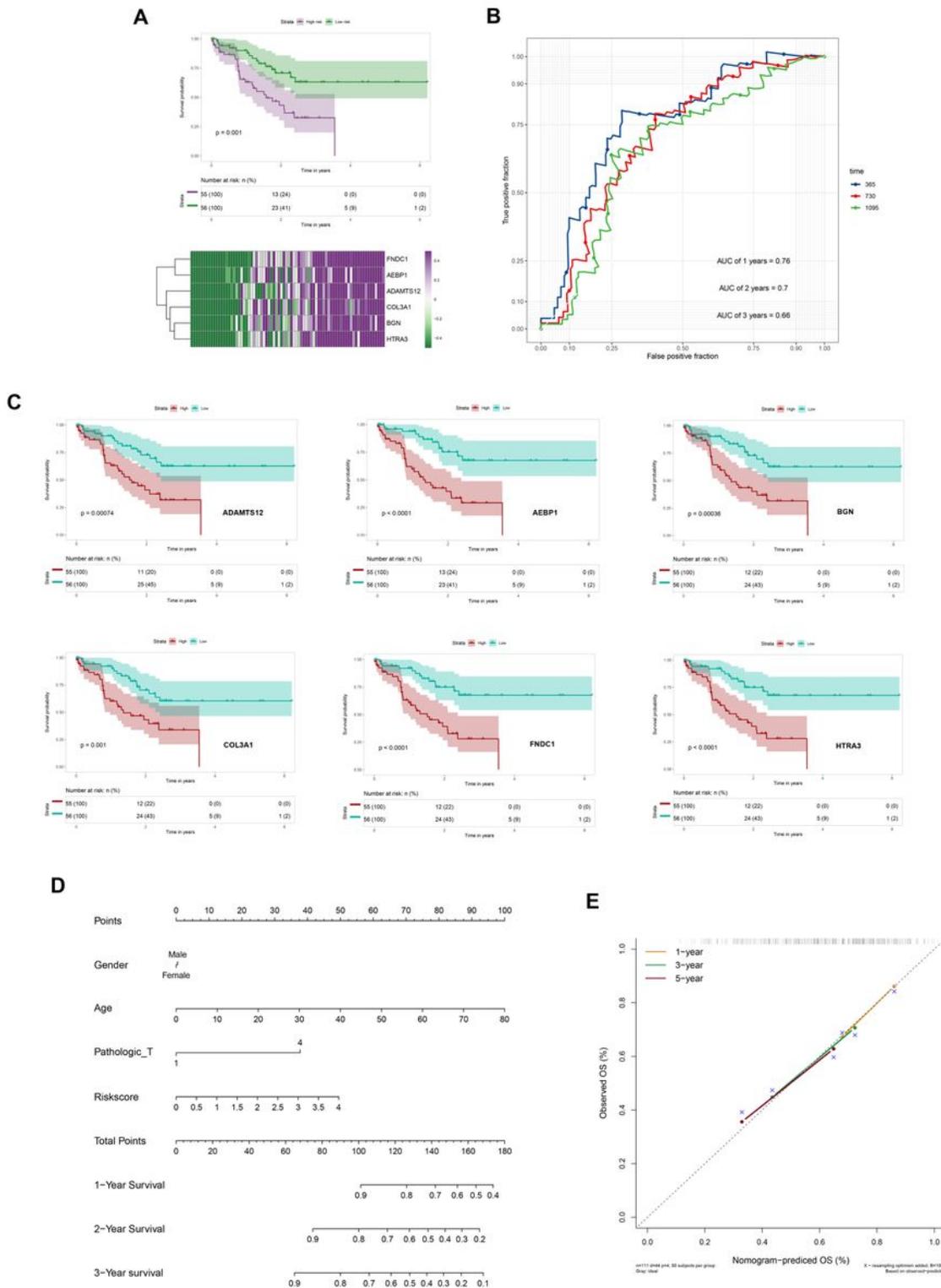


Figure 3

PCGs related to the T stage of STAD. A. Cluster analysis and network topology analysis of tumor samples from 375 patients with STAD. B. Nine different PCG modules. C. The black module was highly correlated with T stage ($Cor=0.89$, $P<0.001$). D. The enrichment results of the black modules were concentrated on the extracellular matrix, complex of collagen trimers, and extracellular space.

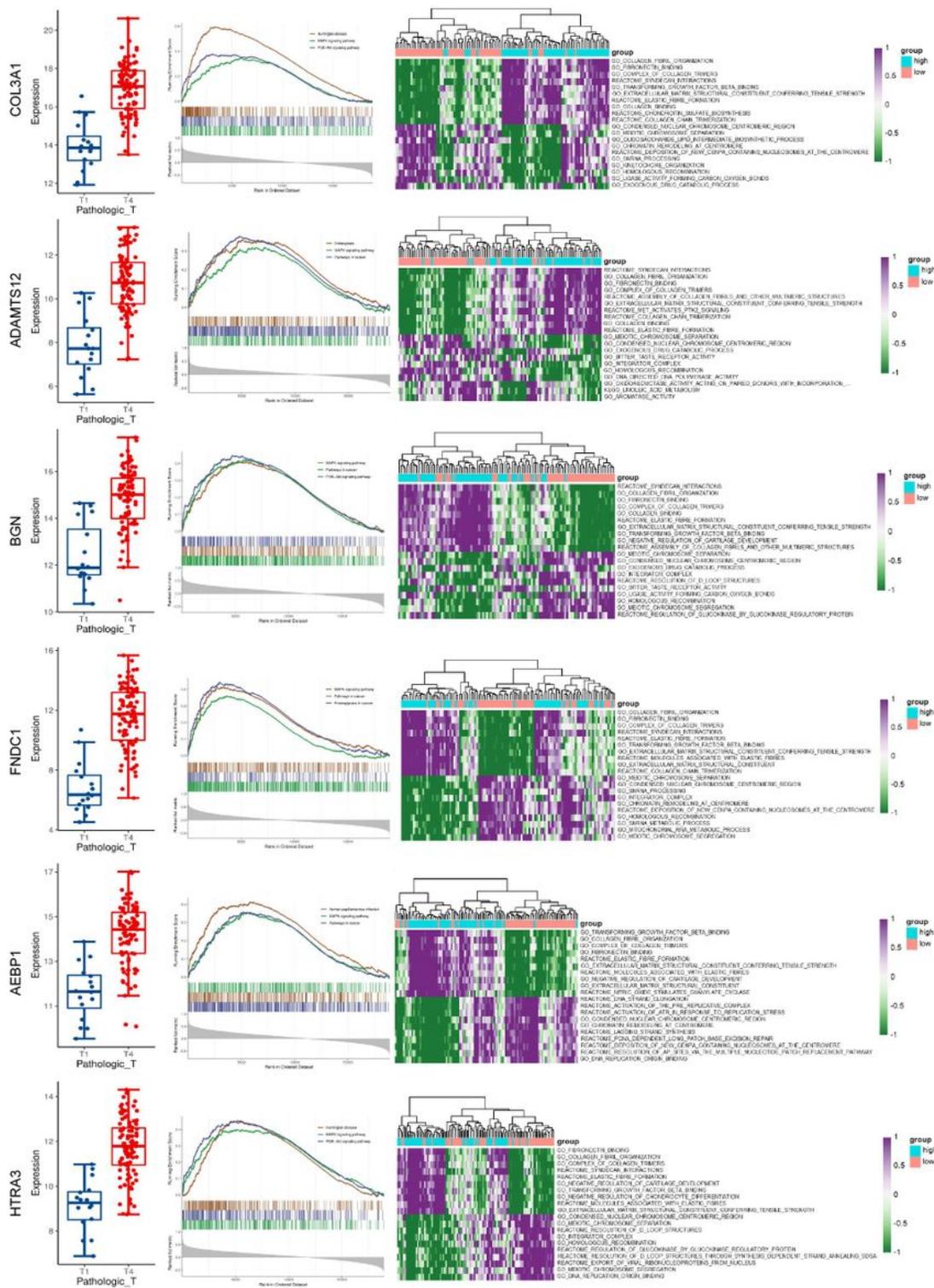


Figure 4

A prognostic model constructed based on WGCNA black module genes. A. A prognostic model was constructed based on six genes to predict the OS of STAD patients. B. The ROC curve showed that the prognostic model has good prediction performance. C. Kaplan-Meier survival curves of patients grouped according to COL3A1, ADAMTS12, BGN, FNDC1, AEBP1 and HTRA3 expression. D. The nomogram was

used to predict the 1-, 2-, and 3-year survival rates of STAD patients. E. The decision curve showed that the survival rates predicted by the nomogram was highly consistent with the actual survival rates.

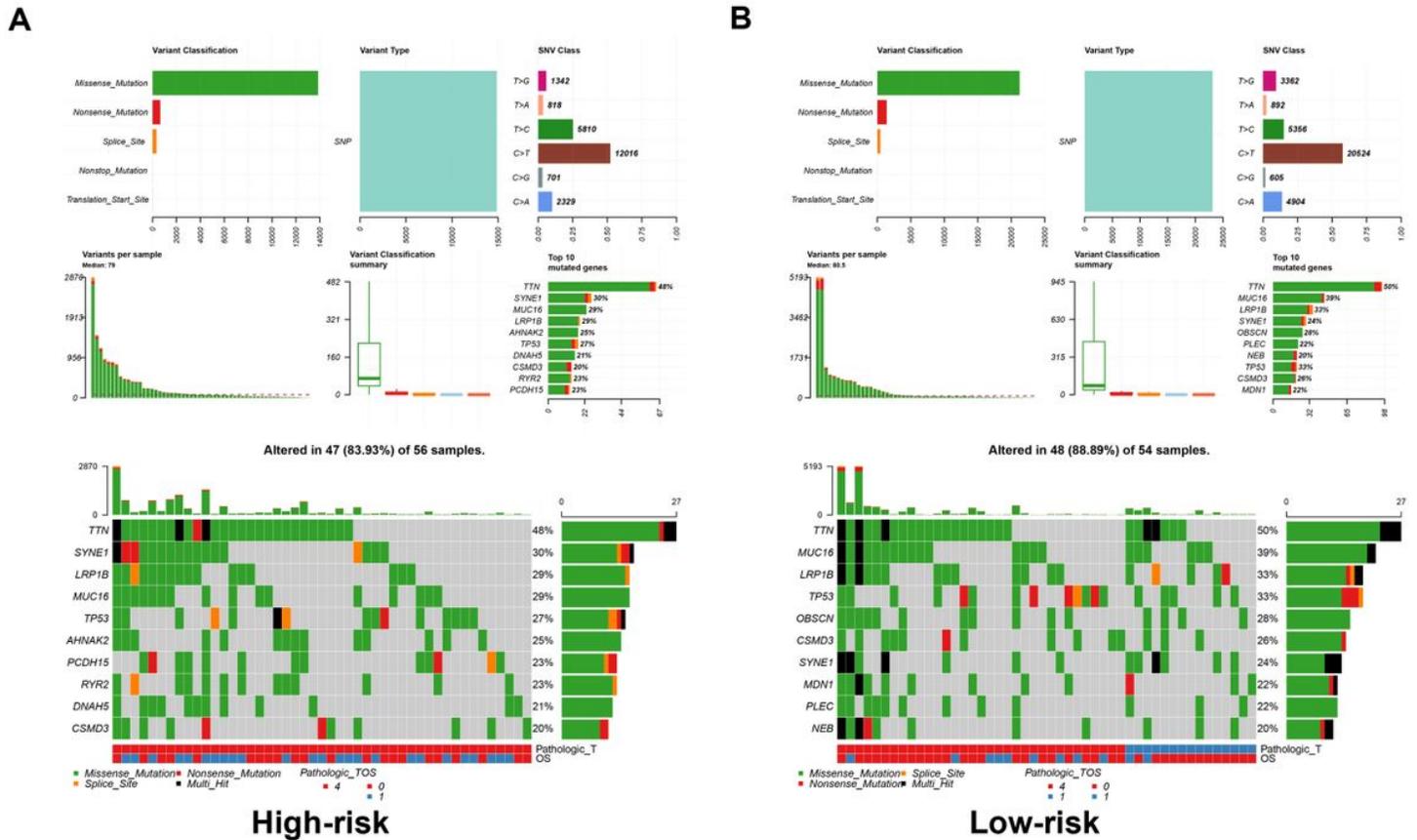


Figure 5

The expression of 6 model genes in STAD with high T stage and GSEA and GSVA results. A-F. COL3A1, ADAMTS12, BGN, FNDC1, AEBP1 and HTRA3 were highly expressed in STAD with high T stage and were analyzed by GSEA and GSVA. Figure 5. Mutation analysis in high-risk and low-risk groups. A. The frequency and type of mutations in high-risk group. B. The frequency and type of mutations in low-risk group.

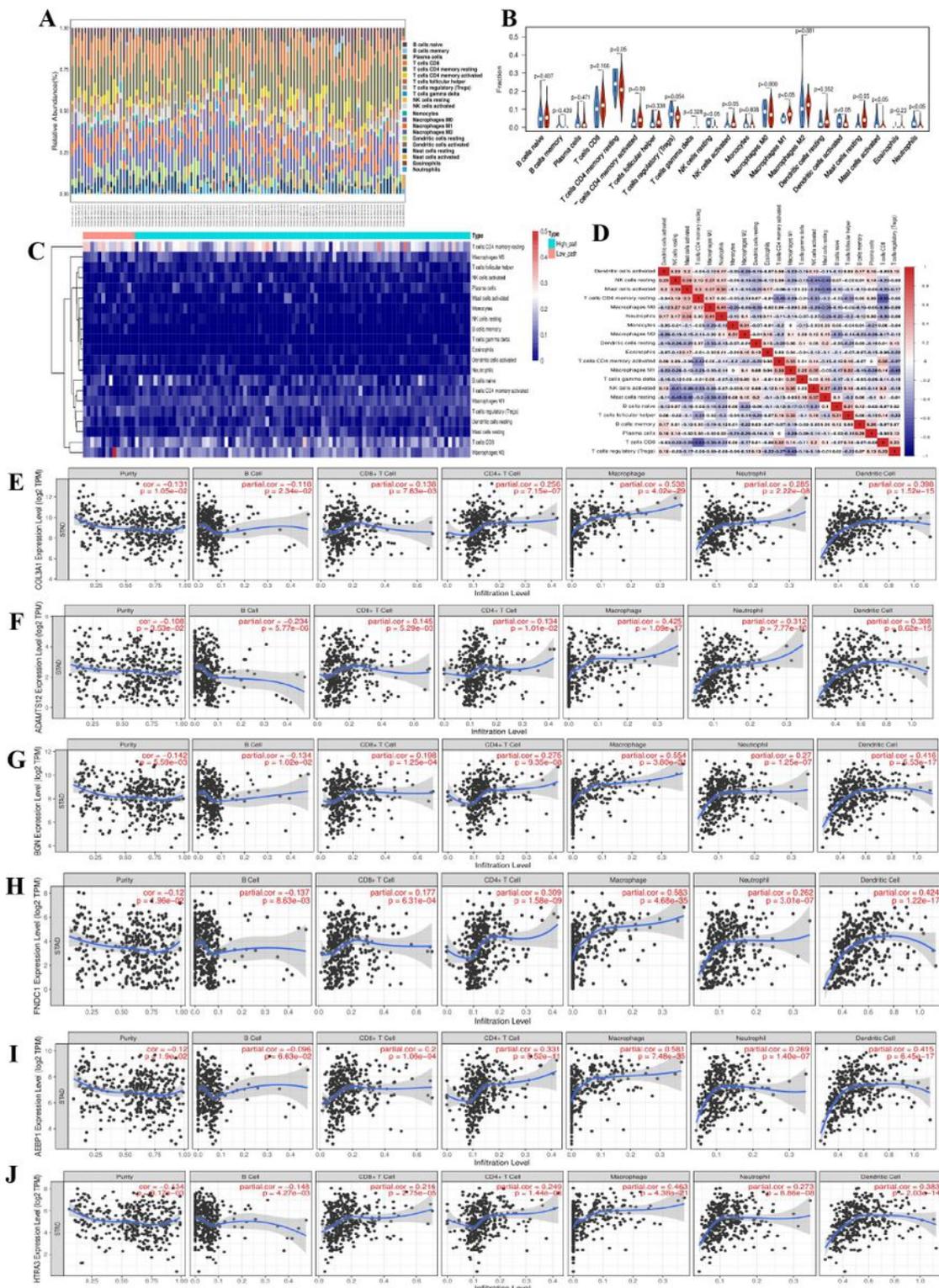


Figure 6

Correlation analysis of gene expression and immune infiltration in STAD. A. The abundance of 21 kinds of immune cells in STAD. B. Differences in immune cell abundance in high- and low-risk groups. C. Heat map of 21 types of immune cells in STAD samples. D. Correlations among different immune cell subtypes in STAD samples. E-I. The correlation of 6 model genes with specific immune cells.

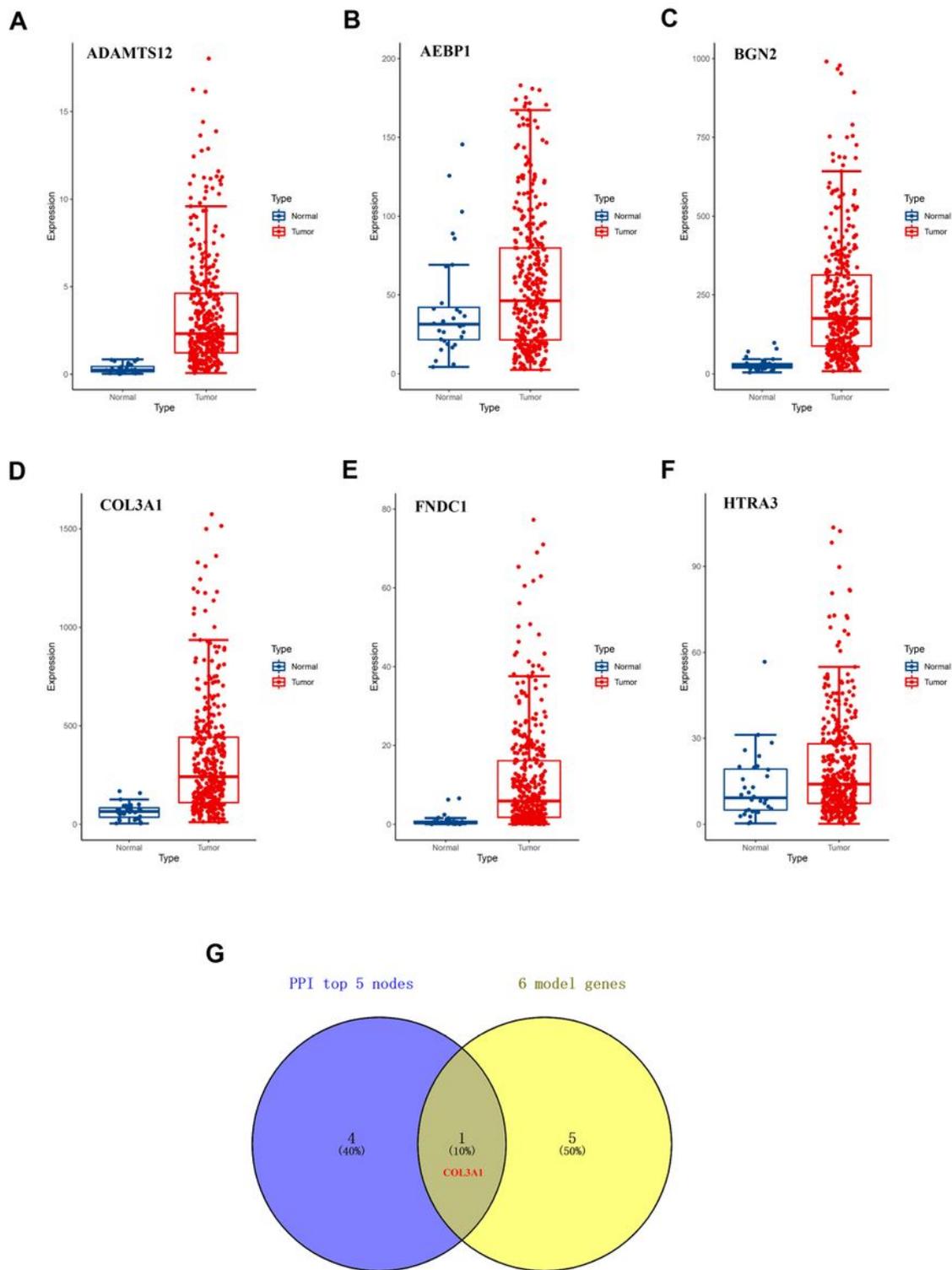


Figure 7

The expression level of 6 model genes in cancer tissues and normal tissues. A-F. The differential expression of COL3A1, ADAMTS12, BGN, FNDC1, AEBP1 and HTRA3 in STAD tissues and normal tissues. G. The Venn diagram showed that the top 5 nodes of the PPI network and the model genes only intersect with COL3A1.

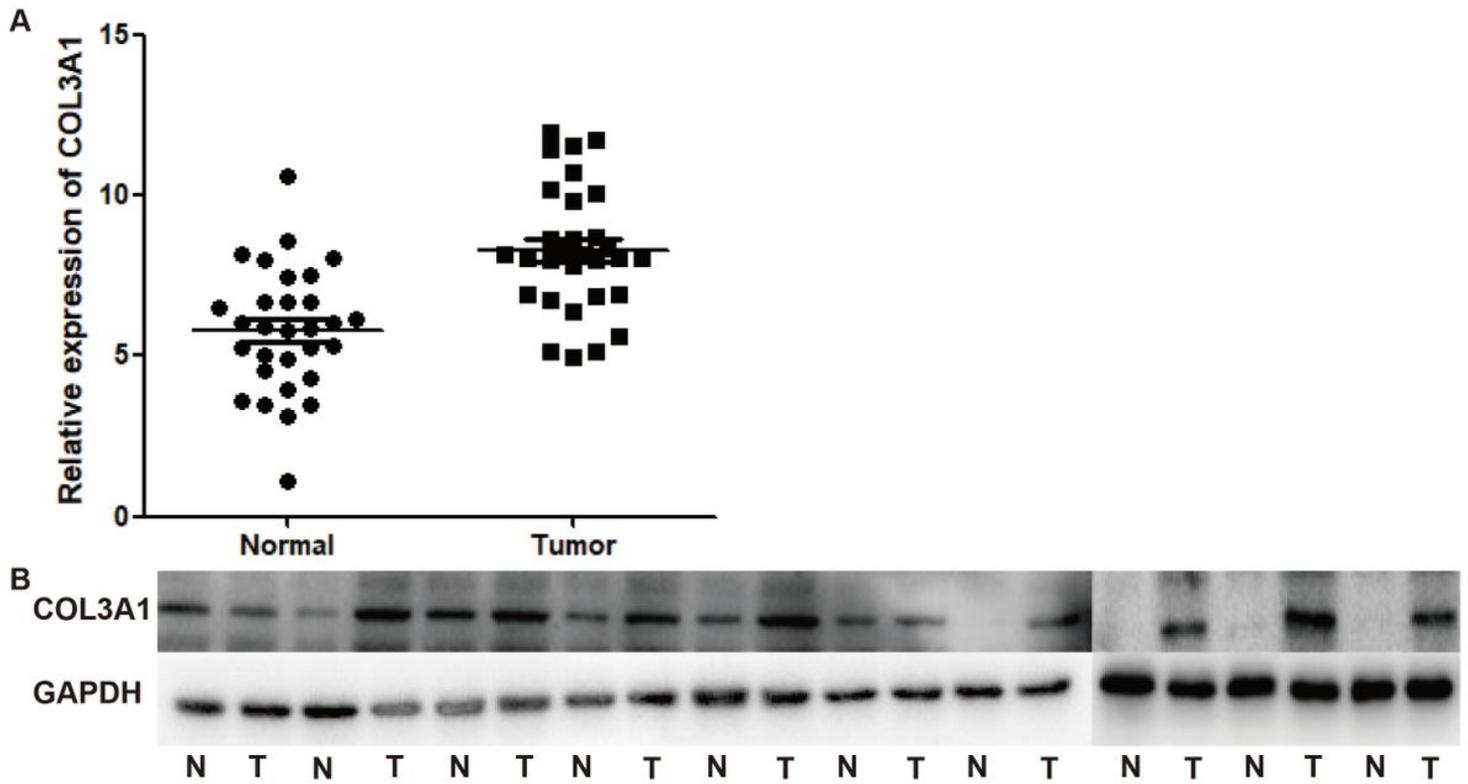


Figure 8

A. RT-qPCR showed the differential expression of COL3A1 mRNA between STAD and normal tissues. B. WB showed the differential expression of COL3A1 protein between STAD and normal tissues.

Supplementary Files

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

- [SupplementaryTable1.docx](#)