BMC Medical Genomics

RESEARCH

Open Access

COL5A2 is a prognostic-related biomarker and correlated with immune infiltrates in gastric cancer based on transcriptomics and single-cell RNA sequencing

Meiru Chen[1,2], Xinying Zhu[3], Lixian Zhang[2] and Dongqiang Zhao[1*]

Abstract

Background There is still a therapeutic challenge in treating gastric cancer (GC) due to its high incidence and poor prognosis. Collagen type V alpha 2 (COL5A2) is increased in various cancers, yet it remains unclear how it contributes to the prognosis and immunity of GC.

Methods The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) datasets were used to download transcriptome profiling (TCGA-STAD; GSE84437), single-cell RNA sequencing (scRNA-seq) data (GSE167297) and clinical information. COL5A2 expression and its relationship with clinicopathological factors were analyzed. We conducted survival analysis and Cox regression analysis to evaluate the prognosis and independent factors of GC. Co-expressed analysis was also performed. To identify the underlying mechanism, we conducted analyses of differentially expressed genes (DEGs) and functional enrichment. The correlations between COL5A2 expression and immune cell infiltration levels and immune infiltrate gene marker sets were further explored. Additionally, we analyzed the association of COL5A2 expression with immunological checkpoint molecules. Furthermore, the relationship between COL5A2 expression and immunotherapy sensitivity was also investigated.

Results COL5A2 expression was elevated in GC. More than this, the scRNA-seq analysis revealed that COL5A2 expression had a spatial gradient. The upregulated COL5A2 was associated with worse overall survival. A significant correlation was found between COL5A2 overexpression and age, T classification and clinical stage in GC. COL5A2 was found to be an independent factor for the unfortunate outcome in Cox regression analysis. The co-expressed genes of COL5A2 were associated with tumor stage or poor survival. Enrichment analysis revealed that the DEGs were mainly associated with extracellular matrix (ECM)-related processes, PI3K-AKT signaling pathway, and focal adhesion. GSEA analyses revealed that COL5A2 was associated with tumor progression-related pathways. Meanwhile, COL5A2 expression was correlated with tumor-infiltrating immune cells. Moreover, immunophenoscore (IPS) analysis and PRJEB25780 cohorts showed that patients with low COL5A2 expression were highly sensitive to immunotherapy.

Conclusions COL5A2 might act as a prognostic biomarker of GC prognosis and immune infiltration and may provide a therapeutic intervention strategy.

Keywords COL5A2, Gastric cancer (GC), Prognosis, Immune cell infiltration, Immunotherapy

*Correspondence:
Dongqiang Zhao
hbzdq1998@163.com
Full list of author information is available at the end of the article

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

GC is one of the most common malignancies that has a poor prognosis. The incidence and mortality are fifth and fourth globally [1]. According to the 2020 global statistics, the number of new cases of GC has exceeded a million, and about 769,000 patients died from GC [1]. Despite major advances in surgery, chemical therapy, radiotherapy, and neoadjuvant therapy, patients with GC (particularly those with advanced GC) do not have a good prognosis [2–5]. Recently, along with the discovery of monoclonal antibodies and small molecule inhibitors, targeted therapy interventions have become more accessible, and have been an important therapy for patients with GC [6]. Hence, a sensitive and novel biomarker that can accurately predict the prognosis and treatment for GC is an absolute necessity.

Collagen is the most plentiful protein in human tissue and is the basal component of the extracellular matrix (ECM) architecture [7]. Researchers have determined that the ECM contributes to the initiation, progression, and migration of cancer [7]. The collagen type V alpha 2 (COL5A2) belongs to the collagen type 5 family, locates at 2q32.2. A low abundance fibrillar collagen alpha chain is encoded by this gene. This gene's mutations were linked to the Ehlers-Danlos syndrome and a genetically complex autoimmune disease [8, 9]. COL5A2 shows abnormal expression in tumors, which affects malignancy and progression [10–12]. Nevertheless, the clinical significance of COL5A2 in GC and its molecular mechanism are not well understood.

COL5A2 was tested for its prognostic value in GC in this study. The TCGA and GEO databases were used to acquire RNA-seq and microarray data. We examined COL5A2 expression in human stomach adenocarcinoma (STAD) samples. In addition, the scRNA-seq analysis was also conducted to detect the expression of COL5A2. COL5A2 expression level and clinicopathological characteristics, along with prognosis, were comprehensively analyzed. We conducted differential expression analysis based on the expression of COL5A2, along with Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of these DEGs, as well as Gene Set Enrichment Analysis (GSEA) to fully study the biological functions of COL5A2 in GC. Moreover, the further investigation suggested that COL5A2 expression was associated with tumor-infiltrating immune cells, immune marker sets and immune checkpoint molecules, and may affect the immunotherapy response. These findings may provide an underlying mechanism by which COL5A2 contributes to GC progression and immune tolerance. Hence, for patients with GC, COL5A2 can be a promising biomarker and therapeutic target.

Materials and methods

Data extraction from TCGA and GEO database

Transcriptome profiling and clinical data of STAD patients were downloaded from the TCGA website (https://portal.gdc.cancer.gov). The Perl (http://www.perl.org/) script was used to merge the RNA-seq data files into a matrix file. Then 375 samples of GC samples and 32 adjacent nontumor samples were obtained. A Perl script was used to convert Ensemble IDs (http://ensembl.org/index.html) into gene symbols. We searched the GEO database (https://www.ncbi.nlm.nih.gov/geo/) with the keywords "gastric cancer survival" and "homo sapiens", and downloaded "GSE84437" which had the maximum dataset.

A set of raw scRNA-seq data, GSE167297, was downloaded from the GEO database for scRNA-seq analysis, which contained 5 deep layer samples, 5 superficial layer samples and 4 normal gastric samples from 5 patients. Seurat package (V4.1.1) in R was used to perform the quality control (QC) process. The raw count matrix was converted into a Seurat object. With the "PercentageFeatureSet" function, the percentage of mitochondrial genes, ribosomal genes and haemoglobin genes were calculated. We excluded cells expressing less than 300 genes, and genes expressed in fewer than three cells, as well as noncoding genes. Mitochondrial gene content > 15%, ribosomal gene content > 3% and haemoglobin gene content > 1% were also excluded. We normalized the filtered data using "LogNormalize" with 10,000 scales. Then, based on the top 2000 highly variable genes, an analysis of principal components (PCA) was conducted. A batch effect was corrected using the harmony R package. To identify cell types, Uniform Manifold Approximation and Projection (UMAP) and marker gene analyses were carried out [13]. The UMAP plots identified 17 cell clusters with a resolution of 0.8. We curated the cell clusters based on known lineage markers, such as T cell (CD3D, CD3E, PTPRC), B cell (MS4A1), myeloid cell (CD68, CD163, CD14, CD86, TPSAB1), plasma (MZB1, SDC1, CD79A), epithelial cell (EPCAM, KRT19, PROM1, ALDH1A1), endothelial cell (VWF) and fibroblasts (FGF7, ACTA2). The expression of COL5A2 was visualized with ggplot2 and clustree package in R. A chi-square test with rcompanion package in R was used to determine whether statistically significant differences existed.

COL5A2 differential expression in patients with GC

Based on TCGA expression profile data, 33 cancers were evaluated for their COL5A2 mRNA levels via Tumor Immune Estimation Resource (TIMER) 2.0 (http://timer.cistrome.org/) [14]. We analyzed the expression of COL5A2 in normal and GC samples

along with 27 pairs of GC and adjacent samples were conducted via limma and ggplot2 packages in R software.

Correlation analysis of COL5A2 and clinicopathological factors

To investigate whether the mRNA expression of COL5A2 correlates with the clinicopathological factors of GC patients, we explored clinical information of 375 GC samples extracted from TCGA and verified the results with 433 GC samples using GSE84437 via limma package in R software.

Immunohistochemistry (IHC)

A total of 30 GC and 10 normal gastric tissues as controls were selected from The Second Hospital of Hebei Medical University, Hebei Province, China, between June 2022 and June 2023. Ethics approval for the study was obtained from the institution. Pathological examinations confirmed the diagnosis of gastric cancer in all patients. No prior treatment was given to any of the patients before undergoing surgical resection. Samples from patients were prepared as formalin-fixed paraffin-embedded (FFPE) sections and the IHC staining against COL5A2 was performed as previously described [15]. Anti-COL5A2 (1:50, OriGene, #TA313657) was then incubated overnight at 4°C with the samples. After incubation with the secondary antibodies (ZSGB-BIO, PV-9000), the sections were stained with DAB (ZSGB-BIO, ZLI-9017) the following day. We evaluated the staining results as described previously [16]. There were four scores for staining intensity in each field: 0 was no staining; 1 was weakly positive; 2 was moderately positive; 3 was strongly positive. In addition, positive cells were divided into five grades (percentage scores): 0–10% (0), 11–25% (1), 26–50% (2), 51–75% (3), and > 75% (4). An overall staining score was calculated by multiplying the intensity score by the percentage score of protein staining. High COL5A2 expression was defined as a final score of ≤ 4, and low COL5A2 expression was defined as a final score of > 5. The immunostaining was evaluated by two independent pathologists who were blind to the study's findings.

Survival and significant prognostic marker analyses of COL5A2

The samples were classified as high-expression and low-expression groups according to COL5A2 expression levels above or below the median value. In this study, based on the TCGA datasets and GSE84437, we conducted survival analyses to evaluate the impact of COL5A2 expression on GC survival. Univariate and multivariate Cox regression analyses were performed to determine independent prognostic factors. The prognostic characteristics of COL5A2 and clinical characteristics were assessed by calculating the area under the receiver operating characteristic (ROC) curve (AUC) for GC patients in TCGA-STAD. The statistical analyses were conducted using the R software. The limma, survival, survminer and timeROC packages in R software were used appropriately to conduct the statistical analyses. A P-value < 0.05 was considered significant.

Analysis of COL5A2 co-expressed genes in GC

For further exploration of the COL5A2-associated molecular mechanism, COL5A2 co-expressed genes were identified with the cBioPortal database (https://www.cbioportal.org/). The top six significant COL5A2 co-expressed genes were used for further analysis, and the TIMER2.0 database was used to verify their correlation with COL5A2 expression. In addition, by using Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/index.html), we explored COL5A2 co-expressed genes in GC and the relationship between these genes and tumor stages. Moreover, we analyzed the prognosis significance of the six co-expression genes via Kaplan–Meier plotter online database (http://kmplot.com/analysis/).

Analysis of differentially expressed genes and functional enrichment

We conducted differential expression analysis with the limma R package. Statistical significance was determined based on adjusted P-value < 0.05 and logFC > 1 [17]. To define the biological functions of DEGs and COL5A2-related genes, GO and KEGG [18] enrichment analyses were conducted via clusterProfiler R package [19]. The package was also used to perform GSEA analysis for the investigation of COL5A2 potential regulatory mechanisms. We downloaded the "h.all.v7.4.symbols.gmt" gene sets from the GSEA website. The terms with a p-value < 0.05 were selected.

Relationship between COL5A2 and tumor immunoreactivity

To investigate the correlation between COL5A2 and 28 tumor-infiltrating lymphocytes (TILs), we employed the TISIDB online platform (http://cis.hku.hk/TISIDB/) [20]. Additionally, the relationship between COL5A2 and immune cell infiltration levels, including CD4 + and CD8 + T cells, Treg cells, mast cells, dendritic cells (DCs), macrophages, monocytes, and neutrophils was conducted via the TIMER2.0 database. Moreover, the relationships between COL5A2 expression and tumor-infiltrating immune cell gene markers were analyzed through the TIMER2.0 database. The markers of CD8 + T cells, T cells (general), B cells, monocytes,

tumor-associated macrophages (TAMs), M1 and M2 macrophages, neutrophils, natural killer (NK) cells, DCs and T cells (Th1, Th2, Tfh, Th17, Tregs, and exhausted T cells) with different function were contained. The immune gene marker sets were quoted in prior studies [21–23].

The Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/index.html) [24] was used to further confirm the significant correlations found in TIMER2.0. We explored the gene expression correlations via Spearman's correlation analysis with both tumor and normal tissues.

Moreover, the association between COL5A2 expression and the 47 common immune checkpoint molecules was examined via R software with limma, reshape2, ggplot2, ggpubr, and corrplot packages.

Predict the sensitivity to immunotherapy

According to previous research, IPS can predict the efficacy of immunotherapy [25]. IPS for GC patients are available through The Cancer Immunome Atlas (TCIA) (https://tcia.at/home). The limma and ggpubr packages in R software were used for analysis. Additionally, as a part of our analysis of the predictive value of COL5A2 expression in immunotherapy response, we downloaded RNA-Seq data of the immunotherapy cohort PRJEB25780 from European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena/browser/home). In this cohort, researchers examined 45 patients with metastatic or recurrent gastric cancer who received anti-programmed cell death protein 1 (PD-1) therapy [26]. The immunotherapy responses were classified into four types: complete response (CR), partial response (PR), stable disease (SD) and progressive disease (PD). We compared COL5A2 expression levels between the group of patients with the confirmed response (CR/PR) and those without clinical response (PD/SD) through limma and ggpubr packages in R software. Furthermore, we derived the ROC curve of COL5A2 expression to predict the response to anti-PD-1 therapy for GC patients through pROC packages in R software.

Statistical analysis

In this study, scRNA-seq analysis was conducted with R software (version 4.1.3), and the rest statistical analyses were conducted using R software (version 4.1.2) and Perl (version 5.30.0). Throughout the study, the appropriate R packages and statistical methods were described. Statistical analyses involved Chi-square tests with SPSS version 26.0 (IBM Corp, Armonk, NY). We considered $P < 0.05$ to be statistically significant.

Results

Abnormally overexpression of COL5A2 in GC

Firstly, to fully evaluate the significance of COL5A2 in various solid tumors, pan-cancer analysis was conducted to compare COL5A2 mRNA expression levels between tumor samples and normal samples of the TCGA-STAD through the TIMER2.0 database. According to the analysis, COL5A2 was significantly upregulated in cancerous tissues, such as STAD, cholangiocarcinoma, thyroid cancer and colon adenocarcinoma. It suggested that COL5A2 in these cancers could be an oncogene (Fig. 1a). Secondly, the TCGA-STAD dataset allowed us to compare COL5A2 expression in 32 paracancerous and 375 GC samples. Based on the scatter plot, we found that COL5A2 expression in GC samples was significantly higher than that in normal tissues ($p < 0.001$, Fig. 1b). Following this, the data from 407 samples was then used to find 27 pairs of GC and adjacent normal samples. It was revealed that COL5A2 was significantly upregulated in GC samples compared to matched paracancerous samples ($p < 0.001$, Fig. 1c). In addition, COL5A2 protein expression was examined in GC tissues and peritumoral tissues via IHC (Fig. 1d-e). High COL5A2 expression was detected in GC tissues (42.5%, 17/30), but not detected in adjacent non-tumor tissues (0%, 0/10). Chi-square tests and p values were calculated (chi-square = 4.685, $P = 0.03$). Overall, the findings revealed significantly higher levels of COL5A2 expression in GC tissues than in normal tissues.

Expression of COL5A2 based on scRNA-seq analysis

After QC, we obtained 21,909 high-quality cells, including 9854 deep layer cells, 9341 superficial layer cells and 2714 normal cells. Based on the known markers, we identified 7 kinds of cell types, including T cells ($n = 8947$), B cells ($n = 4733$), myeloid cells ($n = 2594$), plasma cells ($n = 2591$), epithelial cells ($n = 1498$), endothelial cells ($n = 861$) and fibroblasts ($n = 685$) (Fig. 2a). COL5A2 was expressed on the fibroblast cell type (Fig. 2b). There were 78 COL5A2-expressing cells in superficial layer cells, 236 COL5A2-expressing cells in deep layer cells, and only 16 COL5A2-expressing cells in normal gastric cells (Fig. 2c). It revealed that COL5A2-expressing cells were abundant in GC, and the number of cells expressing COL5A2 in the deep layer was more than that in the superficial layer ($p < 0.05$, Table 1). The results confirmed that the expression of COL5A2 was upregulated in GC and may have an association with the invasion of GC.

Relationship between COL5A2 and clinicopathological traits of GC patients

The association between COL5A2 expression and clinicopathological features in the TCGA clinical data

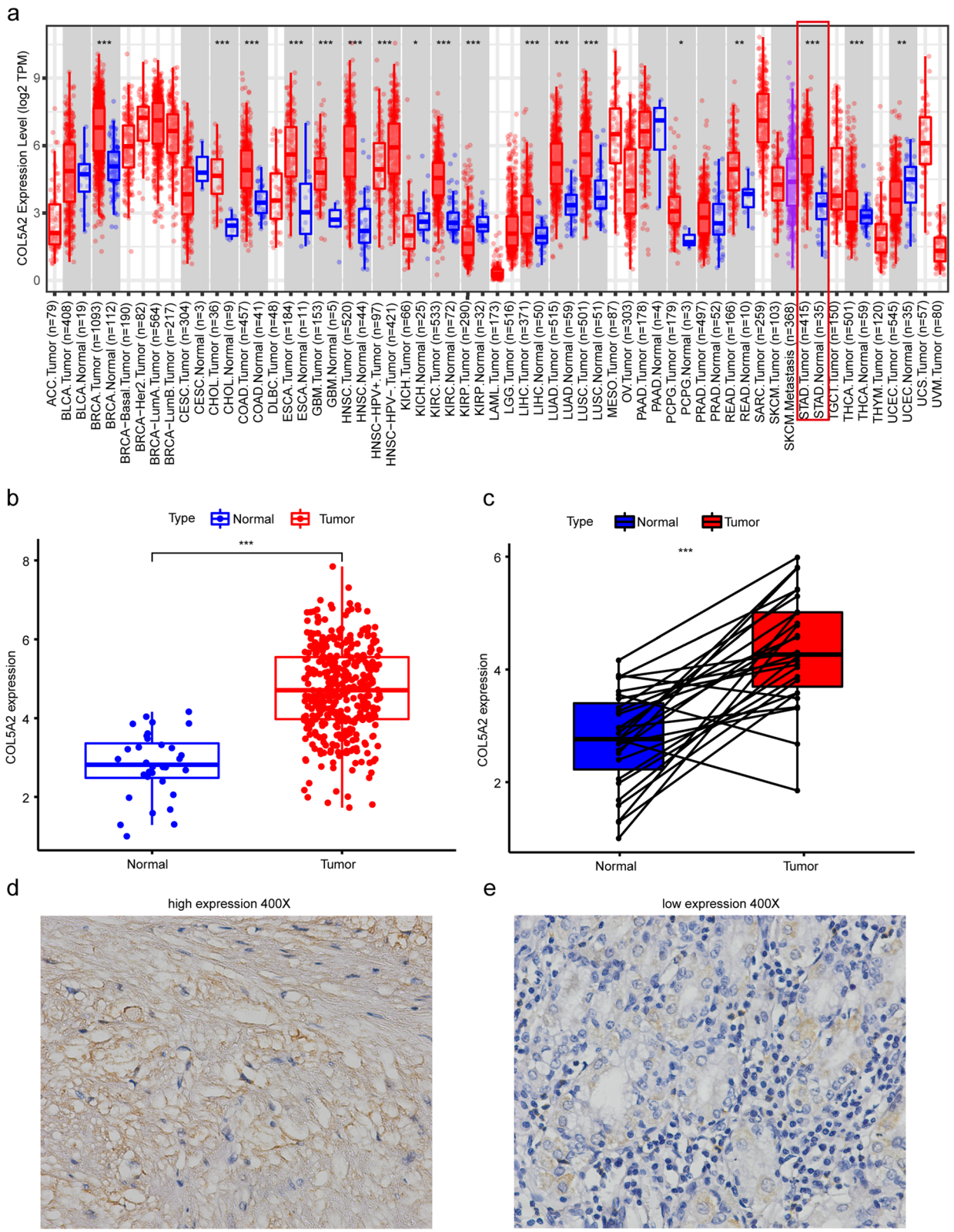


Fig. 1 The expression level of COL5A2 was upregulated in gastric cancer (GC). **a** The level of COL5A2 expression in different tumor types analyzed in TIMER2.0. **b** COL5A2 mRNA levels in GC samples and normal gastric samples. **c** The expression of COL5A2 in GC samples was significantly higher than that of 27 pairs of non-cancerous adjacent samples ($p < 0.001$). **d** High expression of COL5A2 in GC tissues. **e** Low expression of COL5A2 in normal tissues. The magnification was 400 fold. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$

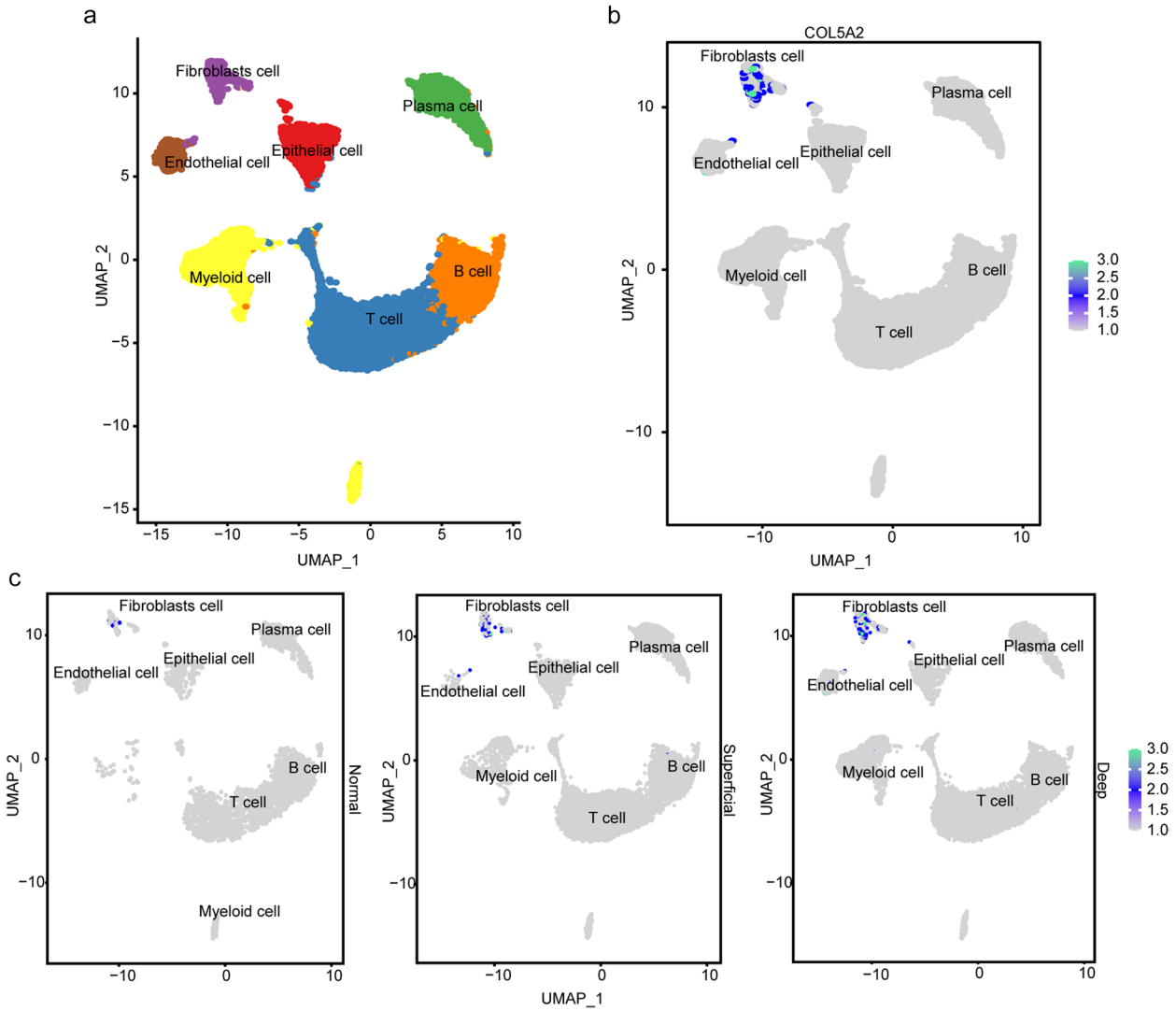


Fig. 2 The expression of COL5A2 in cells with scRNA-seq analysis. **a** The eight major cell types in the UMAP plot. **b** COL5A2 was expressed in the fibroblasts cell type. **c** The expression of COL5A2 in normal, superficial layer and deep layer tissues

Table 1 The number of COL5A2-expressing cells in normal, superficial layer and deep layer tissue in GC with chi-square test

Comparison	p.Chisq	p.adj.Chisq
N: S	2.48e-01	2.48e-01
N: D	4.52e-09	6.78e-09*
S: D	2.70e-17	8.10e-17*

N Normal tissue, *S* Superficial layer tissue, *D* Deep layer tissue

* $P < 0.05$

was analyzed. The clinicopathological characteristics included age, gender, histologic grade, clinical stage, tumor, node and metastasis (TNM) classification, survival status and survival time (Table 2). Increased expression of COL5A2 showed significant association with age ($p < 0.05$, Fig. 3a), T classification (T1 vs. T2, T1 vs. T3, T1 vs. T4) ($P < 0.001$, Fig. 3b) and clinical stage (stageI vs. stageII, stageI vs. stageIII) ($p < 0.05$, Fig. 3c). Similarly, we verified the data in 433 additional instances in GSE84437. The upregulated COL5A2 was significantly related to T classification (T1 vs. T4, T2 vs. T4) ($P < 0.05$, Fig. 3d) and N classification (N1 vs. N0, N2 vs. N0) (N1 vs. N0 $P < 0.05$, N2 vs. N0 $P < 0.01$, Fig. 3e), according to the results. Additionally, based on IHC analysis, clinical samples were divided into low- and high-expression groups to determine the relationship between COL5A2 protein expression and clinical characteristics. As shown in Table 3, the expression of COL5A2 was significantly associated with T classification ($P < 0.01$), N classification ($P = 0.03$), and TNM stage ($P = 0.001$).

Table 2 Clinicopathological characteristics of gastric cancer patients from TCGA database

Clinicopathological Characteristics	Number of Gastric Cancer Patients (*N* = 443)
Age (years)	
≤65	197 (44.5%)
>65	241 (54.4%)
Unknown	5 (1.1%)
Gender	
Male	285 (64.3%)
Female	158 (35.7%)
Grade	
G1	12 (2.7%)
G2	159 (35.9%)
G3	263 (59.4%)
Unknown	9 (2.0%)
Stage	
Stage I	59 (13.3%)
Stage II	130 (29.3%)
Stage III	183 (41.3%)
Stage IV	44 (9.9%)
Unknown	27 (6.1%)
T classification	
T1	23 (5.2%)
T2	93 (20.9%)
T3	198 (44.7%)
T4	119 (26.9%)
Unknown	10 (2.3%)
N classification	
N0	132 (29.8%)
N1	119 (26.9%)
N2	85 (19.2%)
N3	88 (19.9%)
Unknown	19 (4.3%)
M classification	
M0	391 (88.3%)
M1	30 (6.8%)
Unknown	22 (4.9%)
Survival status	
Death	171 (38.6%)
Survival	272 (61.4%)

COL5A2 was an independent adverse prognostic factor in GC

To investigate the relationship between COL5A2 expression and GC prognosis, the survival rate of high and low COL5A2 expression groups was analyzed through the Kaplan–Meier risk estimation method. Notably, survival rates were significantly lower in the group with high COL5A2 expression ($p = 0.013$, Fig. 4a). Then GSE84437 ($n = 433$) was employed to verify the result. Similarly, the high COL5A2 expression group had a worse survival rate ($p = 0.025$) (Fig. 4b). In a word, COL5A2 expression was elevated and associated with an unfavourable outcome in GC. In addition, the survival of GC was significantly associated with age, stage, T stage, N stage, and COL5A2 expression in univariate Cox regression analysis (Table 4). Age, stage, and COL5A2 expression were independent predictors of poor survival of GC in Multivariate Cox regression analysis (Table 4). Subsequently, we assessed the survival rates of GC patients on TCGA-STAD at 1, 3, and 5 years, and AUC values for time-dependent ROC curves were as follows: 0.594, 0.593, and 0.744 (Fig. 4c). The AUC value at 5 years of COL5A2 expression was higher than that of age (Fig. 4d) and tumor stage (Fig. 4e). Taken together, the above findings indicated that COL5A2 was found to be an unfavourable prognostic factor as well as an independent prognostic marker.

COL5A2 co-expression networks in GC

To gain insight into COL5A2's biological meaning in GC, we identified genes that are positively co-expressed with COL5A2 via cBioPortal. We selected a total of six genes with high significance correlations with COL5A2 (absolute Pearson's $r > 0.89$) for further analysis, as shown in Fig. 5a-f. Furthermore, TIMER2.0 was used to verify the correlations between COL5A2 and the six genes. According to the results, COL1A2 ($r = 0.94$, $P = 1.74e\text{-}194$), COL3A1 ($r = 0.926$, $P = 2.19e\text{-}177$), COL5A1 ($r = 0.918$, $P = 1.09e\text{-}167$), COL1A1 ($r = 0.906$, $P = 2.41e\text{-}156$), SPARC ($r = 0.906$, $P = 5.17e\text{-}156$), COL12A1 ($r = 0.898$, $P = 1.2e\text{-}149$) were strongly positively correlated with COL5A2 as shown in Fig. 5g-l. Additionally, in GC we observed a remarkable upregulation of COL5A2-related genes (Additional file 1: Figure S1). There were significant associations between these genes and GC tumor stage, except for COL12A1 (Additional file 2: Figure S2). Furthermore, COL5A2-related genes, except for COL3A1, had significant associations with poor overall survival (Additional file 3: Figure S3).

The biological correlation of COL5A2 in GC

Differentially expressed genes (DEGs) analysis was employed to study COL5A2 and its potential pathogenicity in GC. We identified 521 DEGs between high and low COL5A2 expression datasets. There were 125 downregulated genes and 396 upregulated genes among these DEGs. The 50 DEGs were shown in the heatmap (Fig. 6a). Then, the DEGs were analyzed via GO and KEGG analyses. Following were GO terms with the greatest enrichment: BP (biological process), including extracellular matrix organization, extracellular structure organization, external encapsulating structure organization, CC

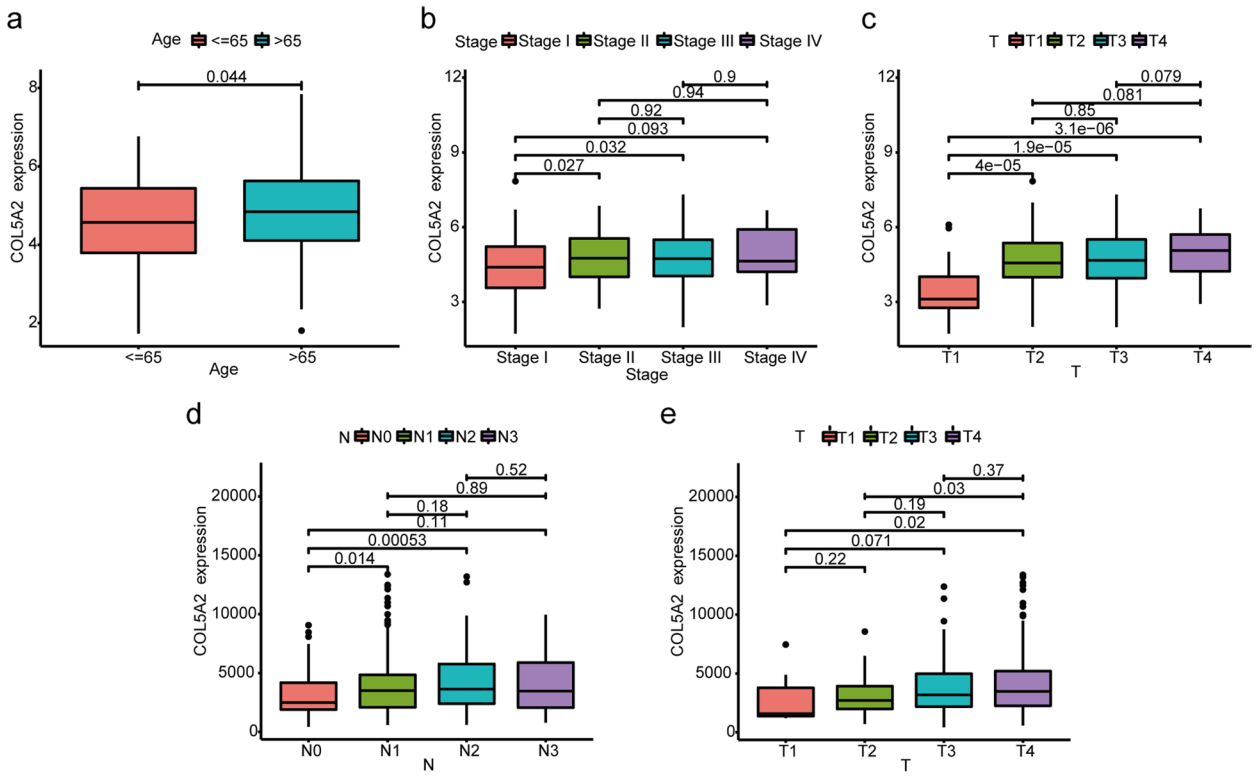


Fig. 3 Association between COL5A2 expression and clinicopathological characteristic of gastric cancer patients. **a**-**c** The relationship between COL5A2 expression and age, stage and T classification based on TCGA data. **d**, **e** The relationship between COL5A2 expression and N and T classification based on GEO data

(cellular component) such as collagen-containing extracellular matrix, endoplasmic reticulum lumen, collagen trimer, and MF (molecular function) such as extracellular matrix structural constituent, heparin binding, glycosaminoglycan binding (Fig. 6b). Meanwhile, there were significant enrichments in the PI3K-Akt signaling pathway, focal adhesion, and protein digestion and absorption according to KEGG pathway analysis (Fig. 6c). Since COL5A2 expression was related to tumor stage as well as the prognosis of GC, we hypothesised that upregulated COL5A2 could promote tumor progression. Moreover, GSEA was conducted and found that there was a dynamic correlation between high COL5A2 expression and hallmarks of tumor such as "angiogenesis", "IL6-JAK-STAT3 signaling" and "NOTCH signaling". As opposed to "KRAS signaling Dn", there was a significant enrichment in the low COL5A2 expression group (Fig. 7a, b). The results may provide a potential mechanism of tumor progression.

COL5A2 affects infiltrating immune cells in GC

Tumor progression, treatment responses, and prognosis are affected by immune cell infiltration. Hence, an analysis of the relationship between immune cell infiltration and COL5A2 expression in GC was conducted via TISIDB as well as TIMER2.0. There was a significant association between COL5A2 expression and the abundance of TILs, as revealed by the results (Fig. 8a). For example, COL5A2 expression was positively correlated with the infiltrating levels of central memory CD8 T cell (rho = 0.359, $p = 6.03e\text{-}14$), central memory CD4 T cell (rho = 0.403, $p < 2.2e\text{-}16$), Th1 (rho = 0.428, $p < 2.2e\text{-}16$), Treg (rho = 0.538, $p < 2.2e\text{-}16$), NK (rho = 0.555, $p < 2.2e\text{-}16$), plasmacytoid dendritic cell (pDC) (rho = 0.377, $p = 1.59e\text{-}15$), macrophage (rho = 0.437, $p < 2.2e\text{-}16$), mast cell (rho = 0.454, $p < 2.2e\text{-}16$). In addition, we employed EPIC, QUANTISEQ, COUNTER, CIBERSORT-ABS, TIMER and XCELL tools from TIMER2.0 to confirm the relationship between immune cell infiltration and COL5A2 expression (Fig. 8b-j). As a result, COL5A2 expression affected the levels of immune infiltration.

COL5A2 expression was associated with immune marker sets

To investigate the correlation between COL5A2 expression and multifarious immune infiltrating cells, the relationship between COL5A2 expression and

Table 3 The expression of COL5A2 and the clinical characteristics of patients with gastric cancer

Characteristics	No. of cases ($N=40$)	COL5A2 expression		Chi-square	*P* value
		Low	High		
Age				0.102	0.749
<65	20	12	8		
≥65	20	11	9		
Gender				1.671	0.196
Male	30	15	15		
Female	10	8	2		
Grade				1.436	0.488
G1-G2	6	4	2		
G3	33	18	15		
Signet ring cell or mucinous	1	1	0		
T classification				15.920	0.000*
T1-T2	14	14	0		
T3-T4	26	9	17		
N classification				4.682	0.030*
N0	12	10	2		
N1-N3	28	13	15		
M classification				0.024	0.878
M0	39	23	16		
M1	1	0	1		
TNM stage				13.606	0.001*
I	8	8	0		
II	12	9	3		
III-IV	20	6	14		
HER2 expression				0.395	0.53
Negative	28	17	11		
Positive	12	6	6		

* $P<0.05$ was considered significant

immune marker sets of diverse immune cells of STAD was analyzed in the TIMER2.0 and GEPIA databases. The immune cells included T cells (general), CD8 + T cells, monocytes, M1 and M2 macrophages, B cells, TAMs, NK cells, DCs and neutrophils. Moreover, we examined functional T cells, such as Th1, Th2, Treg, Th17, Tfh and exhausted T cells. Using purity-adjusted correlations, a link between COL5A2 expression and immune cells can only be established when COL5A2 was associated with all immune cells' markers. we found that COL5A2 expression was significantly related to M1 and M2 macrophages, TAMs, monocytes, Th1, Th2, Treg, and T cell exhaustion (Table 5). Furthermore, correlations between COL5A2 expression and immune markers of monocytes, M1 and M2 macrophages, and TAMs were examined via GEPIA. As with TIMER2.0, the results were similar (Additional file 4: Table S1). All the above findings indicated that COL5A2 expression affected immune infiltration levels in GC.

Expression of COL5A2 was associated with immune checkpoint genes

It is possible to impair the antitumor immune response by overexpressing immune checkpoint molecules. This can cause tumor immune escape and lead to tumor progress [27, 28]. Hence, immune checkpoint molecules were analyzed for their association with COL5A2 expression. According to the results, in GC, many immune checkpoint molecules, such as CD276, PDCD1LG2, HAVCR2, CD200, and CD274 were positively correlated with COL5A2 (Fig. 9). These findings suggested that high COL5A2 expression may cause immune tolerance and escape through multiple immune checkpoint pathways, and lead to worse outcomes for patients with GC. However, the discoveries deserve further clinical verification.

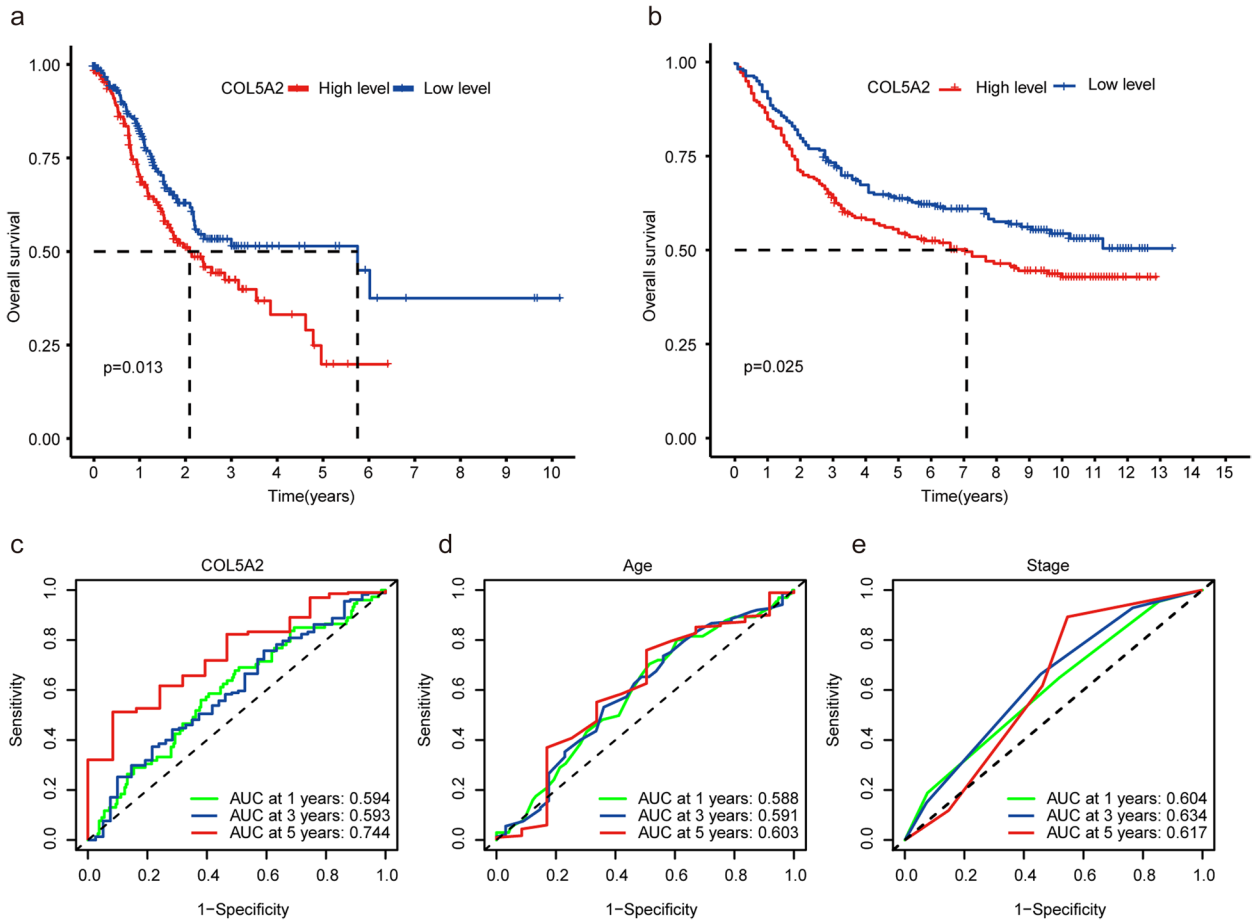


Fig. 4 COL5A2 impact on the overall survival of patients with GC. **a** Kaplan–Meier curve was used to analyze the relationship between COL5A2 mRNA expression and prognosis of GC patients based on TCGA data ($p = 0.013$). **b** Kaplan–Meier curve was used to analyze the relationship between COL5A2 mRNA expression and prognosis of GC patients based on GEO data ($p = 0.025$). **c**-**e** Time-dependent ROC curves of patients' OS at 1, 3, and 5 years depending on COL5A2 expression (**c**), age (**d**) and tumor stage (**e**) in the TCGA-STAD datasets

Table 4 Univariate and multivariate analysis of COL5A2 mRNA levels and clinical parameters in TCGA gastric cancer

Variables	Univariate Cox regression analysis				Multivariate Cox regression analysis			
	HR	HR.95L	HR.95H	*P*-value	HR	HR.95L	HR.95H	*P*-value
Age	1.024	1.006	1.041	0.010**	1.027	1.009	1.047	0.004**
Gender	1.403	0.959	2.054	0.081	NA	NA	NA	NA
Grade	1.330	0.941	1.883	0.107	NA	NA	NA	NA
Stage	1.549	1.245	1.929	9.13E-05**	1.559	1.108	2.193	0.011*
T	1.255	1.002	1.573	0.049*	0.937	0.697	1.260	0.667
M	1.806	0.971	3.358	0.062	NA	NA	NA	NA
N	1.327	1.132	1.555	4.71E-04**	1.116	0.907	1.374	0.301
COL5A2	1.250	1.066	1.465	5.98E-03**	1.211	1.026	1.430	0.024*

* $P < 0.05$; ** $P < 0.01$

The ability of COL5A2 expression to predict immunotherapy sensitivity

Nowadays, immunotherapy is a main option for cancer therapy [29, 30], particularly immune checkpoint blockade (ICB) targeting cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) and PD-1 [31]. Based on COL5A2

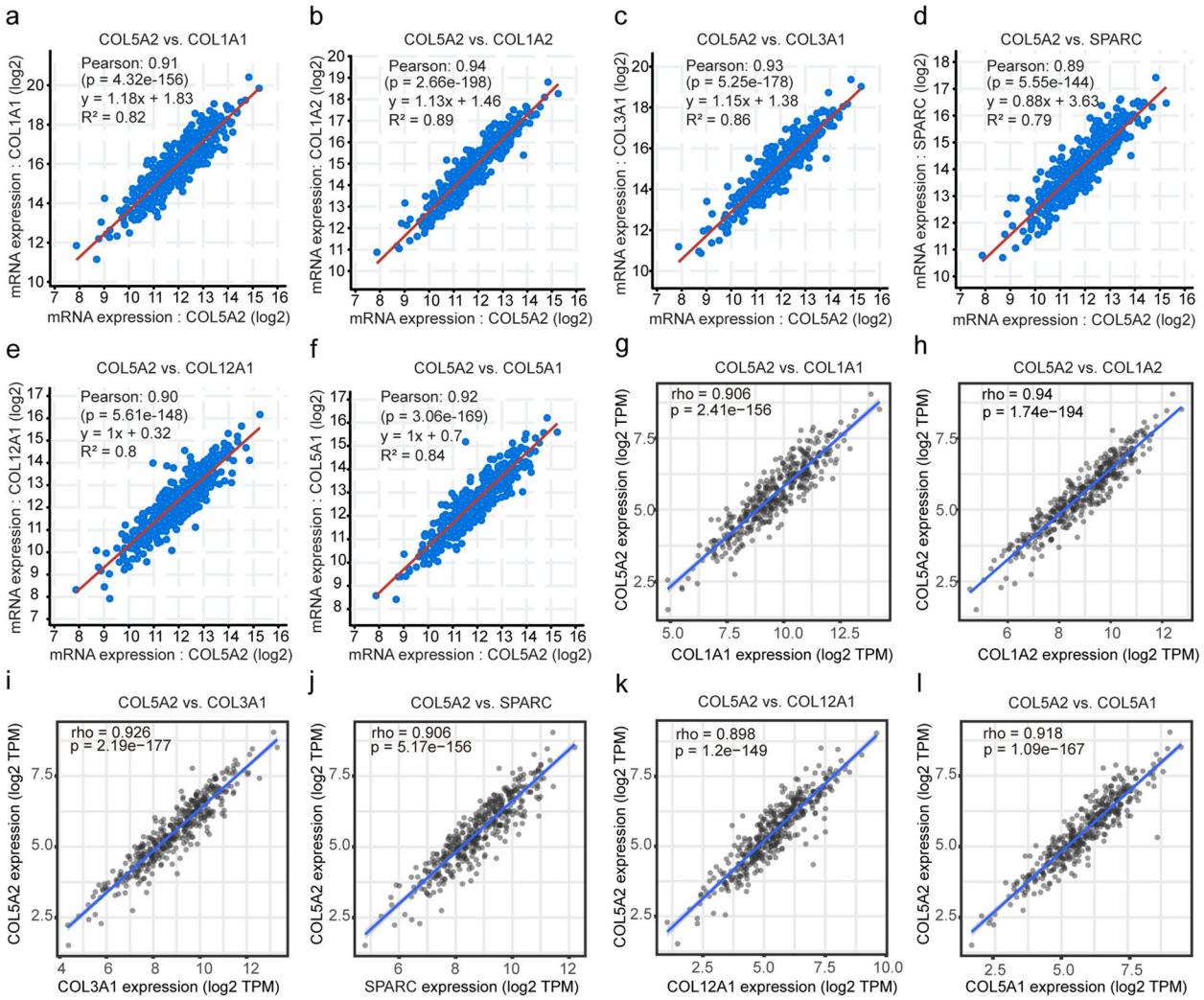


Fig. 5 Co-expressed genes of COL5A2 in GC. **a**-**f** The genes co-expressed with COL5A2 in GC (absolute Pearson's $r \geq 0.89$) were assessed with the cBioPortal database. **g**-**l** COL5A2 was significantly correlated with COL1A2 ($r = 0.94$, $P = 1.74e\text{-}194$), COL3A1 ($r = 0.926$, $P = 2.19e\text{-}177$), COL5A1 ($r = 0.918$, $P = 1.09e\text{-}167$), COL1A1 ($r = 0.906$, $P = 2.41e\text{-}156$), SPARC ($r = 0.906$, $P = 5.17e\text{-}156$), COL12A1 ($r = 0.898$, $P = 1.2e\text{-}149$) in GC with the TIMER2.0 database

expression, we assessed the immunogenicity of each group via IPS analysis. As a result, low COL5A2 expression was linked to higher scores for ips_ctla4_pos_pd1_pos (CTLA-4 positive response and PD-1 positive response), ips_ctla4_neg_pd1_neg (CTLA-4 negative response and PD-1 negative response), ips_ctla4_pos_pd1_neg, as well as ips_ctla4_neg_pd1_pos (Fig. 10a-d). Furthermore, we validated the ability of COL5A2 expression to predict the immunotherapy response with the datasets, which included 45 advanced GC patients treated with pembrolizumab monotherapy. Results showed that patients responding to pembrolizumab expressed significantly lower levels of COL5A2 than those who did not ($p < 0.01$, Fig. 10e). AUC of 0.763 was obtained for COL5A2 expression level predicting the response of gastric cancer patients to pembrolizumab (Fig. 10f). Hence, the patients with low COL5A2 expression may benefit more from CTLA-4 or anti-PD-1 immunotherapy, and COL5A2 expression may be used to predict the immunotherapy response.

Discussion

Collagens, the main protein in the ECM, have been widely reported in diverse cancers [32]. The migration of GC cells was promoted when COL5A2 was overexpressed and inhibited when COL5A2 was silenced in vitro and in vivo [33]. According to previous researchers, COL5A2 was overexpressed in GC

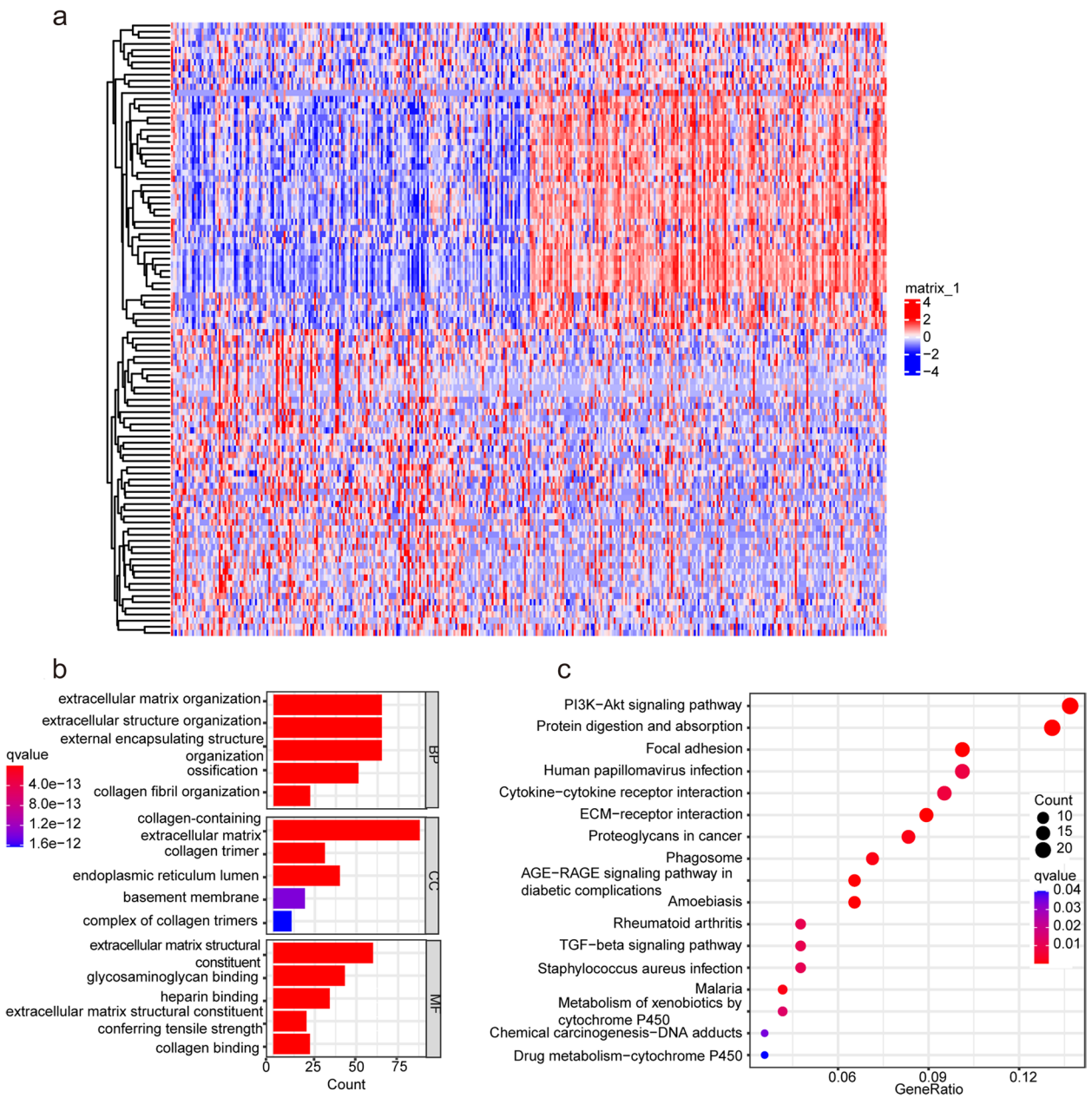


Fig. 6 DEGs and functional analysis. **a** Heatmap of 50 DEGs. **b** GO enrichment analysis of DEGs. **c** KEGG enrichment analysis of DEGs

via RT-qPCR and immunohistochemistry methods and was related to GC prognosis and renal metastasis [34]. A COL3A1/FBN1/COL5A2/SPARC-mir-29a-3p-H19 ceRNA network in GC was revealed by Shen et al. as a prognostic indicator [35]. Mao et al. according to the expression of M2 macrophages, COL5A2, EGF and FLT1 together with clinicopathological characteristics established a nomogram which was an important factor in determining prognosis [36]. Nevertheless, Cao et al. [37] showed that COL5A2 was not related to lower overall survival in GEO datasets. Therefore, the effect of COL5A2 expression on GC is worth to be further explored. In this study, we found that COL5A2 was overexpressed in GC. Based on the single-cell sequencing analysis, a spatial gradient of COL5A2 expression was observed in diffuse-type gastric cancer. Additionally, the high expression of COL5A2 was associated with advanced TNM stage, patients' prognosis and immune infiltration, and may be used as a predictor of the effect of immunotherapy. Therefore, COL5A2 may

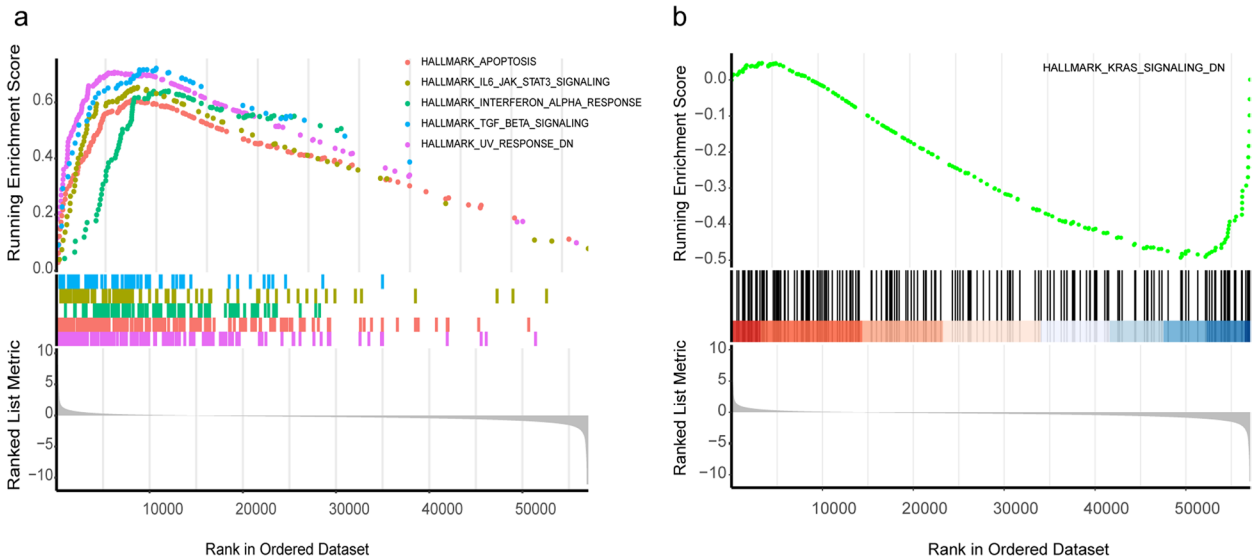


Fig. 7 Gene set enrichment analysis of COL5A2. **a** correlation between high COL5A2 expression and hallmarks of tumor. **b** correlation between low COL5A2 expression and hallmarks of tumor

promote the invasiveness of GC and may serve as a biomarker for GC prognosis and treatment.

The functions of co-expressed genes are generally similar. The co-expressed analysis revealed that the top six co-expressed genes, including COL1A2, COL3A1, COL5A1, COL1A1, SPARC and COL12A1, had a strong positive correlation with COL5A2. Most of them were ECM molecules. They were all upregulated in GC, and the majority of them served as risk factors for advanced stage and poor survival of GC. Co-expressed genes have been shown to promote tumor progression or worse survival in cancers [38–43]. Taken together, we speculated that COL5A2 and its co-expressed genes might be good prognosis markers of GC.

Functional enrichment analyses including GO and KEGG enrichment analyses showed that COL5A2 primarily regulated ECM- related processes, PI3K-AKT signaling pathway and focal adhesion. Normal tissue phenotype and homeostasis are regulated by the ECM, and dysfunction of ECM promotes the development of the tumor [44]. Focal adhesions, connect points between cells and the extracellular matrix [45] and provide a necessary site for adhesion to the matrix during cancer cell migration [46]. Besides, PI3K/AKT signaling pathway is an important factor contributing to tumor progression and poor survival of GC [47, 48]. In addition, GSEA analyses revealed that in the high-COL5A2 expression group, tumor progression-related pathways, including angiogenesis, IL6-JAK-STAT3 signaling as well as Notch signaling were enriched. The occurrence, development, and tumor characteristics of GC are influenced by these signaling pathways. The IL6-JAK-STAT3 pathway, for instance, plays an important role in the pathogenesis of various human malignancies [49]. There is an association between STAT3 and the advanced TNM stage as well as an unfortunate outcome of GC [50]. The Notch signaling pathway also takes participate in regulating many aspects of cancer biology and regulates the crosstalk between the different compartments of the tumor microenvironment (TME) [51]. In GC, NOTCH3 contributes to immune tolerance and promotes tumor development [52]. These results revealed that COL5A2 acted as a regulator of key functions and signaling pathways in GC, leading to an unfortunate prognosis.

Fibroblasts, endothelial cells, immune cells, and their secreted ECM form TME [53]. TME contributes to the progression of tumors and can reduce the resistance of tumor cells to chemotherapy and immunotherapy [54]. The TME and tumor cells may interact through epithelial-mesenchymal transition (EMT), which is an important regular of tumor metastasis [55]. It reported that EMT processes and the immunosuppressive microenvironment in gliomas are regulated by COL5A2 [56]. In addition, the tumor-associated ECM can regulate immune effects as well as the migration and localization of immune cells [57, 58]. COL5A2 encodes extracellular proteoglycans, glycoproteins, and other elements of ECM, contributing to determining ECM composition [59]. Therefore, for GC immunotherapy, COL5A2 could be a possible target.

Immune cells infiltrating tumor tissues not only disrupt cytokine signals but also play an important role

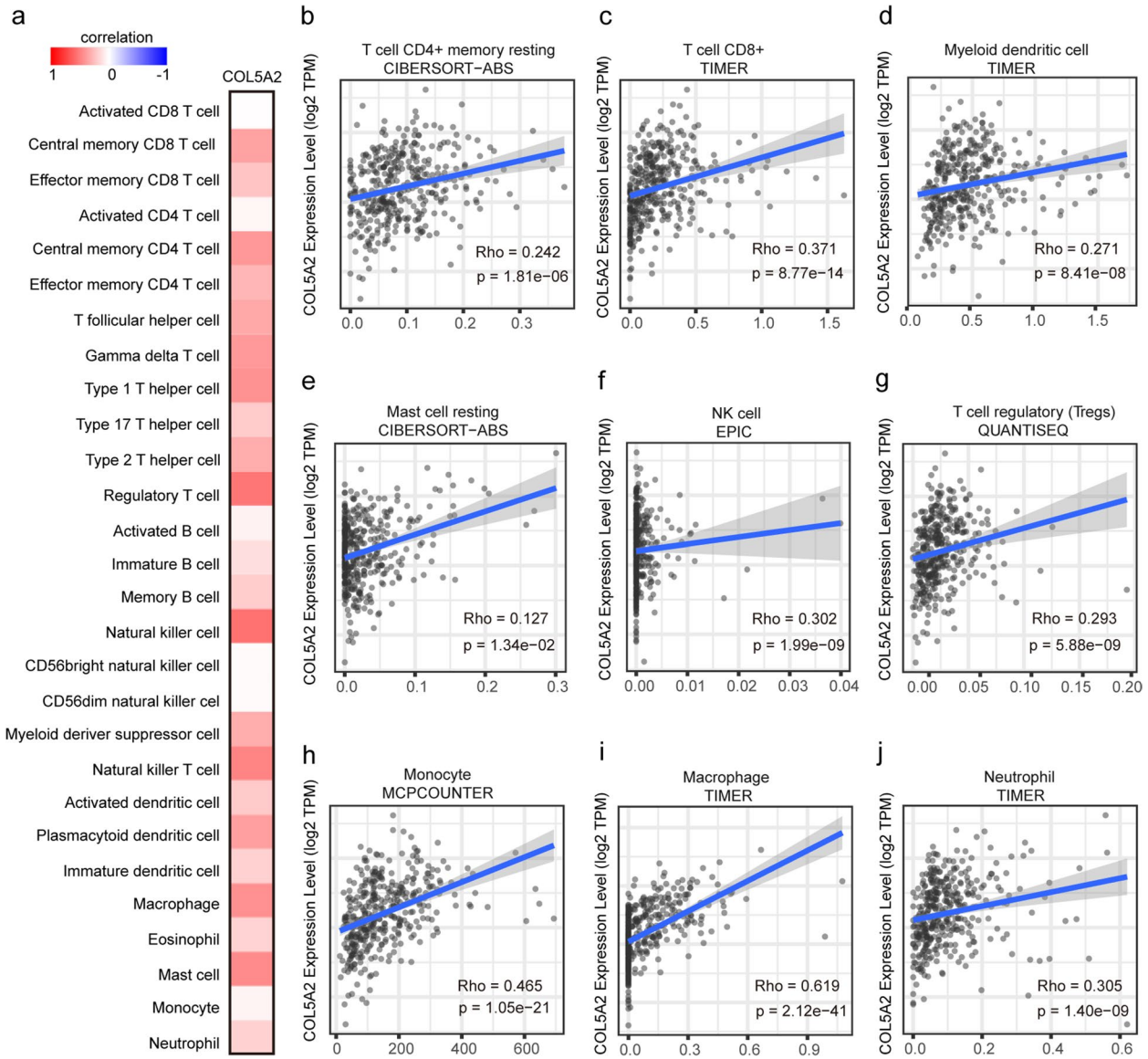


Fig. 8 Correlation between COL5A2 expression and immune infiltration in GC. **a** Correlation between the expression of COL5A2 and the abundance of TILs in GC available at TISIDB database. **b**-**j** Correlation of COL5A2 expression with infiltration levels of CD8 + T cell, CD4 + T cell, Treg cell, mast cell resting, neutrophil, macrophage, dendritic cell, natural killer cell, and monocyte in gastric cancer available at TIMER2.0 database

in cancer biology [60]. In this study, the association between COL5A2 expression with immune infiltration levels in GC was analyzed. It revealed that COL5A2 expression was positively related to immune infiltration levels, such as central memory CD8 and CD4 T cells, Th1, Treg, NK, pDC, macrophages, and mast cells. In addition, we found that COL5A2 expression correlated with immune cell marker genes, which indicated that COL5A2 was involved in regulating tumor immunology. Firstly, COL5A2 expression was weakly correlated with macrophage M1 markers like inducible nitric oxide synthase (iNOS) (NOS2) and interferon regulatory factor 5 (IRF5), while strongly correlated with macrophage M2 gene markers. The findings indicated that COL5A2 may be responsible for polarizing TAMs. The majority of TAMs exhibit the M2 phenotype, which indicates poor outcomes in solid tumors in TME [61]. It has been shown that M2 macrophages can contribute to GC progression and metastasis [62]. Therefore, COL5A2 expression may be related to immunosuppressive activity in

Table 5 Correlation analysis between COL5A2 and immune marker sets of immune cells in TIMER2.0

Description	Gene markers	STAD			
		None		Purity	
		Cor	*P*	Cor	*P*
CD8 + T cell	CD8A	0.132	*	0.106	0.0392
	CD8B	0.001	0.977	-0.014	0.78
T cell (general)	CD3D	0.098	0.0454	0.066	0.202
	CD3E	0.013	0.0359	0.067	0.195
	CD2	0.179	**	0.156	*
B cell	CD19	0.023	0.634	0.013	0.805
	CA79A	0.039	0.432	0.002	0.969
Monocyte	CD86	0.45	***	0.436	***
	CA115(CSF1R)	0.451	***	0.438	***
TAM	CCL2	0.394	***	0.368	***
	CD68	0.31	***	0.285	***
	IL10	0.433	***	0.423	***
M1 Macrophage	INOS (NOS2)	0.128	*	0.138	*
	IRF5	0.216	***	0.214	***
	COX2(PTGS2)	0.385	***	0.39	***
M2 Macrophage	CD163	0.513	***	0.489	***
	VSIG4	0.497	***	0.484	***
	MS4A4A	0.451	***	0.441	***
Neutrophils	CD66b (CEACAM8)	0.012	0.811	0.034	0.507
	CD11b (ITGAM)	0.447	***	0.435	***
	CCR7	0.107	0.0295	0.079	0.127
Natural killer cell	KIR2DL1	0.133	*	0.138	*
	KIR2DL3	0.121	0.0136	0.124	0.0159
	KIR2DL4	0.086	0.0813	0.053	0.305
	KIR3DL1	0.096	0.0508	0.076	0.137
	KIR3DL2	0.126	*	0.12	0.0191
	KIR3DL3	-0.038	0.437	-0.023	0.655
	KIR2DS4	0.106	0.031	0.092	0.0726
Dendritic cell	HLA-DPB1	0.16	*	0.13	0.0114
	HLA-DQB1	0.107	0.0299	0.078	0.13
	HLA-DRA	0.167	**	0.146	*
	HLA-DPA1	0.148	*	0.121	0.0186
	BDCA-1(CD1C)	0.092	0.0609	0.07	0.175
	BDCA-4(NRP1)	0.606	***	0.592	***
	CD11c (ITGAX)	0.492	***	0.497	***
Th1	T-bet (TBX21)	0.15	*	0.136	*
	STAT4	0.234	***	0.223	***
	STAT1	0.218	***	0.211	***
	IFN-γ (IFNG)	0.119	0.0151	0.11	0.0316
	TNF-α (TNF)	0.214	***	0.194	**
Th2	GATA3	0.172	**	0.157	*
	STAT6	0.126	*	0.127	0.0131
	STAT5A	0.351	***	0.354	***
	IL13	0.138	*	0.179	**
Tfh	BCL6	0.315	***	0.279	***
	IL21	0.088	0.0744	0.074	0.152
Th17	STAT3	0.4	***	0.389	***
	IL17A	-0.04	0.422	-0.051	0.319
Treg	FOXP3	0.216	***	0.248	***
	CCR8	0.361	***	0.362	***
	STAT5B	0.361	***	0.367	***
	TGFβ (TGFB1)	0.508	***	0.498	***
T cell exhaustion	PD-1 (PDCD1)	0.131	*	0.123	0.0163
	CTLA4	0.2	***	0.19	**
	LAG3	0.15	*	0.121	0.0189
	TIM-3 (HAVCR2)	0.482	***	0.471	***
	GZMB	0.199	***	0.166	*

STAD Stomach adenocarcinoma, *TAM* Tumor-associated macrophage

* $P < 0.01$; ** $P < 0.001$; *** $P < 0.0001$

GC. Secondly, COL5A2 expression was positively correlated with Treg and T cell exhaustion gene markers. One of the reasons for immune escape is the increase in Tregs in the TME [63]. FOXP3 + Tregs suppress abnormal immune responses against self-antigens and maintain immune homeostasis [64]. Tim3 and PD-1 control T cell responses and are closely linked to T cell exhaustion [65, 66], which is characterized by poor proliferation, reduced cytotoxicity, and loss of effector function [67]. Accordingly, COL5A2 may promote immunosuppression by enhancing Treg differentiation and T cell exhaustion. Thirdly, COL5A2 expression showed significant correlations with the cell markers of T helper cells (Th1, Th2). It may be the possible mechanism that COL5A2 regulates the function of T cells. These results all together revealed that COL5A2 expression may impact immune responses by influencing immunocytes infiltration in GC immune microenvironment, which contributes to a worse prognosis.

Immune checkpoints act on stimulatory and inhibitory pathways and are essential to maintain self-tolerance and regulate the type, magnitude, and duration of immune responses [68]. As a result of activating immune checkpoint pathways, tumors prevent the immune system from recognizing them, thereby inhibiting the immune response [69]. We found a positive correlation between COL5A2 expression and immune checkpoint inhibitors. This indicated that high COL5A2 expression may promote the immune escape of tumor cells in GC. Nevertheless, research is needed to understand the mechanism by which COL5A2 interacts with immune checkpoint molecules.

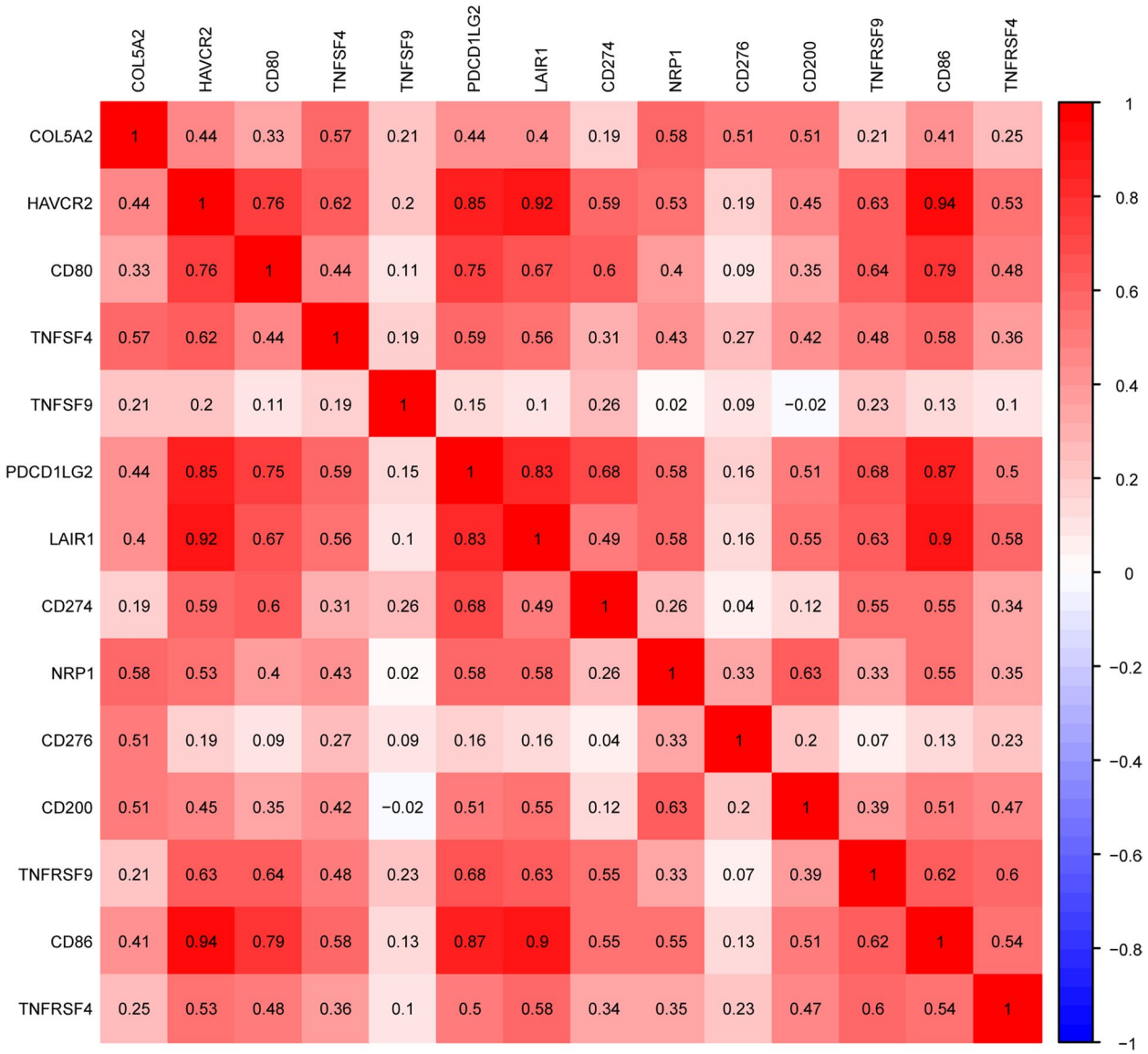


Fig. 9 Correlation analysis of COL5A2 with those of 22 common immune checkpoint genes in gastric cancer

By blocking immune-regulatory receptors like PD-1 and CTLA-4, T cells are restored to function and proliferation, thereby reinvigorating the antitumor immune response [68]. Even though cancer immunotherapy has advanced considerably, most patients do not respond to or benefit from ICB treatment. Immune biomarkers are also important factors that influence the immunotherapy response [70]. Hence, a well-effect of clinically personalized immunotherapy needs comprehensive and precision pre-biomarkers. Considering that COL5A2 expression affects TME immune cell infiltration and may promote tumor cell immune escape, we further analyzed whether COL5A2 expression could assess the response to immunotherapy. Our study demonstrated that COL5A2 expression level may be used to identify GC patients' sensitivity to PD-1 or CTLA-4. Given the associations between COL5A2 expression and the immune landscape, a greater understanding of the mechanisms of COL5A2 for the development of effective treatment strategies is crucial.

Nevertheless, limitations remain in our study. The analysis bias caused by the present retrospective study

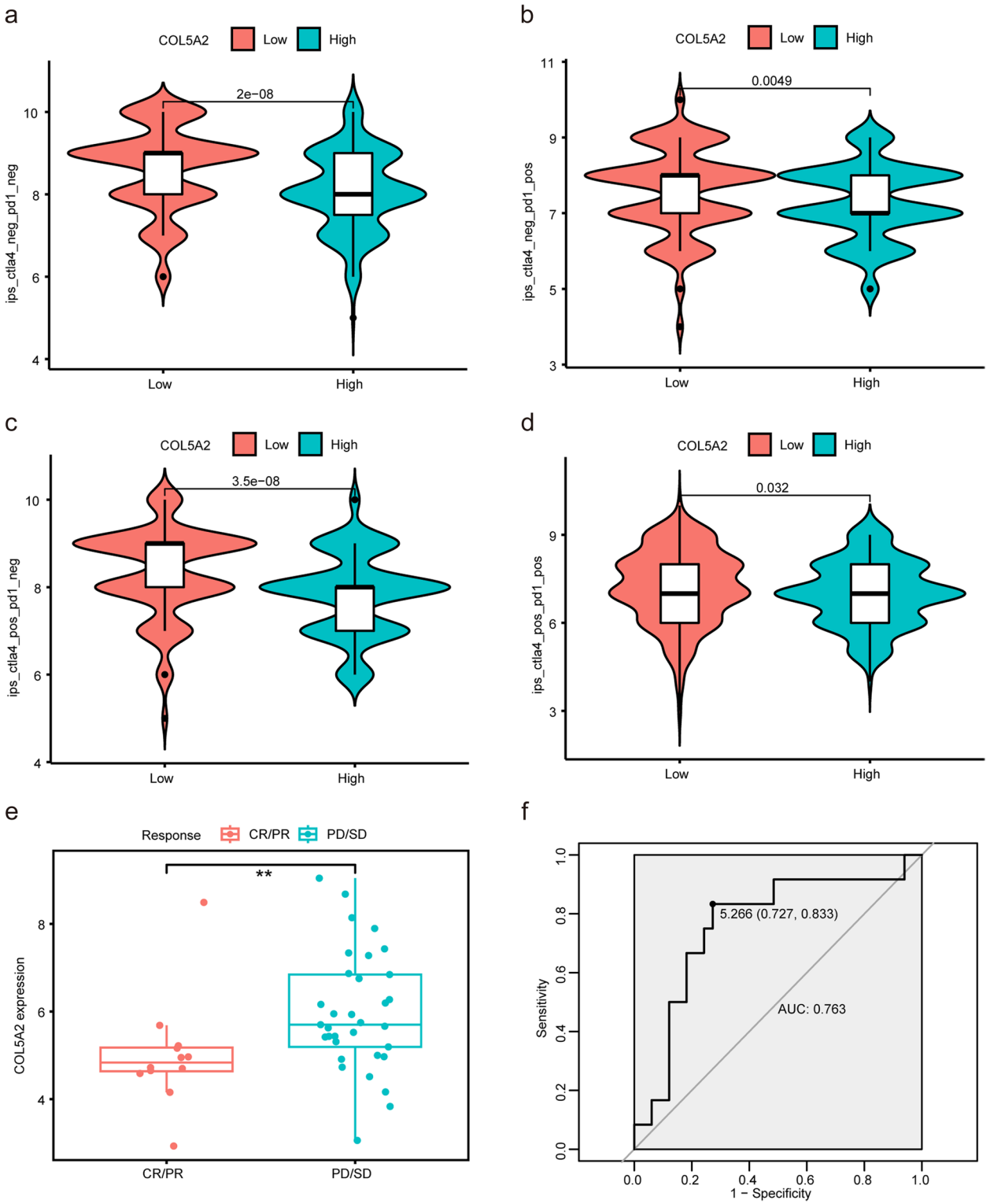


Fig. 10 Correlation between COL5A2 expression and sensitivity to immunotherapy. **a**-**d** The IPS in high and low expression groups of COL5A2. **e** Comparison of COL5A2 expression between responders and non-responders to immunotherapy in PRJEB25780 cohort. **f** ROC curve for COL5A2 expression in predicting response to immunotherapy. $^{**}p < 0.01$; CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease

is inevitable. We will be consistent to perform forward-looking studies further to avoid bias. In addition, to investigate and certify the character and regulatory mechanisms of COL5A2, further experimental studies in vivo and in vitro are necessary.

Conclusions

COL5A2 was upregulated in GC, and the expression had a spatial gradient. COL5A2 had an association with clinicopathological characteristics and terrible outcomes in GC. Additionally, in GC patients, COL5A2 was an independent prognostic factor. Moreover, we also found that COL5A2 expression may be related to immune mechanisms and may be used to predict the clinical immunotherapy response. Accordingly, the results suggested that COL5A2 acted as an important role in GC and might be able to contribute to GC immunotherapy.

Abbreviations

GC	Gastric cancer
COL5A2	Collagen type V alpha 2
TCGA	The cancer genome database
STAD	Stomach adenocarcinoma
GEO	Gene Expression Omnibus
ENA	European Nucleotide Archive
IPS	Immunophenoscore
scRNA-seq	Single-cell RNA sequencing
QC	Quality control
PCA	Principal components
UMAP	Uniform Manifold Approximation and Projection
DEGs	Differential expression genes
AUC	Area under the receiver operating characteristic curve
ROC	Receiver operating characteristic
BP	Biological process(s)
GO	Gene Ontology
CC	Cellular component
MF	Molecular function
KEGG	Kyoto Encyclopedia of Genes and Genomes
GSEA	Gene Set Enrichment Analysis
OS	Overall survival
ECM	Extracellular matrix
IPS	Immunophenoscore
TIMER2.0	Tumor Immune Estimation Resource 2.0
FFPE	Formalin-fixed paraffin-embedded
TILs	Tumor-infiltrating lymphocytes
TAMs	Tumor-associated macrophages
DCs	Dendritic cells
pDC	Plasmacytoid dendritic cell
NK	Natural killer
Th	T helper cell
Tfh	Follicular helper T cell
Treg	Regulatory T cell
GEPIA	The Gene Expression Profiling Interactive Analysis
TCIA	The Cancer Immunome Atlas
ICB	Immune checkpoint blockade
CTLA-4	Cytotoxic T-lymphocyte-associated antigen 4
PD-1	Programmed cell death protein 1
CR	Complete response
PR	Partial response
SD	Stable disease
PD	Progressive disease
EMT	Epithelial-mesenchymal transition
iNOS	Inducible nitric oxide synthase
IRF5	Interferon regulatory factor 5

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-023-01659-9.

Additional file 1: Figure S1. Analysis of the expression of COL5A2 co-expressed genes in gastric cancer using GEPIA. (a-f) the expression of COL1A2, COL3A1, COL5A1, COL1A1, SPARC and COL12A1 mRNA levels in GC tissues and normal gastric tissues. *, $P<0.05$.

Additional file 2: Figure S2. Analysis of the correlations between COL5A2 co-expressed genes and gastric cancer stage using GEPIA. (a-f) The association between COL1A2, COL3A1, COL5A1, COL1A1, SPARC and COL12A1 expression and gastric cancer stage.

Additional file 3: Figure S3. Analysis of the correlations between COL5A2 co-expressed genes and the survival of gastric cancer patients using Kaplan-Meier plotter online database. (a-f) The association between COL1A2, COL3A1, COL5A1, COL1A1, SPARC and COL12A1 expression and gastric cancer survival.

Additional file 4: Table S1. Analysis of the correlations between COL5A2 expression and immune marker sets through GEPIA.

Acknowledgements

The authors would like to thank the TCGA, GEO and ENA databases for the availability of the data. An earlier version of the manuscript has been presented as a preprint [71].

Authors' contributions

DZ and MC designed the study. XZ and LZ collected the data. MC and XZ wrote the manuscript. All authors contributed to manuscript revision and read and approved the submitted manuscript.

Funding

No funding was received.

Availability of data and materials

The data generated and analyzed during this study are described in the following data record: 10.6084/m9.figshare.23647089 [72]. The datasets used and analyzed during the current study are available from the cancer genome database (TCGA-STAD) https://portal.gdc.cancer.gov and NCBI Gene Expression Omnibus (GEO: GSE84437) https://www.ncbi.nlm.nih.gov/geo/. Single-cell RNA sequencing (scRNA-seq) data GSE167297 was downloaded from GEO https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE167297. RNA-Seq data PRJEB25780 were downloaded from ENA (https://www.ebi.ac.uk/ena/data/view/PRJEB25780). These data are free and publicly available.

Declarations

Ethics approval and consent to participate

The present study was reviewed and approved by the research ethics committee of the second hospital of Hebei Medical University (ethics approval number: 2023-R222), and all the patients or their families provided written informed consent. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Gastroenterology, The Second Hospital of Hebei Medical University, Shijiazhuang, Hebei Province 050000, China. [2]Department of Gastroenterology, Hengshui People's Hospital, Hengshui, Hebei Province 053000,

China. [3]Department of Gastroenterology, The Third Hospital of Hebei Medical University, Shijiazhuang, Hebei Province 050000, China.

Received: 23 January 2023 Accepted: 9 September 2023
Published online: 18 September 2023

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49.
2. Joharatnam-Hogan N, Shiu KK, Khan K. Challenges in the treatment of gastric cancer in the older patient. Cancer Treat Rev. 2020;85:101980.
3. Mehta R, Kommalapati A, Kim RD. The impact of ramucirumab treatment on survival and quality of life in patients with gastric cancer. Cancer Manag Res. 2020;12:51–7.
4. Rajabnejad A, Vaida F, Valasek M, Razzaque S, Fanta P, Horgan S, Bouvet M, Lowy AM, Kelly KJ. Predictors and significance of histologic response to neoadjuvant therapy for gastric cancer. J Surg Oncol. 2021;123(8):1716–23.
5. Tan Z. Recent advances in the surgical treatment of advanced gastric cancer: a review. Med Sci Monit. 2019;25:3537–41.
6. Johnston FM, Beckman M. Updates on management of gastric cancer. Curr Oncol Rep. 2019;21(8):67.
7. Walker C, Mojares E, Del Rio Hernandez A. Role of extracellular matrix in development and cancer progression. Int J Mol Sci. 2018;19(10):3028.
8. Mak ACY, Tang PLF, Cleveland C, Smith MH, Kari Connolly M, Katsumoto TR, Wolters PJ, Kwok PY, Criswell LA. Brief report: whole-exome sequencing for identification of potential causal variants for diffuse cutaneous systemic sclerosis. Arthritis Rheumatol. 2016;68(9):2257–62.
9. Cortini F, Villa C. Ehlers-Danlos syndromes and epilepsy: an updated review. Seizure. 2018;57:1–4.
10. Chen HC, Tseng YK, Shu CW, Weng TJ, Liou HH, Yen LM, Hsieh IC, Wang CC, Wu PC, Shiue YL, et al. Differential clinical significance of COL5A1 and COL5A2 in tongue squamous cell carcinoma. J Oral Pathol Med. 2019;48(6):468–76.
11. Meng XY, Shi MJ, Zeng ZH, Chen C, Liu TZ, Wu QJ, Li S, Li S. The role of COL5A2 in patients with muscle-invasive bladder cancer: a bioinformatics analysis of public datasets involving 787 subjects and 29 cell lines. Front Oncol. 2018;8:659.
12. Wang J, Jiang YH, Yang PY, Liu F. Increased collagen type V alpha2 (COL5A2) in colorectal cancer is associated with poor prognosis and tumor progression. Onco Targets Ther. 2021;14:2991–3002.
13. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018. https://doi.org/10.1038/nbt.4314.
14. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, Li B, Liu XS. TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res. 2020;48(W1):W509–14.
15. Gao B, Lin J, Jiang Z, Yang Z, Yu H, Ding L, Yu M, Cui Q, Dunavin N, Zhang M, et al. Upregulation of chemokine CXCL10 enhances chronic pulmonary inflammation in tree shrew collagen-induced arthritis. Sci Rep. 2018;8(1):9993.
16. Xiang Y, Wang Q, Guo Y, Ge H, Fu Y, Wang X, Tao L. Cx32 exerts anti-apoptotic and pro-tumor effects via the epidermal growth factor receptor pathway in hepatocellular carcinoma. J Exp Clin Cancer Res. 2019;38(1):145.
17. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
18. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587–92.
19. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.
20. Ru B, Wong CN, Tong Y, Zhong JY, Zhong SSW, Wu WC, Chu KC, Wong CY, Lau CY, Chen I, et al. TISIDB: an integrated repository portal for tumor-immune system interactions. Bioinformatics. 2019;35(20):4200–2.
21. Cersosimo F, Lonardi S, Bernardini G, Telfer B, Mandelli GE, Santucci A, Vermi W, Giurisato E. Tumor-associated macrophages in osteosarcoma: from mechanisms to therapy. Int J Mol Sci. 2020;21(15):5207.
22. Danaher P, Warren S, Dennis L, D'Amico L, White A, Disis ML, Geller MA, Odunsi K, Beechem J, Fling SP. Gene expression markers of tumor infiltrating leukocytes. J Immunother Cancer. 2017;5:18.
23. Siemers NO, Holloway JL, Chang H, Chasalow SD, Ross-MacDonald PB, Voliva CF, Szustakowski JD. Genome-wide association analysis identifies genetic correlates of immune infiltrates in solid tumors. PLoS ONE. 2017;12(7):e0179726.
24. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45(W1):W98.
25. Liu J, Meng H, Nie S, Sun Y, Jiang P, Li S, Yang J, Sun R, Cheng W. Identification of a prognostic signature of epithelial ovarian cancer based on tumor immune microenvironment exploration. Genomics. 2020;112(6):4827–41.
26. Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JI, Kim K, Liu XQ, Sher X, Jung H, Lee M, et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. Nat Med. 2018;24(9):1449–58.
27. Orabona C, Mondanelli G, Puccetti P, Grohmann U. Immune checkpoint molecules, personalized immunotherapy, and autoimmune diabetes. Trends Mol Med. 2018;24(11):931–41.
28. Darvin P, Toor SM, Sasidharan Nair V, Elkord E. Immune checkpoint inhibitors: recent progress and potential biomarkers. Exp Mol Med. 2018;50(12):1.
29. Johnson DB, Sullivan RJ, Menzies AM. Immune checkpoint inhibitors in challenging populations. Cancer. 2017;123(11):1904–11.
30. Atkins MB, Clark JI, Quinn DI. Immune checkpoint inhibitors in advanced renal cell carcinoma: experience to date and future directions. Ann Oncol. 2017;28(7):1484–94.
31. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep. 2017;18(1):248–62.
32. Nissen NI, Karsdal M, Willumsen N. Collagens and cancer associated fibroblasts in the reactive stroma and its relation to Cancer biology. J Exp Clin Cancer Res. 2019;38(1):115.
33. Tan Y, Chen Q, Xing Y, Zhang C, Pan S, An W, Xu H. High expression of COL5A2, a member of COL5 family, indicates the poor survival and facilitates cell migration in gastric cancer. Biosci Rep. 2021;41(4):BSR20204293.
34. Ding YL, Sun SF, Zhao GL. COL5A2 as a potential clinical biomarker for gastric cancer and renal metastasis. Medicine (Baltimore). 2021;100(7):e24561.
35. Shen H, Wang L, Chen Q, Xu J, Zhang J, Fang L, Wang J, Fan W. The prognostic value of COL3A1/FBN1/COL5A2/SPARC-mir-29a-3p-H19 associated ceRNA network in gastric cancer through bioinformatic exploration. J Cancer. 2020;11(17):4933–46.
36. Mao D, Xu R, Chen H, Chen X, Li D, Song S, He Y, Wei Z, Zhang C. Crosstalk of focal adhesion-related gene defines prognosis and the immune microenvironment in gastric cancer. Front Cell Dev Biol. 2021;9:716461.
37. Cao L, Chen Y, Zhang M, Xu DQ, Liu Y, Liu T, Liu SX, Wang P. Identification of hub genes and potential molecular mechanisms in gastric cancer by integrated bioinformatics analysis. PeerJ. 2018;6:e5180.
38. Ma HP, Chang HL, Bamodu OA, Yadav VK, Huang TY, Wu ATH, Yeh CT, Tsai SH, Lee WH. Collagen 1A1 (COL1A1) is a reliable biomarker and putative therapeutic target for hepatocellular carcinogenesis and metastasis. Cancers (Basel). 2019;11(6):786.
39. 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. J Enzyme Inhib Med Chem. 2019;34(1):420–8.
40. Zhang SW, Zhang N, Wang N. Role of COL3A1 and POSTN on pathologic stages of esophageal cancer. Technol Cancer Res Treat. 2020;19:1533033820977489.
41. Gao ZW, Liu C, Yang L, He T, Wu XN, Zhang HZ, Dong K. SPARC overexpression promotes liver cancer cell proliferation and tumor growth. Front Mol Biosci. 2021;8:775743.
42. Wu Y, Xu Y. Integrated bioinformatics analysis of expression and gene regulation network of COL12A1 in colorectal cancer. Cancer Med. 2020;9(13):4743–55.

43. Zhang J, Zhang J, Wang F, Xu X, Li X, Guan W, Men T, Xu G. Overexpressed COL5A1 is correlated with tumor progression, paclitaxel resistance, and tumor-infiltrating immune cells in ovarian cancer. J Cell Physiol. 2021;236(10):6907–19.
44. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. EMBO Rep. 2014;15(12):1243–53.
45. Rosen GD, Dube DS. ADHESION, CELL–MATRIX | focal contacts and signaling. In: Laurent GJ, Shapiro SD, editors. Encyclopedia of respiratory medicine. Oxford: Academic Press; 2006. p. 41–7.
46. Paluch EK, Aspalter IM, Sixt M. Focal adhesion-independent cell migration. Annu Rev Cell Dev Biol. 2016;32(1):469–90.
47. Liu JY, Jiang L, He T, Liu JJ, Fan JY, Xu XH, Tang B, Shi Y, Zhao YL, Qian F, et al. NETO2 promotes invasion and metastasis of gastric cancer cells via activation of PI3K/Akt/NF-κB/Snail axis and predicts outcome of the patients. Cell Death Dis. 2019;10(3):162.
48. Hu M, Zhu S, Xiong S, Xue X, Zhou X. MicroRNAs and the PTEN/PI3K/Akt pathway in gastric cancer (Review). Oncol Rep. 2019;41(3):1439–54.
49. Tan MSY, Sandanaraj E, Chong YK, Lim SW, Koh LWH, Ng WH, Tan NS, Tan P, Ang BT, Tang C. A STAT3-based gene signature stratifies glioma patients for targeted therapy. Nat Commun. 2019;10(1):3601.
50. Pan YM, Wang CG, Zhu M, Xing R, Cui JT, Li WM, Yu DD, Wang SB, Zhu W, Ye YJ, et al. STAT3 signaling drives EZH2 transcriptional activation and mediates poor prognosis in gastric cancer. Mol Cancer. 2016;15(1):79.
51. Meurette O, Mehlen P. Notch signaling in the tumor microenvironment. Cancer Cell. 2018;34(4):536–48.
52. Cui Y, Li Q, Li W, Wang Y, Lv F, Shi X, Tang Z, Shen Z, Hou Y, Zhang H, et al. NOTCH3 is a prognostic factor and is correlated with immune tolerance in gastric cancer. Front Oncol. 2020;10:574937.
53. Kobayashi H, Enomoto A, Woods SL, Burt AD, Takahashi M, Worthley DL. Cancer-associated fibroblasts in gastrointestinal cancer. Nat Rev Gastroenterol Hepatol. 2019;16(5):282–95.
54. Russi S, Verma HK, Laurino S, Mazzone P, Storto G, Nardelli A, Zoppoli P, Calice G, La Rocca F, Sgambato A, et al. Adapting and surviving: intra and extra-cellular remodeling in drug-resistant gastric cancer cells. Int J Mol Sci. 2019;20(15):3736.
55. Hsieh CH, Tai SK, Yang MH. Snail-overexpressing cancer cells promote M2-like polarization of tumor-associated macrophages by delivering MiR-21-abundant exosomes. Neoplasia. 2018;20(8):775–88.
56. Yin W, Zhu H, Tan J, Xin Z, Zhou Q, Cao Y, Wu Z, Wang L, Zhao M, Jiang X, et al. Identification of collagen genes related to immune infiltration and epithelial-mesenchymal transition in glioma. Cancer Cell Int. 2021;21(1):276.
57. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. Nat Immunol. 2013;14(10):1014–22.
58. Salmon H, Franciszkiewicz K, Damotte D, Dieu-Nosjean MC, Validire P, Trautmann A, Mami-Chouaib F, Donnadieu E. Matrix architecture defines the preferential localization and migration of T cells into the stroma of human lung tumors. J Clin Invest. 2012;122(3):899–910.
59. Andrikopoulos K, Liu X, Keene DR, Jaenisch R, Ramirez F. Targeted mutation in the col5a2 gene reveals a regulatory role for type V collagen during matrix assembly. Nat Genet. 1995;9(1):31–6.
60. Wu T, Dai Y. Tumor microenvironment and therapeutic response. Cancer Lett. 2017;387:61–8.
61. DeNardo DG, Ruffell B. Macrophages as regulators of tumour immunity and immunotherapy. Nat Rev Immunol. 2019;19(6):369–82.
62. Chen Y, Zhang S, Wang Q, Zhang X. Tumor-recruited M2 macrophages promote gastric and breast cancer metastasis via M2 macrophage-secreted CHI3L1 protein. J Hematol Oncol. 2017;10(1):36.
63. Workman CJ, Szymczak-Workman AL, Collison LW, Pillai MR, Vignali DA. The development and function of regulatory T cells. Cell Mol Life Sci. 2009;66(16):2603–22.
64. Chaudhary B, Elkord E. Regulatory T cells in the tumor microenvironment and cancer progression: role and therapeutic targeting. Vaccines (Basel). 2016;4(3):28.
65. Kamphorst AO, Wieland A, Nasti T, Yang S, Zhang R, Barber DL, Konieczny BT, Daugherty CZ, Koenig L, Yu K, et al. Rescue of exhausted CD8 T cells by PD-1-targeted therapies is CD28-dependent. Science. 2017;355(6332):1423–7.
66. Acharya N, Sabatos-Peyton C, Anderson AC. Tim-3 finds its place in the cancer immunotherapy landscape. J Immunother Cancer. 2020;8(1):e000911.
67. Pauken KE, Wherry EJ. Overcoming T cell exhaustion in infection and cancer. Trends Immunol. 2015;36(4):265–76.
68. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer. 2012;12(4):252–64.
69. Beatty GL, Gladney WL. Immune escape mechanisms as a guide for cancer immunotherapy. Clin Cancer Res. 2015;21(4):687–92.
70. Kaderbhaï C, Tharin Z, Ghiringhelli F. The role of molecular profiling to predict the response to immune checkpoint inhibitors in lung cancer. Cancers (Basel). 2019;11(2):201.
71. Chen M, Zhang L. COL5A2 is a prognostic-related biomarker and correlated with immune infiltrates in gastric cancer. Research Square. 2022. [Preprint]. Cited 2023 Jan 21.
72. Chen M, Zhu X, Zhang L, Zhao D. Metadata record for the manuscript: COL5A2 is a prognostic-related biomarker and correlated with immune infiltrates in gastric cancer based on transcriptomics and single-cell RNA sequencing. 2023. https://doi.org/10.6084/m9.figshare.23647089.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC