***COL6A2/3* can serve as potential therapeutic targets and prognostic biomarkers for clear cell renal cell carcinoma**

Wingkeung Yiu

First Affiliated Hospital of Jinan University

Can-Xuan Li (✉ china20170713@163.com)

First Affiliated Hospital of Jinan University

Jie Chen

First Affiliated Hospital of Jinan University

Research Article

Keywords: COL6As, clear cell renal cell carcinoma, biomarkers, immune cell infiltrates.

Posted Date: December 10th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-119445/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Growing evidence has shown that the type VI collagen alpha chain (*COL6A*) family involved in the tumorigenesis and progression of diverse malignancies; however, its biological roles and potential mechanisms in clear cell renal cell carcinoma (ccRCC) remain unknown. The study was designed to explore the potential mechanisms and functions of *COL6As* in ccRCC.

Methods: ONCOMINE and GEPIA databases were used to compare the transcriptional expression data of *COL6As* in ccRCC samples and normal renal samples. UALCAN database was utilized to determine the association between clinicopathological features and *COL6As* expression. Kaplan–Meier method was employed to determine the prognostic value of *COL6As* mRNA expression in ccRCC. CBioPortal database was used to investigate the genetic alterations of *COL6As* in ccRCC. Co-expression analyses, functional enrichment analyses, and gene set enrichment analysis (GSEA) were utilized to explore the potential action mechanisms of *COL6As* in ccRCC. Finally, we estimated the relationship between *COL6As* expression with immune cell infiltrates.

Results: Upregulated transcriptional *COL6A2/COL6A3* expression was observed in ccRCC specimens by comparison with noncancerous renal specimens. Patients with increased *COL6A2/COL6A3* mRNA expression have a poor clinical outcome and unfavorable prognosis. Gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and GSEA analyses showed that *COL6A2/COL6A3* might promote the tumorigenesis and progression of ccRCC by involving in several cancer-related pathways, such as axon guidance, focal adhesion, ECM receptor interaction. Besides, we found that *COL6A2/COL6A3* expression was significantly associated with immune infiltration levels in ccRCC.

Conclusions: *COL6A2* and *COL6A3* could act as candidate prognostic biomarkers and therapeutic targets in ccRCC. However, further experimental work was required to validate the conclusions.

Introduction

Clear cell renal cell carcinoma (ccRCC) is one of the most common urological malignancies worldwide^[1]. Over 30% of the patients are diagnosed with advanced disease because of the lack of specific clinical symptoms in the early stage^[2]. However, the crucial question worried us is that these patients have missed the opportunity for radical surgery and responded poorly to routine chemoradiotherapy^[3, 4]. Furthermore, ccRCC has been clearly considered as a heterogeneous disease, leading to distinct tumor biologic traits, treatment response, and clinical outcomes^[5]. Although improvement has been achieved in the clinical treatment of ccRCC, patients with recurrence and metastasis still show poor survival rates^[3, 6]. Therefore, it is particularly urgent to find novel diagnostic and prognostic biomarkers and targets for ccRCC patients.

The type VI collagen alpha chain (*COL6A*) family is a major structural component of microfibrils that contain six family members, including alpha1(VI) to alpha6(VI), also known as *COL6A1* to *COL6A6*. *COL6A4* is a pseudogene that does not encode protein among them. All other members encode primary extracellular matrix proteins that could prohibit cell invasion and metastasis^[7]. It is thus not surprising that aberrantly deregulated *COL6As* contribute to the tumorigenesis and progression of human tumors.

Accumulating evidence has revealed that *COL6As* were involved in the occurrence and development of various cancers; for example, *COL6A1* was highly expressed in pancreatic cancer samples than in adjacent control tissues and upregulation of *COL6A1* indicates a poor clinical outcome of predicts unfavorable prognosis in patients with

pancreatic cancer^[8]. Evidence derived from bioinformatics analyses showed that collagen members *COL6A1*, *COL6A2*, and *COL6A3* were increased in bladder cancer and negatively associated with patient prognosis^[9]. High *COL6A3* expression in gastric cancer promotes proliferation, migration, invasion and prohibits apoptosis of tumor cells through the PI3K-Akt signaling pathway^[10]. A recent research showed that *COL6A5* rs13062453 and rs1497305 were associated with the susceptibility of lung adenocarcinoma^[11]. Besides, a previous study showed that *COL6A6* was upregulated in pituitary adenoma tissues than in normal tissues^[12].

However, the associated with the *COL6A* gene family and ccRCC have rarely been recorded. A systematic functional study of *COL6As* will help us clearly understand their roles in malignancies. The study was aimed to probe into the roles of *COL6As* in ccRCC. In the current study, several available databases were employed to perform the expression patterns, clinicopathological analysis, prognosis value assessment, gene alterations exploration, functional enrichment annotate, gene set enrichment analysis and immune infiltration of *COL6As* in ccRCC. In conclusion, our study indicates that *COL6A2* and *COL6A3* could serve as potential biomarkers for diagnosis and prognosis of ccRCC patients.

Materials And Methods

ONCOMINE database

The gene expression array datasets of ONCOMINE (<https://www.oncomine.org>) are an online cancer microarray database^[13]. In this study, the ONCOMINE database was used to analyze mRNA levels of *COL6As* in different malignancies. The mRNA expressions of *COL6As* in diverse cancer samples were compared with that in normal controls, using a students' t-test to obtain a P value. The cut-off of P value and fold change were defined as 0.001 and 2, respectively.

GEPIA Dataset

The Gene Expression Profiling Analysis Interaction Analysis (GEPIA) (<http://gepia.cancer-pku.cn>) is an interactive web that includes 9,736 tumors and 8,587 normal samples from TCGA and the Genotype-Tissue Expression (GTEx) projects^[14].

GEPIA provides customizable functions such as tumor/normal differential expression analysis, profiling according to cancer types or pathological stages, patient survival analysis, similar gene detection, correlation analysis and dimensionality reduction analysis. In the study, GEPIA was used to validate *COL6As* mRNA expression in ccRCC specimens. Furthermore, GEPIA database was employed to assess the association between *COL6As* mRNA expression and the survival rate of ccRCC patients. The 516 ccRCC patients with available overall survival (OS) and disease-free survival (DFS) time data were split into low- and high- expression groups by the median transcripts per kilobase million (TPM), a logarithmic rank $P < 0.05$ was considered significant.

UALCAN database

UALCAN (<http://ualcan.path.uab.edu>) includes TCGA grade 3 RNA-seq and clinical data of 31 human cancer types, which the platform can be used to analyze the relationship between gene expression and clinicopathological features^[15]. In our study, UALCAN was utilized to compare differential mRNA levels of *COL6As* between paired AJCC stages or ISUP grades.

cBioPortal database

The cBioPortal for cancer genomics (<http://www.cbioportal.org/>) is an open-source database, which can provide multidimensional cancer genomics and clinical data retrieved from the TCGA database^[16]. Here, three kidney renal clear cell carcinoma (KIRC) dataset, including data from 1,485 cases with pathology reports, was selected for further analyses of *COL6As* using cBioPortal database. The cBioPortal Cancer Genomics resource was used to assess the alteration frequency of the *COL6As* in ccRCC cases.

STRING and GeneMANIA database

The Search Tool for the Retrieval of Interacting Genes (<http://string-db.org/>)^[17] online database was utilized to identify ten co-regulated proteins; an interaction with a combined score > 0.4 was regarded as statistically significant. The GeneMANIA (<http://genemania.org/>)^[18] online database was utilized to identify co-regulated partners at the gene level. Furthermore, a protein-protein and gene–gene interaction networks for the *COL6As* and their functional partners were constructed using the STRING and GeneMANIA database.

Correlation and Functional Enrichment analyses

LinkedOmics (www.linkedomics.org/login.php) is a publicly available portal that provides a unique platform for biologists and clinicians to access, analyze and compare cancer multi-omics data within and across tumor types^[19]. The *COL6As* co-expression analysis was statistically using Pearson's correlation coefficient through the LinkedOmics database. Correlated genes with a Pearson's correlation coefficient $\geq |0.5|$ and FDR < 0.05 were selected to perform functional enrichment analysis. Metascape (<http://metascape.org>) is a web-based portal designed to perform functional enrichment, interactome analysis, gene annotation, and membership search^[20]. In the current study, Metascape was used to analyze the enrichment of Gene ontology (GO) for *COL6As* and its correlated genes, gene terms with P-value < 0.01, minimum count of 3, and enrichment factor of > 1.5 were regarded significantly. Additionally, the clusterProfiler package, which could automate biological terminology classification and gene cluster enrichment analysis processes^[21], was utilized to analyze the Kyoto Encyclopedia of Genes and Genomes (KEGG) for *COL6As* and its correlated genes, signal pathway enrichment analysis was conducted with the criterion both P and FDR < 0.05.

Gene set enrichment analysis

To better understand the potential functions of *COL6As* in ccRCC, we conducted GSEA using transcriptional sequences from the TCGA database. In this study, gene expression profiles of 530 ccRCC patients were sorted into two subgroups (high-expression subgroup and low-expression subgroup) based on the median value of expression of *COL6As*. The “c2.cp.kegg.v7.2.symbols.gmt” gene sets from the Molecular Signatures Database (MSigDB) were analyzed using GSEA 3.0 software. Enrichment analysis was carried out using the default weighted enrichment statistics method. The number permutations of was set to 1,000 times for each analysis. Normalized enrichment scores (NES) were utilized to identify the significance of signaling pathways enriched in the high KIF4A expression subgroup. Gene sets with both a P-value of < 0.05 and a false discovery rate (FDR) of < 0.05 were considered as significantly enriched gene sets.

TIMER database

The Tumor Immunity Resource (TIMER) (<https://cistrome.shinyapps.io/timer/>) database contains mRNA sequencing data for 10,897 samples of 32 human cancers from the Cancer Genome Atlas (TCGA)^[22]. It is an ideal tool for systematic analysis of immune infiltration in various cancer types. In this study, we assessed the *COL6As* expression with the abundance of different immune infiltrating cells.

Results

Increased COL6A2 and COL6A3 expression in ccRCC

Using ONCOMINE database, we examined the differences of *COL6As* expression between ccRCC samples and normal renal samples. Results showed that *COL6A1*, *COL6A2* and *COL6A3* mRNA expression was highly expressed in neoplastic samples by comparison with normal renal specimens in the datasets (Fig. 1A). As indicated in Fig. 1B-C, in Beroukhi Renal dataset, *COL6A2* mRNA expression was significantly upregulated in ccRCC tissues compared with the noncancerous tissues (Hereditary ccRCC vs. Normal, $p = 4.90E-11$, fold change = 3.482; Non-Hereditary ccRCC vs. Normal, $p = 6.46E-8$, fold change = 3.209), in Gumz Renal dataset, *COL6A2* mRNA expression was highly expressed in ccRCC tissues than in noncancerous tissues (ccRCC vs. Normal, $p = 1.93E-5$, fold change = 3.356). In addition, increased *COL6A3* mRNA expression was observed in ccRCC specimens based on data from Gumz Renal dataset (ccRCC vs. Normal, $p = 3.06E-4$, fold change = 4.420) (Fig. 1D). GEPIA database was used to further validate *COL6As* mRNA expression in ccRCC. Elevated transcription levels of *COL6A2* and *COL6A3* were found in ccRCC tissues compared to adjacent normal tissues based on the RNA-seq data from TCGA and the Genotype-Tissue Expression (GTEx) projects (Fig. 2).

Elevated COL6A2 and COL6A3 expression significantly associated with advanced clinicopathological features and unfavorable prognosis in ccRCC patients

Using the UALCAN database, we found that *COL6A2/COL6A3* expression was significantly associated with clinicopathological parameters of patients with ccRCC. As can be seen from Fig. 3A-D, patients who were in more advanced stages or more advanced grade scores tended to express higher mRNA expression of *COL6A2/COL6A3*. Given the relationship between *COL6A2/COL6A3* expression and clinicopathological features, survival analysis in Kaplan–Meier method was conducted further to evaluate the prognostic value of *COL6A2/COL6A3* in ccRCC patients. Results revealed that *COL6A2/COL6A3* expression was remarkably correlated with the overall survival (OS) and disease-free survival (DFS). As indicated in Fig. 4A-D, patients with the upregulated *COL6A2/COL6A3* mRNA expression were likely to be predicted as having weak OS and DFS.

COL6A2/COL6A3 genetic alteration analysis and co-regulated network construction

Alteration frequency of *COL6A2/COL6A3* mutations in ccRCC was analyzed using cBioPortal database. A total of 1,485 patients from three datasets of KIRC were analyzed. The analysis of genetic variations in *COL6A2/COL6A3* reported in different datasets was showed in Fig. 5A, genetic alterations were 4.06 4.31% and 4.9%, respectively. The variations frequency included amplification (red), deep deletions (blue), mutation (green). Additionally, overview of the analyses of genetic variations in *COL6A2/COL6A3* were showed in Fig. 5B, *COL6A2/COL6A3* were altered in 66/1,485 (4.5%) of ccRCC patient samples. Furthermore, in our study, protein-protein network and gene-gene network were constructed using the STRING and GeneMANIA database, respectively (Fig. 5C-5D).

Correlation and Functional Enrichment analyses of COL6A2/COL6A3

Genes co-expressed with *COL6A2/COL6A3* were analyzed using the linkedomics database. As shown in Fig. 6A and Fig. 6D, the dark red dots show a significantly positive correlation with *COL6A2/3*, whereas dark green dots reveal a significantly negative correlation with *COL6A2/3*. The top 50 significant genes positively and negatively correlated with *COL6A2/3*, as shown in the heat maps (Figs. 6B, C, E, F). According to the criteria: Pearson's correlation coefficient $\geq |0.5|$ and $FDR < 0.05$.

We found that 580 genes were positively associated with *COL6A2*. Conversely, 137 genes were negatively correlated with *COL6A2*. Meanwhile, we found that *COL6A2* expression showed a strong correlation with the expression of

COL6A1 (Pearson's correlation = 8.730e-01, FDR = 1.047e-163), *FHL3* (Pearson's correlation = 8.242e-01, FDR = 2.180e-129), *ECM1* (Pearson's correlation = 8.106e-01, FDR = 7.979e-122), *MEIS3* (Pearson's correlation = 8.100e-01, FDR = 1.500e-121) and *COL5A1* (Pearson's correlation = 8.018e-01, FDR = 2.840e-117)(Fig. 6G-6K). Furthermore, there were 417 genes having a remarkably positive connection with *COL6A3*, whereas 23 genes having an obviously negative correlation with *COL6A3*. Additionally, *COL6A3* expression showed a strong association with the expression of *COL5A1* (Pearson's correlation = 8.917e-01, FDR = 6.158e-181), *COL3A1* (Pearson's correlation = 8.794e-01, FDR = 1.960e-169), *COL1A2* (Pearson's correlation = 8.471e-01, FDR = 3.363e-144), *COL1A1* (Pearson's correlation = 8.462e-01, FDR = 1.056e-143) and *FAP* (Pearson's correlation = 8.248e-01, FDR = 4.414e-130)(Fig. 6L-6P). Furthermore, GO and KEGG analyses of *COL6A2/COL6A3* and their co-expressed genes in KIRC was carried out to explore the potential mechanisms and biological functions of *COL6A2/COL6A3* in ccRCC. As illustrated in Fig. 7, functional enrichment analyses of *COL6A2* and its co-expressed genes were mainly participated in extracellular matrix, collagen fibril organization, collagen metabolic process, supramolecular fiber organization and adherens junction. In addition, we found that extracellular matrix organization, collagen fibril organization, blood vessel development, skeletal system development and collagen binding were significantly regulated by *COL6A3* and its co-expressed genes in ccRCC (Fig. 8). Furthermore, the top 20 KEGG pathways for *COL6A2/COL6A3* and their correlated genes were showed in Fig. 9A-9B, respectively. Results showed that several signal pathways contributed to the occurrence and progression of malignancies, including Focal adhesion, ECM – receptor interaction, Proteoglycans in cancer, Relaxin signaling pathway and Axon guidance were involved in the tumorigenesis and progression of ccRCC.

Identification of COL6A2/COL6A3-related signaling pathways by GSEA

To better illustrate the biological mechanisms and potential functions of *COL6A2/COL6A3* in ccRCC. GSEA was conducted to identify the signaling pathways enriched in the high *COL6A2/COL6A3* expression ccRCC group. As shown in Fig. 10A-10B, genes in high expression groups of *COL6A2/COL6A3* were mainly enriched in axon guidance, focal adhesion, ECM receptor interaction, and dilated cardiomyopathy. Meanwhile, vascular smooth muscle contraction gene sets were enriched in the high *COL6A2*-expression group. Whereas arrhythmogenic right ventricular cardiomyopathy arvc gene sets were enriched in the high *COL6A3*-expression group. The details are shown in **Table 1** and **Table 2**.

Table 1. Gene sets enriched in the high *COL6A2* expression phenotype.

Gene set	NES	NOM p-val	FDR q-val
KEGG_AXON_GUIDANCE	2.259	0.000	0.001
KEGG_FOCAL_ADHESION	2.241	0.000	0.002
KEGG_DILATED_CARDIOMYOPATHY	2.215	0.000	0.002
KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	2.110	0.000	0.009
KEGG_ECM_RECEPTOR_INTERACTION	2.110	0.000	0.007
KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM	2.084	0.000	0.007
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	2.075	0.000	0.007
KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC	2.074	0.000	0.006
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE	2.045	0.000	0.008
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION	1.972	0.006	0.015
KEGG_GAP_JUNCTION	1.963	0.000	0.015
KEGG_DORSO_VENTRAL_AXIS_FORMATION	1.931	0.000	0.019
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_KERATAN_SULFATE	1.930	0.002	0.017
KEGG_NOTCH_SIGNALING_PATHWAY	1.905	0.002	0.021
KEGG_CALCIUM_SIGNALING_PATHWAY	1.859	0.000	0.032
KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION	1.850	0.004	0.032
KEGG_SMALL_CELL_LUNG_CANCER	1.810	0.000	0.044
KEGG_MAPK_SIGNALING_PATHWAY	1.807	0.000	0.042

COL6A2, collagen type VI alpha 2 chain; NES, normalized enrichment score; NOM, nominal; FDR, false discovery rate.

Table 2. Gene sets enriched in the high *COL6A3* expression phenotype.

Gene set	NES	NOM p-val	FDR q-val
KEGG_FOCAL_ADHESION	2.392	0.000	0.000
KEGG_ECM_RECEPTOR_INTERACTION	2.312	0.000	0.000
KEGG_AXON_GUIDANCE	2.169	0.000	0.002
KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC	2.147	0.000	0.002
KEGG_DILATED_CARDIOMYOPATHY	2.135	0.000	0.002
KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM	2.135	0.000	0.002
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	2.125	0.000	0.002
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE	2.098	0.000	0.003
KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION	2.096	0.002	0.002
KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	2.084	0.000	0.003
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION	2.075	0.000	0.003
KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS	1.987	0.000	0.008
KEGG_VEGF_SIGNALING_PATHWAY	1.952	0.000	0.012
KEGG_TGF_BETA_SIGNALING_PATHWAY	1.945	0.002	0.012
KEGG_RENAL_CELL_CARCINOMA	1.937	0.000	0.012
KEGG_PATHWAYS_IN_CANCER	1.899	0.000	0.019
KEGG_CALCIUM_SIGNALING_PATHWAY	1.888	0.000	0.020
KEGG_WNT_SIGNALING_PATHWAY	1.876	0.000	0.022
KEGG_GAP_JUNCTION	1.876	0.002	0.021
KEGG_MAPK_SIGNALING_PATHWAY	1.860	0.000	0.023
KEGG_TIGHT_JUNCTION	1.847	0.000	0.026
KEGG_COLORECTAL_CANCER	1.843	0.004	0.026
KEGG_NEUROTROPHIN_SIGNALING_PATHWAY	1.834	0.000	0.027
KEGG_ADHERENS_JUNCTION	1.831	0.000	0.026
KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY	1.808	0.012	0.032
KEGG_SMALL_CELL_LUNG_CANCER	1.800	0.000	0.034
KEGG_CHRONIC_MYELOID_LEUKEMIA	1.780	0.002	0.038
KEGG_JAK_STAT_SIGNALING_PATHWAY	1.780	0.004	0.037
KEGG_LONG_TERM_DEPRESSION	1.769	0.002	0.040
KEGG_NOTCH_SIGNALING_PATHWAY	1.762	0.008	0.040
KEGG_AMYOTROPHIC_LATERAL_SCLEROSIS_ALS	1.760	0.000	0.040
KEGG_LONG_TERM_POTENTIATION	1.758	0.002	0.039
KEGG_CELL_CYCLE	1.746	0.022	0.042
KEGG_DORSO_VENTRAL_AXIS_FORMATION	1.742	0.006	0.043
KEGG_CHEMOKINE_SIGNALING_PATHWAY	1.737	0.024	0.043
KEGG_PANCREATIC_CANCER	1.722	0.004	0.047
KEGG_PROGESTERONE_MEDIATED_OOCYTE_MATURATION	1.716	0.000	0.048
KEGG_GLIOMA	1.714	0.004	0.047
KEGG_PROSTATE_CANCER	1.707	0.004	0.049

COL6A3, collagen type VI alpha 3 chain; NES, normalized enrichment score; NOM, nominal; FDR, false discovery rate.

COL6A2/COL6A3 expression significantly correlated with immune infiltrates in ccRCC

Accumulating evidence demonstrated that tumor-infiltrating lymphocytes could utilize to predict sentinel lymph node status and survival in malignancies^[23]. Hence, we further analyzed the correlation between *COL6A2/COL6A3* expression with tumor purity and tumor-infiltrating immune cells. As illustrated in Fig. 11A-11B, elevated *COL6A2* was negatively correlated with tumor purity ($r=-0.28$, $p = 8.73e-10$) in ccRCC and significantly associated with infiltrating levels of CD8 + T cells ($r = 0.105$, $p = 2.85e-02$), CD4 + T ($r = 0.43$, $p = 4.39e-22$), macrophages ($r = 0.117$, $p = 1.31e-02$),

neutrophils ($r = 0.147$, $p = 1.66e-03$), and dendritic cells ($r = 0.132$, $p = 4.69e-03$). Furthermore, increased *COL6A3* was negatively correlated with tumor purity ($r = -0.303$, $p = 2.7e-11$) in ccRCC and significantly associated with infiltrating levels of CD8 + T cells ($r = 0.193$, $p = 4.58e-05$), CD4 + T ($r = 0.37$, $p = 2.07e-16$), macrophages ($r = 0.229$, $p = 9.80e-07$), neutrophils ($r = 0.258$, $p = 2.86e-08$), and dendritic cells ($r = 0.23$, $p = 7.45e-07$). However, this study did not find significant correlations between *COL6A2/COL6A3* expression and infiltrating levels of B cells in ccRCC. These findings suggested that *COL6A2/COL6A3* may play crucial role in the ccRCC microenvironment.

Discussion

Here, we analyzed the expression patterns and biological functions of six *COL6A* gene members in ccRCC. Our results show that *COL6A2/COL6A3* was significantly elevated in ccRCC. Further research suggests that upregulated *COL6A2/COL6A3* expression was markedly correlated with advanced clinical stage and higher tumor grade. Additionally, we found that patients with *COL6A2/COL6A3* overexpression have an unfavorable prognosis. All of this evidence indicates that *COL6A2/COL6A3* might serve as oncogene in ccRCC. Thus, we carry out functional explorations, including GO, KEGG and GSEA analyses, to explore the functional mechanism of *COL6A2/COL6A3* in ccRCC. Mechanically, *COL6A2* and *COL6A3* were both found to involve in several tumorigenesis and tumor progression pathways, such as Focal adhesion, ECM – receptor interaction, Proteoglycans in cancer, Relaxin signaling pathway and Axon guidance. These pathways have been proven to associate with the occurrence and progression of ccRCC and several other types of cancer. Previous studies have shown that Focal adhesion played a vital role in tumor development, growth, infiltration and spread^[24], and many studies have shown activation of Focal adhesion-related genes in ccRCC^[25–27]. The extracellular matrix (ECM) plays crucial role in the maintenance of normal structure and function of cells^[28]. Deregulation ECM – receptor interaction contributes to the out of control of numerous cellular activities such as adhesion, migration, differentiation, proliferation, and apoptosis, and eventually affecting tumor progress^[29]. proteoglycans have been demonstrated to be critical molecules in the tumor microenvironment, which contributing to the biological behavior of several malignancies, including proliferation, adhesion, angiogenesis and metastasis, resulting in tumor development and progression^[30, 31]. Several proteoglycans have been proven to promote tumor progression and metastasis in ccRCC^[32, 33]. Previous reports have indicated that relaxin could affect cell growth, cell migration, angiogenesis and extracellular matrix remodeling, targeting Relaxin signaling pathway has gradually become a promising therapeutic way in cancer treatment^[34, 35]. Axon guidance is integral to several basic cellular processes, including organogenesis, regeneration, wound healing. A global genomes analysis has uncovered the potential involvement of axon guidance signaling pathway related genes in pancreatic carcinogenesis^[36]. In conclusion, Focal adhesion, ECM – receptor interaction, Proteoglycans in cancer, Relaxin signaling pathway and Axon guidance may be significantly associated with *COL6A2/COL6A3* expression in ccRCC.

Increasing studies have demonstrated that tumor-infiltrating immune cells (TICs) could interact with tumor cells and influence the clinical outcome and immunotherapy^[37, 38].

Recent studies have indicated that RCC tumors are more frequently infiltrated with lymphocytes than other malignancies^[39]. In this work, we first found the relationship between *COL6A2/COL6A3* expression and TICs in ccRCC. The overexpression *COL6A2/COL6A3* resulted in an increased infiltration levels of various immune cells. Among them, we found that the *COL6A2/COL6A3* expression level was remarkably positively associated with infiltrating levels of CD4 + T cells, recent research has shown that CD4 + T cells could promote RCC cell proliferation

by modulating the YBX1/HIF2 α signaling pathway^[40]. These findings indicate that *COL6A2/COL6A3* plays vital role in immune infiltrations in ccRCC, particularly for CD4 + T cells.

In this investigation, we systemically perform an analysis to investigate the roles of *COL6As* in ccRCC. The research shows that the high *COL6A2/COL6A3* expression might significantly affect the tumorigenesis and progression of ccRCC. The potential limitations of our study could not be ignored. The work was carried out based on several databases; clinical samples are required to validate the *COL6A2/3* transcription and protein expression. Furthermore, the biological functions of *COL6A2/COL6A3* and the relationship between *COL6A2/COL6A3* expression with immune infiltrates are unclear. Taken together, *COL6A2/COL6A3* could act as candidate prognostic biomarkers and potential therapeutic targets for ccRCC; however, further in-house experimental studies are needed to prove our conclusions.

Declarations

Ethical approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: The datasets used and analyzed in this study are available from the corresponding websites of the databases, including ONCOMINE (<https://www.oncomine.org>); GEPIA (<http://gepia.cancer-pku.cn>); UALCAN (<http://ualcan.path.uab.edu>); cBioPortal (<http://www.cbioportal.org/>); STRING (<http://string-db.org>); GeneMANIA (<http://genemania.org/>); LinkedOmics (www.linkedomics.org/login.php); Metascape (<http://metascape.org>); TIMER (<https://cistrome.shinyapps.io/timer/>); The results were often aggregated from multiple databases and analyzed online without the need to download the data. The data of the 530 ccRCC patients from TCGA database were downloaded from <https://portal.gdc.cancer.gov/>.

Competing interests: The authors declare that they have no competing interests, and all authors confirm its accuracy.

Funding: Not applicable.

Authors' contributions: (I) Conception and design: Can-Xuan Li; (II) Administrative support: Can-Xuan Li, wingkeung Yiu, Jie Chen; (III) Provision of study materials or patients: Can-Xuan Li; (IV) Collection and assembly of data: Can-Xuan Li, wingkeung Yiu, Jie Chen; (V) Data analysis and interpretation: Can-Xuan Li, wingkeung Yiu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Acknowledgments: Not applicable.

Authors' information: wingkeung Yiu^{1,#}, Can-Xuan Li^{1,&}, Jie Chen¹

¹ Department of Urology, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, 510630, China.

[#]First Author: wingkeung Yiu

Email: wingkeungyiujnu@163.com

[&]Corresponding Author: Can-Xuan Li

Email: china20170713@163.com

References

- [1] SIEGEL R L, MILLER K D, JEMAL A. Cancer statistics, 2019 [J]. CA: A Cancer Journal for Clinicians, 2019, 69(1): 7-34.
- [2] VERA-BADILLO F E, TEMPLETON A J, DURAN I, et al. Systemic therapy for non-clear cell renal cell carcinomas: a systematic review and meta-analysis [J]. European urology, 2015, 67(4): 740-9.
- [3] FERN NDEZ-PELLO S, HOFMANN F, TAHBAZ R, et al. A Systematic Review and Meta-analysis Comparing the Effectiveness and Adverse Effects of Different Systemic Treatments for Non-clear Cell Renal Cell Carcinoma [J]. European urology, 2017, 71(3): 426-36.
- [4] BARATA P C, RINI B I. Treatment of renal cell carcinoma: Current status and future directions [J]. CA: a cancer journal for clinicians, 2017, 67(6): 507-24.
- [5] CAPITANIO U, MONTORSI F. Renal cancer [J]. Lancet (London, England), 2016, 387(10021): 894-906.
- [6] HSIEH J J, PURDUE M P, SIGNORETTI S, et al. Renal cell carcinoma [J]. Nat Rev Dis Primers, 2017, 3(17009-).
- [7] LAMAND S R, BATEMAN J F. Collagen VI disorders: Insights on form and function in the extracellular matrix and beyond [J]. Matrix biology : journal of the International Society for Matrix Biology, 2018, 71-72(348-67).
- [8] OWUSU-ANSAH K G, SONG G, CHEN R, et al. COL6A1 promotes metastasis and predicts poor prognosis in patients with pancreatic cancer [J]. Int J Oncol, 2019, 55(2): 391-404.
- [9] ZHU H, CHEN H, WANG J, et al. Collagen stiffness promoted non-muscle-invasive bladder cancer progression to muscle-invasive bladder cancer [J]. OncoTargets and therapy, 2019, 12(3441-57).
- [10] AO R, GUAN L, WANG Y, et al. Silencing of COL1A2, COL6A3, and THBS2 inhibits gastric cancer cell proliferation, migration, and invasion while promoting apoptosis through the PI3k-Akt signaling pathway [J]. Journal of cellular biochemistry, 2018, 119(6): 4420-34.
- [11] DUAN Y, LIU G, SUN Y, et al. Collagen type VI $\alpha 5$ gene variations may predict the risk of lung cancer development in Chinese Han population [J]. 2020, 10(1): 5010.
- [12] MICHAELIS K A, KNOX A J, XU M, et al. Identification of growth arrest and DNA-damage-inducible gene beta (GADD45beta) as a novel tumor suppressor in pituitary gonadotrope tumors [J]. Endocrinology, 2011, 152(10): 3603-13.
- [13] RHODES D R, YU J, SHANKER K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform [J]. Neoplasia (New York, NY), 2004, 6(1): 1-6.
- [14] TANG Z, LI C, KANG B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses [J]. Nucleic acids research, 2017, 45(W1): W98-w102.
- [15] CHANDRASHEKAR D S, BASHEL B, BALASUBRAMANYA S A H, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses [J]. Neoplasia (New York, NY), 2017, 19(8): 649-58.
- [16] GAO J, AKSOY B A, DOGRUSOZ U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal [J]. Science signaling, 2013, 6(269): pl1.

- [17] SZKLARCZYK D, FRANCESCHINI A, WYDER S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life [J]. *Nucleic acids research*, 2015, 43(Database issue): D447-D52.
- [18] FRANZ M, RODRIGUEZ H, LOPES C, et al. GeneMANIA update 2018 [J]. *Nucleic acids research*, 2018, 46(W1): W60-w4.
- [19] VASAIKAR S V, STRAUB P, WANG J, et al. LinkedOmics: analyzing multi-omics data within and across 32 cancer types [J]. *Nucleic acids research*, 2018, 46(D1): D956-d63.
- [20] ZHOU Y, ZHOU B, PACHE L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets [J]. 2019, 10(1): 1523.
- [21] YU G, WANG L G, HAN Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters [J]. *OMICS*, 2012, 16(5): 284-7.
- [22] LI T, FAN J, WANG B, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells [J]. *Cancer research*, 2017, 77(21): e108-e10.
- [23] AZIMI F, SCOLYER R A, RUMCHEVA P, et al. Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma [J]. *J Clin Oncol*, 2012, 30(21): 2678-83.
- [24] EKE I, CORDES N. Focal adhesion signaling and therapy resistance in cancer [J]. *Seminars in cancer biology*, 2015, 31(65-75).
- [25] BRENNER W, GREBER I, GUDEJKO-THIEL J, et al. Migration of renal carcinoma cells is dependent on protein kinase Cdelta via beta1 integrin and focal adhesion kinase [J]. *Int J Oncol*, 2008, 32(5): 1125-31.
- [26] OGASAWARA N, KUDO T, SATO M, et al. Reduction of Membrane Protein CRIM1 Decreases E-Cadherin and Increases Claudin-1 and MMPs, Enhancing the Migration and Invasion of Renal Carcinoma Cells [J]. *Biological & pharmaceutical bulletin*, 2018, 41(4): 604-11.
- [27] WU P, LIU J L, PEI S M, et al. Integrated genomic analysis identifies clinically relevant subtypes of renal clear cell carcinoma [J]. *BMC cancer*, 2018, 18(1): 287.
- [28] NAJAFI M, FARHOOD B, MORTEZAEE K. Extracellular matrix (ECM) stiffness and degradation as cancer drivers [J]. 2019, 120(3): 2782-90.
- [29] ZHANG H J, TAO J, SHENG L, et al. Twist2 promotes kidney cancer cell proliferation and invasion by regulating ITGA6 and CD44 expression in the ECM-receptor interaction pathway [J]. *OncoTargets and therapy*, 2016, 9(1801-12).
- [30] IOZZO R V, SANDERSON R D. Proteoglycans in cancer biology, tumour microenvironment and angiogenesis [J]. *Journal of cellular and molecular medicine*, 2011, 15(5): 1013-31.
- [31] TZANAKAKIS G, NEAGU M, TSATSAKIS A, et al. Proteoglycans and Immunobiology of Cancer-Therapeutic Implications [J]. *Front Immunol*, 2019, 10(875).

- [32] MITSUI Y, SHIINA H, KATO T, et al. Versican Promotes Tumor Progression, Metastasis and Predicts Poor Prognosis in Renal Carcinoma [J]. Molecular cancer research : MCR, 2017, 15(7): 884-95.
- [33] MIKAMI S, MIZUNO R, KOSAKA T, et al. Expression of TNF- α and CD44 is implicated in poor prognosis, cancer cell invasion, metastasis and resistance to the sunitinib treatment in clear cell renal cell carcinomas [J]. Int J Cancer, 2015, 136(7): 1504-14.
- [34] NESCHADIM A, SUMMERLEE A J, SILVERTOWN J D. Targeting the relaxin hormonal pathway in prostate cancer [J]. Int J Cancer, 2015, 137(10): 2287-95.
- [35] MARDHIAN D F, STORM G, BANSAL R, et al. Nano-targeted relaxin impairs fibrosis and tumor growth in pancreatic cancer and improves the efficacy of gemcitabine in vivo [J]. Journal of controlled release : official journal of the Controlled Release Society, 2018, 290(1-10).
- [36] BIANKIN A V, WADDELL N, KASSAHN K S, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes [J]. Nature, 2012, 491(7424): 399-405.
- [37] GAJEWSKI T F, SCHREIBER H, FU Y X. Innate and adaptive immune cells in the tumor microenvironment [J]. Nature immunology, 2013, 14(10): 1014-22.
- [38] ALLARD B, ASPESLAGH S, GARAUD S, et al. Immuno-oncology-101: overview of major concepts and translational perspectives [J]. Seminars in cancer biology, 2018, 52(Pt 2): 1-11.
- [39] ZHANG S, ZHANG E, LONG J, et al. Immune infiltration in renal cell carcinoma [J]. Cancer science, 2019, 110(5): 1564-72.
- [40] WANG Y, WANG Y, XU L, et al. CD4 + T cells promote renal cell carcinoma proliferation via modulating YBX1 [J]. Exp Cell Res, 2018, 363(1): 95-101.

Figures

Figure 1

COL6As expression levels in different types of human cancers. (A) Increased or decreased COL6As in data sets of different cancers compared with normal tissues in the ONCOMINE database. (B-C) In Gumz Renal and Beroukhim Renal datasets, the COL6A2 mRNA level was found significantly upregulated in ccRCC. (D) The COL6A3 mRNA level was found remarkably elevated in ccRCC based on data from the Gumz Renal dataset. COL6As, the type VI collagen alpha chains; COL6A1, collagen type VI alpha 1 chain; COL6A2, collagen type VI alpha 2 chain; COL6A3, collagen type VI alpha 3 chain; COL6A5, collagen type VI alpha 5 chain; COL6A6, collagen type VI alpha 6 chain ccRCC, clear cell renal cell carcinoma.

Figure 2

The COL6As expression patterns in ccRCC (GEPIA). COL6As, the type VI collagen alpha chains; COL6A1, collagen type VI alpha 1 chain; COL6A2, collagen type VI alpha 2 chain; COL6A3, collagen type VI alpha 3 chain; COL6A5, collagen type VI alpha 5 chain; COL6A6, collagen type VI alpha 6 chain ccRCC, clear cell renal cell carcinoma; GEPIA, Gene Expression Profiling Interactive Analysis.

Figure 3

The association between COL6A2/3 expression with clinicopathological variables in ccRCC patients from the UALCAN database. (A-B) high COL6A2 mRNA level expression was remarkably correlated with advance stages and grade scores; (C-D) Elevated COL6A3 expression was significantly associated with advance stages and grade scores. *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ****, $p \leq 0.0001$. COL6A2, collagen type VI alpha 2 chain; COL6A3, collagen type VI alpha 3 chain; ccRCC, clear cell renal cell carcinoma.

Figure 4

Survival analysis of COL6A2/3 in ccRCC patients. (A) The relationship between COL6A2 expression and OS; (B) The relationship between COL6A2 expression and DFS; (C) The association between COL6A3 expression with OS; (D) The association between COL6A3 expression with DFS; COL6A2, collagen type VI alpha 2 chain; COL6A3, collagen type VI alpha 3 chain; ccRCC, clear cell renal cell carcinoma; OS, overall survival; DFS, disease-free survival

Figure 5

Genetic mutation analyses of COL6A2/COL6A3 in ccRCC and COL6A2/COL6A3 related-interaction network construction. (A-B) COL6A2/COL6A3 expression and mutation analysis in ccRCC (cBioPortal). (C) The protein-protein interaction network for COL6A2/COL6A3 (STRING); (D) The gene-gene interaction network for COL6A2/COL6A3 (GeneMANIA). COL6A2, collagen type VI alpha 2 chain; COL6A3, collagen type VI alpha 3 chain; ccRCC, clear cell renal cell carcinoma.

Figure 6

Differentially expressed genes correlated with differentially expressed COL6A2/3 in ccRCC (LinkedOmics). (A-C) Volcano plots and heat maps indicating genes positively and negatively genes correlated with COL6A2 in ccRCC, respectively (top 50); (D-F) Volcano plots and heat maps indicating genes positively and negatively genes correlated with COL6A3 in ccRCC, respectively (top 50); (G-K) The scatter plots show Pearson's correlation of COL6A2 expression with expression of COL6A1, FHL3, ECM1, MEIS3 and COL5A1; (L-P) The scatter plots show Pearson's correlation of COL6A3 expression with expression of COL5A1, COL3A1, COL1A2, COL1A1 and FAP.

Figure 7

Functional enrichment analyses of COL6A2 and its co-expressed genes (Metascape). (A) Heatmap of the GO enriched terms colored by P-values. (B) Network of GO enriched terms colored by P-value. (C) Network of GO enriched terms colored by clusters. COL6A2, collagen type VI alpha 2 chain; GO, Gene Ontology.

Figure 8

Functional enrichment analyses of COL6A3 and its co-expressed genes (Metascape). (A) Heatmap of the GO enriched terms colored by P-values. (B) Network of GO enriched terms colored by P-value. (C) Network of GO enriched terms colored by clusters. COL6A3, collagen type VI alpha 3 chain; GO, Gene Ontology.

Figure 9

Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of COL6A2/COL6A3 and its co-expressed genes. (A) COL6A2; (B) COL6A3. COL6A2, collagen type VI alpha 2 chain; COL6A3, collagen type VI alpha 3 chain.

Figure 10

Gene sets enriched in the high COL6A2/COL6A3 expression phenotype. (A) COL6A2; (B) COL6A3. COL6A2, collagen type VI alpha 2 chain; COL6A3, collagen type VI alpha 3 chain; ccRCC, clear cell renal cell carcinoma.

A

B

Figure 11

The correlation between COL6A2/COL6A3 expression with immune infiltration levels and in ccRCC. COL6A2, collagen type VI alpha 2 chain; COL6A3, collagen type VI alpha 3 chain; ccRCC, clear cell renal cell carcinoma.