Original Article

Comprehensive analysis of the collagen family members as prognostic markers in clear cell renal cell carcinoma

Lingyu Guo[1,2], Tian An[3], Zhixin Huang[1,2], Ziyan Wan[1,2], Tie Chong[2]

[1]Department of Medicine, Xi'an Jiaotong University, Xi'an, China; [2]Department of Urology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China; [3]Department of Dermatology and Plastic Surgery, The Second Affiliated Hospital of Shaanxi University of Traditional Chinese Medicine, Xianyang, China

Contributions: (I) Conception and design: L Guo, T Chong; (II) Administrative support: T Chong; (III) Provision of study materials or patients: L Guo, T Chong; (IV) Collection and assembly of data: T An; (V) Data analysis and interpretation: Z Huang, Z Wan; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Tie Chong. Department of Urology, The Second Affiliated Hospital of Xi'an Jiaotong University, 157 West Fifth Road, Xi'an 710000, China. Email: chongtie@126.com.

Background: Clear cell renal cell carcinoma (ccRCC) is one of the common malignant tumors worldwide. There is still a lack of effective diagnostic and therapeutic targets for the recurrence and metastasis of ccRCC. In this study, we sought to identify effective diagnostic and therapeutic targets for ccRCC recurrence and metastasis.

Methods: Gene Expression Omnibus (GEO) dataset was used to obtain differentially expressed genes (DEGs) between primary and metastasis ccRCC. We used The Cancer Genome Atlas (TCGA), GeneMANIA, cBioPortal, MethSurv, and TIMER to analyze the expression differences, mutation status, prognostic value, molecular function, and immune infiltration of hub genes in renal cell carcinoma (RCC).

Results: We obtained a total of 35 different gene lists. Six collagen family members were identified as hub genes. The expression level of collagen family members was closely related to ccRCC. Moreover, differences in the expression levels of collagen family members were closely related to the stage and prognosis of ccRCC. Members of the collagen family were responsible for more than 15% of the genetic alterations in ccRCC and are involved in multiple signaling pathways. The expression level of collagen family members was closely related to the infiltration of tumor-associated immune cells. Univariate and multivariate Cox regression identified the prognosis-related genes: COL5A1.

Conclusions: Our study implied that members of the collagen family may serve as a biomarker for ccRCC metastasis and prognosis.

Keywords: Bioinformatics analysis; collagen family; renal clear cell carcinoma; tumor metastasis; prognosis

Submitted Feb 18, 2022. Accepted for publication May 26, 2022.
doi: 10.21037/tcr-22-398
View this article at: https://dx.doi.org/10.21037/tcr-22-398

Introduction

Renal cell carcinoma (RCC) is one of the most common malignancies of the urinary system in the world, accounting for ~2–3% of all malignant tumors (1). Thanks to advances in medical diagnostic technology, a large number of renal cancers are diagnosed at an early stage, but there are still many RCC patients who have a poor prognosis due to distant metastasis and other reasons. Recurrence and metastasis of RCC are the leading causes of death in patients. At present, there is still a lack of accurate and effective targets for recurrence and metastasis of RCC (2). Renal carcinoma has a high degree of morphological heterogeneity, which can be divided into 16 histological subtypes according to the World Health Organization (WHO) classification of tumors in 2016 (3). The most common pathological type is clear cell renal cell carcinoma

(ccRCC), papillary RCC, and chromophobe RCC. ccRCC accounts for about 70–75%. Since there are no typical clinical symptoms or specific diagnostic markers in the early stage of renal cancer, 20–30% of patients have developed distant metastasis or advanced renal cancer at the time of initial diagnosis. The existing treatment methods for metastatic RCC [radiotherapy, chemotherapy, interferons (IFN) immune therapy, etc.] are not sensitive (4). Molecular targeted therapy is one of the main treatment strategies for metastatic RCC, the most common molecular targeted therapy includes mammalian target of rapamycin (mTOR) inhibitors sirolimus, tyrosine kinase inhibitors sunitinib, and vascular endothelial growth factor (VEGF) inhibitor bevacizumab (5). However, most patients develop drug resistance to targeted drugs 6–15 months after targeted therapy, resulting in a ≤10% 5-year survival rate of patients with metastatic ccRCC (6). Therefore, exploring a new diagnosis and treatment of RCC has become an urgent problem to be solved in clinical practice.

Collagen widely exists in various tissues of the human body, with a total of 28 different types encoded by different genes and located in specific tissues of the human body, playing a variety of biological functions (7). Previous study has shown that members of the collagen family can participate in regulating the growth and migration of cancer cells. COL1A1 expression level can be used to predict the prognosis and immunotherapy effect of gastric cancer patients (8). Besides, COL4A1 is an active oncogene in glioma and is associated with tumor stage and prognosis (9). COL6A3 polymorphisms were associated with lung cancer risk (10). COL10A1 can promote the proliferation and migration of breast cancer cells *in vitro* (11). DNA methylation regulates gene transcription and translation, and the methylation level of many genes is closely related to cancer progression. The relationship between collagen gene methylation level and cancer has not been elucidated. In addition, the level of tumor immune cell infiltration significantly affects the progression of cancer, which has attracted widespread attention (12). Collagens can not only directly regulate the proliferation and metastasis of tumor cells, but also affect the function of tumor-associated immune cells such as tumor-associated macrophages and T cells, suggesting that collagens play an important role in tumor immunity and can be used as a target for tumor immunotherapy (13). Study have shown that collagen changes in melanoma can affect the motility of immune cells, thus affecting tumor progression (14). Study *in vitro* has confirmed that collagens can affect the motor ability of T cells and regulate the proportion of CD4 and CD8 in T cells (15). Due to the high heterogeneity of ccRCC, the prognosis of patients with ccRCC varies greatly. Some immune-related genes have been found to be related to the prognosis of patients with ccRCC (16), which can improve the accuracy of the existing prognosis prediction methods such as TNM staging system (17). There is no systematic study on the relationship between collagen and immune cell infiltration in ccRCC.

In this study, we used a series of bioinformatics methods to explore the role of collagen in ccRCC metastasis. First, we analyzed the Gene Expression Omnibus (GEO) data set to find the differentially expressed collagen genes in the process of ccRCC metastasis. We then assessed the relationship between collagen genes expression and ccRCC stage and prognosis. Finally, we explored the methylation levels of collagen genes in ccRCC and their relationship with tumor immune invasion. We believe that this study will contribute to a clearer understanding of the role of the collagen gene family in ccRCC metastasis and provide a basis for screening prognostic markers and therapeutic targets. We present the following article in accordance with the STREGA reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-398/rc).

Methods

Differentially expressed genes (DEGs) screening

We selected two sequencing datasets, GSE22541 and GSE105261, containing gene expression of ccRCC metastasis from the GEO database. GSE22541 from the GPL570 Affymetrix Human Genome U133 Plus 2.0 Array includes 24 primary ccRCC and 44 pulmonary metastases of ccRCC tissues. GSE105261 from the GPL10558 Illumina HumanHT-12 V4.0 expression bead chip includes 9 normal, 9 primary ccRCC, and 26 metastatic ccRCC tissues. GEO2R tool was used for data analysis, analysis parameters were set to $|\log FC| \geq 1$ and adjusted $P<0.05$.

PPI network analysis

GeneMANIA (http://www.genemania.org) uses extensive genomic and proteomic data to find genes with similar functions (18). We used this website to predict protein interactions and to analyze pathways of the common DEGs. Cytohubba is a plug-in for Cytoscape software to identify hub nodes. It is used to analyze the previously obtained DEGS interaction network to search for hub genes.

Gene enrichment analysis

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 (https://david.ncifcrf.gov/) can associate genes from the input list with biological annotations (19). We used the DAVID website to conduct enrichment analysis of gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for DEGs.

Gene expression analysis

Oncomine (https://www.oncomine.org) is a cancer microarray database and integrated data-mining platform (20). We analyzed and compared the mRNA expression of collagen family members in renal cancer tissues and normal tissues, and the screening parameters were set as $P<0.0001$, $|logFC| \geq 2$, and a top 10% gene rank. The Cancer Genome Atlas (TCGA) database includes expression data, miRNA expression data, methylation data, mutation data, and copy number data for 33 tumors. were verified. We used TCGA-KIRC data to analyze collagen family genes' expression levels in ccRCC tissues.

Mutation analysis

The cBioPortal for Cancer Genomics (https://www.cbioportal.org/) provides a visual tool for research and analysis of cancer genetic data (21). CBioPortal helps understand genetics, epigenetics, gene expression, and proteomics from molecular data derived from cancer tissue and cytology studies. In the study, this tool was used to analyze the mutation of collagen family genes.

Identification of differentially expressed and prognosis-related collagens

The survival package was used to perform survival analysis of TCGA data and plot Kaplan-Meier (KM) curves. Subsequently, we performed a univariate regression analysis between collagen family genes and ccRCC overall survival (OS). Then, we selected genes that were statistically significant in univariate regression analysis for multivariate regression analysis and finally obtained genes with significance in both univariate and multivariate analysis were considered as candidates with significant correlation with the prognosis of ccRCC.

Gene set enrichment analysis (GSEA)

LinkedOmics database (http://www.linkedomics.org/) contains multiple omics and clinical data for 32 cancer types (22). We selected ccRCC as the tumor type in the database website and screened genes related to collagen family genes based on Pearson correlation analysis by using the LinkFinder function of the website. Then, the LinkInterpreter functional module of the website was used to conduct GO and KEGG gene enrichment analysis.

DNA methylation analysis

MethSurv (https://biit.cs.ut.ee/methsurv/) is a web-based tool for survival analysis based on cytosine-phosphate-guanine (CpG) methylation patterns (23). We used TCGA methylation data contained in MethSurv to perform survival analysis of CpGs located near collagen family genes.

Immune infiltration and drug response analysis

The TIMER website (https://cistrome.shinyapps.io/timer/) provides a comprehensive and systematic analysis of immune infiltrations across different cancer types (24). We first estimated the relationship between collagen family members' expression level and tumor purity and the level of tumor-associated immune cell invasion including B cells, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, macrophages, and dendritic cells. Subsequently, we used the SCNA module of the website to explore the correlation between tumor immune cell infiltration and gene copy number. GSCALite offers a variety of analytical models including methylation analysis and drug sensitivity analysis (25). In the current study, GSCALite (http://bioinfo.life. hust.edu.cn/web/GSCALite/) is a tumor genome analysis platform that integrates genomic data from the TCGA for 33 tumor types, drug response data from GDSC, CTRP, and normal tissue data from GTEX for genome analysis in a unified data analysis process. GSCALite was used to analyze the correlation between expression of the collagen family and drug sensitivity based on the data of GDSC.

Statistical analysis

Statistical analysis of data was carried out by R software (V4.0.2). We performed Cox regression analysis on collagen family gene expression and OS, obtained hazard ratios (HRs)

and 95% confidence intervals (CIs). The results of statistical analysis were considered to be significant if the P value was less than 0.05.

Ethical statement

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Results

Identification of DEGs in ccRCC

A total of 375 up-regulated genes and 226 down-regulated genes were found in GSE22541. A total of 80 up-regulated genes and 35 down-regulated genes were found in GSE105261 (*Figure 1A*,*1B*). Then, we obtained 29 up-regulated genes and 6 down-regulated genes in both data sets by Venn diagram (*Figure 1C*,*1D*). Based on these lists of DEGs, we performed PPI network analysis (*Figure 1E*). DEGs are mainly involved in biological functions including extracellular matrix (ECM) structural constituent, collagen trimer, etc. Then, we applied the CytoHubba plug-in to obtain hub genes. The results showed that the top ten hub genes include *COL3A1*, *COL1A1*, *COL5A2*, *COL1A2*, *POSTN*, *COL6A3*, *COL5A1*, *LUM*, *DCN*, and *THBS2* (*Table 1*). There were 6 genes in the collagen family. This result suggests that the collagen family plays a key role in the process of kidney cancer metastasis.

Next, we conducted gene enrichment analysis of the DEGs to understand their biological functions, and the results showed that DEGs mainly affected ECM organization, collagen catabolic process, collagen fibril organization, and ECM structural constituent. The main pathways involving DEGs were ECM-receptor interaction, protein digestion and absorption, platelet activation, focal adhesion, amoebiasis, pi3k/Akt signaling pathway, and beta-alanine metabolism (*Figure 2A-2D*).

Expression of the collagen family members

We obtained the mRNA expression of 6 collagen family members in renal carcinoma and normal tissues through the Oncomine database. The collagen family members we screened showed elevated expression levels in various tumor tissues, as well as in renal cancer tissues. These results suggest that members of the collagen family may play a role in cancer progression. Results showed that compared to normal tissues, the expression levels of COL1A2, COL3A1, and COL5A1 were elevated in more kidney cancer datasets, while COL6A3 was decreased (*Figure 3*). Then we combined the data in the TCGA database, and the results were consistent with the previous results. The results showed that COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, and COL6A3 were highly expressed in renal tumor samples (*Figure 4*).

Genetic mutation analysis of collagen expression in ccRCC

By analyzing the ccRCC data of cBioPortal, the results showed the mutation rate of COL3A1 and COL5A2 was 8%, which were the highest among them (Figure S1A). We studied the mutations of collagen family members in different types of renal cancer, and the results showed that high mutation levels of collagen family members were prevalent in different types of renal cancer (Figure S1B-S1H). We also found that altered expression of collagen family genes is also common in renal cancer, suggesting that mutations and altered expression of collagen family members play a role in ccRCC.

Survival analysis of collagen expression in ccRCC

We used RNAseq data from TCGA KIRC database for survival analysis. Patients were divided by the medium value of gene expression. The results showed that elevated expression levels in most collagen family members were associated with shorter survival. Among them, the high expression levels of COL1A1, COL5A1, and COL6A3 were significantly correlated with the OS of ccRCC (log-rank P<0.05) (*Figure 5A-5F*), and the high expression levels of COL1A1, COL1A2, COL5A1, and COL6A3 were significantly correlated with the DSS of ccRCC (log-rank P<0.05) (Figure S2A-S2F). These results suggest that collagen family members play an important role in the progression of ccRCC, significantly affect the survival of patients with ccRCC, and can be used as a prognostic marker of ccRCC.

Subsequently, univariate and multivariate regression analyses were conducted respectively. Univariate analysis showed that COL1A1 (HR: 1.161; P<0.001), COL1A2 (HR: 1.109; P<0.05), COL5A1 (HR: 1.233; P<0.001), and COL6A3 (HR: 1.191; P<0.001) were correlated with ccRCC OS, while multivariate analysis showed that COL1A2 (HR: 0.389; P<0.001) and COL5A1 (HR: 2.308; P<0.001) were correlated with ccRCC prognosis (*Table 2*).

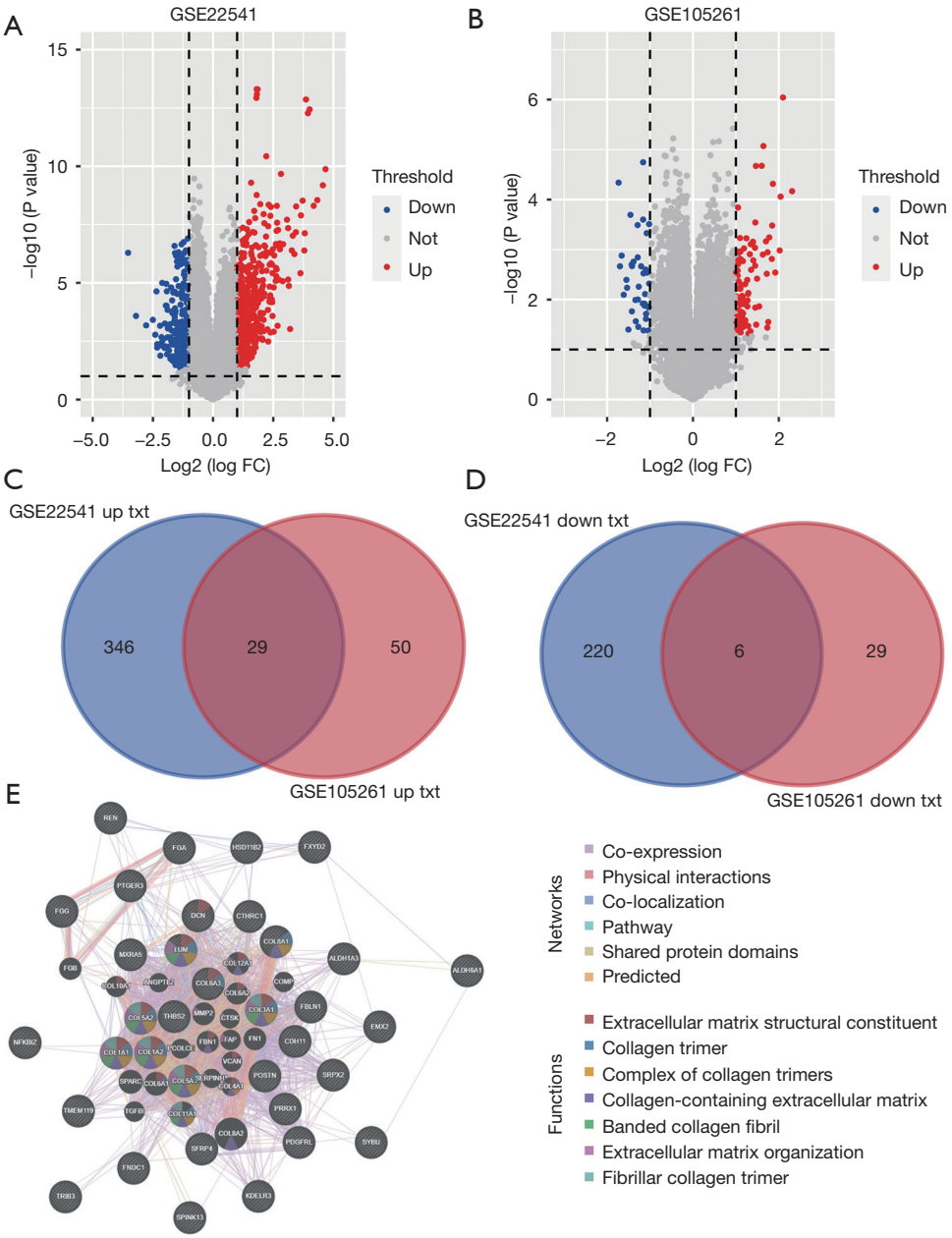

Figure 1 DEGs were identified from two gene expression profiles. (A,B) Volcano plots of upregulated (red) and downregulated (blue) DEGs between metastatic ccRCC samples and primary tumor samples in GSE22541 (A) and GSE105261 (B). (C,D) Venn diagram of upregulated and downregulated DEGs. (E) Protein-protein interaction of DEGs (GeneMANIA). DEGs, differential expression genes; ccRCC, clear cell renal cell carcinoma.

In general, COL5A1 can be used as independent prognostic factors of ccRCC.

GSEA analysis of COL5A1

In order to further understand the COL5A1-related molecular functions and possible molecular mechanisms involved in tumor progression, genes related to COL5A1 expression in tumors were screened and gene enrichment analysis was performed. We screened 7,089 genes that were positively correlated with COL5A1 expression, and 7,689 genes that were negatively correlated with COL5A1

Table 1 Top 10 in network ranked by MCC method

Rank	Name	Score
1	COL3A1	1864806
2	COL1A1	1864802
3	COL5A2	1864800
3	COL1A2	1864800
3	POSTN	1864800
6	COL6A3	1859760
7	COL5A1	1854720
8	LUM	1819440
9	DCN	1088641
10	THBS2	771120

MCC, maximal clique centrality.

expression (*Figure 6A-6C*) (P<0.05; false discovery rate <0.05). The gene heat map shows the genes with the top 50 correlations. Enrichment analysis of relevant genes obtained showed that genes associated with COL5A1 were primarily involved in extracellular structure organization, amoebiasis, ECM-receptor interaction, and valine, leucine and isoleucine degradation (Figure S3A,S3B).

DNA methylation analysis

The methylation level of gene DNA promoter is closely related to tumor survival. We used TCGA KIRC methylation data contained in MethSurv to perform survival analysis of CPGs located near collagen family genes. Our study found that the methylation levels of collagen family members changed in ccRCC, and CpG methylation sites

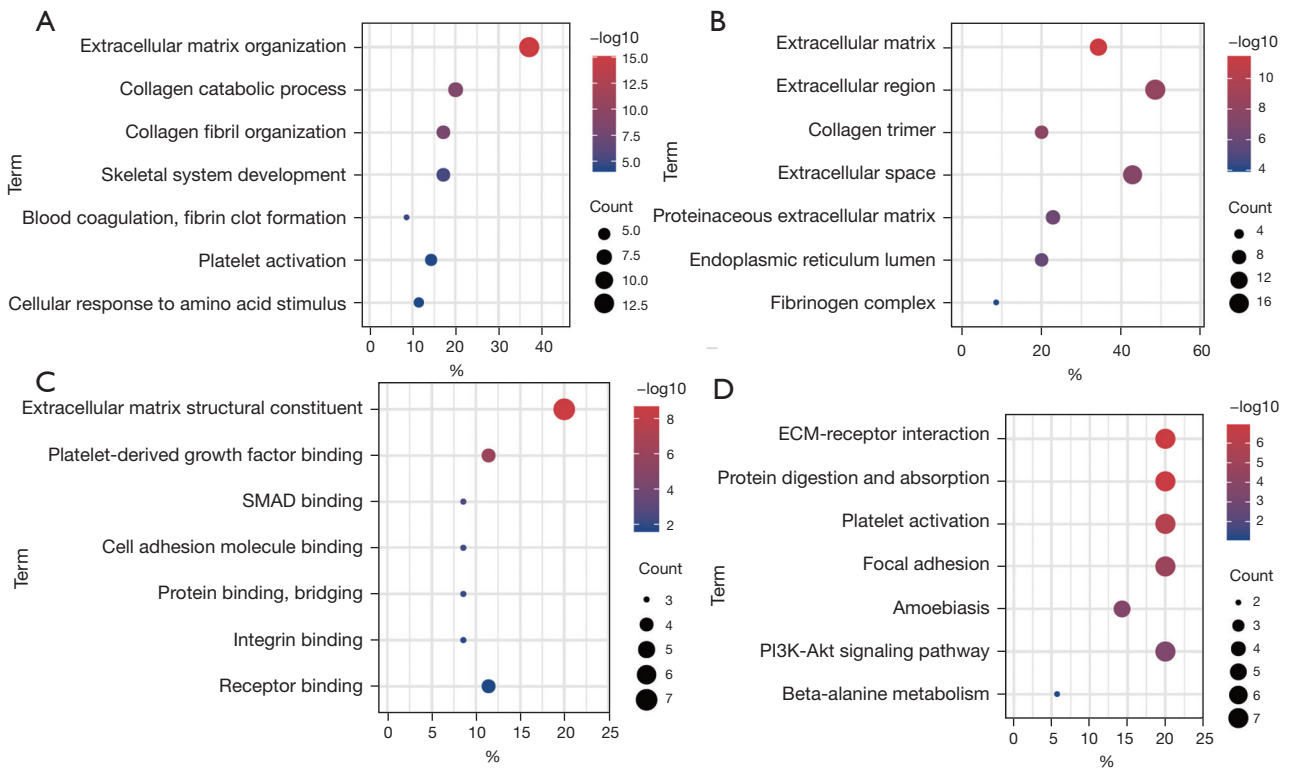

Figure 2 Gene enrichment analysis of DEGs. (A) Biological process; (B) cellular component; (C) molecular function; (D) KEGG pathway analysis. DEGs, differential expression genes; SMAD, smad proteins; ECM, extracellular matrix; KEGG, Kyoto Encyclopedia of Genes and Genomes.

were associated with ccRCC survival. The DNA promoter methylation levels of COL1A1 and COL1A2 were significantly reduced in renal cancer, which to some extent explained the high expression of these two genes in ccRCC. In contrast, the promoter methylation levels of COL6A3 were significantly increased (*Figure* 7). In addition, we found that certain CpG sites in collagen members were associated with ccRCC prognosis, including 14 sites of COL1A1,

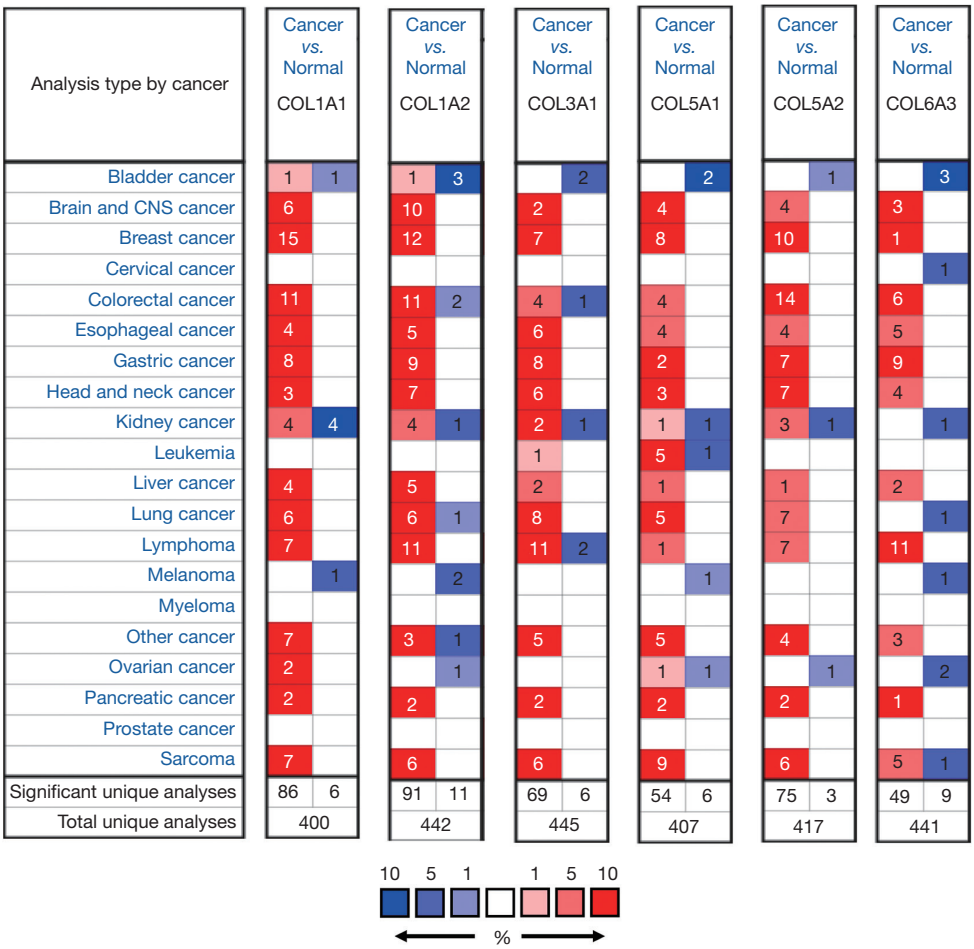

Figure 3 The mRNA expression of collagen family genes (ONCOMINE). The numbers in the figure represent the number of datasets with significant differences in gene expression, red representing up-regulated genes and blue representing down-regulated genes. CNS, central nervous system.

10 sites of COL1A2, 2 sites of COL3A1, 42 sites of COL5A1, 6 sites of COL5A2, and 27 sites of COL6A3 (P<0.05) (*Table 3*). In conclusion, these results suggest that methylation levels in collagen family members influence the prognosis of ccRCC.

Immune infiltration and drug response

We used the ccRCC data from TIMER database to detect the correlation between collagen family members' expression levels and the infiltration levels of tumor-immune infiltrating cells (TIICs). The results showed that collagen family members were positively correlated with detected immune cells, but negatively correlated with tumor purity (*Figure 8A-8F*). Subsequently, we used the SCNA module of the database to detect the somatic copy number alterations of collagen family members, and the results showed that the arm-level deletion, arm-level gain, deep deletion, and high amplification of collagen family members were closely related to the level of immune cell infiltration in ccRCC (Figure S4A-S4F). These results suggest that members of the collagen family may influence the prognosis of ccRCC by regulating the level of tumor immune cell infiltration.

Previous studies have shown that the expression level of collagen family members is correlated with the prognosis of ccRCC, and these gene expression changes may affect the prognosis of the tumor by regulating the level of tumor-associated immune cell infiltration through the regulation of DNA methylation (26,27). Thus, members of the collagen family may have the potential to become targets for ccRCC therapy. Our test results in the GSCALite database showed

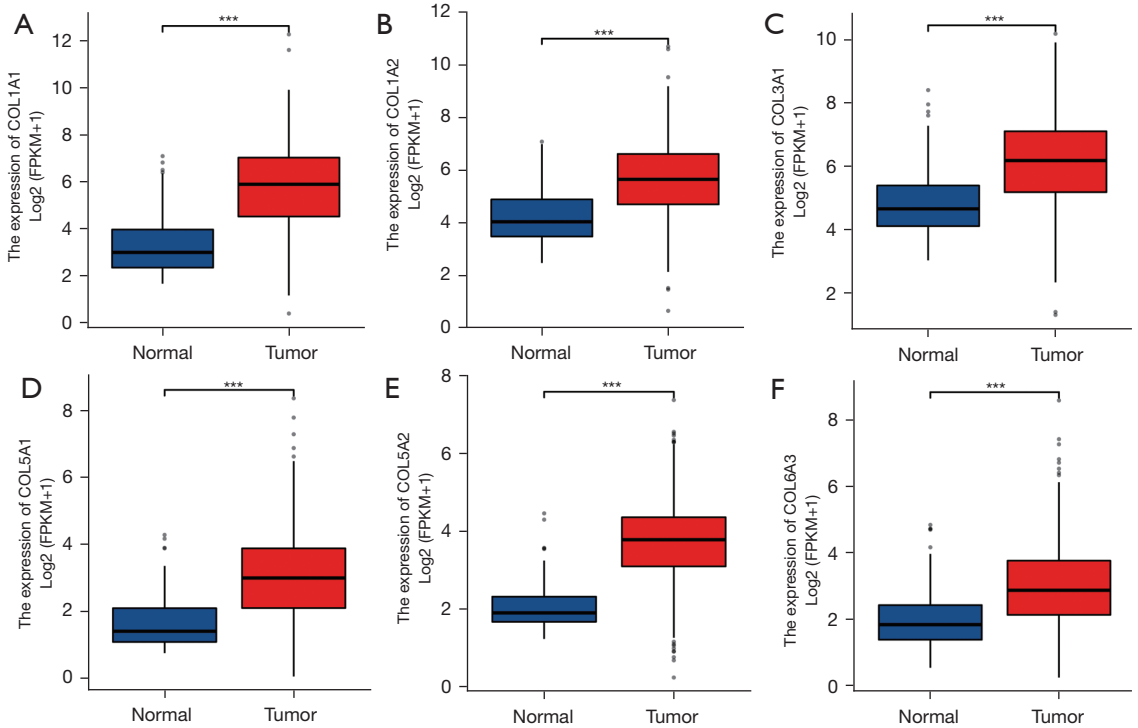

Figure 4 The expression of collagen family members in TCGA KIRC database. (A) COL1A1; (B) COL1A2; (C) COL3A1; (D) COL5A1; (E) COL5A2; (F) COL6A3. ***, P<0.001. TCGA, the Cancer Genome Atlas; KIRC, kidney renal clear cell carcinoma.

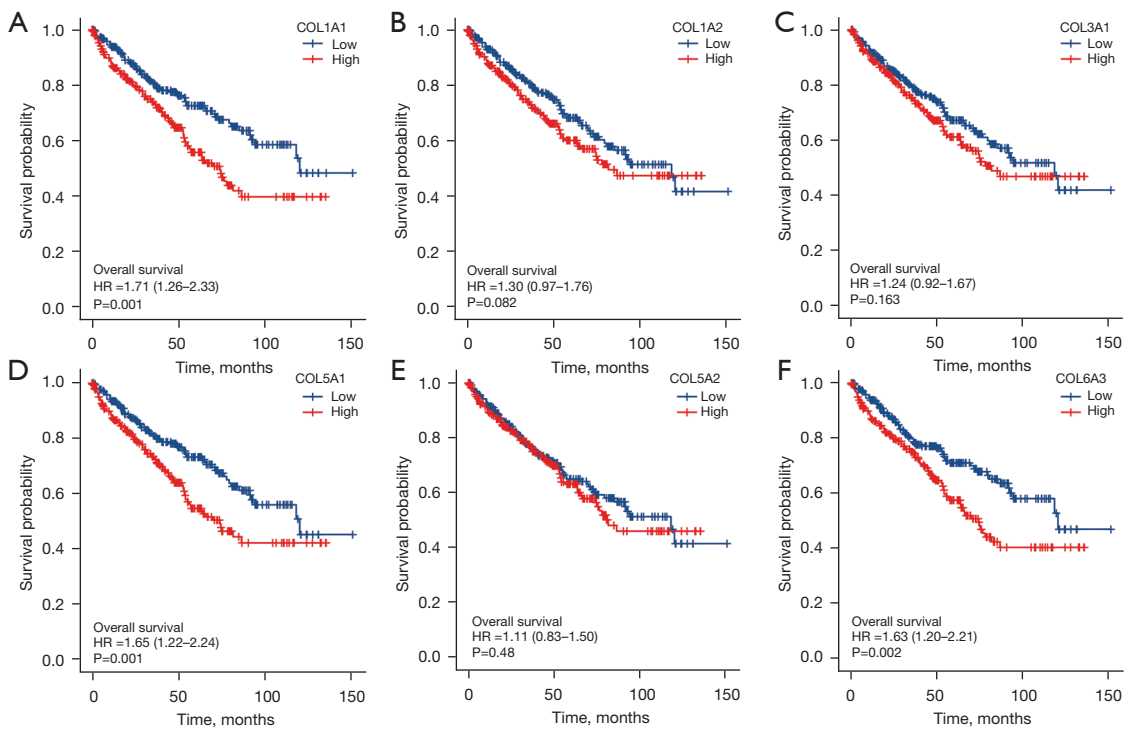

Figure 5 The prognostic value of collagen family in ccRCC (KM plotter). (A) COL1A1; (B) COL1A2; (C) COL3A1; (D) COL5A1; (E) COL5A2; (F) COL6A3. HR, hazard ratio; ccRCC, clear cell renal cell carcinoma; KM, Kaplan-Meier.

Table 2 Cox analysis of collagen family in the TCGA

Name	Total (N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
COL1A1	539	1.161 (1.066–1.263)	<0.001	1.314 (0.930–1.857)	0.122
COL1A2	539	1.109 (1.002–1.227)	0.045	0.389 (0.288–0.524)	<0.001
COL3A1	539	1.063 (0.962–1.176)	0.230		
COL5A1	539	1.233 (1.111–1.369)	<0.001	2.308 (1.497–3.557)	<0.001
COL5A2	539	1.087 (0.957–1.236)	0.200		
COL6A3	539	1.191 (1.075–1.320)	<0.001	0.983 (0.803–1.203)	0.866

TCGA, The Cancer Genome Atlas; CI, confidence interval.

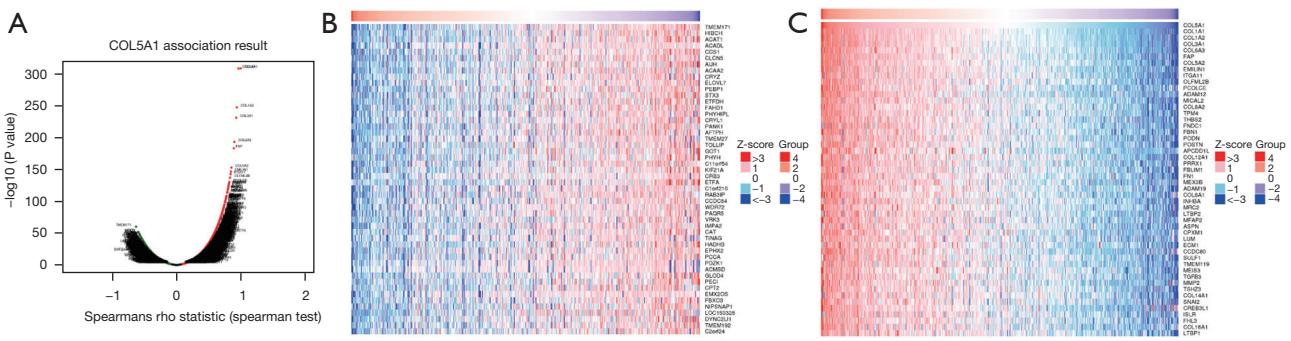

Figure 6 Genes correlated with COL5A1 (LinkedOmics). (A) Volcano maps of top 50 genes correlated with COL5A1. (B) Heat maps of genes negatively correlated with COL5A1. (C) Heat maps of genes positively correlated with COL5A1.

that the expression levels of collagen family members were most closely related to the drug sensitivity of the tumor. The number of related drugs or small molecules from most to least is COL5A1, COL5A2, COL1A1, COL1A2, COL6A3, and COL3A1, which are 14, 11, 9, 8, 7, and 3 respectively (Figure S5). The results may suggest that the collagen family especially COL5A1, and COL5A2 are potential biomarkers for drug screening.

Discussion

ccRCC is a common malignant tumor, which often leads to death due to tumor recurrence and metastasis (28). The treatment has improved with advances in technology, but there is still no effective treatment for recurrent and metastatic tumors. The lack of specific diagnostic and prognostic markers limits the early diagnosis and treatment of ccRCC. Therefore, the development of specific targets for the diagnosis and treatment of ccRCC is crucial. In this study, we identified 6 collagen family genes by analyzing 2 GEO ccRCC metastasis datasets. Further studies showed that collagen family genes were highly expressed in ccRCC tissues and were closely related to the prognosis of ccRCC. Subsequently, we assessed the methylation level of collagen family genes in ccRCC, their relationship with tumor immune cell infiltration, and their responsiveness to therapeutic drugs. The results confirmed that collagen family genes can be used as prognostic markers of ccRCC and help improve the level of diagnosis and treatment of ccRCC.

Previous study has shown the prognostic value of collagens in a variety of tumors. Elevated COL1A2 expression level is a predictor of gastric cancer prognosis (29). m6A methylation-mediated COL3A1 up-regulation promotes metastasis of triple-negative breast cancer (TNBC) (30). Furthermore, CircACAP2 promotes breast cancer proliferation and metastasis by targeting the miR-29a/b-3p-COL5A1 axis (31). COL5A2 acts as a potential clinical biomarker for gastric cancer and renal metastasis (32).

These studies are consistent with our findings. At

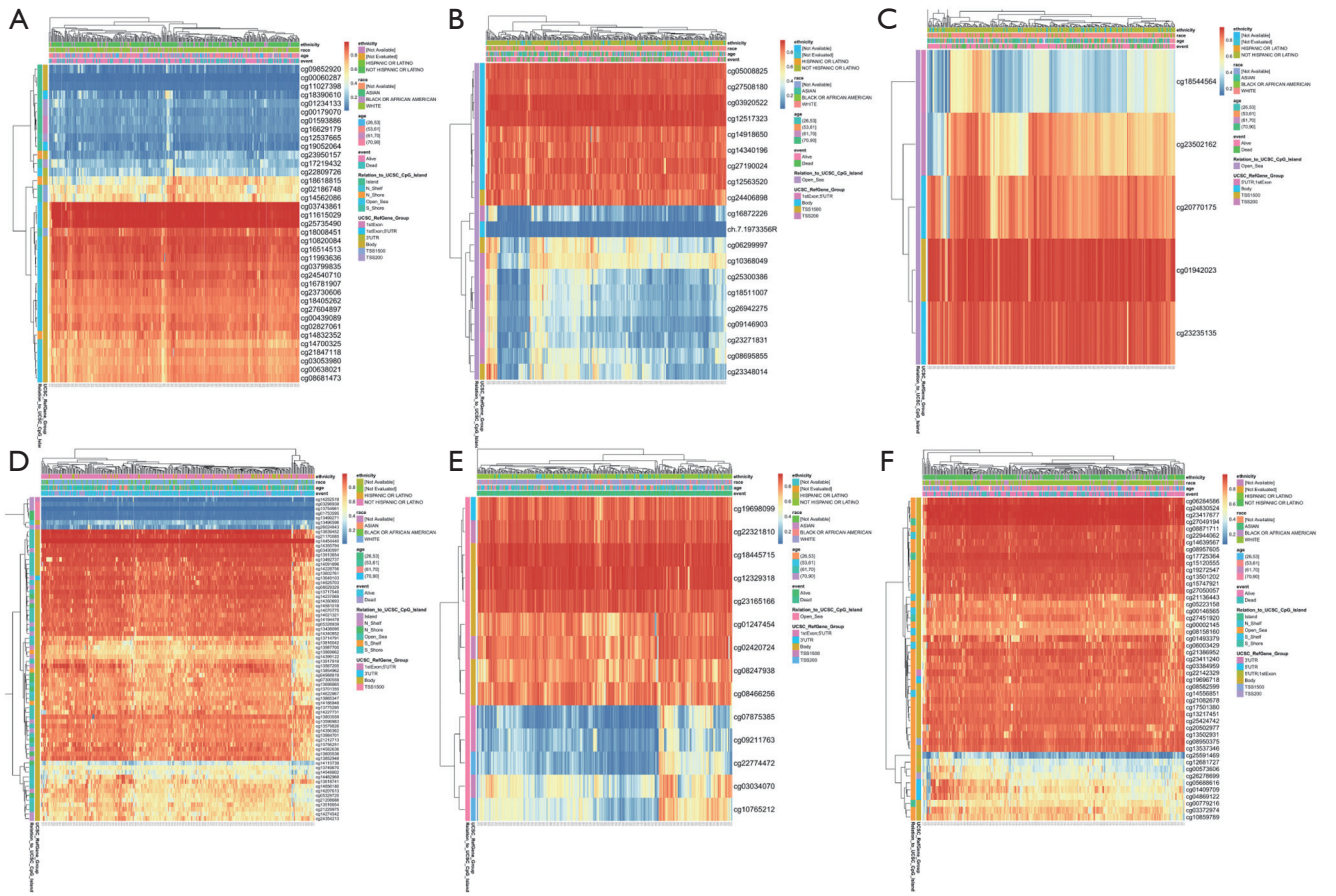

Figure 7 DNA methylation of collagen family members in MethSurv. (A) COL1A1; (B) COL1A2; (C) COL3A1; (D) COL5A1; (E) COL5A2; (F) COL6A3.

present, it is widely believed that DNA methylation is closely related to the prognosis of tumors (33). A high methylation level of gene DNA promoter often leads to gene silencing, and methylation of key genes can affect the progress of the tumor (34). Previous study has shown that DNA methylation of TMEM130 promotes cell migration in breast cancer (35). DIO3OS DNA methylation drives non-small cell lung cancer progression (36). ANGPTL4 DNA methylation promotes colorectal cancer metastasis by activating the ERK pathway (37). We assessed the methylation levels of the collagen family genes in ccRCC and found that the methylation levels of COL1A1 and COL1A2 decreased in ccRCC and COL6A3 was increased. In addition, multiple CpG sites of collagen family genes are associated with the prognosis of ccRCC.

Tumor immunotherapy is now very effective against many tumor types, especially inoperable tumors. The infiltration level of tumor-associated immune cells directly affects the effect of tumor immunotherapy. Previous study has shown that the activation of the programmed death (PD)-1/PD-ligand (PD-L) pathway and regulatory T cells (Tregs) in the tumor microenvironment contributes to the evasion of the transformed cells from the immune surveillance and the suppression of an antitumor immune response (38). In patients with TNBC, tumor-infiltrating lymphocytes (TILs) are associated with improved survival (39). Collagen promotes anti-PD-1/PD-L1 resistance in cancer through LAIR1-dependent CD8(+) T cell exhaustion (40). We found those collagen family genes are closely associated with levels of infiltration of various tumor-associated immune cells. Collagen family genes can be used as potential tumor immunotherapy targets. In addition, the results of drug sensitivity analysis showed that the collagen family genes were associated with multiple chemotherapeutic drug sensitivities in ccRCC, especially COL5A1 and COL5A2. These results suggest that collagen

Table 3 The significant prognostic values of CpG in the collagen family members

Name	CpG name	HR	95% CI	LR test P value	UCSC RefGene Group	Relation to UCSC CpG Island
COL1A1	cg00060287	0.603	(0.37–0.983)	0.0332	Body	Island
	cg02186748	1.761	(1.176–2.636)	0.0049	TSS1500	S_Shore
	cg03799835	0.518	(0.353–0.761)	0.0009	Body	Open_Sea
	cg11027398	0.575	(0.356–0.929)	0.0172	Body	Island
	cg14562086	0.564	(0.343–0.928)	0.0170	TSS1500	S_Shore
	cg14700325	0.546	(0.362–0.824)	0.0056	Body	N_Shelf
	cg16781907	0.591	(0.4–0.873)	0.0076	Body	N_Shelf
	cg18405262	0.47	(0.316–0.701)	0.0002	Body	Open_Sea
	cg18618815	2.973	(1.549–5.705)	0.0002	Body	N_Shore
	cg21847118	1.669	(1.004–2.777)	0.0373	Body	Open_Sea
	cg22809726	2.639	(1.476–4.72)	0.0002	3'UTR	Open_Sea
	cg23950157	2.879	(1.872–4.427)	0.0000	Body	N_Shore
	cg24540710	0.367	(0.247–0.546)	0.0000	Body	Open_Sea
	cg27604897	0.612	(0.416–0.901)	0.0141	Body	Open_Sea
COL1A2	cg03920522	0.537	(0.358–0.805)	0.0037	Body	Open_Sea
	cg08695855	1.864	(1.079–3.221)	0.0165	TSS200	Open_Sea
	cg09146903	1.503	(1.013–2.229)	0.0402	TSS200	Open_Sea
	cg10368049	0.376	(0.216–0.655)	0.0001	TSS200	Open_Sea
	cg14340196	0.585	(0.359–0.952)	0.0231	Body	Open_Sea
	cg16872226	2.472	(1.382–4.419)	0.0007	TSS200	Open_Sea
	cg23348014	2.676	(1.433–4.998)	0.0005	TSS1500	Open_Sea
	cg24406898	0.654	(0.446–0.959)	0.0303	TSS1500	Open_Sea
	cg25300386	0.586	(0.359–0.958)	0.0249	1stExon;5'UTR	Open_Sea
	ch.7.1973356R	2.145	(1.442–3.191)	0.0003	Body	Open_Sea
COL3A1	cg01942023	0.554	(0.337–0.91)	0.0134	TSS1500	Open_Sea
	cg20770175	0.541	(0.325–0.899)	0.0116	Body	Open_Sea
COL5A1	cg01753595	0.6	(0.361–0.997)	0.0376	TSS1500	N_Shore
	cg03298938	0.455	(0.304–0.68)	0.0002	TSS1500	Island
	cg03430597	0.552	(0.332–0.917)	0.0147	Body	Island
	cg05328939	1.69	(1.017–2.809)	0.0324	Body	Island
	cg05329720	2.851	(1.558–5.215)	0.0001	Body	N_Shore
	cg07300559	2.34	(1.332–4.108)	0.0011	Body	N_Shore
	cg08029329	1.858	(1.105–3.125)	0.0125	Body	Island
	cg13438095	1.897	(1.278–2.818)	0.0012	Body	Open_Sea
	cg13492737	1.747	(1.05–2.907)	0.0229	Body	Open_Sea

Table 3 (*continued*)

Table 3 (*continued*)

Name	CpG Name	HR	95% CI	LR test P value	UCSC RefGene Group	Relation to UCSC CpG Island
	cg13496596	1.781	(1.082–2.931)	0.0162	Body	S_Shore
	cg13499271	0.413	(0.241–0.705)	0.0004	TSS1500	N_Shore
	cg13516654	0.559	(0.336–0.929)	0.0170	Body	Open_Sea
	cg13567205	1.95	(1.315–2.892)	0.0007	Body	N_Shelf
	cg13596983	0.603	(0.367–0.992)	0.0361	Body	Island
	cg13605536	2.049	(1.258–3.337)	0.0020	Body	N_Shore
	cg13639452	1.958	(1.319–2.908)	0.0007	Body	Open_Sea
	cg13698865	0.658	(0.446–0.971)	0.0335	Body	Open_Sea
	cg13714791	2.596	(1.475–4.566)	0.0002	Body	S_Shore
	cg13717540	1.82	(1.082–3.063)	0.0161	Body	Open_Sea
	cg13754661	0.511	(0.325–0.804)	0.0023	TSS1500	Island
	cg13775295	0.527	(0.353–0.787)	0.0014	Body	Open_Sea
	cg13854962	2.294	(1.476–3.566)	0.0001	Body	S_Shelf
	cg13865347	2.73	(1.526–4.885)	0.0001	Body	Open_Sea
	cg13913654	2.099	(1.215–3.628)	0.0038	Body	Open_Sea
	cg13917918	1.791	(1.175–2.732)	0.0051	Body	Open_Sea
	cg14070775	1.698	(1.031–2.797)	0.0282	Body	Open_Sea
	cg14091896	0.605	(0.368–0.994)	0.0370	Body	Open_Sea
	cg14194478	0.647	(0.439–0.954)	0.0267	Body	Open_Sea
	cg14207613	1.962	(1.29–2.985)	0.0011	Body	N_Shelf
	cg14227731	0.416	(0.24–0.721)	0.0006	Body	Open_Sea
	cg14228756	1.8	(1.115–2.906)	0.0111	Body	Open_Sea
	cg14237069	1.612	(1.073–2.421)	0.0255	Body	N_Shore
	cg14274542	2.718	(1.519–4.863)	0.0001	Body	Island
	cg14350693	1.627	(1.083–2.443)	0.0228	Body	Island
	cg14355794	1.788	(1.049–3.047)	0.0227	Body	Open_Sea
	cg14356362	0.566	(0.34–0.944)	0.0207	Body	Island
	cg14399122	0.413	(0.235–0.726)	0.0006	Body	Island
	cg14581018	1.909	(1.283–2.839)	0.0020	Body	N_Shore
	cg14622967	1.528	(1.022–2.284)	0.0354	Body	S_Shore
	cg14656180	2.356	(1.363–4.074)	0.0007	Body	Open_Sea
	cg21208686	2.409	(1.606–3.613)	0.0000	Body	S_Shore
	cg24354213	1.866	(1.11–3.138)	0.0118	Body	Island
COL5A2	cg02420724	0.529	(0.318–0.882)	0.0092	TSS1500	Open_Sea
	cg07875385	2.378	(1.33–4.254)	0.0012	1stExon;5'UTR	Open_Sea

Table 3 (*continued*)

Table 3 (*continued*)

Name	CpG Name	HR	95% CI	LR test P value	UCSC RefGene Group	Relation to UCSC CpG Island
	cg08247938	0.596	(0.403–0.881)	0.0086	Body	Open_Sea
	cg09211763	2.544	(1.423–4.55)	0.0004	1stExon;5'UTR	Open_Sea
	cg10765212	1.508	(1.021–2.227)	0.0375	TSS200	Open_Sea
	cg12329318	0.341	(0.187–0.623)	0.0001	Body	Open_Sea
COL6A3	cg00002145	2.265	(1.29–3.978)	0.0017	Body	Open_Sea
	cg00779216	2.361	(1.344–4.149)	0.0010	Body	Island
	cg03372974	1.917	(1.139–3.224)	0.0086	Body	Open_Sea
	cg05223158	0.44	(0.255–0.761)	0.0013	Body	Open_Sea
	cg06284586	2.387	(1.398–4.077)	0.0005	Body	Open_Sea
	cg08871711	1.77	(1.065–2.941)	0.0192	Body	Open_Sea
	cg08950375	0.56	(0.373–0.841)	0.0071	TSS1500	Open_Sea
	cg08957605	2.286	(1.277–4.09)	0.0021	Body	Open_Sea
	cg12681727	2.521	(1.674–3.798)	0.0000	Body	Open_Sea
	cg13502931	0.515	(0.347–0.764)	0.0014	Body	Open_Sea
	cg13537346	0.59	(0.402–0.866)	0.0076	Body	Open_Sea
	cg14556851	1.647	(1.002–2.708)	0.0384	Body	S_Shelf
	cg15747921	2.183	(1.479–3.222)	0.0001	Body	Open_Sea
	cg17725364	2.14	(1.27–3.604)	0.0020	Body	Island
	cg19696718	0.668	(0.454–0.982)	0.0397	5'UTR	Open_Sea
	cg20502977	0.473	(0.27–0.831)	0.0044	Body	Open_Sea
	cg21136443	2.203	(1.27–3.822)	0.0021	Body	N_Shelf
	cg21386952	2.33	(1.548–3.507)	0.0000	Body	Open_Sea
	cg22944062	1.847	(1.242–2.748)	0.0020	Body	N_Shelf
	cg23417677	2.248	(1.316–3.841)	0.0012	Body	Open_Sea
	cg24830524	2.712	(1.484–4.955)	0.0002	Body	Open_Sea
	cg25424742	1.557	(1.037–2.338)	0.0378	Body	Open_Sea
	cg25591469	2.246	(1.319–3.827)	0.0011	5'UTR;1stExon	Open_Sea
	cg26278699	0.593	(0.36–0.976)	0.0304	TSS200	Open_Sea
	cg27049194	2.827	(1.605–4.982)	0.0001	Body	Island
	cg27050057	1.764	(1.061–2.933)	0.0203	Body	Open_Sea
	cg27451920	1.793	(1.2–2.68)	0.0059	Body	S_Shore

CpG, cytosine-phosphate-guanine; HR, hazard ratio; CI, confidence interval; LR, Likelihood ratio; UCSC, University of California Santa Cruz.

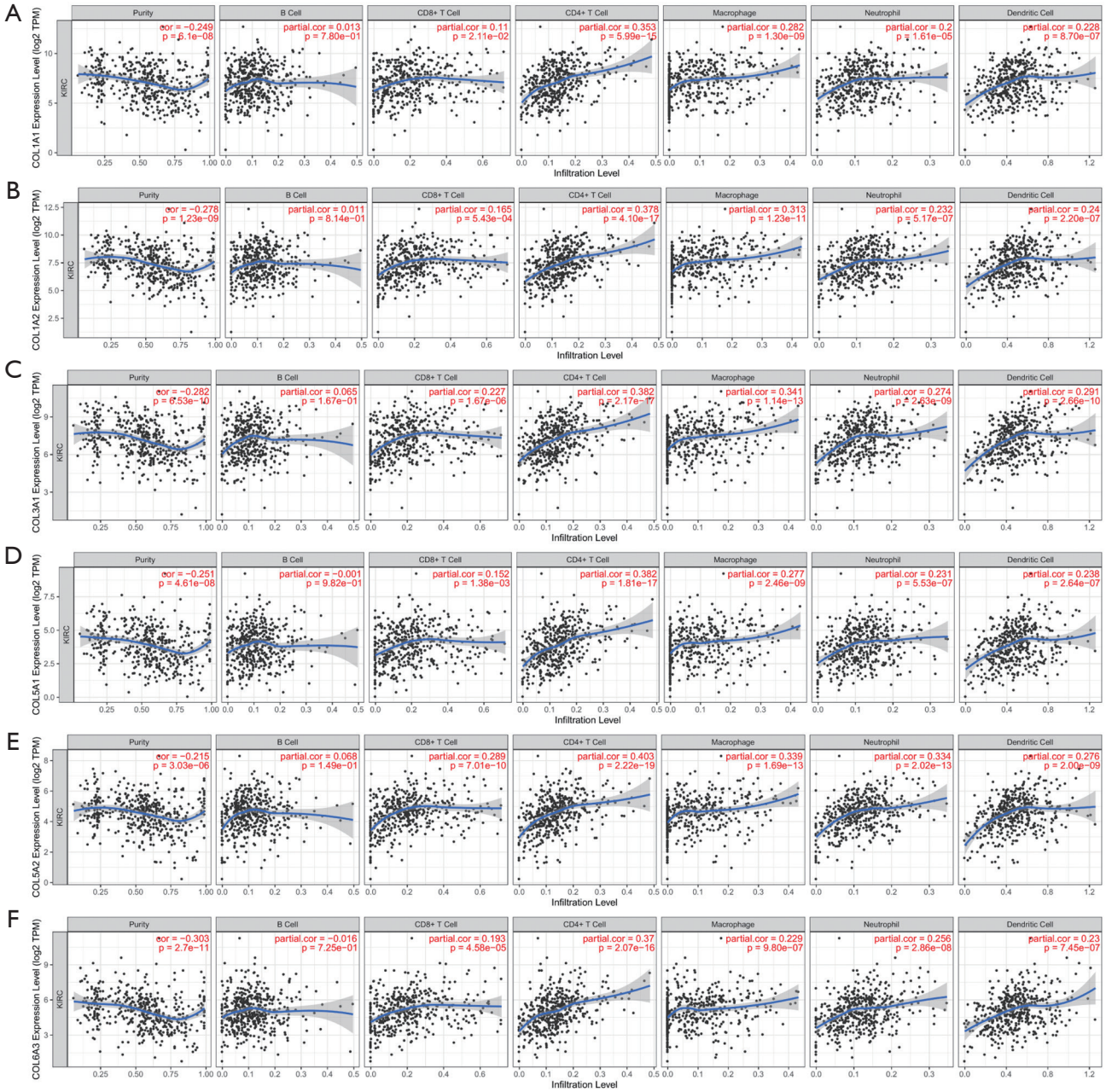

Figure 8 The correlation between collagens and immune cell infiltration in ccRCC (TIMER). (A) COL1A1; (B) COL1A2; (C) COL3A1; (D) COL5A1; (E) COL5A2; (F) COL6A3. ccRCC, clear cell renal cell carcinoma.

family genes are closely associated with ccRCC prognosis and can be used as potential therapeutic targets for ccRCC.

Conclusions

In summary, we found that the collagen family genes are key genes for ccRCC metastasis. The collagen family genes' expression levels and methylation levels both affect the prognosis of ccRCC. In particular, COL5A1 can be used as independent prognostic factors of ccRCC. In addition, collagen expression was also associated with tumor immune cell infiltration level and chemotherapy

drug sensitivity. Therefore, our study suggests that collagen family genes can be used as a prognostic and therapeutic target for ccRCC.

Acknowledgments

We acknowledge the TCGA and GEO databases, as well as the researchers who upload their datasets.
Funding: This work was supported by National Natural Science Foundation of China (No. 82070716).

Footnote

Reporting Checklist: The authors have completed the STREGA reporting checklist. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-398/rc

Peer Review File: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-398/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-398/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Siegel RL, Miller KD, Fuchs HE, et al. Cancer Statistics, 2021. CA Cancer J Clin 2021;71:7-33.
2. D'Aniello C, Berretta M, Cavaliere C, et al. Biomarkers of Prognosis and Efficacy of Anti-angiogenic Therapy in Metastatic Clear Cell Renal Cancer. Front Oncol 2019;9:1400.
3. Moch H, Cubilla AL, Humphrey PA, et al. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. Eur Urol 2016;70:93-105.
4. Trapani D, Curigliano G, Alexandru E, et al. The global landscape of drug development for kidney cancer. Cancer Treat Rev 2020;89:102061.
5. Nikolaou M, Pavlopoulou A, Georgakilas AG, et al. The challenge of drug resistance in cancer treatment: a current overview. Clin Exp Metastasis 2018;35:309-18.
6. Bedke J, Gauler T, Grünwald V, et al. Systemic therapy in metastatic renal cell carcinoma. World J Urol 2017;35:179-88.
7. Kumagai Y, Nio-Kobayashi J, Ishida-Ishihara S, et al. The intercellular expression of type-XVII collagen, laminin-332, and integrin-β1 promote contact following during the collective invasion of a cancer cell population. Biochem Biophys Res Commun 2019;514:1115-21.
8. Wang Y, Zheng K, Chen X, et al. Bioinformatics analysis identifies COL1A1, THBS2 and SPP1 as potential predictors of patient prognosis and immunotherapy response in gastric cancer. Biosci Rep 2021;41:BSR20202564.
9. Wang H, Liu Z, Li A, et al. COL4A1 as a novel oncogene associated with the clinical characteristics of malignancy predicts poor prognosis in glioma. Exp Ther Med 2021;22:1224.
10. Duan Y, Liu G, Sun Y, et al. COL6A3 polymorphisms were associated with lung cancer risk in a Chinese population. Respir Res 2019;20:143.
11. Yang W, Wu X, Zhou F. Collagen Type X Alpha 1 (COL10A1) Contributes to Cell Proliferation, Migration, and Invasion by Targeting Prolyl 4-Hydroxylase Beta Polypeptide (P4HB) in Breast Cancer. Med Sci Monit 2021;27:e928919.
12. Su Y, Fu J, Du J, et al. First-line treatments for advanced renal-cell carcinoma with immune checkpoint inhibitors: systematic review, network meta-analysis and cost-effectiveness analysis. Ther Adv Med Oncol 2020;12:1758835920950199.
13. Rømer AM, Thorseth ML, Madsen DH. Immune Modulatory Properties of Collagen in Cancer. Front Immunol 2021;12:791453.
14. Kaur A, Ecker BL, Douglass SM, et al. Remodeling of the Collagen Matrix in Aging Skin Promotes Melanoma Metastasis and Affects Immune Cell Motility. Cancer

Discov 2019;9:64-81.
15. Sadjadi Z, Zhao R, Hoth M, et al. Migration of Cytotoxic T Lymphocytes in 3D Collagen Matrices. Biophys J 2020;119:2141-52.
16. Zou Y, Hu C. A 14 immune-related gene signature predicts clinical outcomes of kidney renal clear cell carcinoma. PeerJ 2020;8:e10183.
17. Wang Y, Yang J, Zhang Q, et al. Extent and characteristics of immune infiltration in clear cell renal cell carcinoma and the prognostic value. Transl Androl Urol 2019;8:609-18.
18. Franz M, Rodriguez H, Lopes C, et al. GeneMANIA update 2018. Nucleic Acids Res 2018;46:W60-4.
19. Sherman BT, Hao M, Qiu J, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res 2022. [Epub ahead of print]. doi: 10.1093/nar/gkac194.
20. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 2007;9:166-80.
21. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401-4.
22. Vasaikar SV, Straub P, Wang J, et al. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res 2018;46:D956-63.
23. Modhukur V, Iljasenko T, Metsalu T, et al. MethSurv: a web tool to perform multivariable survival analysis using DNA methylation data. Epigenomics 2018;10:277-88.
24. Li T, Fu J, Zeng Z, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res 2020;48:W509-14.
25. Liu CJ, Hu FF, Xia MX, et al. GSCALite: a web server for gene set cancer analysis. Bioinformatics 2018;34:3771-2.
26. Shi X, Zhou X, Yue C, et al. A Five Collagen-Related Gene Signature to Estimate the Prognosis and Immune Microenvironment in Clear Cell Renal Cell Cancer. Vaccines (Basel) 2021;9:1510.
27. Gao S, Yan L, Zhang H, et al. Identification of a Metastasis-Associated Gene Signature of Clear Cell Renal Cell Carcinoma. Front Genet 2021;11:603455.
28. Rossi SH, Klatte T, Usher-Smith J, et al. Epidemiology and screening for renal cancer. World J Urol 2018;36:1341-53.
29. Pan H, Ding Y, Jiang Y, et al. LncRNA LIFR-AS1 promotes proliferation and invasion of gastric cancer cell via miR-29a-3p/COL1A2 axis. Cancer Cell Int 2021;21:7.
30. Shi Y, Zheng C, Jin Y, et al. Reduced Expression of METTL3 Promotes Metastasis of Triple-Negative Breast Cancer by m6A Methylation-Mediated COL3A1 Up-Regulation. Front Oncol 2020;10:1126.
31. Zhao B, Song X, Guan H. CircACAP2 promotes breast cancer proliferation and metastasis by targeting miR-29a/b-3p-COL5A1 axis. Life Sci 2020;244:117179.
32. Ding YL, Sun SF, Zhao GL. COL5A2 as a potential clinical biomarker for gastric cancer and renal metastasis. Medicine (Baltimore) 2021;100:e24561.
33. Zhang X, Gao C, Liu L, et al. DNA methylation-based diagnostic and prognostic biomarkers of nonsmoking lung adenocarcinoma patients. J Cell Biochem 2019;120:13520-30.
34. Fujiwara S, Nagai H, Jimbo H, et al. Gene Expression and Methylation Analysis in Melanomas and Melanocytes From the Same Patient: Loss of NPM2 Expression Is a Potential Immunohistochemical Marker for Melanoma. Front Oncol 2019;8:675.
35. Liu H, Xie HQ, Zhao Y, et al. DNA methylation-mediated down-regulation of TMEM130 promotes cell migration in breast cancer. Acta Histochem 2021;123:151814.
36. Zhang M, Wu J, Zhong W, et al. DNA-methylation-induced silencing of DIO3OS drives non-small cell lung cancer progression via activating hnRNPK-MYC-CDC25A axis. Mol Ther Oncolytics 2021;23:205-19.
37. Zhang K, Zhai Z, Yu S, et al. DNA methylation mediated down-regulation of ANGPTL4 promotes colorectal cancer metastasis by activating the ERK pathway. J Cancer 2021;12:5473-85.
38. Zhulai G, Oleinik E. Targeting regulatory T cells in anti-PD-1/PD-L1 cancer immunotherapy. Scand J Immunol 2022;95:e13129.
39. Harano K, Wang Y, Lim B, et al. Rates of immune cell infiltration in patients with triple-negative breast cancer by molecular subtype. PLoS One 2018;13:e0204513.
40. Peng DH, Rodriguez BL, Diao L, et al. Collagen promotes anti-PD-1/PD-L1 resistance in cancer through LAIR1-dependent CD8+ T cell exhaustion. Nat Commun 2020;11:4520.

Cite this article as: Guo L, An T, Huang Z, Wan Z, Chong T. Comprehensive analysis of the collagen family members as prognostic markers in clear cell renal cell carcinoma. Transl Cancer Res 2022;11(7):1954-1969. doi: 10.21037/tcr-22-398

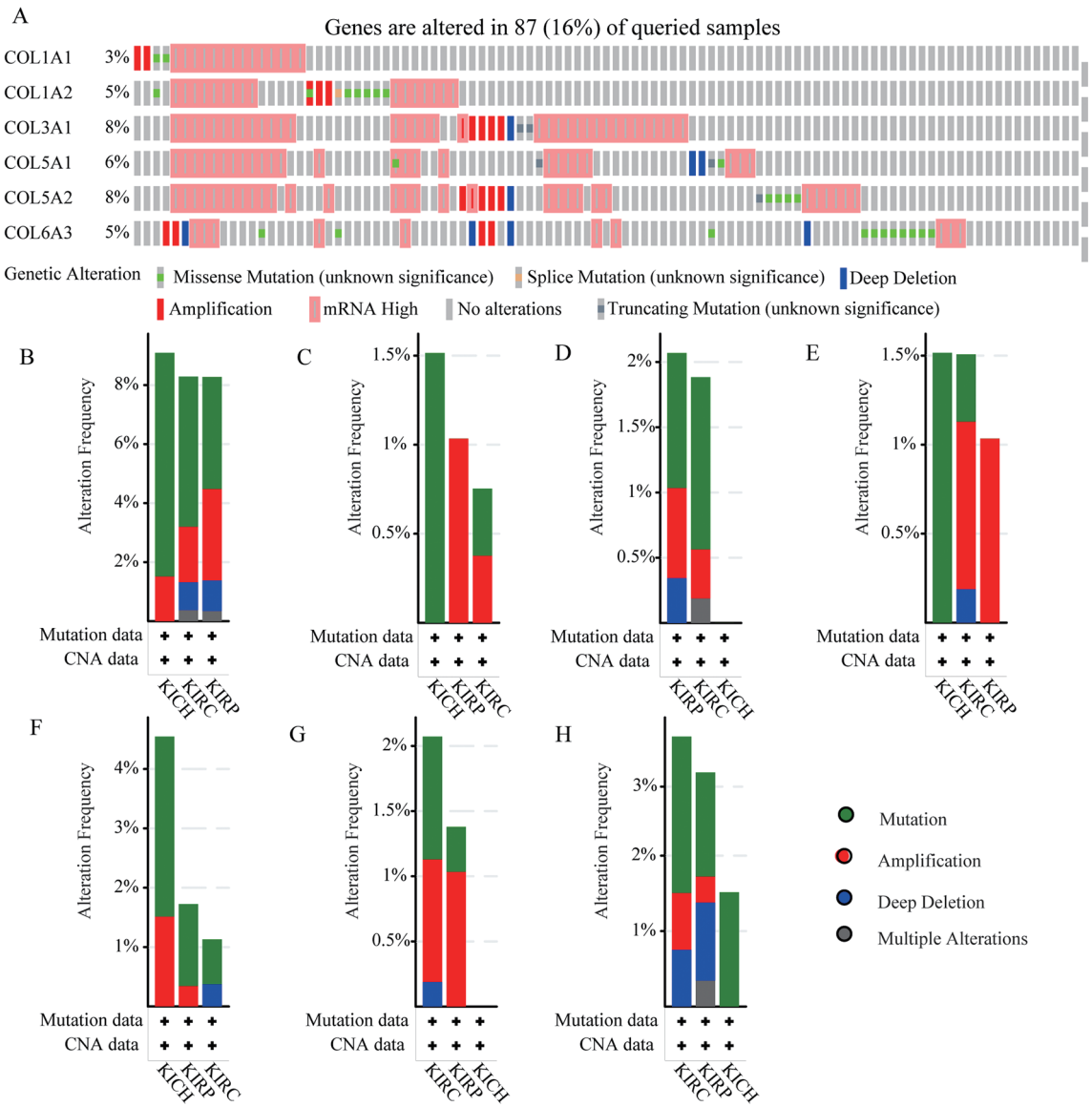

Figure S1 Collagen gene mutation analysis in ccRCC (cBioPortal). (A) Collagen gene mutation in renal cell carcinoma. (B) collagen family mutation. (C-H) Different collagens mutation. (C) COL1A1. (D) COL1A2. (E) COL3A1. (F) COL5A1. (G) COL5A2. (H) COL6A3.

https://dx.doi.org/10.21037/tcr-22-398

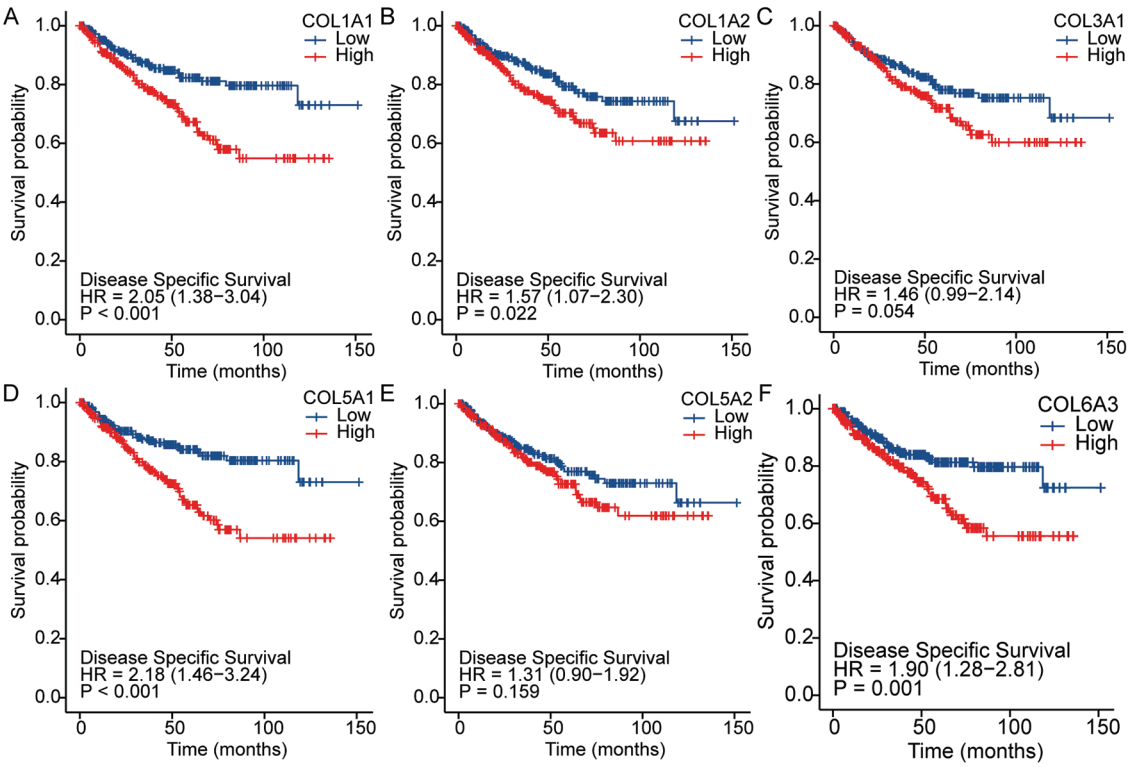

Figure S2 The prognostic value of collagen family in ccRCC (KM plotter). (A) COL1A1. (B) COL1A2. (C) COL3A1. (D) COL5A1. (E) COL5A2. (F) COL6A3. HR, Hazard Ratio; DSS, disease-specific survival.

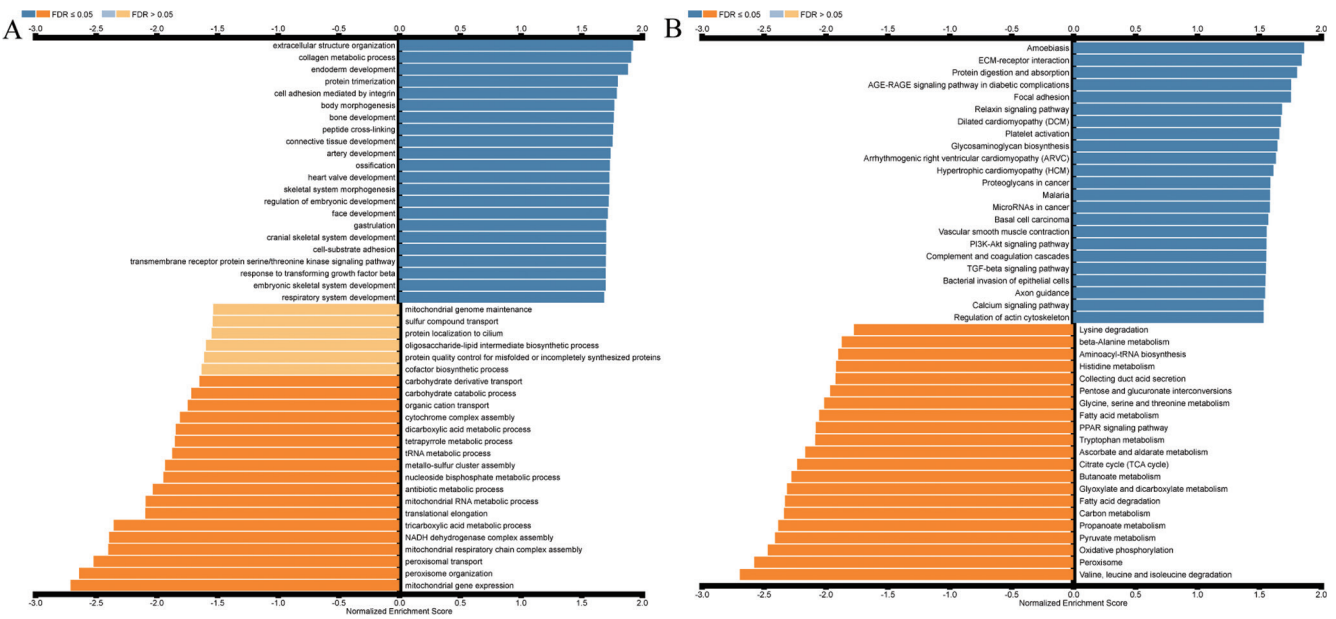

Figure S3 GSEA analysis of COL5A1 (LinkedOmics). (A) GO (Biological Process) analysis of the genes significantly correlated with COL5A1. (B) KEGG analysis of the genes significantly correlated with COL5A1.

https://dx.doi.org/10.21037/tcr-22-398

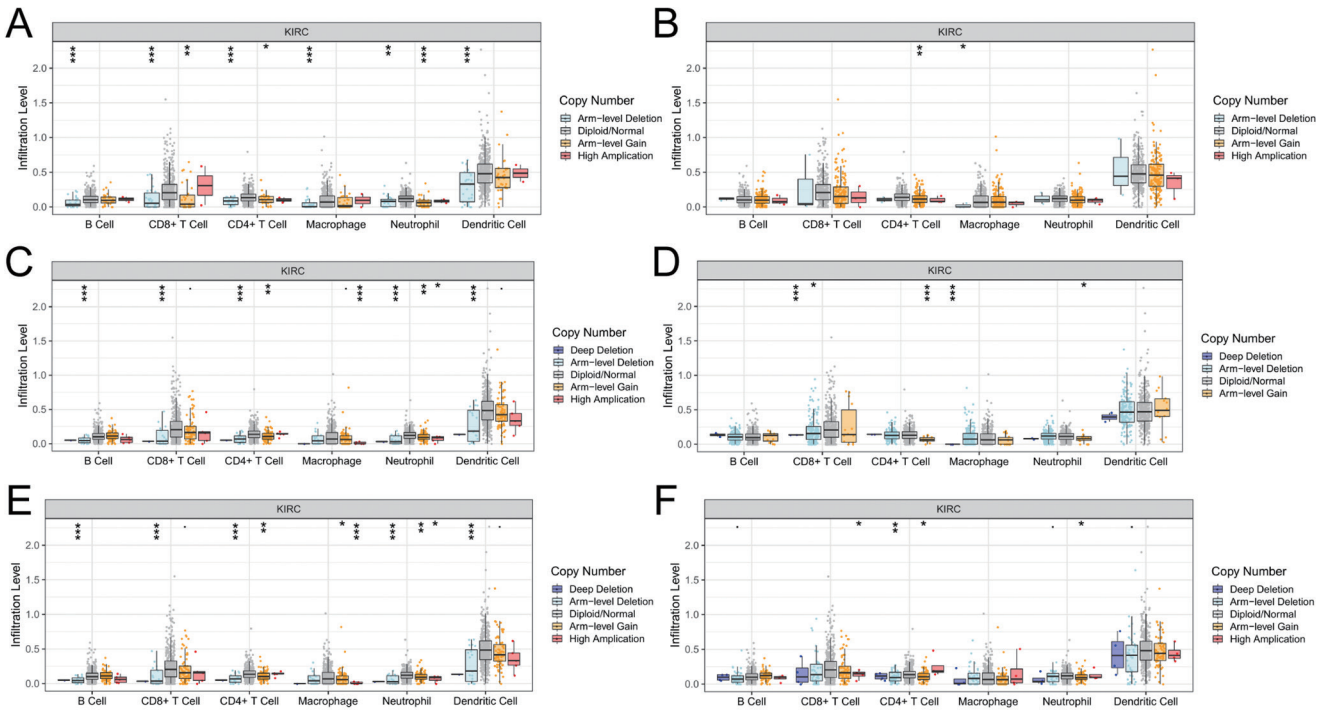

Figure S4 Somatic copy numbers' alterations in the collagen family members (TIMER). (A) COL1A1. (B) COL1A2. (C) COL3A1. (D) COL5A1. (E) COL5A2. (F) COL6A3. *, P<0.05; **, P<0.01; ***, P<0.001.

https://dx.doi.org/10.21037/tcr-22-398

Figure S5 Correlation analysis between collagen family and tumor drug resistance (GSCALite).

https://dx.doi.org/10.21037/tcr-22-398