A Five Collagen-related Gene Signature to Estimate the Prognosis and Immune Microenvironment in Clear Cell Renal Cell Cancer

Xiaokai Shi
Department of Urology, the affiliated changzhou No.2 people's hospital of nanjing medical university
Xiao Zhou
Department of Oncology, the affiliated changzhou No.2 people's hospital of nanjing medical university
Chuang Yue
Department of Urology, the affiliated changzhou No.2 people's hospital of nanjing medical university
Chao Lu
Department of Urology, the affiliated changzhou No.2 people's hospital of nanjing medical university
Shenglin Gao
Department of Urology, the affiliated changzhou No.2 people's hospital of nanjing medical university
Li Zuo (zuoli@njmu.edu.cn)
Department of Urology, the affiliated changzhou No.2 people's hospital of nanjing medical university
Zhiqin Sun
School of clinical medicine, the affiliated changzhou No.2 people's hospital of nanjing medical university
Lifeng Zhang
Department of Urology, the affiliated changzhou No.2 people's hospital of nanjing medical university

Research Article

Keywords: collagen, epithelial-mesenchymal transition (EMT), GSEA, immune microenvironment, prognosis

Posted Date: July 14th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-602578/v1

Version of Record: A version of this preprint was published at Vaccines on December 20th, 2021. See the published version at https://doi.org/10.3390/vaccines9121510.

A five collagen-related gene signature to estimate the prognosis and immune microenvironment in clear cell renal cell cancer

Xiaokai Shi[1,#], Xiao Zhou[2,#], Chuang Yue[1], Chao Lu[1], Shenglin Gao[1], Li Zuo[1,*], Zhiqin Sun[3,*], Lifeng Zhang[1,*]

[1] Department of Urology, The Affiliated Changzhou No.2 People's Hospital of Nanjing Medical University, Changzhou, 213000, China

[2] Department of Oncology, The Affiliated Changzhou No.2 People's Hospital of Nanjing Medical University, Changzhou, 213000, China

[3] School of Clinical Medicine, The Affiliated Changzhou No.2 People's Hospital of Nanjing Medical University, Changzhou, 213000, China

[#]Xiaokai Shi and Xiao Zhou contribute equally to this work and should be considered co-first authors

[*]Corresponding author:

Name: Li Zuo (First corresponding author)

Address: Department of Urology, The Affiliated Changzhou No.2 People's Hospital of Nanjing Medical University, Changzhou, 213000, China

E-mail: zuoli1978@hotmail.com

TEL: +8615051957700

Name: Zhiqin Sun

Address: School of clinical medicine, The Affiliated Changzhou No. 2 People's Hospital of Nanjing Medical University, Changzhou, Jiangsu, China

E-mail: 1257509030@qq.com

TEL: +8613861285688

Name: Lifeng Zhang

Address: Department of Urology, The Affiliated Changzhou No.2 People's Hospital of Nanjing Medical University, Changzhou, 213000, China

E-mail: nj-likky@163.com

TEL: +8613651505691

Abstract

Background: Collagen is the main component of Extracellular matrix (ECM) and might play an important role in tumor microenvironment. However, the relationship between collagen and clear cell renal cell cancer (ccRCC) still not fully clarified. Hence, we aimed to establish a collagen-related signature to predict the prognosis and estimate the tumor immune microenvironment in ccRCC patients.

Results: In this study, we established a five collagen-related gene signature to estimate the immune microenvironment and predict the prognosis of ccRCC patients. Patients with high-risk score were often correlated with unfavorable overall survival (OS) and immunosuppressive microenvironment. In addition, the collagen-related genetic signature was highly correlated with clinical pathological features and can be considered as independent prognostic factors in ccRCC patients. Besides, GSEA results also show that patients with high-risk grade tend to be associated with epithelial-mesenchymal junctions (EMT) and immune responses.

Conclusion: In this study, we developed a collagen-related gene signature, which might possess the potential to predict the prognosis and immune microenvironment of ccRCC patients and function as an independent prognostic factor in ccRCC.

Keywords: collagen; epithelial-mesenchymal transition (EMT); GSEA; immune microenvironment; prognosis.

Background

Renal cell carcinoma is the most common kidney malignant tumor, accounting for about 2% to 3% of adult malignant tumors. Among them, clear cell renal cell carcinoma is the most common type of renal cell carcinoma with its morbidity reached up to 70% - 80% [1]. In recent years, the

incidence and fatality rate of renal clear cell carcinoma have been increasing over the world [2]. Nowadays, due to the widespread use of imaging technology, the early diagnosis rate of renal clear cell carcinoma has increased significantly. However, about 1/3 of renal clear cell patients have been accompanied by distant metastasis at the time of initial diagnosis [3]. Currently, surgical resection is the main treatment for early localized clear cell renal carcinoma, but even with radical or partial nephrectomy, local or distant metastases still occur in 16% of patients [4]. For advanced metastatic clear cell renal cell carcinoma, because patients are not sensitive to radiotherapy and chemotherapy, the main treatment is targeted therapy and immunotherapy. However, 30% of patients with metastatic clear cell renal cell carcinoma have primary drug resistance toward molecularly targeted drugs and some patients will develop secondary drug resistance about 1 year after receiving treatment, which ultimately leads to poor prognosis of patients [5]. Therefore, it is of great significance to explore the molecular mechanisms related to the progression of renal clear cell carcinoma for improving the diagnosis and prognosis of patients with renal clear cell carcinoma.

Over the past century, researches of malignant tumor have been focused on the mechanism of abnormal proliferation of tumor cells. Therefore, many genes and pathways related to cancer cell growth and metabolism have been elucidated. In recent years, the seed-soil theory was proposed and elucidate the development process of tumors which is a dynamic process of interaction between tumor cells and their microenvironment [6–8]. The tumor microenvironment is involved in the re-differentiation of tumor cells and extracellular matrix (ECM), which play an important role in tumor development. [9]. These researches opened up a new horizon for exploring the complex mechanisms of tumor progression. Collagen is the main component of ECM. As the attachment and scaffold for cell growth, it can induce the proliferation, differentiation and migration of epithelial cells, and plays

an important role in maintaining intercellular adhesion, tissue integrity and repairing as well as supporting organs. Accumulating evidence showed that the expression of collagen I, III, IV, V, VI, and X in gastric cancer and bladder cancer tissues is with significantly difference [10–14], implying that collagen is significantly correlated with tumorigenesis and development.

In this study, we intended to establish a collagen-related gene signature which might serve as a prognostic biomarker to assist the diagnosis and prediction of prognosis in clear renal cell cancer. This risk model might also used to predict the immune microenvironment in ccRCC groups. These results will deepen our understanding between collagen and ccRCC, and immune microenvironment.

Results

1. Characteristics of patients separated into training and test cohort

Clinical and pathological information of 537 ccRCC patients and 72 normal tissue samples were obtained from the cancer genome atlas (TCGA, https://portal.gdc.cancer.gov/) database. All patients were randomly separated into train cohort (n=320) and test cohort (n=217). Then, we summarized the clinical information including age at diagnosis, gender, OS, survival state, histological stage, pathological stage and clinical stage in **Table 1.**

Table 1. Clinical characteristics of the RCC patients in Train and Test cohorts

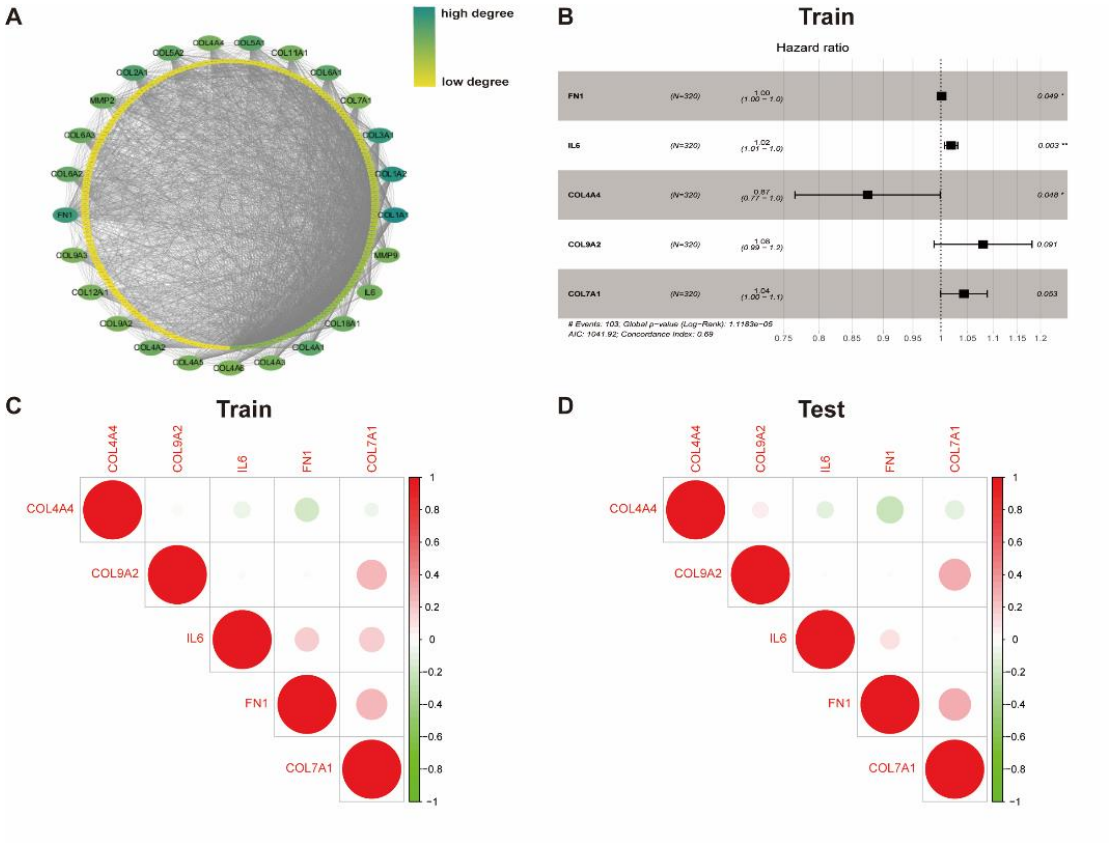
Variables	Train (n=320)	Test (n=217)	Overall (n=537)
Age			
Mean	60.03	61.42	61.52
Median[min,max]	60.00[29,90]	61[26,90]	61[26,90]
Gender			
MALE	209(65.31%)	137(63.13%)	346(64.43%)
FEMALE	111(34.69%)	80(36.87%)	191(35.57%)
Overall Survival time			
Mean	1130.68	1146.57	1147.79
Median[min,max]	1026.50[3,3431]	1106.00[2,3668]	1091[2,3668]
Unknow	3	2	5
Survival State			
Alive	217 (67.81%)	150(69.12%)	367(68.34%)

Dead	103(32.19%)	67(30.88%)	170(31.66%)
Histologic grade			
G1	10(3.13%)	4(1.84%)	14(2.61%)
G2	133(41.56%)	97(44.70%)	230(42.83%)
G3	127(39.69%)	80(36.87%)	207(38.55%)
G4	45(14.06%)	33(15.21%)	78(14.53%)
Unknow	5(1.56%)	3(1.38%)	8(1.49%)
T-stage			
T1	159(49.69%)	116(53.46%)	275(51.21%)
T2	43(13.44%)	26(11.98%)	69(12.85%)
T3	111(34.69%)	71(32.72%)	182(33.89%)
T4	7(2.19%)	4(1.84%)	11(2.05%)
Unknow	0(0%)	0(0%)	0(0%)
N-stage			
N1	12(3.75%)	5(2.30%)	17(3.17%)
N0	134(41.88%)	106(48.85%)	240(44.69%)
Unknow	174(54.38%)	106(48.85%)	280(52.14%)
M-stage			
M1	48(15.00%)	31(14.29%)	79(14.71%)
M0	256(80.00%)	170(78.34%)	426(79.33%)
Unknow	16(5.00%)	16(7.37%)	32(5.69%)
Clinical stage			
Stage I	155(48.44%)	114(52.53%)	269(50.09%)
Stage II	34(10.63%)	23(10.60%)	57(10.61%)
Stage III	79(24.69%)	46(21.20%)	125(23.28%)
Stage IV	50(15.63%)	33(15.21%)	83(15.46%)
Unknow	2(0.63%)	1(0.46%)	3(0.56%)

2. Establishment of a Collagen-Related risk model

The collagen-related gene set, which contained 257 genes **(Table S1)**, was adopted from the Molecular Signature Database (MsigDB, http://www.gsea-msigdb.org/gsea/msigdb/index.jsp). For better comprehension, we embedded 257 collagen-related genes into STRING online database (https://string-db.org/) to construct a protein-protein interaction network. Based on the interaction degrees, the Cytoscape was used to re-visualize the interaction network and screen out the top 25 hub genes with highest degree (**Figure1 A**), suggesting their critical relationship with collagen.

We sought to construct a collagen-related risk model which can predict the prognosis of ccRCC patients using single and multiple stepwise regression analyses based on top 25 collagen-related



genes (10 percent of total genes) in the training cohort. In the single factor regression, 16 genes with statistic significant were correlated with patients' OS **(Table S2).** Subsequently, we performed a multivariate cox regression with these genes and generated a five collagen-related signature containing IL6, FN1 and three genes encoding collagen (COL4A4, COL9A2, COL7A1) to predict the prognosis of ccRCC patients **(Figure 1B)**. Each patient in this study obtained a risk-score which was calculated with the following formula:

Risk-score=(0.0015×FN1) + (0.019×IL6) + (-0.1338×COL4A4) + (0.0772×COL9A2) + (0.0422×COL7A1).

All six FRGs were not significantly correlated with each other in both Train and Test cohorts, indicating this risk model avoided the overfitting caused by collinearity (**Figure 1C,D**).

Figure1. Identification of collagen-related signature to predict prognosis of ccRCC. **(A)** 25 hub genes dragged

out from 257 collagen-related genes based on interaction degrees; **(B)** Establishment of a collagen-related risk model by univariate and multivariate cox regression; **(C,D)** Spearman correlation analysis of five collagen-related gene in train and test cohort.

Then, we searched in Tumor IMmune Estimation Resource (TIMER, https://cistrome.shinyapps.io/timer/) database to preview the expression of these five collagen-related genes in pan-cancer. **Figure 2A-E** demonstrated that three collagen-related genes were found to be differently expressed in genitourinary tumors like bladder cancer (BLCA), clear cell renal cell cancer (ccRCC/KIRC) and prostate cancer (PRAD). Besides, the expression of FN1 were found to be significantly differently expressed in KIRC and PRAD while IL6 were significantly differently expressed in BLCA, KIRC and PRAD (P<0.1 was considered statistically significant according to the TIMER database).

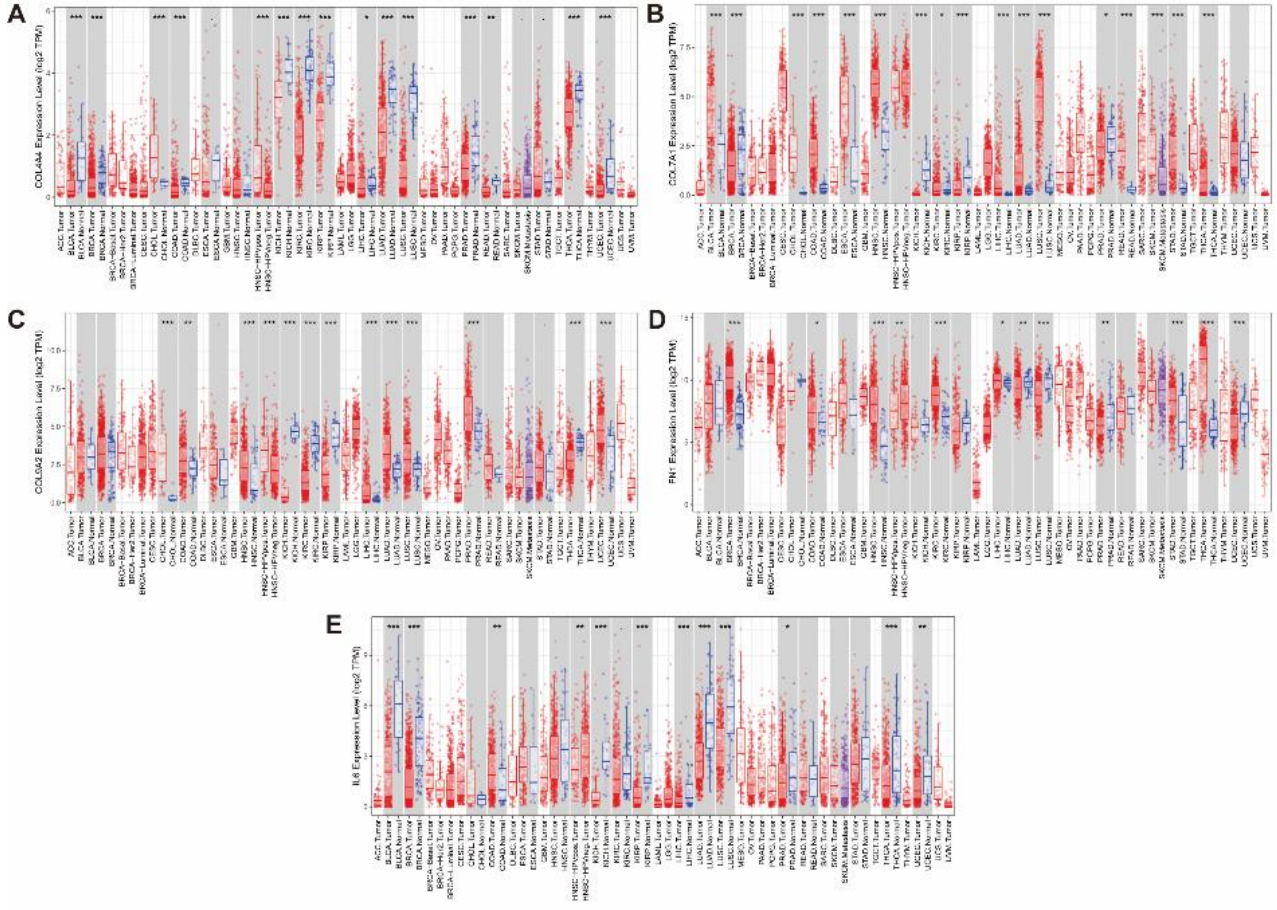


Figure2. The gene expression of five collagen-related genes in TIMER database.

3. Prognostic Value of the Collagen-Related Signature in ccRCC Patients

Having developed the five collagen-related risk model, samples in the training cohort and test cohort were assigned a risk score. Then, the medium value of risk scores in the training cohort was set as a cutoff to judge and separate patients into high-risk and low-risk groups in both training and test cohort **(Figure 3A,B)**.

As the main component of extracellular matrix, collagen is closely related to the degree of tumor malignancy, whether it increases or decreases. Hence, the prognostic significance of the collagen-related signature was further detected. As demonstrated in the heatmap **(Figure 3C,D)**, the expression level of COL4A4 was increased in the low-risk group while the expression level of remaining four genes were increased in the high-risk group in both Train and Test cohort. Compared to the low-risk group, the fatality rate in the high-risk group was significantly higher in both Training and Testing cohort **(Figure 3E,F)**. Furthermore, the implications for prognosis of the collagen-related risk model in ccRCC were assessed with Kaplan-Meier analysis. According to **Figure 3G**, patients with high-risk scores tend to obtain unfavorable overall survival in training cohort and this result were confirmed by test cohort **(Figure 3H).**

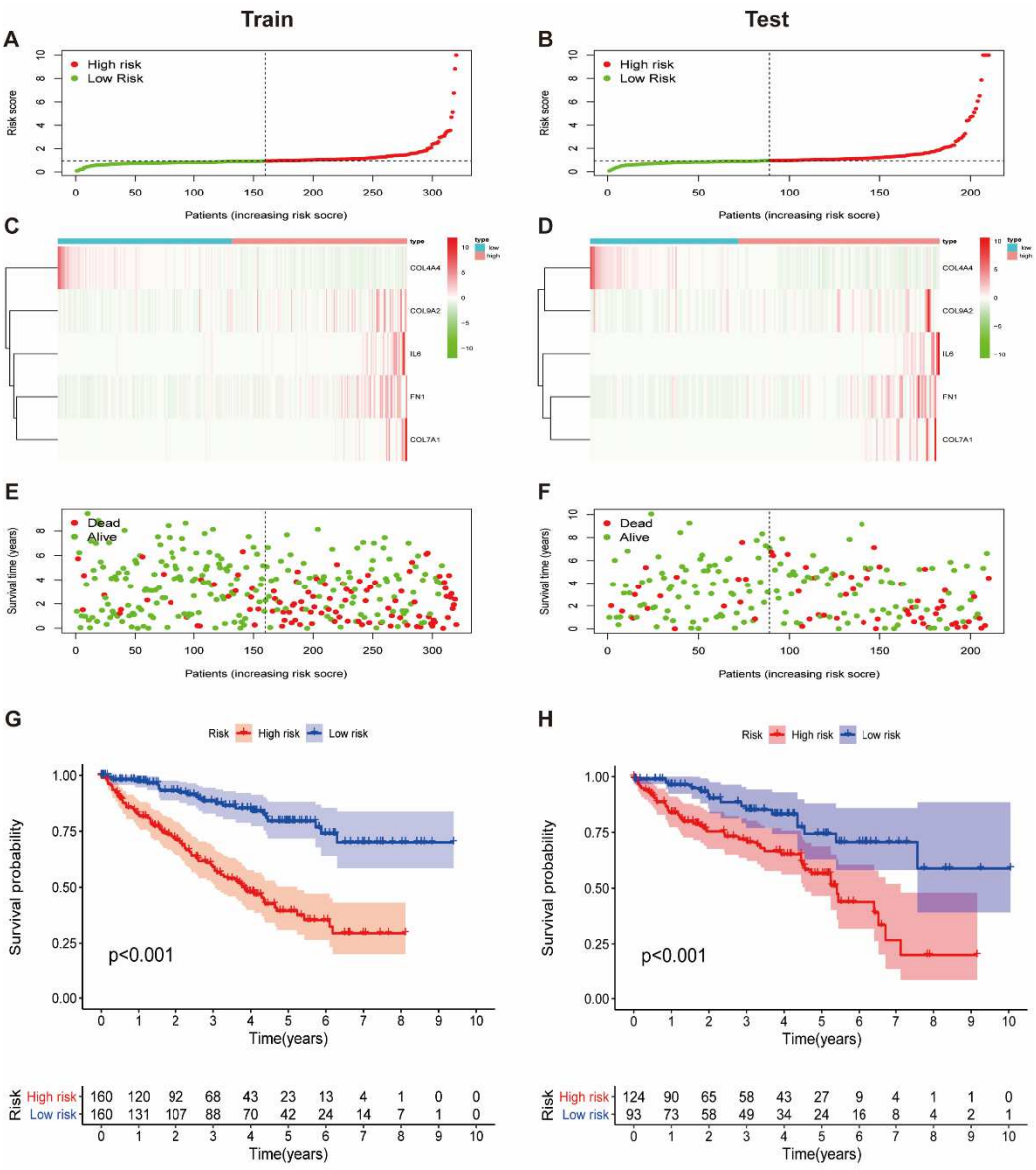
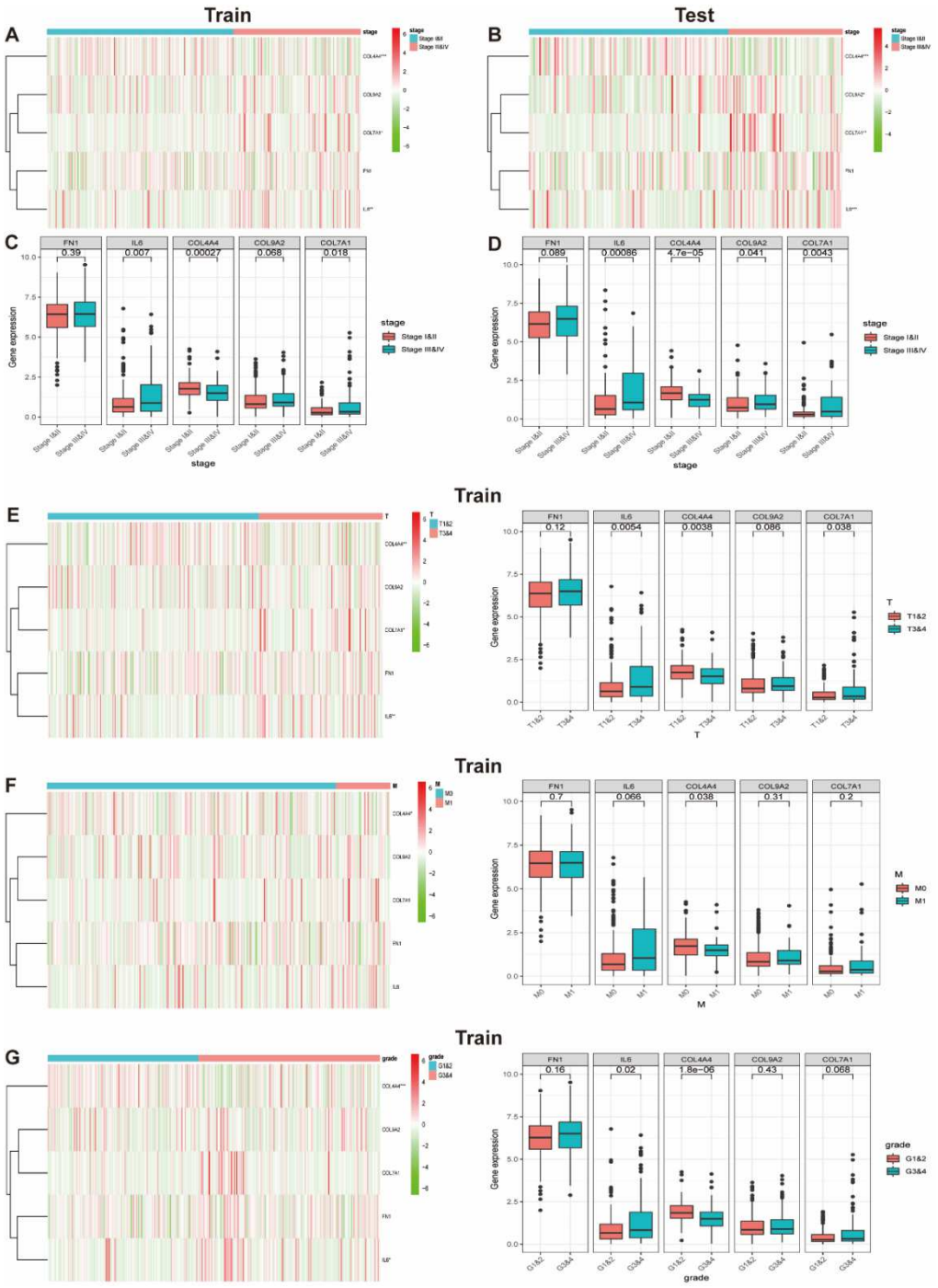


Figure3. Prognosis value of collagen-related signature in train and test cohort. **(A,B)** Risk curve of ccRCC patients in train and test cohort. The median risk score in Train cohort was set as the cutoff value to separate patients into high and low risk groups. The cutoff value in train cohort was also used to calculate the risk score of patients in Test cohort; **(C,D)** The heatmap showing five hub gene expression profiles in high and low risk groups from Train and Test cohort; **(E,F)** Patient status distribution in high and low risk groups; **(G,H)** The Kaplan-Meier overall survival curves for patients assigned to high and low risk groups based on the risk score.

4. The expression of Collagen-related genes is associated with clinical and pathological characteristics.

In view of the significant biological function of collagen in the occurrence and development of tumors, we comprehensively analyzed the relationship between the five collagen-related genes and the clinicopathological characteristics of ccRCC, including clinical stage and WHO grades. In both training and test cohort, the gene expression level of COL4A4 is increased in low clinical stage groups while the expression level of COL7A1 and IL6 are increased in high clinical stage groups (**Figure 4A,B**), implying that COL4A4 might act as a critical protective factor while another two genes are risk factors in tumor development. Quantitative analysis were also performed and confirmed a significant association between tumor stage and mRNA expression in Training **(Figure 4C)** and Test **(Figure 4D)** cohort. Furthermore, we investigated the association between gene expression and T, M stage and WHO grade in training cohort, separately. The findings revealed that the mRNA expression of COL4A4 was stably increased in low T and M stage group as well as low WHO grade group **(Figure 4E-G)**, and were further validated in the test cohort **(Figure S1A-C)**,

159 implying COL4A4 might act be an important predictive factor in tumor development.



160 **Figure4.** Collagen-related gene expression is correlated with clinicopathological features of ccRCC. **(A,B)** The

161 heatmap showing five collagen-related gene expression profiles in different clinical stages from train and test cohorts;

162 **(C,D)** The expression levels of five collagen-related genes in ccRCC with different clinical stages; **(E-G)** The

163 heatmap and expression levels of five collagen-related genes in different T-stage, M-stage and WHO grades from

train and test cohorts; *P<0.05,**P<0.01 and ***P<0.001

5. The Accuracy of Collagen-Related Signature for prognosis evaluation

To estimate the accuracy in overall survival prediction of the collagen-related signature, the received operating characteristic (ROC) curves were analyzed based on datasets from the training and test cohorts. The area under the ROC curve at 1 year and 5 years was up to 0.758 and 0.729, separately, indicting a relatively high accuracy of prognosis prediction **(Figure 5A)**. Results were subsequently confirmed in the test cohort **(Figure 5B)**.

Moreover, the single and multiple stepwise regression analyses were further adopted to estimate the independent prognostic value of collagen-related signature. The single factor regression analysis revealed that patients with high-risk scores were associated with unfavorable overall survival **(Figure 5C)** as well as other evaluating indicators including age, WHO grade, clinical stage and pathological stage. This was further validated in test cohort **(Figure 5D)**. Moreover, multiple stepwise regression analysis demonstrated that high risk score was tended to be independently correlated with unfavorable overall survival in ccRCC patients **(Figure 5E)**, indicting an

independent prognostic value for ccRCC. This was also validated in the test cohort **(Figure 5F)**.

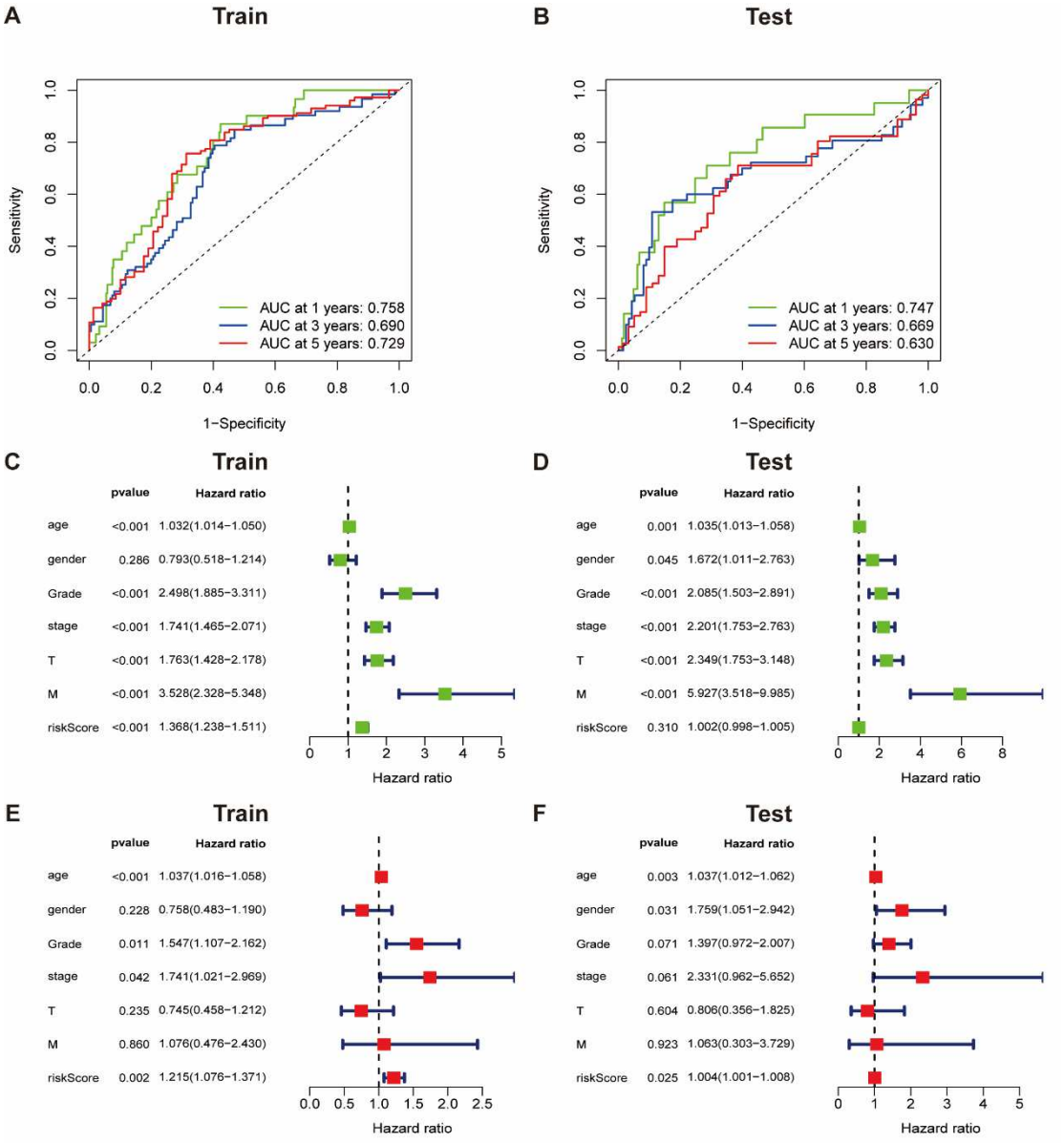
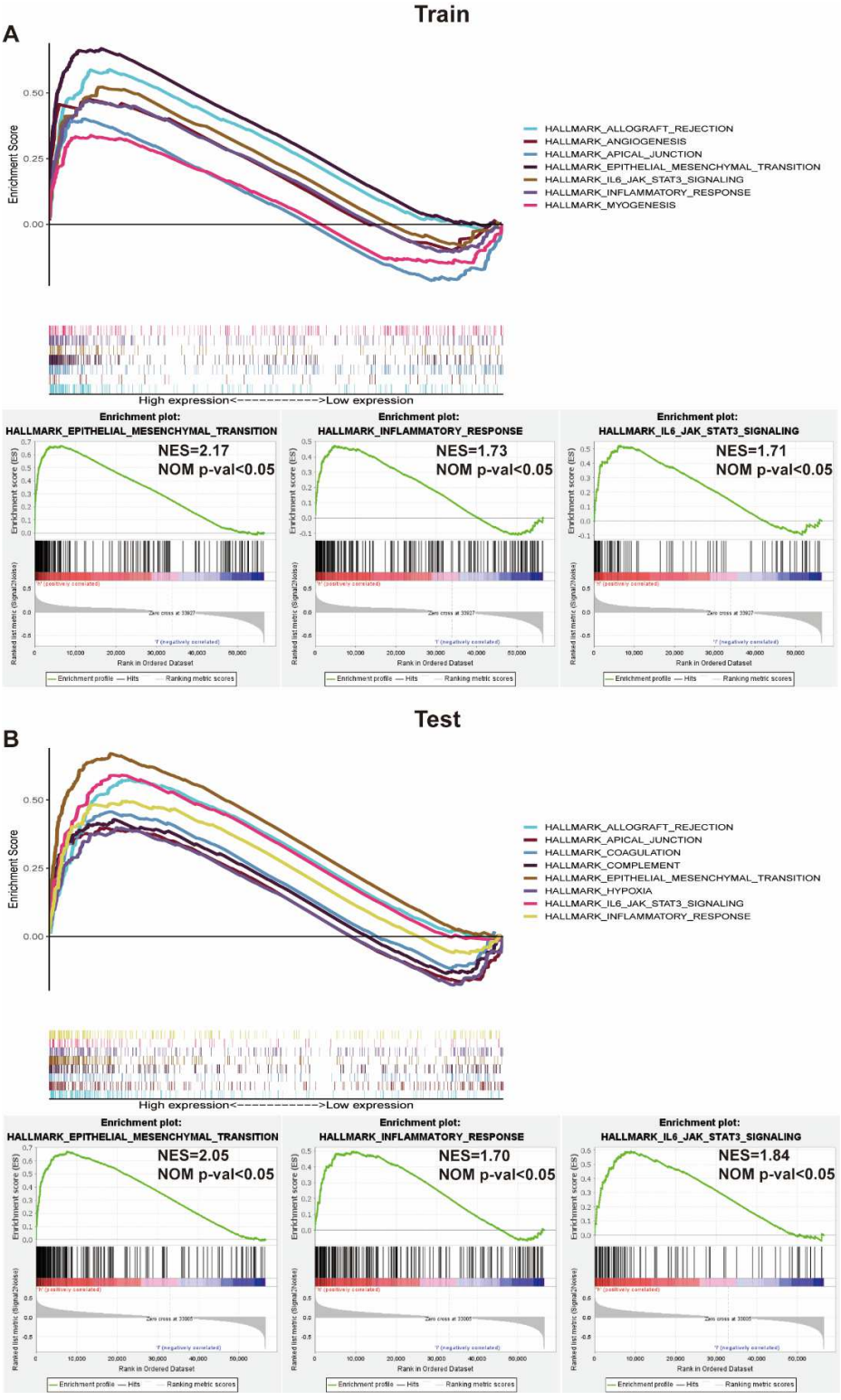


Figure5. Prognostic value of the collagen-related signature in ccRCC. **(A,B)** ROC curves showing the predictive efficiency of the collagen-related signature on the 1-,3- and 5-years survival rate; **(C-F)** Univariate and multivariate cox regression analysis evaluating the independent prognostic value of collagen-related signature in terms of OS in ccRCC patients.

6. GSEA Identifies Potential Signaling Pathways

GSEA was adopted to investigate the potential signaling pathways activated in high-risk group. Genes were differently concentrated in high-risk groups based on data from Training cohort, as they

were related to tumorigenesis and immune response, such as epithelial-mesenchymal transition, inflammatory response and IL6-JAK-STAT3 signaling pathway **(Figure 6A).** Similarly**,** this was also verified with test cohort (**Figure 6B**).



Figrue6. GSEA enrichment analysis between high and low risk groups. **(A)** The hallmark enrichment of high

and low risk groups by GSEA method; GSEA revealing that genes in the high-risk group were enriched for hallmarks of epithelial mesenchymal transition (EMT) and immune response in train cohort**; (B)** The results were further validated by the test cohort.

7. Identification of Immune cells Infiltrated in Patients with different Risk Scores

As immune response was identified activated in the high-risk group, we explored the possibility of a collagen-related signature in assessing the immune microenvironment. The CIBERSORT method was utilized to evaluate the discrepancies in the immune infiltration of 22 immune cell types between variable risk groups. **Figure 7A** summarized the findings acquired from 320 ccRCC patients in training cohort and 217 patients in test cohort. Patients in high-risk group possessed relatively higher ratio of regulatory T cells (Tregs) **(Figure 7B)** but lowers proportions of Macrophages M1 **(Figure 7C)**, indicating that patients with high-risk score might generate an immunosuppressive environment.

Moreover, GSEA was utilized to probe the connection between biological processes associated with immunity and collagen-related signature. Results suggested that high risk ccRCC samples were remarkably correlated with negative regulation of immunity pathway, such as negative regulation of immune response, $CD4^+$ αβ T cell activation, αβ T cell activation and T cell mediated immunity **(Figure 7D)**.

Consequently, to improve the efficiency of immunotherapy, collagen-related genes might be a critical target to focus on.

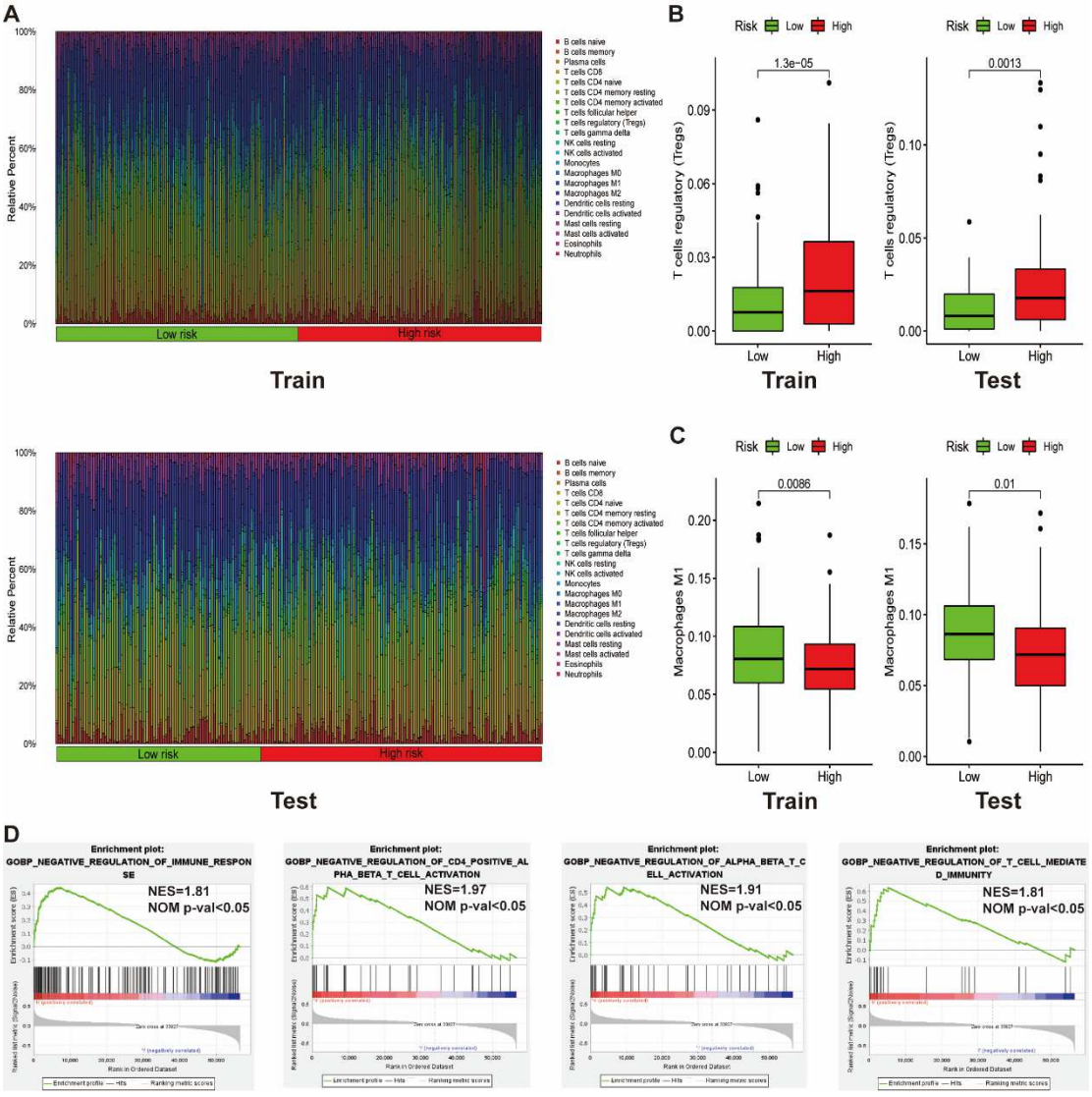


Figure7. Immune landscape between low and high risk ccRCC patients. **(A)** Relative proportion of immune infiltration in high and low risk patients. **(B-C)** Box plots visualizing significantly different immune cells between high-risk and low-risk patients. **(D)** GSEA demonstrating that collagen-related signature correlated with immune-related biological function.

8. The immune microenvironment of patients with high-risk scores tend to be suppressed

We compared expression of genes negatively regulating the processed in cancer immunity cycle between low-risk and high-risk cohorts. Gene symbols were obtained from Tracking Tumor

Immunophenotype website. According to **Figures 8A, B,** the majority genes which negatively regulate the tumor immune circulation were overexpressed in high-risk group, suggesting that the activity of immune response in these patients were suppressed.

As our previous findings revealed that the proportions of Tregs are increased in the high-risk group, the expression of molecules correlated with immune checkpoints were analyzed in both low and high-risk groups. Findings demonstrated that the expression of LAG3 and CTLA-4, which positively associated with collagen-related risk score, were upregulated in the high-risk group (**Figure 8C-F**). In addition, the mRNA level of other significant immune checkpoints such as PD1 and TIGIT were significantly upregulated in high-risk groups **(Figure 8G-H)** while PD-L1 was downregulated in high-risk groups (**Figure 8I**). Besides, two immunosuppressive cytokines

including IL10 and TGFB1 were also upregulated in high-risk groups **(Figure 8J)**.

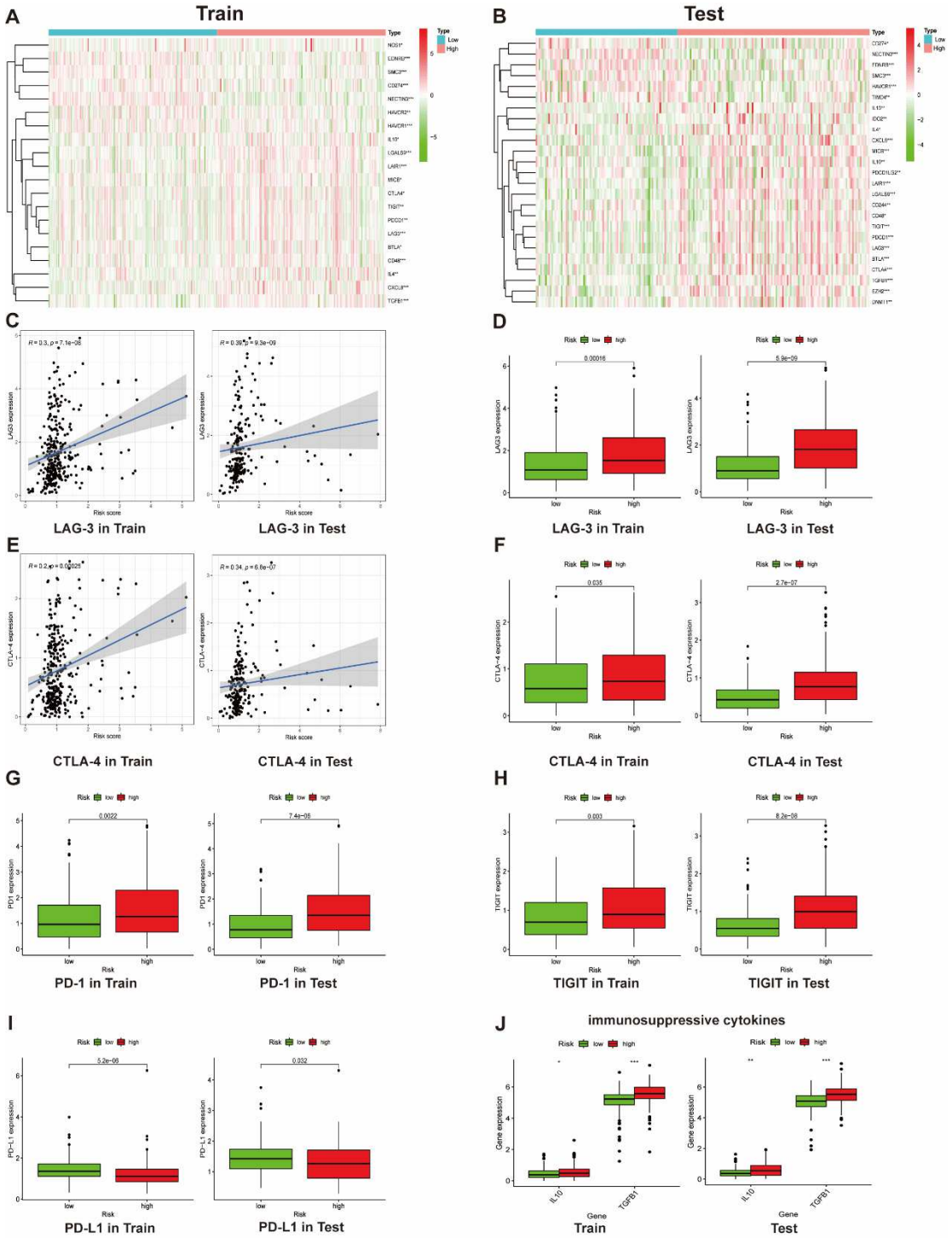


Figure8. High collagen-related risk score indicates an immunosuppressive microenvironment. **(A,B)** Heatmap of gene profiles involved in the negative regulation of the Cancer-immunity Cycle in high and low risk groups in train and test cohorts; **(C)** Correlation between LAG3 expression and risk score; **(D)** LAG3 expression in high and low risk groups; **(E)** Correlation between CTLA-4 expression and risk score; **(F)** CTLA-4 expression in high and

low risk groups; **(G-I)** PD1,TIGIT and PD-L1 expression in high and low risk groups; **(J)** Tumor immunosuppressive cytokines expression in high and low risk groups; $^{*}P<0.05$,$^{**}P<0.01$ and $^{***}P<0.001$.

Discussion

Evidence from previous researches have showed that collagen-related genes are aberrantly expressed in variable tumors [10, 12, 13, 15, 16]. However, the expression and roles of collagen-related genes in ccRCC were with limited information. Hence, firstly, in this study, we sorted out 257 collagen-related genes from the MSigDB and listed them in **Table S1**. Then, a PPI network was constructed and top 10 percent of these genes with supreme degrees of interaction were screened out. Subsequently, the single and multiple stepwise regression analyses were adopted to analyze the prognostic impact on ccRCC patients. Finally, a risk model which could predict ccRCC patients' prognosis based on five hub collagen-related genes were constructed. These findings promote the identification of novel biomarkers for the prediction of diagnosis and prognosis of ccRCC.

As is reported that several hub collagen-related genes including FN1, IL6, COL4A4 and COL7A1 were involved in the progression and development of variable cancers. Fibronectin 1 (FN1), an extracellular matrix glycoprotein, plays major roles in cell adhesion, migration, and differentiation [17, 18]. Importantly, FN1 is also the key mediator of carcinoma genesis and tumor metastasis, including in lung adenocarcinoma, gastric cancer, and brain glioblastoma [19]. Interleukin-6 (IL-6) which plays a significant role in cancer progression, is a pleiotropic factor that belongs to a cytokine subfamily [20]. Evidence showed that IL-6 could mediate down-regulation of type II collagen through JAK/STAT pathway [21]. Collagen Type IV Alpha 4 Chain (COL4A4), encodes one of the six subunits of type IV collagen, the major structural component of basement membranes. Mutations in this gene are associated with type II autosomal recessive Alport syndrome

(hereditary glomeruli nephropathy) and with familial benign hematuria (thin basement membrane disease) [22]. Also, COL4A4 was also found to be down-regulated in esophageal tumor tissues [23, 24]. COL7A1 gene encodes for collagen type VII, and was found aberrantly expressed in esophageal squamous cell carcinoma [25]. Besides, high levels of type VII collagen expression was found to be correlated with the migration and invasion of recessive dystrophic epidermolysis bullosa cutaneous squamous cell carcinoma keratinocytes [26]. Nevertheless, these genes were rarely reported in ccRCC. In this research, COL4A4, COL7A1 and IL6 were found with extraordinary expression and were found to be associated with clinicopathological features with statistic significant. In contrast with IL6 and COL7A1, the lower expression of COL4A4 was often correlated with lower clinical stage, pathological stage and WHO grade, indicting COL4A4 is a protecting factor while IL6 and COL7A1 are risk factors of ccRCC.

Moreover, based on the expression of five hub genes related with collagen, a risk model was established to forecast prognosis of ccRCC patients in both training and testing groups, KM curves demonstrated that high-risk group was tremendously correlated with unfavorable OS. Furthermore, ROC curves revealed that the signature of five collagen-related hub genes exist a significant prognosis value for discriminating ccRCC patients with unfavorable OS.

Furthermore, to investigate whether the molecular biology mechanism of the five collagen-related genes promoted clear cell renal carcinoma genesis and progression, ccRCC patients in train cohort were separated into high and low risk groups based on the median risk score. Results revealed that patients in high-risk groups were tend to correlate with epithelial-mesenchymal transition (EMT) and immune response. Recent studies have found that Tumor epithelial cells and their adjacent normal epithelial cells can be transformed into cancer-associated fibroblasts(CAFs) through EMT,

which can induce the invasion and migration of Tumor cells and promote the development of tumors [27, 28]. Activated CAFs can promote migration by secreting extracellular matrix components such as collagen glycotenin. Through expressing a series of growth factors and cytokines, VEGF and monocyte chemotactic protein1 (MCP1), it can further activate tumor matrix and promote the formation of microenvironment needed for tumor development [29].

Additionally, several researches revealed that Epithelial-to-Mesenchymal Transition contributes to generation of immunosuppressive microenvironment [30, 31]. Moreover, the high collagen density environment also tends to generate a immunosuppressive microenvironment [32]. Hence, from this perspective, we compared the proportion of immune cell infiltrated in high-risk and low-risk groups. Findings revealed that the content of regulatory T cells (tregs) in high-risk groups was relatively high while the proportion of M1 macrophages that contribute to the antitumor response was relatively low, indicating that the immune microenvironment of people with high-risk scores are likely to be suppressed. Results of GSEA based on the gene ontology gene set also revealed that patients in high-risk groups are negatively correlated with regulation of immune pathways, for example, negative regulation of immune response, $CD4^+$ αβ T cell activation, αβ T cell activation and T cell mediated immunity. These results implying that our risk model might have the potential to predict the immune microenvironment.

As cytokines function as a critical role in regulatory tumor immunity, we investigated the expression of two immunosuppressive cytokines including IL-10 and TGFB1. IL-10 helps maintain the expression of Foxp3 and TGF-β, thereby stabilizing the phenotype and function of Treg [33]. Besides, TGF-β contribute to the inhibition of NK cell activity and dendritic cell maturation as well as decreasing of cytokine production [34, 35]. Corresponding to these researches, the expression

297 level of IL-10 and TGF-β were found overexpressed in high risk group patients.

298 Except for immunosuppressive cytokines, immune checkpoints also play an important part in
299 tumorigenesis through promotion of tumor immunosuppressive effects. Tumors have the capability
300 to stimulate immune checkpoint targets such as LAG-3, CTLA-4, PD1, PD-L1, TIGIT and TIM-3
301 to protect themselves from attacking. It is reported that high expression of collagen could exhaust
302 the proportion of $CD8^+$ T cells through activation of LAIR1 receptor, which is upregulated following
303 CD18 interaction with collagen. Targeting impression of interaction between collagen and CD18
304 could enhance the sensitivity of Anti-PD-1 to lung cancer [36]. In our study, we analyzed the
305 correlation between these immune checkpoints and risk score between high and low risk groups.
306 Among them LAG-3 and CTLA-4 were found with relatively high positive correlation. Besides, the
307 expression of these genes was also found over expressed in high-risk groups.

308 Collectively, our risk model based on five collagen-related genes has a better prognostic value
309 for ccRCC patients and could also predict the immune environment of these patients. Targeting
310 immune checkpoints such as LAG-3 and CTLA-4 might contribute to the treatment of ccRCC.
311 Nevertheless, there are several shortcomings worth mentioning. At first, our risk signature was
312 merely constructed based on the data from TCGA database and was only validated with internal
313 data, which need to be further validated with external data and clinical patient cohort as well as
314 multi-center study. In the second place, further investigations, including in vivo and in vitro
315 experiments are requisite to elucidate the molecular biological mechanisms.

316 # Conclusion

317 To sum up, a prognostic risk signature consisting of five collagen-related genes was established,
318 which were closely related with clinicopathology and immune response. Our results demonstrated

new perspective for the individualized diagnosis of ccRCC patients.

Methods

1. Data obtaining and pre-processing

The transcriptome RNA sequence data (FPKM value) and corresponding clinical and pathological information of 537 ccRCC patients and 72 normal tissue samples can be possessed from the TCGA (https://portal.gdc.cancer.gov/) database. All patients were randomly separated into train cohort (n=320) and test cohort (n=217). Then, we summarized the clinical information including age at diagnosis, gender, OS, survival state, histological stage, pathological stage and clinical stage in **Table 1.** In addition, the Tumor IMmune Estimation Resource (TIMER, https://cistrome.shinyapps.io/timer/) database was used to obtain the expression of genes in pan-cancer. All parameters in this section were default.

2. Establishment of a Protein-protein interaction network

Firstly, we established a Protein-Protein Interaction Network (PPI) with STRING (https://stringdb.org) Online Database. Then, Cytoscape software (V3.7.2, https://cytoscape.org) was adopted to analyze the interaction degree of collagen-related genes and re-visualize the interaction network. To screen hub genes of PPI, the plug-in named Network analyzer was adopted to count nodes degree which characterized as the interaction degree.

3. Collagen-related risk model establishment.

The univariate cox regression was performed to figure out the relationship between overall survival (OS) and hub collagen-related genes. Hub collagen-related genes screened with statistic significant were then embedded into multiple regression to acquire the coefficients, the risk-score was acquired based on the following formula: $\sum_{i=1}^{N}(Expi * Coefi)$, where N=5, Exp_i was the

expression level of each collagen-related gene, $Coef_i$ was the corresponding coefficient obtained from multiple regression.

4. GSEA analysis and immune cell type fraction estimation

GSEA was used to investigate the discrepancy in the set of genes expressed between the high-risk and low-risk groups in the enrichment of the MSigDB Collection (h.all.v7.4.symbols.gmt; c5.go.v7.4.symbol.gmt). The Risk scores were used as the phenotype label. All other parameters are default. CIBERSORT (https://cibersort.stanford.edu/) is a useful instrument which have the capability to distinguish 22 immune cell types in a mixed cell population based on gene expression data. Here, the CIBERSORT was adopted to assess the fractions of 22 immune cell types.

5. Statistics

We used the R language software (V.4.0.2, https://cran.r-project.org) to handle this analysis in this study. Wilcoxon rank sum test was used to compare the mRNA levels of collagen related genes between ccRCC samples and normal samples. The hub genes in risk model were identified based on the univariate and multivariate cox regression. We adopted the 'survival' R package and the log-rank test to plot and assess the Kaplan-Meier (KM) curve. To verify the precision of our risk signature in forecasting the prognosis (OS) of ccRCC patients, ROC curves were generated using 'survivalROC' R package.

List of abbreviations

ECM: Extracellular matrix

ccRCC: clear cell renal cell cancer

OS: overall survival

EMT: epithelial-mesenchymal junctions

GSEA: gene set enrichment analysis

MsigDB: Molecular Signature Database

ROC: received operating characteristic

Tregs: regulatory T cells

CAFs : cancer-associated fibroblasts

MCP1: monocyte chemotactic protein1

KM: Kaplan-Meier

Declarations

Ethics approval and consent to participate

This study, as a bioinformatics analysis is exempt from any requirement for Institutional Review Board approval.

Consent for publication

Not applicable

Availability of data and materials

The datasets analysed during the current study are available in the [TCGA and STRING] repository, [https://portal.gdc.cancer.gov/; https://stringdb.org]

Competing interests

The authors declare that they have no competing interests

Funding

This study was supported by National Natural Science Foundation (No. 81902565), funding from Young Talent Development Plan of Changzhou Health Commission (No. CZQM2020065), Young Scientists Foundation of Changzhou No.2 People's Hospital (2019K008), Changzhou Sci & Tech program (CJ20190100).

Author Contributions

Li Zuo, Lifeng Zhang: Project development

Xiaokai Shi: Data collection; Writing- Original draft

Xiao Zhou: software; visualization

Chuang Yue, Chao Lu: Investigation; Validation

Zhiqin Sun, Shenglin Gao: Supervision; Writing- review & editing

All authors have access to the data.

Acknowledgments:

The authors have no acknowledgements.

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: a cancer journal for clinicians 2021. doi:10.3322/caac.21660.

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: a cancer journal for clinicians. 2020;70:7–30. doi:10.3322/caac.21590.

3. Li F, Jin Y, Pei X, Guo P, Dong K, Wang H, et al. Bioinformatics analysis and verification of gene targets for renal clear cell carcinoma. Computational biology and chemistry. 2021;92:107453. doi:10.1016/j.compbiolchem.2021.107453.

4. Win AZ, Aparici CM. Clinical effectiveness of (18)f-fluorodeoxyglucose positron emission tomography/computed tomography in management of renal cell carcinoma: a single institution experience. World J Nucl Med. 2015;14:36–40. doi:10.4103/1450-1147.150535.

5. Molina AM, Lin X, Korytowsky B, Matczak E, Lechuga MJ, Wiltshire R, Motzer RJ. Sunitinib objective response in metastatic renal cell carcinoma: analysis of 1059 patients treated on clinical trials. Eur J Cancer. 2014;50:351–8. doi:10.1016/j.ejca.2013.08.021.

6. García-Mendoza MG, Inman DR, Ponik SM, Jeffery JJ, Sheerar DS, van Doorn RR, Keely PJ. Neutrophils drive accelerated tumor progression in the collagen-dense mammary tumor microenvironment. Breast Cancer Res. 2016;18:49. doi:10.1186/s13058-016-0703-7.

7. Hu M, Polyak K. Microenvironmental regulation of cancer development. Curr Opin Genet Dev. 2008;18:27–34. doi:10.1016/j.gde.2007.12.006.

8. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nat Med. 2013;19:1423–37. doi:10.1038/nm.3394.

9. Fullár A, Dudás J, Oláh L, Hollósi P, Papp Z, Sobel G, et al. Remodeling of extracellular matrix by normal and tumor-associated fibroblasts promotes cervical cancer progression. BMC Cancer. 2015;15:256. doi:10.1186/s12885-015-1272-3.

10. Vázquez-Villa F, García-Ocaña M, Galván JA, García-Martínez J, García-Pravia C, Menéndez-Rodríguez P, et al. COL11A1/(pro)collagen 11A1 expression is a remarkable biomarker of human invasive carcinoma-associated stromal cells and carcinoma progression. Tumour Biol. 2015;36:2213–22. doi:10.1007/s13277-015-3295-4.
11. Chen S, Wei Y, Liu H, Gong Y, Zhou Y, Yang H, Tang L. Analysis of Collagen type X alpha 1 (COL10A1) expression and prognostic significance in gastric cancer based on bioinformatics. Bioengineered. 2021;12:127–37. doi:10.1080/21655979.2020.1864912.
12. Zeng X-T, Liu X-P, Liu T-Z, Wang X-H. The clinical significance of COL5A2 in patients with bladder cancer: A retrospective analysis of bladder cancer gene expression data. Medicine (Baltimore). 2018;97:e0091. doi:10.1097/MD.0000000000010091.
13. Miyake M, Hori S, Morizawa Y, Tatsumi Y, Toritsuka M, Ohnishi S, et al. Collagen type IV alpha 1 (COL4A1) and collagen type XIII alpha 1 (COL13A1) produced in cancer cells promote tumor budding at the invasion front in human urothelial carcinoma of the bladder. Oncotarget. 2017;8:36099–114. doi:10.18632/oncotarget.16432.
14. Brooks M, Mo Q, Krasnow R, Ho PL, Lee Y-C, Xiao J, et al. Positive association of collagen type I with non-muscle invasive bladder cancer progression. Oncotarget. 2016;7:82609–19. doi:10.18632/oncotarget.12089.
15. Banyard J, Bao L, Hofer MD, Zurakowski D, Spivey KA, Feldman AS, et al. Collagen XXIII expression is associated with prostate cancer recurrence and distant metastases. Clin Cancer Res. 2007;13:2634–42. doi:10.1158/1078-0432.CCR-06-2163.
16. Chen P, Cescon M, Bonaldo P. Collagen VI in cancer and its biological mechanisms. Trends Mol Med. 2013;19:410–7. doi:10.1016/j.molmed.2013.04.001.

17. Ritzenthaler JD, Han S, Roman J. Stimulation of lung carcinoma cell growth by fibronectin-integrin signalling. Mol Biosyst. 2008;4:1160–9. doi:10.1039/b800533h.
18. Cai X, Liu C, Zhang T-N, Zhu Y-W, Dong X, Xue P. Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration, and invasion. J Cell Biochem. 2018;119:4717–28. doi:10.1002/jcb.26651.
19. Liang H, Yu M, Yang R, Zhang L, Zhang L, Di Zhu, et al. A PTAL-miR-101-FN1 Axis Promotes EMT and Invasion-Metastasis in Serous Ovarian Cancer. Mol Ther Oncolytics. 2020;16:53–62. doi:10.1016/j.omto.2019.12.002.
20. Kumari N, Dwarakanath BS, Das A, Bhatt AN. Role of interleukin-6 in cancer progression and therapeutic resistance. Tumour Biol. 2016;37:11553–72. doi:10.1007/s13277-016-5098-7.
21. Legendre F, Dudhia J, Pujol J-P, Bogdanowicz P. JAK/STAT but not ERK1/ERK2 pathway mediates interleukin (IL)-6/soluble IL-6R down-regulation of Type II collagen, aggrecan core, and link protein transcription in articular chondrocytes. Association with a down-regulation of SOX9 expression. J Biol Chem. 2003;278:2903–12. doi:10.1074/jbc.M110773200.
22. Wu Y, Hu P, Xu H, Yuan J, Yuan L, Xiong W, et al. A novel heterozygous COL4A4 missense mutation in a Chinese family with focal segmental glomerulosclerosis. J Cell Mol Med. 2016;20:2328–32. doi:10.1111/jcmm.12924.
23. Chattopadhyay I, Phukan R, Singh A, Vasudevan M, Purkayastha J, Hewitt S, et al. Molecular profiling to identify molecular mechanism in esophageal cancer with familial clustering. Oncol Rep. 2009;21:1135–46. doi:10.3892/or_00000333.

24. Li J, Wang X, Zheng K, Liu Y, Li J, Wang S, et al. The clinical significance of collagen family gene expression in esophageal squamous cell carcinoma. PeerJ. 2019;7:e7705. doi:10.7717/peerj.7705.

25. Kita Y, Mimori K, Tanaka F, Matsumoto T, Haraguchi N, Ishikawa K, et al. Clinical significance of LAMB3 and COL7A1 mRNA in esophageal squamous cell carcinoma. Eur J Surg Oncol. 2009;35:52–8. doi:10.1016/j.ejso.2008.01.025.

26. Pourreyron C, Chen M, McGrath JA, Salas-Alanis JC, South AP, Leigh IM. High levels of type VII collagen expression in recessive dystrophic epidermolysis bullosa cutaneous squamous cell carcinoma keratinocytes increases PI3K and MAPK signalling, cell migration and invasion. Br J Dermatol. 2014;170:1256–65. doi:10.1111/bjd.12715.

27. Gloushankova NA, Zhitnyak IY, Rubtsova SN. Role of Epithelial-Mesenchymal Transition in Tumor Progression. Biochemistry (Mosc). 2018;83:1469–76. doi:10.1134/S0006297918120052.

28. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. Nat Rev Mol Cell Biol. 2019;20:69–84. doi:10.1038/s41580-018-0080-4.

29. Jing Y, Han Z, Zhang S, Liu Y, Wei L. Epithelial-Mesenchymal Transition in tumor microenvironment. Cell Biosci. 2011;1:29. doi:10.1186/2045-3701-1-29.

30. Dongre A, Rashidian M, Reinhardt F, Bagnato A, Keckesova Z, Ploegh HL, Weinberg RA. Epithelial-to-Mesenchymal Transition Contributes to Immunosuppression in Breast Carcinomas. Cancer Res. 2017;77:3982–9. doi:10.1158/0008-5472.CAN-16-3292.

31. Dongre A, Rashidian M, Eaton EN, Reinhardt F, Thiru P, Zagorulya M, et al. Direct and

Indirect Regulators of Epithelial-Mesenchymal Transition (EMT)-mediated Immunosuppression in Breast Carcinomas. Cancer Discov 2020. doi:10.1158/2159-8290.CD-20-0603.

32. Larsen AMH, Kuczek DE, Kalvisa A, Siersbæk MS, Thorseth M-L, Johansen AZ, et al. Collagen Density Modulates the Immunosuppressive Functions of Macrophages. J Immunol. 2020;205:1461–72. doi:10.4049/jimmunol.1900789.

33. Barbi J, Pardoll D, Pan F. Treg functional stability and its responsiveness to the microenvironment. Immunol Rev. 2014;259:115–39. doi:10.1111/imr.12172.

34. Colak S, Dijke P ten. Targeting TGF-β Signaling in Cancer. Trends Cancer. 2017;3:56–71. doi:10.1016/j.trecan.2016.11.008.

35. Haque S, Morris JC. Transforming growth factor-β: A therapeutic target for cancer. Hum Vaccin Immunother. 2017;13:1741–50. doi:10.1080/21645515.2017.1327107.

36. Peng DH, Rodriguez BL, Diao L, Chen L, Wang J, Byers LA, et al. Collagen promotes anti-PD-1/PD-L1 resistance in cancer through LAIR1-dependent CD8+ T cell exhaustion. Nat Commun. 2020;11:4520. doi:10.1038/s41467-020-18298-8.

Supplementary Files

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

- TableS1.xls
- TableS2.xls
- tableS3trainGroup.xls
- tableS4testGroup.xls
- FigureS1.tif