OPEN ACCESS

GAVIN PUBLISHERS

Original Article

COL5A2-Dependent Cancer-associated Fibroblasts (CAF) Reveals Regulation of the Tumor Microenvironment and Response to Immunotherapy in Lung Adenocarcinoma

Xiao-Qin Shi[1], Xiao-Lan Wang[1], Xiu-Jun Chang[2], Kang-An Li[3], Yiran Cai[1]*

[1]Pathology Center, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[2]Department of Thoracic Surgery, Beijing Chest Hospital, Beijing, China

[3]Department of Radiology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

*Corresponding author: Yi-Ran Cai, Pathology Center, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine. 100 Haining ST. Hongkou Distirct, Shanghai, China

Abstract

Background Cancer-associated fibroblasts (CAFs) are crucial for tumor microenvironment (TME) remodeling and correlated with tumor progression. Dynamic interactions between CAFs and tumor cells and immune cells in lung adenocarcinoma (LAC) are still not clear. Method The role of CAFs in LAC and potential novel mediators of their functions were investigated. Hallmark signals associated with CAFs and immune components in LAC were analyzed in cohorts from TCGA and GEO databases. These cohorts were analyzed by bioinformatic method with R and Bioconductor packages. Twenty LAC patients who were treated with anti-PD-1 drug were involved to evaluate their pathological response. Result Genes based on CAF markers in the literature were clustered and sieved in LAC to find representative biomarkers which reflect TME and predict the effect of immunotherapy. Most of the cancer hallmark signaling pathways were enriched in extracellular matrix organization-related GO terms. COL5A2 were upregulated in CAFs compared to normal tissue. The expression of COL5A2 as negatively correlated with CD8+ T cells. COL5A2 indicated poor prognostic outcomes and might be correlated with the immunosuppressive tumor microenvironment (TME). LACs with COL5A2 overexpression had better clinical outcomes after anti-PD-1 inhibitor in twenty LAC with neoadjuvant therapy. Conclusion CAFs play an essential role in tumor progression and the TME. We identified an extracellular protein, COL5A2, as a prognostic marker and potential therapeutic target in LACs.

Introduction

Lung cancer is the most common malignant tumor worldwide. Lung adenocarcinoma (LAC) is an important histological type in non-small cell lung cancer. Recently, the management of LAC and the understanding of its biology have been the subject of great progresses. Thus, different histological subtypes have been identified, characterized by distinct genetic and molecular alterations, corresponding to different pathways of oncogenesis, inhibited by targeted therapies. The World Health Organization classification, which was revised in 2021 distinguishes subtypes, including lepidic, acinar, papillary, micropapillary and solid LAC. Of these, micropapillary and solid LACs are regarded poorly differentiated and associated with poor prognosis [1]. Anti-EGFR tyrosine kinase inhibitors (EGFR-TKI, such as Gefitinib and Ecotinib) and immunotherapies (such as nivolumab and ipilimumab) have successively led to a radical change in patient management and represent significant therapeutic options in the treatment of LAC [2, 3]. To further enhance the benefits of these treatments, different combinations, and sequences of TKI and immunotherapies are being investigated in clinical trials [3]. Patients who relapse on these therapies are left with very few options.

Tumor microenvironment (TME) is a complex ecosystem composed of tumor cells, infiltrating immune cells, and stromal cells intertwined with non-cellular components. The co-evolvement and dynamic interplay within and between these components shape the tumor's distinct biology and influence its response to cancer therapies. Cancer-associated fibroblasts (CAFs) constitute vastly heterogenous stromal cells and are prominent components of the microenvironment in solid tumors. Functionally, CAFs can contribute to the malignant development and progression by diverse mechanisms, including supporting tumor cell growth by secreting growth factors, extracellular matrix remodeling, promoting angiogenesis, and by mediating tumor-promoting inflammation [4,5]. The crucial role of TME, which serves as the soil for seeds (cancer cells), has been proven in many studies [6-8]. Cells in the TME mainly include stromal cells and immune cells. Recently, increasing evidence has high-lighted that appropriate stromal cells are crucial for the development of tumors [9,10] Among them, CAFs represent the main fraction, and accumulating evidence has indicated their role in cancer proliferation, progression and invasion [4,11]. Although various clinical trials targeting CAFs have been performed in recent years, such as targeting surface markers, reducing CAF infiltration and normalizing CAF functions, most of them are still ongoing [4]. Previous studies have identified many CAF markers, but few of them have moved into clinical practice.

This may be due to the internal heterogeneity of CAFs. The CAFs seemed to originate from diverse cell types, such as fibrocytes, stellate cells, endothelial cells, and mesenchymal stem cells [12]. It is well accepted that most activated fibroblasts are derived from fibroblasts of adjacent normal tissues and induced by oxidative stress or specific cytokines and chemokines from cancer cells [13]. Hence, distinct subclusters have been identified by previous studies. Therefore, focusing on the function and mechanism of fibroblasts in the tumor microenvironment may provide a strategy for LAC treatment, especially immunotherapy.

In this study, we explored the relative infiltration level of fibroblasts in LAC and the correlation between CAFs and immune components in the TME. We further explored upregulated secreted proteins, which could be used to predict CAF function in LAC. To establish the relevance of the role of fibroblasts and the upregulated protein in LAC using publicly available datasets and clinical samples raises the probability that targeting biomarkers may yield clinical utility.

Methods

Datasets and tissue specimens

The Cancer Genome Atlas (TCGA) dataset was obtained using the TCGA biolinks and analyzed by packages in R software. The transcripts per million (TPM) value was estimated at the transcript level. Patients who were diagnosed with LAC histologically and available for transcriptomic data were included. A total of 501 LAC patients were enrolled in the TCGA cohort. Overall survival (OS) was assessed using vital status and days from diagnosis to death or the last follow-up date. Only patients with active follow-up information were included in survival analysis. Patients diagnosed with LAC and available for active follow-up information were included. For other LAC datasets, GSE68465 with the expression matrices were downloaded from the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). The probes were mapped using the corresponding annotation platforms. The expression values were further normalized by the limma R package if necessary. The cell types were annotated using the Single R package if necessary. No identification information of participant was involved during download and analysis.

Tissues for immunohistochemistry and primary cell isolation were obtained at Beijing chest hospital and Shanghai general hospital between January 2010 and June 2021. Informed consent was obtained from participants. The study was conducted according to the principles stated in the Declaration of Helsinki.

Molecular markers, CAF score and CAF clustering

Representative immune-related genes (IRGs) and CAF signatures included the T cell signature (IFNG1, STAT1, CXCL10, IDO1, CXCL9), Myeloid dendritic cells (CD1A, CD1B, CD1E,

CLEC10A, CLIC2), inhibitory immune ligands/receptors (HAVCR2, CTLA4, LAG3, PDCD1, CD274), immune modulators (ENTPD1, NT5E), T cells associated-immune receptors (CD28, CD3D, CD3G, CD5, CD6, CHRM3-AS2, CTLA4, FLT3LG, ICOS, MAL, MGC40069, PBX4, SIRPG, THEMIS, TNFRSF25, TRAT1, CD8B, CD8A, EOMES, FGFBP2, GNLY, KLRC3, KLRC4, KLRD1) were analyzed. CAF scoring was evaluated through Tumor Immune Dysfunction and Exclusion (TIDE, http://tide.dfci.harvard.edu) based on tumor pre-treatment expression profiles, this TIDE module can estimate multiple published transcriptomic biomarkers to predict patient response. The edgeR and limma packages were used to calculate the fold change of genes between groups. Clustering was performed in individual datasets, and the samples were further classified into high-, medium-, and low-infiltration groups using the Complex Heatmap package. To further confirm the clustering results, the principal components analysis (PCA) method was applied as previously described. Inflammatory and myofibroblastic CAF features were also included to assess the internal characteristics. Comparisons of biological markers among different CAF infiltration groups are shown by the ggheatmap and Complex Heatmap packages.

Tumor microenvironment estimation

The Estimate the Proportion of Immune and Cancer cells (EPIC), xCell, and Microenvironment Cell Populations-counter (MCP-counter) algorithms were applied to calculate the cancer-associated fibroblast scores in datasets [14,15]. To analyze the correlation among fibroblasts and immune cells, fractions of 22 immune cells were estimated using the Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) algorithm [16,17]. The estimation of stromal and immune cells in malignant tumor tissues using expression data (ESTIMATE) was applied to calculate the overall stromal and immune scores in cancer [18].

Functional analysis

The GSVA package was used for gene set variation analysis (GSVA) [19]. The GSVA results were compared between the high- and low-CAF infiltration groups and are displayed. Gene set enrichment analysis (GSEA) was used to explore the biological functions and performed using GSEA 4.1.0. Hallmark and gene ontology gene sets were obtained from the MSigDB Collections (http://www.gsea-msigdb.org/gsea).

Immunohistochemistry

First, Formalin-fixed, paraffin-embedded tissue was cut into 4-μm sections and mounted on glass slides and then were baked in a 63 °C oven for one hour. The tissues were de-vaxed and rehydrated with a sequential procedure: dimethylbenzene 15 min × 2 times; water-free alcohol, 7 min × 2 times; 90% alcohol, 7 min;

80% alcohol, 7 min; 70% alcohol, 7 min; triple-rinsed with water for three minutes each time. Antigen recovery was performed for 5 min in boiling citric acid solution. Next, the tissues were incubated in blocking solution (76.8% methanol and 7.2% H_2O_2) for 10 min, and then triple-rinsing with 1 × PBS for 5 min each time. Rabbit polyclonal COL5A2 antibody (OriGene) was diluted with Dako REAL™ Antibody Diluent (Dako S3022) at a ratio of 1:500 and incubated overnight in a humid container in a 4 °C refrigerator. Tissues were triple-rinsed with 1 × PBS for 5 min each time before a further incubation with Dako EnVision™+/HRP second antibody reagent (Dako SM802) for 30 min. The tissues were washed with PBS and stained with substrate 3,3′-diaminobenzidine (DAB) (Dako EnVision™) DAB+ Chromogen (Dako DM827) in EnVision™ Substrate Buffer (Dako SM803) for 5 min. The nuclei were counterstained with Dako Hematoxylin. IHC staining was assessed by scores based on the percentage of positive cells (0: <5%; 1: 5%–25%; 2: 25%–50%; 3: 50%–75%; 4: >75%) multiplied by scores based on the intensity of staining, (0: colorless; 1: light yellow; 2: brown; 3: dark brown), with 6–12 considered high expression and 0–4 considered low expression. The primary antibody against COL5A2 used in IHC testing was purchased from LifeSpan BioSciences, lnc (Seattle, WA, U.S.A.).

Western blot

For LAC tumors, cells were laser captured from tumor bulks to isolate CAF, tumor cells as our former study [20]. These cells were lysed using RIPA buffer (Beyotime) at 4 °C directly. Phenylmethanesulfonyl fluoride (PMSF, Beyotime) was added to reduce protein degradation during extraction. Proteins were separated in SDS-polyacrylamide gels and transferred to PVDF membranes, and nonfat milk was used to block the nonspecific binding sites on the membrane. The membranes were incubated with primary antibodies against AKT1, COL5A2 (1:1000, Proteintech) and GAPDH (1:3000, SAB) at 4 °C overnight. The secondary antibody (1:5000, Bioss) was applied on the following day, and the reaction was detected using enhanced chemiluminescence solution (ECL, Affinity).

Pathologic Assessment of Response in LACs based on COL5A2 in resected tumors after ICI therapy

Twenty LAC patients, 18 to 65 years of age, were then recruited for immune checkpoint inhibitor therapy (ICIT) by Camrelizumab. Tumor size and standardized uptake value at baseline were evaluated by Positron-emission tomography (PET) plus contrast-enhanced CT. These patients had histologically confirmed NSCLC (stage IB–IIIA, American Joint Committee on Cancer, 8th edition) that was surgically resectable. All patients had treatment-naive primary tumor and adequate organ function. Exclusion criteria include as previous study [21]. Operation was performed between day 26 and 41 after two cycles treatment

of Camrelizumab (200 mg, intravenously, day 1 out of 22). All patients provided written informed consent for the use of their tumor specimens. Sufficient fresh specimens were used for the following assays.

A standardized procedure was used to evaluate the percentages of (1) viable tumor, (2) necrosis, and (3) stroma (including inflammation and fibrosis). The tumors with diameter < 3 cm were entirely sampled. If a tumor was larger than 3 cm, an approximately 0.5 cm thick cross-section of tumor was separated in its maximum dimension. True tumor bed consisted of viable tumor, necrosis and stroma which included both fibrosis and inflammation confined to the tumor bed.

The percentages of viable tumor, stromal tissue, and necrosis were estimated based on the microscopic sections. Each component was assessed in 10% increments unless the amount was less than 5%. A semi quantitative approach was done on the two stromal tissue components: fibrosis and inflammation. The final pathologic responses were determined based on the histologic features correlated with the gross findings.

Statistical methods

The best cutoff values for specific markers in each cohort were determined using the survminer package. The survival package was used for Kaplan-Meier overall survival analysis, and the log-rank test was applied for comparison. The hazard ratio (HR) was calculated via univariate Cox regression. Immune signatures were divided into two groups according to the median value and calculated by Cox regression in. Student's t-test or Wilcoxon rank-sum test were used for comparison of normally and non-normally distributed variables in unpaired groups, respectively. The paired Student's t-test was performed for paired samples. Chi-square test and Fisher's exact test were applied for comparison of clinical features. The Spearman method was applied for correlation

analysis. All P values were two-tailed. Statistical analysis was performed using R software (Version 4.3.1, https://www.r-project.org).

Results

CAF scores Are Correlated To A Poor Prognosis In LAC

CAF scoring was evaluated through EPIC, xCell, MCPcounter R packages and Tumor Immune Dysfunction and Exclusion (TIDE) on 111 CAF-related genes from MCPcounter (Table 1). We hypothesized that CAFs remodel tumor microenvironment and promote tumor progression and influence the prognosis. To explore these possibilities, we correlated clinical data to the CAF scores based on the expression levels of CAF-related genes (Table 1).

These CAF scores were correlated with overall survivals in TCGA cohort by EPIC, MCPcounter, xCell and TIDE (Figure 1A-C, OS, p=0.025, HR=1.456, 95% CI 1.048-2.025; p=0.022, HR= 1.631, 95% CI 1.068-2.493; p=0.006, HR=1.55, 95% CI 1.129-2.127, respectively). The stromal scores for TCGA cohort were also analyzed, but we found that patients with high stromal score had better prognosis than those with low stromal scores (Figure 1D, p=0.004, HR=0.651, 95% CI 0.484-0.876).

Similarly, the prognosis of patients from GSE68465 correlated to CAFs were evaluated by xCell and TIDE, and found that the overall survival of patients with low-CAF scores had better prognoses than those with high-CAF scores by xCell analysis (Figure 1E, p= 0.022, HR= 1.644, 95% CI 1.069-2.529), but it was contradicted by TIDE analysis showing a better clinical outcome in patients with high-CAF scores than with low-CAF scores (Figure 1F, p= 0.013, HR= 0.72, 95% CI 0.554-0.934). Patients with high stromal scores had better prognoses than those with low stromal scores (Fig. 1G, p=0.032, HR=0.73, 95% CI 0.546-0.974). It was consistent with that from TCGA cohorts.
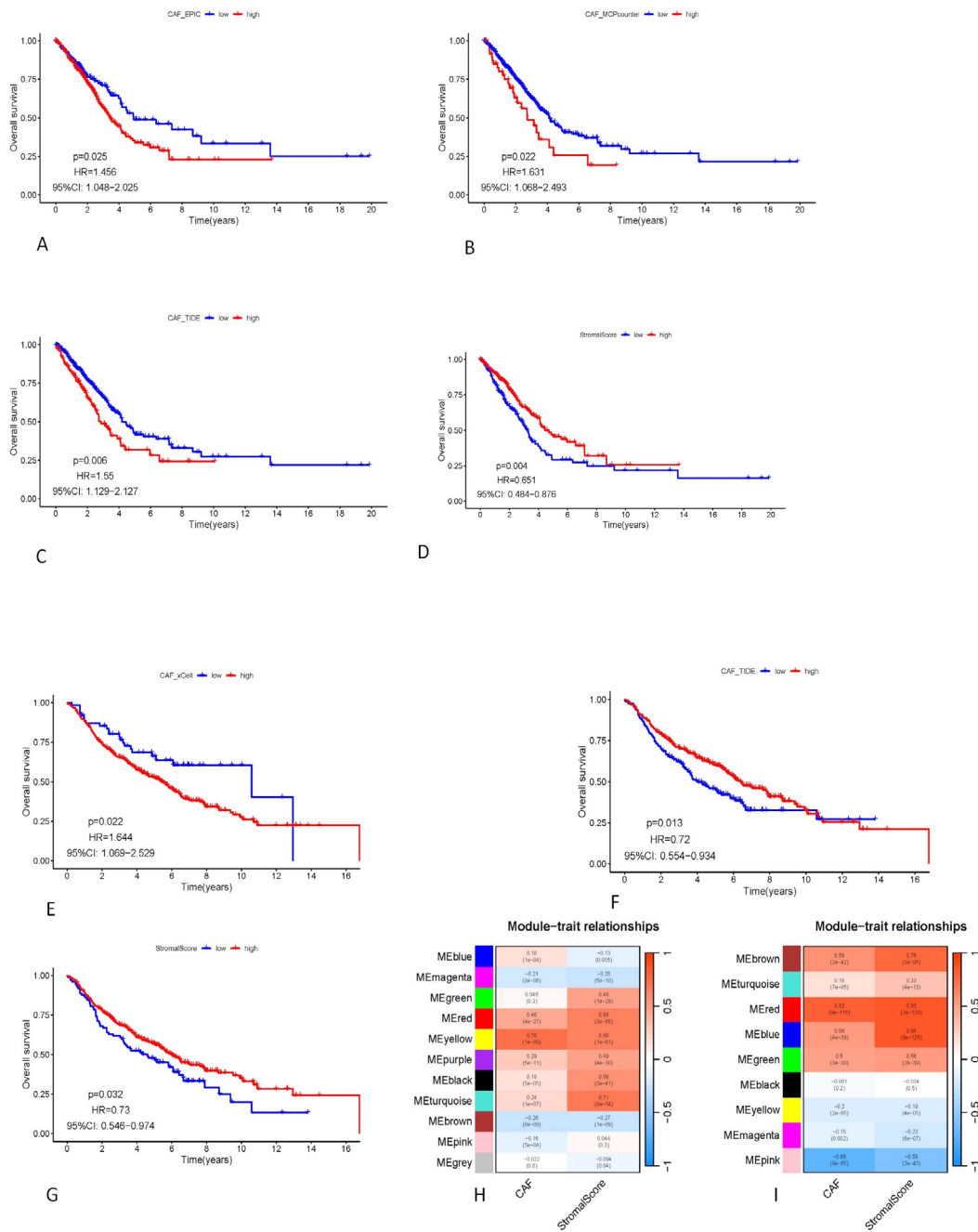
Figure 1: The expression levels of CAF-related genes were correlated with overall survivals in TCGA cohort by different analyses of EPIC, MCPcounter, xCell and TIDE. (A-C) the overall survivals of high- and low- CAF-associated genes expression groups from TCGA cohort were compared by EPIC, MCPcounter and TIDE. (D) the overall survivals of patients from TCGA with high- and low-stromal scores were analyzed. (E-F) the overall survivals of high- and low- CAF-associated genes expression groups from GSE68465 were compared by xCell and TIDE. (G) the overall survivals of patients from GSE68465 with high- and low- stromal scores were analyzed. (H) Heatmap showed the results of correlation analyses of different modules for TCGA cohort. (I) Heatmap showed the results of correlation analyses of different modules for GSE68465 cohort.

Functional Enrichment Analysis of CAF-Associated Differentially Expressed Genes in LACs of TCGA and GSE68465 cohorts

The relative abundance of fibroblasts in LACs was estimated by expression profile clustering using classic CAF markers. The intersection analysis displayed by Venn plot of the common genes from TCGA and GSE68465. These genes were SPARC, GLT8D2, ANGPTL2, MMP2, COL6A3, AEBP1, INHBA, COL5A2, COL5A1, COL1A2, MXRA5, and THBS2 (Figure 2A). Results from gene ontology (GO) enrichment analysis indicated that the DEGs mapped to the extracellular matrix organization-related GO terms, such as extracellular structure organization, endodermal cell differentiation and collagen fibril organization (Figure 2B). The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis also displayed the enrichment of protein digestion and absorption and ECM-receptor interaction pathways (Figure 2C). Thus, the overall functions of DEGs seemed to map on extracellular matrix remodeling-related activities, which implied that the involvement of CAFs was a predominant feature of TME in LACs.
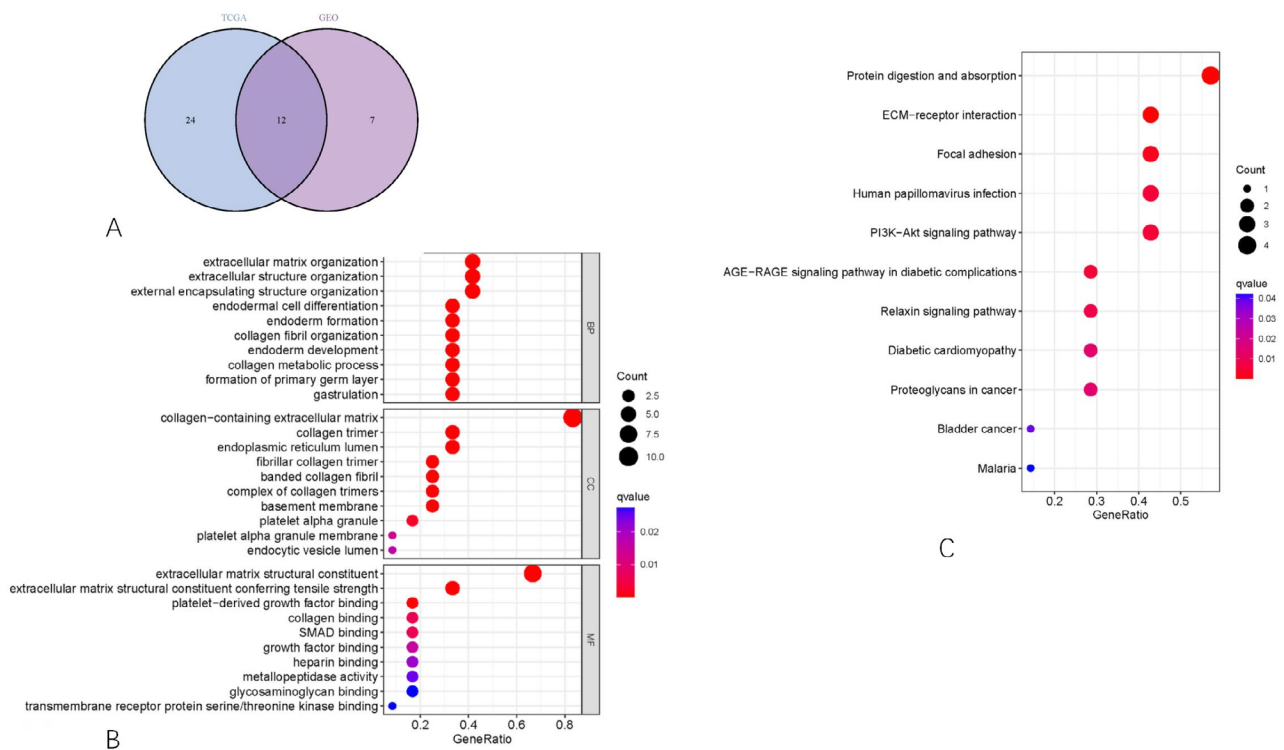

Figure 2: Common genes from TCGA and GEO cohorts and GO and KEGG analyses. (A) Venn plots showed common CAF-associated genes from TCGA and GSE68465 cohorts. (B) Bubble GO analysis chart identified molecular functions with genes enriched in each module. (C) Bubble KEGG plot analysis identified cancer-related pathways with genes enriched in each module.

Construction of a degenerated TME subtyping method by WGCNA-LASSO

To investigate effective biomarkers derived from CAFs and their corresponding prognostic value and to increase the accuracy and simplicity of the signature, WGCNA analysis was performed to divide the DEGs into different modules and the mostly prognosis-related genes were clustered by lasso regression in each module. The purpose of this step was to enhance the orthogonality of the genes involved in the signature. Cluster analysis on gene expression of TCGA cohort by WGCNA and the result were shown in Figure 1H. The MEyellow group was mostly associated with CAF score and stromal score. Likely, the cluster analysis for GSE68465 cohort was explored and the result was shown in Figure 1I. The MEred group was significantly associated with CAF score and stromal scores.

As shown in Figure 1H-I, these different modules were correlated with different microenvironmental factors and different clinical characteristics (Figs. 1A-G). Subsequently, we performed a Pearson's correlation analysis on the prognostic genes. Based on these results, the genes that we selected exhibited different micro environmental statuses and a high orthogonality. Lastly, two modules from TCGA cohort and 9 modules from GSE68465 were selected for lasso regression, which included UNC5B, CALD1, FSTL1, LAMA4, PRRX1, SPARC, TIMP2, GLT8D2, THY1, ANGPTL2, OLFML2B, COL6A2, ADAMTS2, MMP14, FBN1, FAP, ANTXR1, ADAMTS12, MMP2, COL6A3, AEBP1, INHBA, COL5A2, COL5A1, COL1A2, COL15A1, POSTN, COL3A1, ITGA11, MXRA5, ADAM12, SULF1, VCAN, THBS2, COL1A1 and COL12A1 out of MEyellow module from TCGA cohort and GLT8D2, SPARC, COL6A3, ANGPTL2, ACTA2, COL1A2, CTSK, MMP2, BGN, ISLR, CDH11, AEBP1, COL5A1, DCN, THBS2, COL5A2, TAGLN, INHBA, MXRA5 out of MEred module from GSE68465 cohort. Next, a prognostic model consisting of two genes, COL5A2 and COL5A1 was generated (Figure 3A-B)), and the risk score of each patient as the sum of each gene multiplied by the corresponding coefficient 0.128 for COL5A2 and 0.003 for COL5A1.

To avoid overfitting effects, we evaluated the the correlation of CAF and risk scores of samples from TCGA cohort by GGally algorithm using EPIC, MCPcounter, xCell and TIDE. All the results from different algorithms were positively correlated. Based on the classification, patients in the TCGA cohort were divided into high- and low-risk groups by the median risk score. According to the expression of COL5A2 and COL5A1, patients were classified in the high-risk group (n=235) and low-risk group (n=236) with mean risk score= 1.1, range 0.59-1.55) (Figure 3D). We found that with increased risk, COL5A2 and COL5A1 were up-regulated accordingly. Notably, the 2-gene panel integrated with all of the algorithms turned out to represent the CAF-related genes panel (Figure 3E). CAF is an important component in tumor and may influence therapeutic effects. Therefore, this signature was testified by tumor immune dysfunction and exclusion (TIDE) to predict tumor treatment effects based on tumor pre-treatment expression profiles, this TIDE module can estimate multiple published transcriptomic biomarkers to predict patients' response (Figure 3F). 54% versus 18% of patients respond to anti-PD-1 or anti-CTLA-4 in low-risk group compared to high-risk group (Fig. 3G-H, p <0.001). Noticeably, The accuracy of the signature was validated by the AUCs, which were 0.757 (95% CI 0.708-0.801) (Figure 3I). These results indicated that this signature had a promising application in the prediction of clinical outcomes of immune therapy.
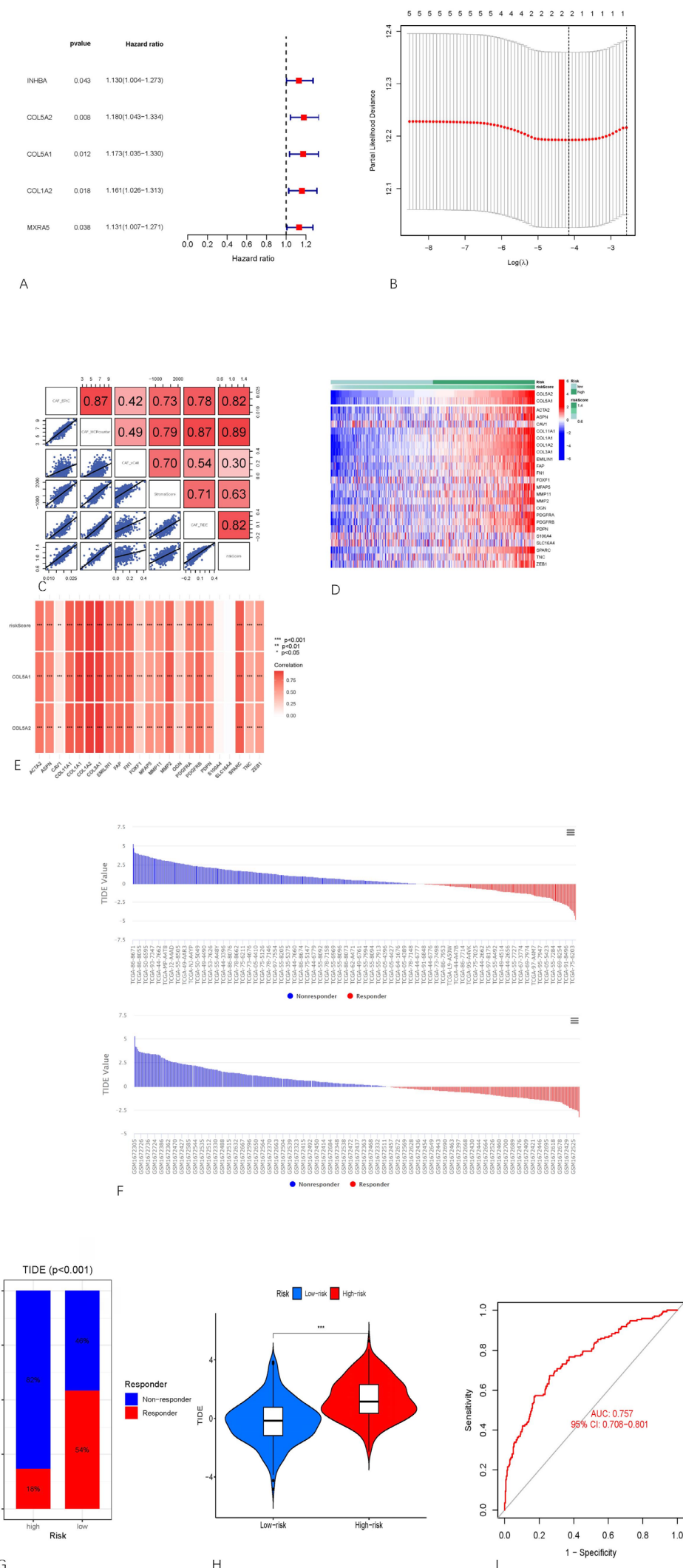

Figure 3: CAF-related genes were analysed with patients' clinical traits and a model were explored to predict immunotherapeutic response of LAC patients. (A) Univariate Cox regression analysis identifying CAF-associated genes correlating with overall survival. (B) Partial likelihood deviance revealed by LASSO regression in the 10-fold cross-validation to establish a model to predict risk of CAF-associated genes. The optimal values were shown within the two dotted vertical lines. (C) Correlation of the expression of CAF-associated genes, stromal score and risk score were analyzed by EPIC, MCPcounter, xCell and TIDE. (D) Heatmap for CAF-associated genes generated by comparison of the high score group vs. the low score group in risk. Row name of heatmap is the gene name, and column name is the ID of samples which not shown in plot. (E) Correlation of the genes in predicting model with the CAF-associated genes. (F) the upper figure showed the response of patients from TCGA cohort to immune checkpoint inhibitor. The lower figure showed the response of patients from GSE68465 cohort to immune checkpoint inhibitor. (G) Boxplot showed the responsive percentage of patients from high-risk and low-risk groups to anti-PD-1 therapy. (H) Violin plot showed the significant difference of response to immune checkpoint inhibitor between high-risk and low-risk groups. (I) ROC curve of the response to immune checkpoint inhibitor.

Functional Analysis to CAF High- and Low-Risk Groups

The GSVA package was used for gene set variation analysis (GSVA). The GSVA results were compared between the high- and low-CAF infiltration groups and are displayed. Gene set enrichment analysis (GSEA) was used to explore the biological functions of COL5A2 and COL5A1 using GSEA 4.3.2 Hallmark; and gene ontology gene sets were obtained from the MSigDB Collections (http://www.gsea-msigdb.org/gsea). The GO analyses demonstrated the activation cytosolic ribosome, large ribosomal subunit, ribosomal subunit, ribosome and structure constituent of ribosome were present in low-risk group (Figure 4A). In contrast, nucleosome assembly, DNA packaging complex, nucleosome, extracellular matrix structural constituent and structural constituent of chromatin were activated in high-risk group (Figure 4B). The KEGG analysis showed that glycine serine and threonine metabolism, linoleic acid metabolism, oxidative phosphorylation, parkinsons disease and ribosome pathways were involved in low-risk group (Figure 4C). However, cytokine-cytokine receptor interaction, ECM receptor interaction, focal adhesion, pathways in cancer and systemic lupus erythematousus pathways were upregulated in high-risk group (Figure 4D). These results suggested that extracellular matrix-associated pathways and function were active in COL5A2-upregualted LACs. Correlation analyses revealed that vatomical structure formation (R=0.65, p <0.001), fibroblast migration (R=0.66, p <0.001) and fibroblast proliferation (R=0.64, p <0.001) were the significant risk factors in LACs (Figure 4E-G).

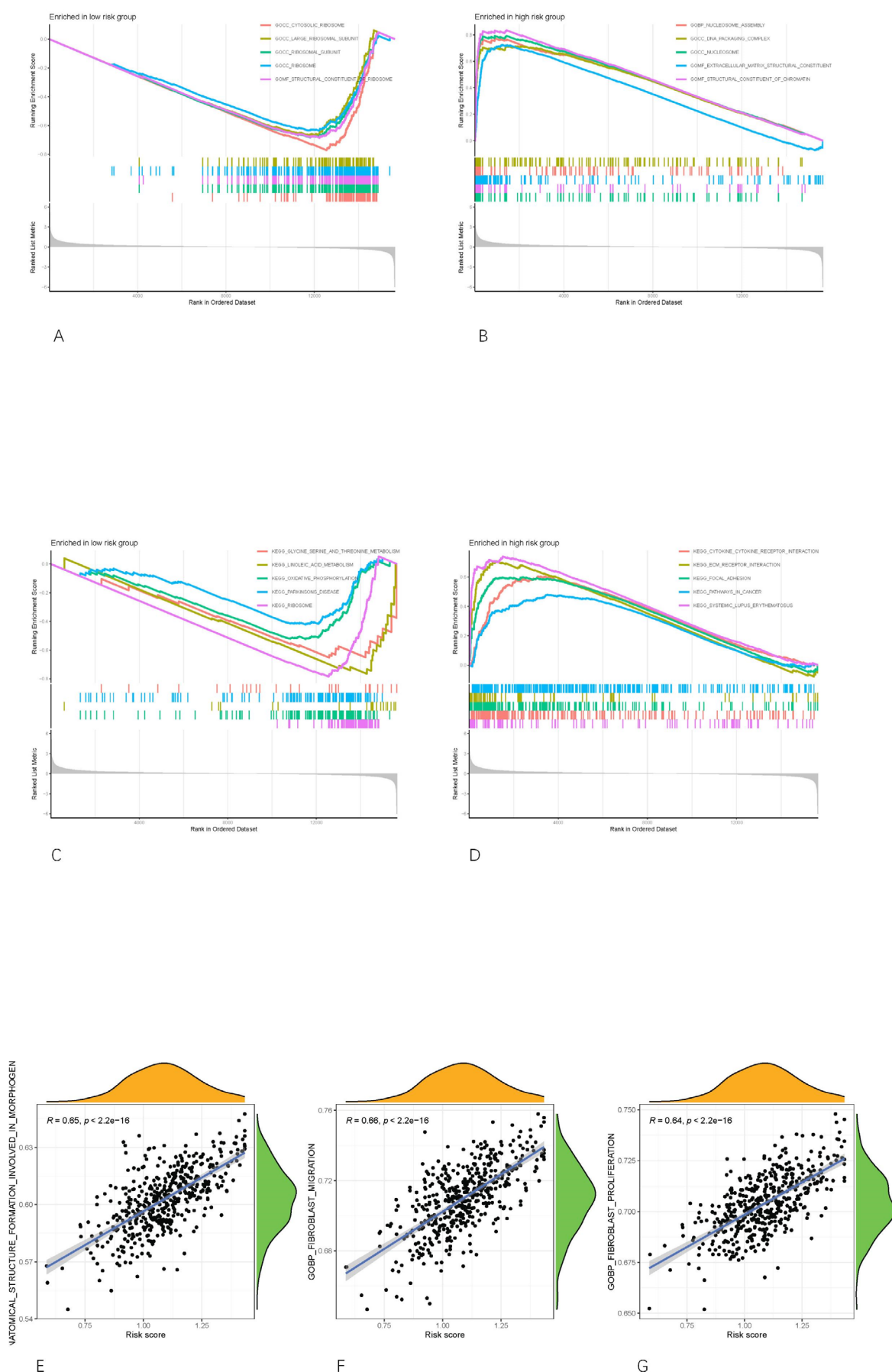**Figure 4: GO and KEGG analyses on LACs from low- and high-risk groups.** A-B, GO analyses of CAF-associated genes are shown in low-and high risk groups. C-D, KEGG signaling pathways based on CAF-associated genes are shown in low-and high- risk groups. E-G, Scatter plots showing the correlation of structure formation, fibroblast migration and proliferation with risk score. p and R values from Spearman correlation analyses.

COL5A2 Expression Were Associated With the Clinicopathological Staging of LAC Patients

For determining the relationship between the proportion of immune and stromal components with the clinicopathological characteristics of LAC cases from TCGA database. We firstly analyzed the corresponding the expression of COL5A2 in normal and tumor tissues. Whether or not in paired normal and tumor tissue, the expression level was higher in tumor than in normal tissues (Figure 5A-B). We also analyzed the expression level of COL5A2 to clinical information. COL5A2 expression was not correlated with patient's age, gender (Fig. 5C-D). We noticed that COL5A2 expression was associated with T stage. Tumor in T1 had low level of COL5A2 expression than T2 and T3 (Figure 5E). Similar tendency was found in stage I versus stage II (p <0.01, Figure 5H). However, with in advanced stage of LACs, the expression of COL5A2 was not positively increased with tumor progression (Figure 5F-G). This suggested that COL5A2 expression increased with tumor enlargement. Heatmap plot showed other clinical characteristics in the groups of high- and low-expresion of COL5A2 (Figure 5I).

As shown in Figure 5J, A nomogram model was constructed that included T stages, tumor status, pathologic stages, and COL5A2 expression levels as parameters. The nomogram showed a significantly high clinical value in predicting the 1 -, 3-, and 5-year survival probability of the LAC patients (Figure 5J).

Univariate COX regression analysis on clin-pathological factors, and the results indicated that the expression of COL5A2 and clinical stage were the significant prognostic factors to patients with HR=1.174 (95% CI 1.05-1.313) and HR=1.626 (95% CI 1.413-1.871), respectively (Figure 5K). Similar results were concluded by multivariate regression (Fig. 5L).
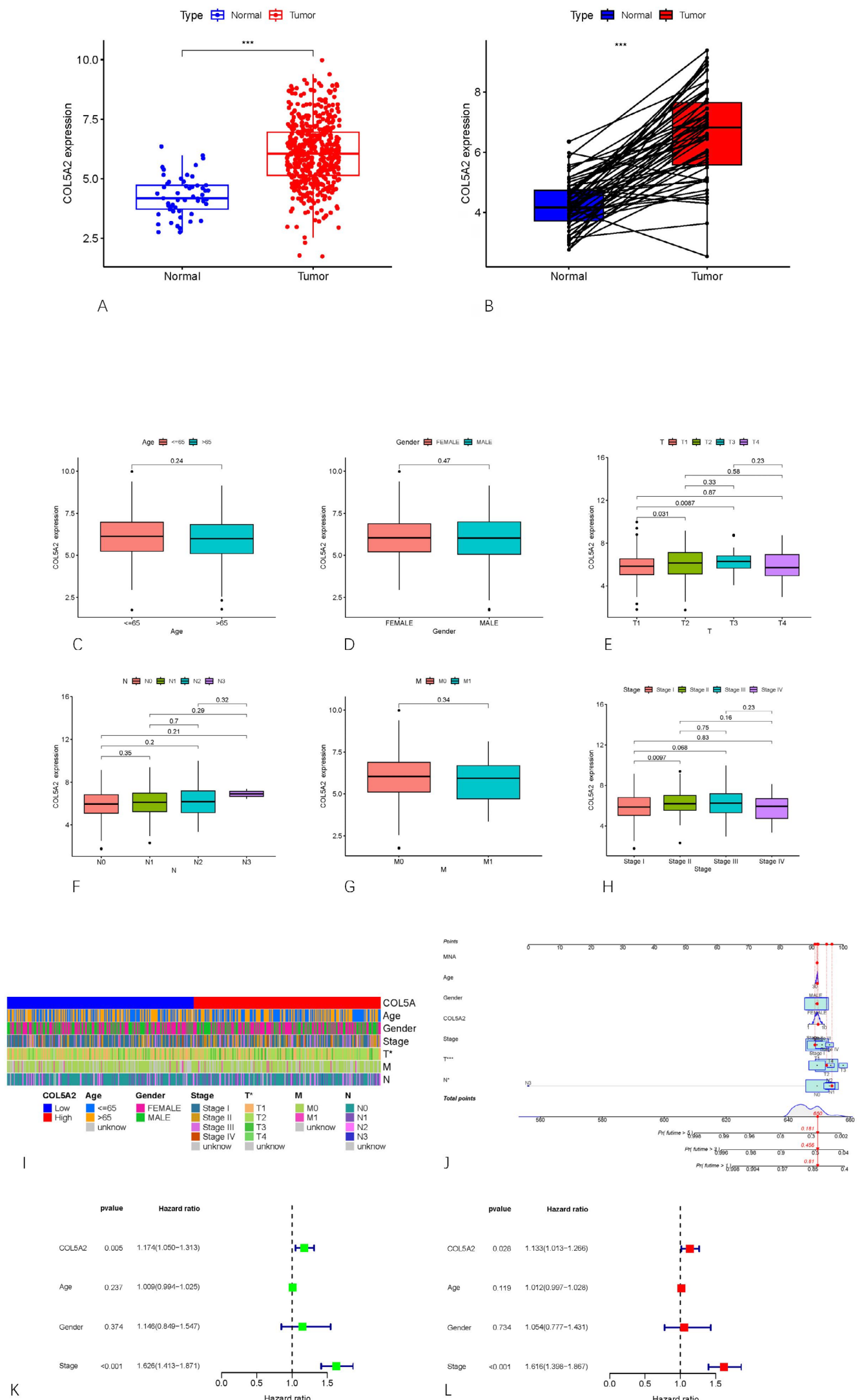

Figure 5: COL5A2 expression in LACs and compared among the groups with different clinical traits. (A) TCGA database analysis shows the COL5A2 expression levels in LAC tissues and their corresponding adjacent normal tissues. *p < 0.05; ***, p < 0.001. (B) COL5A2 expression levels were significantly higher in the LACs tissues compared to the adjacent peritumoral tissues. (C-H) COL5A2 expression levels were significantly lower in T1 LAC patients compared to T2 and T3 LACs (p < 0.05). Similar difference occurrence when compared in patients with different clinical stage (p <0.05). There were no significance in patients according to age, gender, N staging and M staging. (I) Heatmap showing the association of clinical characteristics in low- and high expression level groups. (J) Nomogram model analysis to evaluate the prediction efficacy the nomogram model that includes clinicopathological factors (T stages, tumor status, and pathologic stages) and COL5A2 expression levels to predict the 1-, 3-, and 5-year survival rates of LAC patients. (K-L) uniCox and multiCox variates analyses show that COL5A2 expression is the crucial factor associated with patients' prognosis.

COL5A2 Had Potential to Be an Indicator of tumor progression

Given the levels of COL5A2 were negatively correlated with the survival of LAC patients, to ascertain the exact alterations of gene profile with COL5A2 expression, the correlation analysis between COL5A2 and other genes were carried out in TCGA cohort. The total 8204 correlated genes were obtained and the top 5 upregulated and down-regulated genes by COL5A2 were plotted in Figure 6A. There were 890 DEGs based on the COL5A2 median expression level by fdr cutoff=0.05. 706 genes were upregulated and 184 genes were downregulated. The top 50 genes related to COL5A2 expression in low- and high-level were shown in the heatmap (Figure 6B).

To perform a functional annotation of the COL5A2-associated DEGs in the LAC patients using the "clusterProfiler" R package. The GO enrichment analysis results consisting of the highly enriched biological processes, cellular components, and molecular functions (p <0.05) were shown in Figure 6C. KEGG analysis exhibited that DEGs were clustered in "PI3K-AKT", "neuroactive ligand-receptor interaction", "protein digestion and absorption" pathways. GSEA was implemented to show the COL5A2-associated DEGs were significantly aggregated in clusters, which were dysfunctioned in "mitochondrial respiratory chain complex assembly", "nuclear transcribed mRNA catabolic process", and "cytosolic ribosome"; enriched in "arachidonic acid metabolism", "glutathione metabolism" and "oxidative phosphorylation" pathways (Figure 6D). The top biological processes enriched by GO analysis included "extracellular matrix organization", "extracellular structure organization", and "external encapsulating structure organization". The most enriched cellular components were "collagen-containing extracellular organization", "endoplasmic reticulum lumen", and "collagen trimer". The most enriched molecular functions were "extracellular matrix structural constituents", "receptor ligand activity", and "glycosaminoglycan binding" (Figure 6E).

We further verified our results using Broad Institute Cancer Cell Line Encyclopedia (CCLE) database to testify the expression level in lung cancer cells and CAFs. We found that COL5A2 expression are significantly increased in CAFs than in tumor cells (Figure 7 A-B). In addition, we also explored to CAF, tumor cells and tumor bulk of LAC at the protein level by Western blotting and observed a significantly higher level of pAKT1 protein in CAFs, lower level in tumor cells (Figure 7C), which is consistent with the result from+ public dataset.
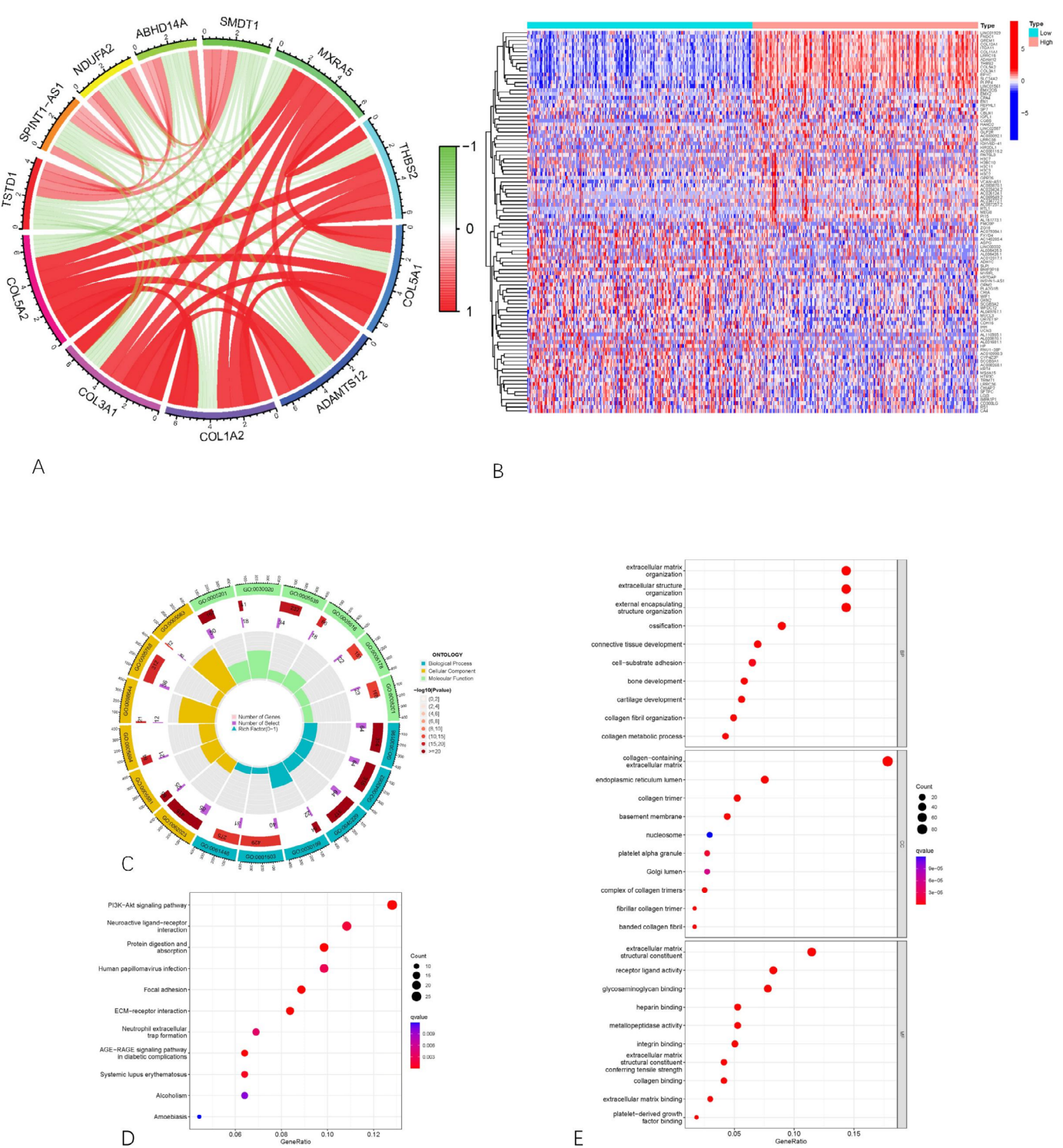

Figure 6: Different expression genes and pathways associated with COL5A2 expression. (A) Coexpression circling plot of 10 genes with COL5A2 showed the genes were upregulated and downregulated. (B) Heatmap showed the DEGs in the groups with COL5A2 low- and high-level expression. (C) Circling plot showed the results of GO enrichment clustered genes. (D) KEEG analysis on the associated pathways with COL5A2. (E) GO function enrichment clustering showed the pathways are activated with COL5A2.
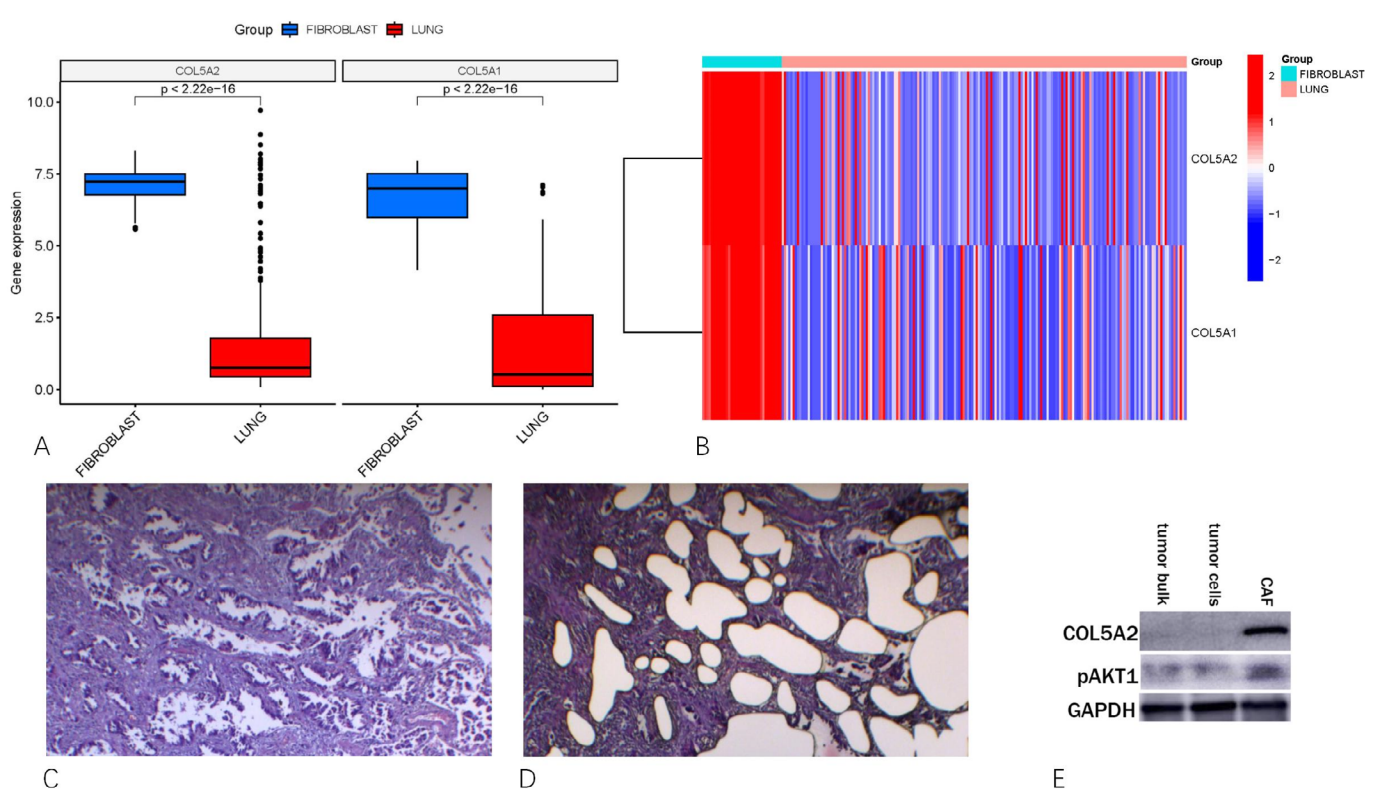

Figure 7: COL5A2 and COL5A1 expression in lung cancer cells and fibroblasts. (A) Upregulation of COL5A2 and COL5A1 in fibroblasts compared to lung cancer cells (LUNG) from Broad Institute Cancer Cell Line Encyclopedia (CCLE) database. (B) Heatmap exhibited the upregulation of COL5A2 and COL5A1 in fibroblast and lung cancer cells (LUNG) from Broad Institute Cancer Cell Line Encyclopedia (CCLE) database. (C) LAC tumor tissue section was stained by hematoxylin-eosin staining. (D) LAC tumor tissues were micro dissected and classified into cancer cells and tumor stroma groups. (E) Western blot showed the expression of COL5A2 and pAKT1 proteins in fibroblasts isolated from normal and cancer tissues of LAC.

Correlation of COL5A2 With the Proportion of TICs and drug sensitivity to different risk groups.

To further confirm the correlation of COL5A2 expression with the immune microenvironment, the proportion of tumor infiltrating immune subsets was analyzed using CIBERSORT algorithm, immune scores were significantly increased in the group with high-level COL5A2 expression vs the group with low-level COL5A2 (Figure 8A). 22 kinds of immune cell profiles in LAC samples were constructed (Figure 8B). The results from the difference and correlation analyses showed that a total of 11 kinds of immune cells were correlated with the expression of COL5A2 (Figure 8B). Among them, five kinds of TICs were positively correlated with COL5A2 expression, including macrophage M0, CD4+ memory activated T cells, NK resting cells, activated mast cells and neutrophil; Six kinds of immune cells were negatively correlated with COL5A2 expression, including T cells follicular helper, activated NK cells, resting mast cells, monocytes, CD8+ T cells and resting dendritic cells (Figure 8C). These results further supported that the levels of COL5A2 affected the immune activity of TME. To predict immune therapy effect, we analyzed the correlations between COL5A2 and immune checkpoint inhibitor-associated genes listing in Figure 8D. These results indicated that COL5A2 expression were positively related to these genes except TNFSF15.

CAFs are important components in tumor and may influence chemical therapy. We analyzed potential impacts on drugs through Bioconductor package: "TxDb.Hsapiens.UCSC.hg19.knownGene" and "oncoPredict" R package. Tumors with high risk were sensitive to alisertib (p <0.001), docetaxel (p <0.001), cisplatin (p <0.001), cytophosphamide (p <0.001) and AT13148 (p <0.001) (Figure 9 A-E).

In this study, we observed CD4+, CD8+ T cells and COL5A2 expression in CAFs in twenty LACs with neoadjuvant immunotherapy. We found that CD4+ T cells infiltrating both in tumor parenchyma and in stroma (Figure 9?). However, CD8+ T cells infiltrated in stroma and at the edge of tumor parenchyma. This distribution resulted in different TIME. In addition, we also evaluated the expression of COL5A2 in Table 4. 9 out of 20 (45%) LACs treated by ICI had high level of COL5A2 expression, whereas 11 LACs (55%) had low levels of COL5A2 expression.
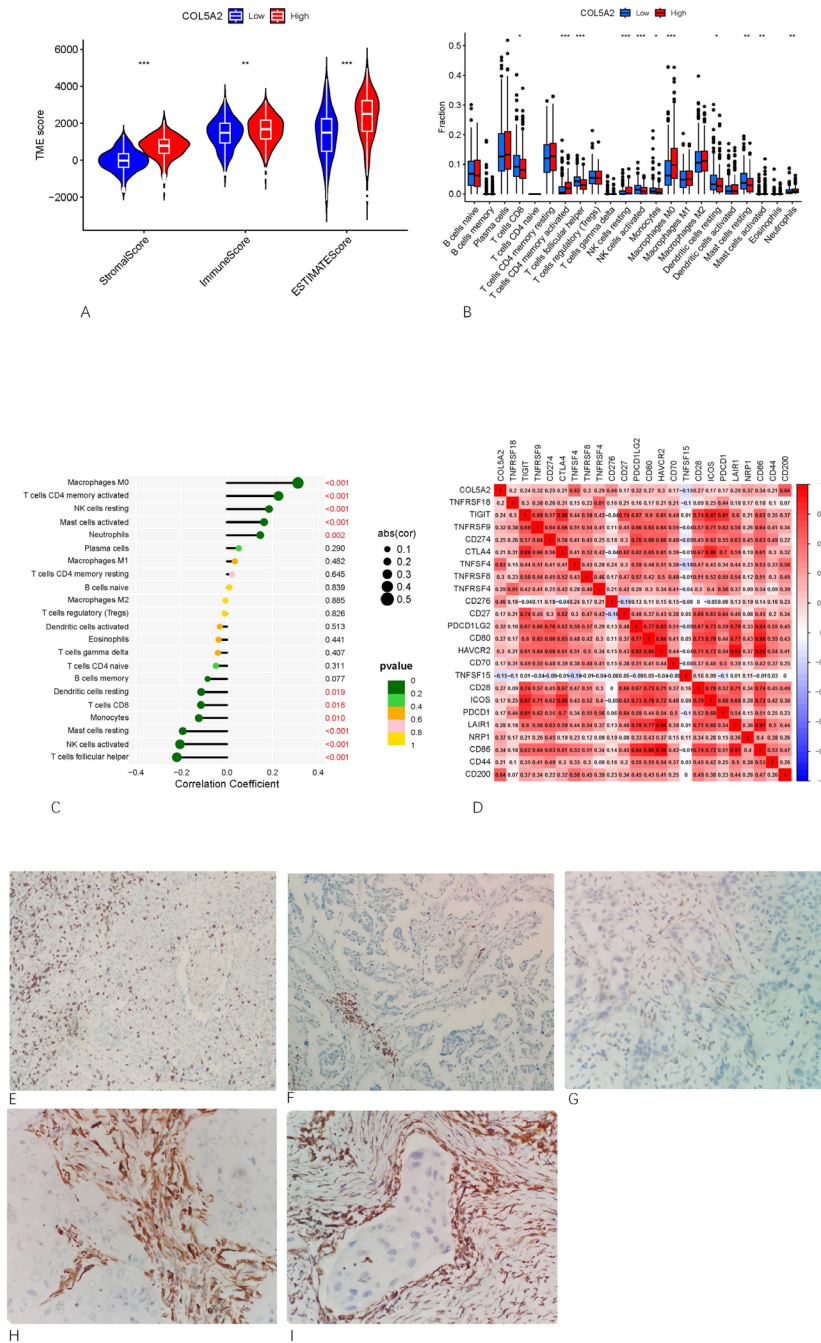
Figure 8: TME scores and immune cells infiltrating in different groups with COL5A2. (A) Violin plot showed TME scores for LAC patients in different subgroups. (B) Boxplot showed the expression profiles of 22 immune cells in different subgroups. (C) Lillipop plot exhibited the immune cells are positively and negatively associated with COL5A2 expression. (D) Heatmap showed relationship between the immune checkpoint inhibitor-associated genes with COL5A2 expression. (E-I) Immunohistochemistry was used to stain specific biomarkers. (E) CD4+ T cells were stained by immunohistochemistry. These cells infiltrated in both tumor parenchyma and stroma. (F) CD8+ T cells were showed infiltrating in stroma only. (G) COL5A2 protein was expressed in low level. (H) COL5A2 protein expressed in high level in stroma. (I) COL5A2 protein was overexpressed by CAFs in LAC after ICIT.

Figure 9: Drug sensitivity of IC50 with COL5A2 expression. (A-E) alisertib, docetaxel, cisplatin, cytophosphamide and AT13148 had better effects in the subgroup of high-level COL5A2 expresssion (p < 0.001).

Pathological response to immune checkpoint inhibitor therapy (ICIT)

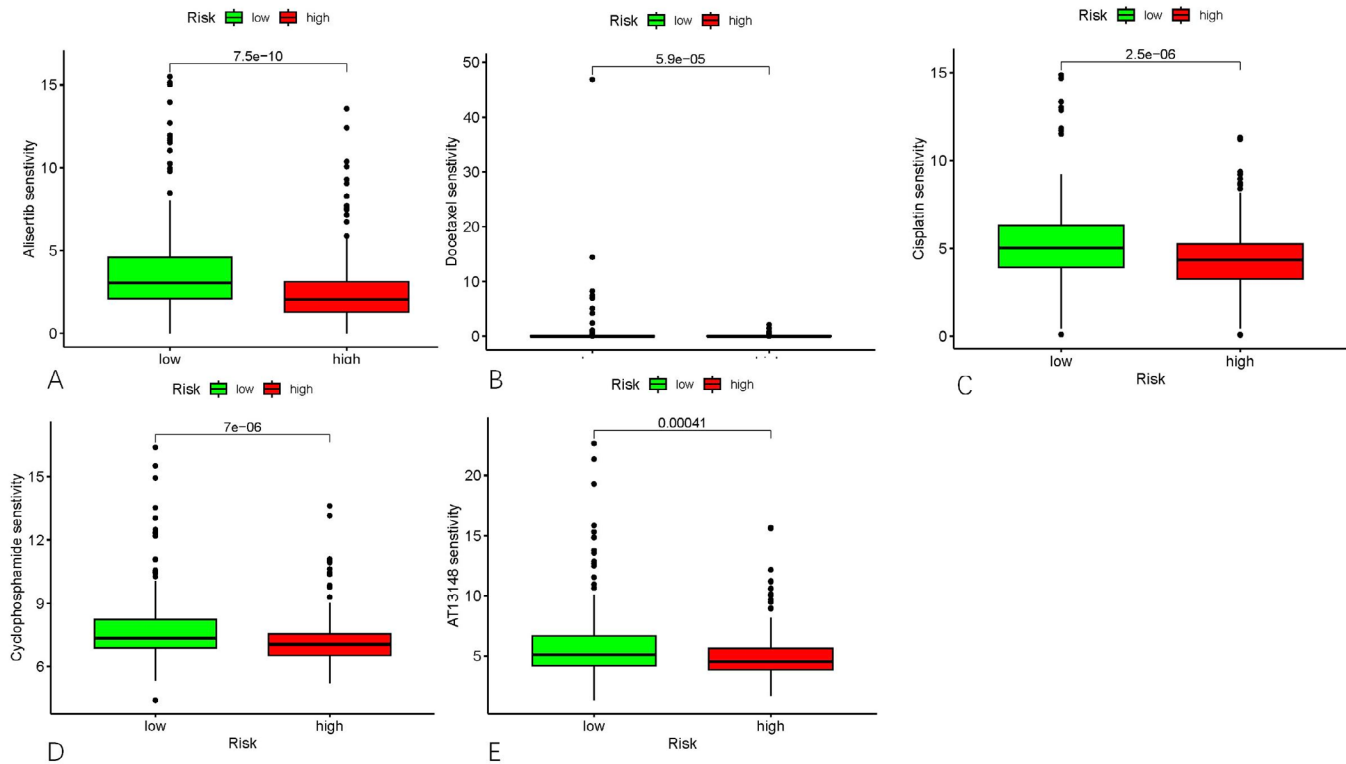In this study, twenty LAC patients were examined before and after ICIT by computerized tomography (Figure 10A-B). Pathological responses were summarized in Table 3. The histological types present in these LACs were acinar, papillary and solid patterns. A total of 4 LACs obtained complete pathological response (CPR) and all these LACs had uniform histological components. Two of the 4 LACs with CPR were found at IB and two at IIA with acinar and solid patterns, respectively. Among the 4 LACs with major pathological response (MPR), three LACs (75%) contained uniform solid or acinar pattern. Additionally, these four LACs with MPR were found at stage IIA. As far as the 12 LACs with less pathological response (LPR) were concerned, 10 LACs (83.3%) had heterogeneous histological constituents and were diagnosed at IIB and IIIA. Pathological responses of LACs were correlated with their clinical stages and histological constituents (P <0.05). Among the 9 LACs with COL5A2 overexpression, 3 LACs (75%) had CPR, 3 LACs (75%) had MPR, and 3 LACs

(25%) had LPR (p =0.098). We also noticed that more than 90% tumor retraction occurred in LACs with COL5A2 overexpression (p= 0.025). It suggested that LACs with uniform pattern and at early clinical stage had better pathological response. Tumors with CPR were found with more necrosis compared to those with MPR and LPR. Fibrosis and inflammation were not associated with different pathological responses.

Pathological responses included fibrosis, inflammation and necrosis occurring differently in LACs. (Fig. 10 C-H, Table 4). The mean proportions of necrosis were 11%, 18% and 30% in LPR, MPR and CPR, respectively. Necrosis was the obvious change within different pathological responses (p =0.024), compared to fibrosis and inflammation. Inflammation and fibrosis were common in tumor bed and the proportions were not significantly associated with effectiveness of ICIT (Fig. 10H). Further observation on LACs with MPR and LPR indicated that survived tumor cells were entrapped in fibrotic stroma and were isolated from immune cells by proliferated collagens and the immune cells aggregated at the rim of tumor nests (Fig. 10E). In addition, immune cells infiltrated

tumor bed less in those tumors with LPR compared to those with MPR and CPR (Table 4).
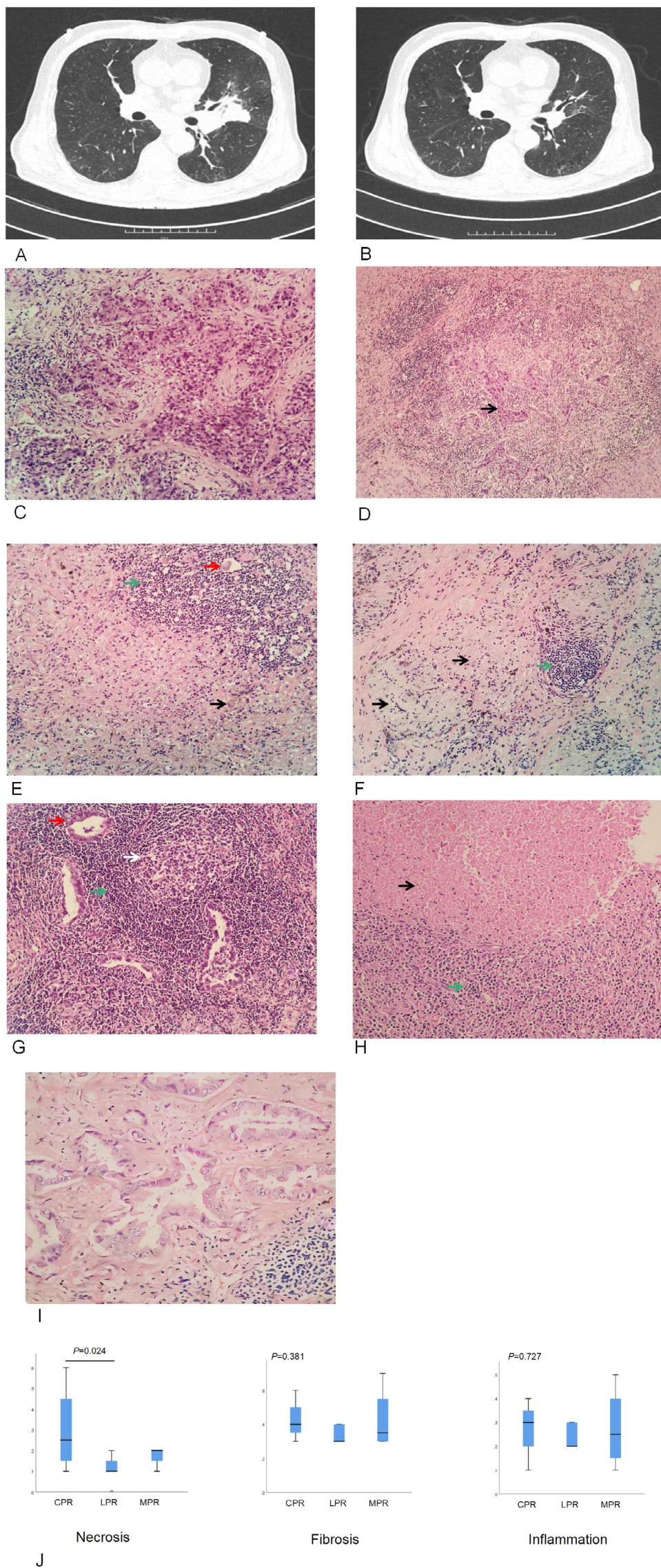

Figure 10: **A representative patient was used as an example of LAC to show the clinicopathological features.** (A) A tumour mass was present in the inferior lobe of the left lung, which was determined using computed tomography (CT) before immune checkpoint inhibitor therapy (ICIT). (B) The tumor mass was scanned by CT after three-week ICIT before surgical operation. (C) Acinar component was obtained by fine needle aspiration. (D), The pathological response with major pathological remission was evaluated on resected tumor. Residule tumors (black arrow) are entrapped in inflammatory stroma. (E) A single tumor cell (black arrow) was entrapped alive in fibrosed tumor bed and scattered tumor cells (red arrow) were surrounded by lymphocytes (green arrow) (magnification ×200). (F) Tumor bed (black arrow) was fibrosed. Lymphocytes (green arrow) were aggregated in stroma to form an immature germinal center (magnification ×200). (G) Residual tumor cells (red arrow) were surrounded by lymphocytes (green arrow), which aggregated in stroma to form a mature germinal center (magnification ×200). (H) Obvious necrosis in tumor bed and surrounded by diffuse infiltrating lymphocytes (magnification ×200). (I) survived tumor cells were entrapped in fibrotic stroma and were isolated to immune cells by proliferated collagen. The immune cells aggregated at the rim of tumors (black arrow). (magnification ×400). (J) Pathological responses were compared within different subgroups. Necrosis occurred more frequently in the LACs with CPR than those with MPR and LPR after ICIT (p=0.024). Fibrosis and inflammation were not significantly different in subgroups.

Discussion

Recent studies have indicated the crucial roles of cancer-associated fibroblasts. Many of them focus on the heterogeneity and corresponding biological features of different CAF clusters in cancer [22, 23]. However, the detailed mechanism underlying how CAFs influence the TME has been a prominent theme in recent years.

Tumor stroma is made up of diverse populations of cells of mesenchymal origin included in the extracellular matrix. These cells are fibroblasts, endothelial, inflammatory, and mesenchymal stem cells. The tumor microenvironment is qualitatively and quantitatively variable depending on the organ in which the tumor develops and depending on the type of tumor. This tumor stroma will change over time and during the process of tumor development, according to the interactions that can occur between tumor cells and stromal cells. Treatments can also modify the tumor's stroma. Among the stromal cells appearing during these mechanisms, CAFs are of particular importance. CAFs produce numerous pro-tumoral cytokines (including IL6, IL8, IL10, TNFα, TGF-β) and generate a collagen matrix which hinders the action of T lymphocytes within the tumor.

Hence, it is more important to explore the expression profile of CAFs. As CAF-derived proteins represent the most common way for intercellular crosstalk and might serve as biomarkers for cancer and identified representative coding gene. Here, we have explored the expression profile of the tumor bulk to analyze the characteristics of fibroblasts in LACs. We evaluated the impact of fibroblasts on patients' prognoses and further identified an extracellular secreted protein, COL5A2, as a biomarker for CAFs and a predictor for poor prognosis in LAC and a potential predictor for chemical therapy.

Clinically, activated CAFs have been associated with worse prognosis, resistance to therapies, and disease recurrence in multiple cancers [24,25]. It is significant to investigate the

CAF-associated genes and sieve representative biomarkers to predict the potential application in clinical evaluation. Therefore, in this study, we explored CAF-associated genes in two cohorts from TCGA dataset and GSE68465 and found that poorer clinical outcomes emerged in patients with more CAF-associated stromal scores. We also noticed that there were paradoxical results from different algorithms (Fig.1A-C, E-F). The cause may result from the constituents of these two cohorts because the patients had different clinical traits. However, higher stromal scores of LACs from both TCGA and GEO datasets all correlated to poorer patients' prognoses. This result suggests that tumor stroma is crucial to worse clinical outcomes.

To further analyze the inconsistent clinical outcomes from TCGA and GSE68465, we determined to extract their common genes and pathways to find the crucial CAF-related genes. Consequently, 12 genes were obtained and the result of GO and KEGG analyses showed that "protein digestion and absorption", "ECM-receptor interaction" pathways and extracellular matrix remodeling-related activities were upregulated in LACs. These outcomes were consistent with previous studies, which showed that fibroblasts exerted their function by producing excreted factors, remodeling the extracellular matrix, influencing cancer cell metabolism and direct cell-cell interactions. Furthermore, CAFs serve as leading cells for cancer cells during cell migration [26]. Fibroblasts pave the way for subsequent malignant cells and lead to tumor invasion by direct cell-to-cell contact.

CAFs are one of the components in tumor stroma and especially prominent due to abundance of CAFs and a complex TME, which is of significant impact on T cells recruitment, infiltration, and cytotoxic function within the tumor[27]. More attention should be focused on the specific activation of CAFs, therefore, we need to sieve and find the most representative genes that could accurately predict the response to clinical treatment. To degenerate and concentrate the representative genes, WGCNA and LASSO regression were used and a model consisting of two genes, COL5A2 and COL5A1, were constituted to predict the complex LAC stroma. Consequently, we testified the integrity of this model to CAF-related genes. The results demonstrated that COL5A2 and COL5A1 were all positively correlated to other CAF-related genes (Fig.3E). This result indicated that COL5A2 and COL5A1 were worthwhile to represent CAFs' traits.

Patients have different clinical outcomes when using a same treatment regimen. It is crucial to find characteristic markers to predict the possible outcomes, especially in immunotherapies. Cancer immunotherapies have rapidly changed the therapeutic landscape for cancer. Although impressive efficacy demonstrated in subsets of patients, most of the patients show innate or acquired resistance to these therapies [28-30]. A better understanding of the mechanisms that impede immune activation may thus enhance the

potential of cancer immunotherapy. An emerging role of CAFs have been highlighted in shaping the tumor immune microenvironment (TIME) and influencing response to cancer immunotherapies [31]. Extensive crosstalk between CAFs and cellular components of the immune system has been shown to contribute to immune escape and an immunosuppressive milieu of tumors via both biochemical and biomechanical mechanisms [31,32]. Hence, we subgrouped the patients from TCGA cohort in low-risk and high-risk groups based on COL5A2 and COL5A1expression. We then observed the anti-PD-1 response and found that more patients in high-risk group were beneficial from anti-PD-1 immunotherapy than those in low-risk group (54% vs 18%). This result prompt that our model is worthwhile to improve clinical prediction to immunotherapy. ROC curve showed the promising sensitivity and specificity. Future CAF-targeting therapeutic strategies particularly may be used in the context of optimizing the success of immunotherapies.

According to low- and high-risk groups, we further explored the related molecular pathways and function by GSEA analysis. ECM receptor interaction, focal adhesion pathways were involved in ECM remodeling. ECM remodeling is associated with fibroblast migration and proliferation. The promising yet limited success of cancer immunotherapy prompt intensified efforts in developing novel combination therapies to overcome resistance by targeting additional mechanisms that impede immune activation in TME. However, such efforts require a better explanation to the complex composition and diverse biology of TME from different tumor types, different histology of the tumors, different metastatic sites, or even within the same tumor [33,34]. We observed in this study that COL5A2 was the biomarker which is more concordant to the increase of risk score than COL5A1 (Fig. 3D). Therefore, we further evaluate it to identify its role in TME regulation. Firstly, we compared the expression of COL5A2 in tumor and normal lung tissues. And find that its expression is significantly upregulated in LACs. When compared in clinical characteristics, we noticed that COL5A2 expression increased with tumor size and closely associated with T stage. During the progression of tumor, COL5A2 keep higher level of expression except T4. These results indicate that COL5A2 participates in the construction of tumor at earlier stage. Overexpression of COL5A2 was also demonstrated as a hazardous factor for patients' prognoses whether in univariate and multivariate analysis. COL5A2 is a promising biomarker to predict clinical outcomes. COL5A2 can upregulate MXRA5, THBS2, COL5A1, ADAMTS12, COL1A2 and COL3A1. The activated molecular pathways were "PI3K-AKT", "neuroactive ligand-receptor interaction", "protein digestion and absorption" pathways. All these pathways influence extracellular matrix and structure organization of TME, which further impact on extracellular matrix structural constituents. Furthermore, we revealed that COL5A2 were mainly produced from fibroblasts whereas from LAC cells.

As known that TME consists of tumor parenchyma and stroma. The latter component comprises of CAF, immune cells, and extracellular matrix. To define the immune presence in TME, the concept of hot and cold tumors has initially emerged based on the presence or absence of T cells in the tumor, respectively [35]. Additional categories have been integrated to take into consideration the localization of the T cells within the tumor and stroma [36,37]. It is important to note that these classifications are an important framework to better understand the different tumor immune microenvironment (TIME) but also to help to tailor effective immunotherapies. Herein, we investigated the stromal score, immune score and the whole score of tumor bulk, respectively. The immune-associated scores were increased in LACs with high level of COL5A2 expression. Meanwhile, immune cells varied in TME with more infiltrating macrophages M0, CD4$^+$ memory activated T cells, resting NK cells than follicular helper T cells, activated NK cells, resting mast cells, and CD8$^+$ T cells. All these immune cells construct a useful framework not only by the quantity and spatial distribution of T cells, in particular, CD8$^+$ T cells in the TIME [28,38]. Historically, this classification of tumor-immune phenotypes is derived from CD3 or CD8 immunohistochemistry (IHC) analysis in solid tumors. However, such classification presents great challenges to the pathologists due to the continuous nature of T cell infiltration and high tumor heterogeneity. To address these challenges, the with low or absence of T cells both in the tumor epithelium and stroma [38,39]. With the increased understanding of CAFs and their multifaceted role in mediating immune suppression and shaping the tumor immunity continuum. The CAF-derived ECM is composed of a complex mixture of macromolecules, including collagen fibers, ECM-degrading proteases, glycosaminoglycans, and glycoproteins [40]. ECM proteins provide structural signals and support for tumor cells to grow and migrate. More importantly, over-production of ECM increases tissue stiffness and matrix rigidity and serves as a physical barrier that inhibits the access of antitumor immune cells and impedes the delivery of therapeutic drugs [41-43]. These may account for the lower level expression of COL5A2 in T4-staged tumor. New approaches targeting ECM components emerged to reduce the physical barrier of ECM in order to increase intratumoral permeability of antitumor immune cells and therapeutic agents. We predicted drug efficacy and find that alisertib, docetaxel, cisplatin, cytophosphamide and AT13148 had substantially therapeutic effect in LACs with high-level expression of COL5A2. These results suggest that cytotoxic drug could kill both tumor cells and CAFs to destroy extracellular matrix and increase its permeability. Alisertib is reported exhibiting various regulatory effects on the PI3K/Akt and mitogen-activated protein kinase (MAPK) pathways [44]. In addition, AT13148 is a first-in-class multi-AGC kinase inhibitor. AT13148 treatment in gastric cancer cells dramatically suppressed activation of multiple AGC kinases, including Akt (at p-Thr-308),

p70S6 kinase (p70S6K), glycogen synthase kinase 3beta (GSK-3beta) and p90 ribosomal S6 kinase (RSK) [45]. All these results prompt that LAC with COL5A2 overexpression will be beneficial from chemotherapy combined with conventional and targeting drugs.

In this study, owing to COL5A2 and COL5A1 high-risk group having better response to immunotherapy of anti-PD-1. Twenty LACs were then involved to evaluate the response to immune checkpoint inhibitor therapy (ICIT). We found that more than 90% tumor retraction occurred in LACs with COL5A2 overexpression. Further observation on the LACs with MPR and LPR revealed that survived tumor cells were encompassed in fibrotic stroma and were isolated from immune cells by proliferated collagens. In addition, the immune cells aggregated at the rim of remnant tumor nests. All these suggest that ICIT could have an effective tumor suppression, however, the tumor cell may potentially escape from immune surveillance by the isolation owing to reactive proliferative fibrous stroma after ICIT. It should pay more attention to tumor-specific increase in macromolecular permeability and enhanced the intratumoral delivery of the chemotherapeutic agents to inhibit tumor growth and prolong survival during ICIT.

We also realize the limitations of our study. First, the relative abundance of fibroblasts was evaluated by a clustering method using classic CAF markers with higher specificity instead of all well-known fibroblast markers. Besides, the results are generally comparable in LACs without considering specific gene status, such as EGFR, ALK or KRAS. Second, sampling bias might occur because the fraction of the stromal part varies in different samples due to the internal heterogeneity of the tumor bulk. Limited cases of LAC administrated with ICI and further stratification could not be achieved. Finally, the mechanism underlying how COL5A2 affects the TME, especially after ICIT was not explored in this study. Further investigations about how COL5A2 affects tumor cells and immune components, such as CD8 + T cells and macrophages after ICIT, will advance our understanding of the roles of CAFs in LACs.

Acknowledgments

Contributors

All authors participated in the study planning and analysis or laboratory experiments. X-Q.S and Y-R.C designed the study and

collected, processed and analyzed the datasets. X-L. W. evaluated pathological response of ICIT. X-Q.S performed the experiments. X-J.C.and K-A. L. collected and analyzed the clinical information. Y-R.C. performed the statistical analysis.

Data sharing statement

All datasets involved in this study can be viewed in Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) or data availability part of the corresponding articles. All the transcriptomic data could be obtained and analyzed according to the requirements and guidance of the providers.

References

1. Board WCoTE. (2021) Thoracic tumours / edited by WHO Classification of Tumours Editorial Board. World Health Organization classification of tumours., Fifth edition edition. Lyon: International Agency for Research on Cancer.

2. Lynch TJ, Bell DW, Sordella R et al. (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. The New England journal of medicine 350 : 2129-2139.

3. Haratani K, Hayashi H, Tanaka T et al. (2017) Tumor immune microenvironment and nivolumab efficacy in EGFR mutation-positive non-small-cell lung cancer based on T790M status after disease progression during EGFR-TKI treatment. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 28: 1532-1539.

4. Sahai E, Astsaturov I, Cukierman E et al. (2020) A framework for advancing our understanding of cancer-associated fibroblasts. Nature reviews Cancer 20: 174-186.

5. Denton AE, Roberts EW, Fearon DT (2018) Stromal Cells in the Tumor Microenvironment. Adv Exp Med Biol 1060: 99-114.

6. Thorsson V, Gibbs DL, Brown SD et al. (2018) The Immune Landscape of Cancer. Immunity 48: 812-830 e814.

7. Zheng S, Zou Y, Xie X et al. (2020) Development and validation of a stromal immune phenotype classifier for predicting immune activity and prognosis in triple-negative breast cancer. International journal of cancer Journal international du cancer 147: 542-553.

8. Tamborero D, Rubio-Perez C, Muinos F et al. (2018) A Pan-cancer Landscape of Interactions between Solid Tumors and Infiltrating Immune Cell Populations. Clinical cancer research : an official journal of the American Association for Cancer Research 24: 3717-3728.

9. Lambrechts D, Wauters E, Boeckx B et al. (2018) Phenotype molding of stromal cells in the lung tumor microenvironment. Nature medicine 24: 1277-1289.

10. Bertero T, Oldham WM, Grasset EM et al. (2019) Tumor-Stroma Mechanics Coordinate Amino Acid Availability to Sustain Tumor Growth and Malignancy. Cell Metab 29: 124-140 e110.

11. Boyle ST, Poltavets V, Kular J et al. (2020) ROCK-mediated selective activation of PERK signalling causes fibroblast reprogramming and tumour progression through a CRELD2-dependent mechanism. Nat Cell Biol 22 (7): 882-895.

12. Chen X, Song E (2019) Turning foes to friends: targeting cancer-associated fibroblasts. Nat Rev Drug Discov 18 : 99-115.

13. Chen Y, McAndrews KM, Kalluri R (2021) Clinical and therapeutic relevance of cancer-associated fibroblasts. Nature reviews Clinical oncology 18: 792-804.

14. Becht E, Giraldo NA, Lacroix L et al. (2016) Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol 17: 218.

15. Racle J, de Jonge K, Baumgaertner P et al. (2017) Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. Elife: 6.

16. Newman AM, Liu CL, Green MR et al. (2015) Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12 : 453-457.

17. Li T, Fu J, Zeng Z et al. (2020) TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res 48: W509-W514.

18. Yoshihara K, Shahmoradgoli M, Martinez E et al. (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 4: 2612.

19. Hanzelmann S, Castelo R, Guinney J (2013) GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 14: 7.

20. Cai YR, Dong YJ, Wu HB et al. (2016) Micropapillary: A component more likely to harbour heterogeneous EGFR mutations in lung adenocarcinomas. Sci Rep 6: 23755.

21. Gao S, Li N, Gao S et al. (2020) Neoadjuvant PD-1 inhibitor (Sintilimab) in NSCLC. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer 15: 816-826.

22. Costa A, Kieffer Y, Scholer-Dahirel A et al. (2018) Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. Cancer cell 33: 463-479 e410.

23. Kieffer Y, Hocine HR, Gentric G et al. (2020) Single-Cell Analysis Reveals Fibroblast Clusters Linked to Immunotherapy Resistance in Cancer. Cancer discovery 10: 1330-1351.

24. Calon A, Lonardo E, Berenguer-Llergo A et al. (2015) Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. Nature genetics 47: 320-329.

25. Fiori ME, Di Franco S, Villanova L et al. (2019) Cancer-associated fibroblasts as abettors of tumor progression at the crossroads of EMT and therapy resistance. Molecular cancer 18: 70.

26. Gaggioli C, Hooper S, Hidalgo-Carcedo C et al. (2007) Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells. Nat Cell Biol 9: 1392-1400.

27. Baker AT, Abuwarwar MH, Poly L et al. (2021) Cancer-Associated Fibroblasts and T Cells: From Mechanisms to Outcomes. J Immunol 206: 310-320.

28. Hegde PS, Chen DS (2020) Top 10 Challenges in Cancer Immunotherapy. Immunity 52: 17-35.

29. van Elsas MJ, van Hall T, van der Burg SH (2020) Future Challenges in Cancer Resistance to Immunotherapy. Cancers 12.

30. Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A (2017) Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. Cell 168: 707-723.

31. Barrett R, Pure E (2020) Cancer-associated fibroblasts: key determinants of tumor immunity and immunotherapy. Curr Opin Immunol 64: 80-87.

32. Henke E, Nandigama R, Ergun S (2019) Extracellular Matrix in the Tumor Microenvironment and Its Impact on Cancer Therapy. Front Mol Biosci 6: 160.

33. Jimenez-Sanchez A, Memon D, Pourpe S et al. (2017) Heterogeneous Tumor-Immune Microenvironments among Differentially Growing Metastases in an Ovarian Cancer Patient. Cell 170: 927-938 e920.

34. Jamal-Hanjani M, Quezada SA, Larkin J, Swanton C (2015) Translational implications of tumor heterogeneity. Clinical cancer research : an official journal of the American Association for Cancer Research 21: 1258-1266.

35. Gajewski TF, Schreiber H, Fu YX (2013) Innate and adaptive immune cells in the tumor microenvironment. Nat Immunol 14: 1014-1022.

36. Binnewies M, Roberts EW, Kersten K et al. (2018) Understanding the tumor immune microenvironment (TIME) for effective therapy. Nature medicine 24: 541-550.

37. Galon J, Bruni D (2019) Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. Nat Rev Drug Discov 18: 197-218.

38. Hegde PS, Karanikas V, Evers S (2016) The Where, the When, and the How of Immune Monitoring for Cancer Immunotherapies in the Era of Checkpoint Inhibition. Clinical cancer research : an official journal of the American Association for Cancer Research 22: 1865-1874.

39. Joyce JA, Fearon DT (2015) T cell exclusion, immune privilege, and the tumor microenvironment. Science 348: 74-80.

40. Botti G, Cerrone M, Scognamiglio G et al. (2013) Microenvironment and tumor progression of melanoma: new therapeutic prospectives. J Immunotoxicol 10: 235-252.

41. De Jaeghere EA, Denys HG, De Wever O (2019) Fibroblasts Fuel Immune Escape in the Tumor Microenvironment. Trends Cancer 5: 704-723.

42. Desbois M, Udyavar AR, Ryner L et al. (2020) Integrated digital pathology and transcriptome analysis identifies molecular mediators of T-cell exclusion in ovarian cancer. Nat Commun 11: 5583.

43. Ziani L, Chouaib S, Thiery J (2018) Alteration of the Antitumor Immune Response by Cancer-Associated Fibroblasts. Front Immunol 9: 414.

44. Ren B, Geng Y, Chen S et al. (2023) Alisertib exerts KRAS allele-specific anticancer effects on colorectal cancer cell lines. Exp Ther Med 25: 243.

45. Xi Y, Niu J, Shen Y et al. (2016) AT13148, a first-in-class multi-AGC kinase inhibitor, potently inhibits gastric cancer cells both in vitro and in vivo. Biochem Biophys Res Commun 478: 330-336.

Table 1: Genes from MCPcounter algorithm clustering analysis.

HUGO symbols	Cell population	ENTREZID	ENSEMBL ID
CD28	T cells	940	ENSG00000178562
CD3D	T cells	915	ENSG00000167286
CD3G	T cells	917	ENSG00000160654
CD5	T cells	921	ENSG00000110448
CD6	T cells	923	ENSG00000013725
CHRM3-AS2	T cells	100506915	ENSG00000233355
CTLA4	T cells	1493	ENSG00000163599
FLT3LG	T cells	2323	ENSG00000090554
ICOS	T cells	29851	ENSG00000163600
MAL	T cells	4118	ENSG00000172005
MGC40069	T cells	348035	NA
PBX4	T cells	80714	ENSG00000105717
SIRPG	T cells	55423	ENSG00000089012
THEMIS	T cells	387357	ENSG00000172673
TNFRSF25	T cells	8718	ENSG00000215788
TRAT1	T cells	50852	ENSG00000163519
CD8B	CD8 T cells	926	ENSG00000172116
CD8A	Cytotoxic lymphocytes	925	ENSG00000153563
EOMES	Cytotoxic lymphocytes	8320	ENSG00000163508
FGFBP2	Cytotoxic lymphocytes	83888	ENSG00000137441
GNLY	Cytotoxic lymphocytes	10578	ENSG00000115523
KLRC3	Cytotoxic lymphocytes	3823	ENSG00000205810
KLRC4	Cytotoxic lymphocytes	8302	ENSG00000183542
KLRD1	Cytotoxic lymphocytes	3824	ENSG00000134539
BANK1	B lineage	55024	ENSG00000153064
CD19	B lineage	930	ENSG00000177455
CD22	B lineage	933	ENSG00000012124
CD79A	B lineage	973	ENSG00000105369
CR2	B lineage	1380	ENSG00000117322
FCRL2	B lineage	79368	ENSG00000132704
IGKC	B lineage	3514	ENSG00000211592
MS4A1	B lineage	931	ENSG00000156738
PAX5	B lineage	5079	ENSG00000196092
CD160	NK cells	11126	ENSG00000117281
KIR2DL1	NK cells	3802	ENSG00000125498
KIR2DL3	NK cells	3804	ENSG00000243772
KIR2DL4	NK cells	3805	ENSG00000189013
KIR3DL1	NK cells	3811	ENSG00000167633
KIR3DS1	NK cells	3813	ENSG00000275037
NCR1	NK cells	9437	ENSG00000189430

PTGDR	NK cells	5729	ENSG00000168229
SH2D1B	NK cells	117157	ENSG00000198574
ADAP2	Monocytic lineage	55803	ENSG00000184060
CSF1R	Monocytic lineage	1436	ENSG00000182578
FPR3	Monocytic lineage	2359	ENSG00000187474
KYNU	Monocytic lineage	8942	ENSG00000115919
PLA2G7	Monocytic lineage	7941	ENSG00000146070
RASSF4	Monocytic lineage	83937	ENSG00000107551
TFEC	Monocytic lineage	22797	ENSG00000105967
CD1A	Myeloid dendritic cells	909	ENSG00000158477
CD1B	Myeloid dendritic cells	910	ENSG00000158485
CD1E	Myeloid dendritic cells	913	ENSG00000158488
CLEC10A	Myeloid dendritic cells	10462	ENSG00000132514
CLIC2	Myeloid dendritic cells	1193	ENSG00000155962
WFDC21P	Myeloid dendritic cells	645638	ENSG00000261040
CA4	Neutrophils	762	ENSG00000167434
CEACAM3	Neutrophils	1084	ENSG00000170956
CXCR1	Neutrophils	3577	ENSG00000163464
CXCR2	Neutrophils	3579	ENSG00000180871
CYP4F3	Neutrophils	4051	ENSG00000186529
FCGR3B	Neutrophils	2215	ENSG00000162747
HAL	Neutrophils	3034	ENSG00000084110
KCNJ15	Neutrophils	3772	ENSG00000157551
MEGF9	Neutrophils	1955	ENSG00000106780
SLC25A37	Neutrophils	51312	ENSG00000147454
STEAP4	Neutrophils	79689	ENSG00000127954
TECPR2	Neutrophils	9895	ENSG00000196663
TLE3	Neutrophils	7090	ENSG00000140332
TNFRSF10C	Neutrophils	8794	ENSG00000173535
VNN3	Neutrophils	55350	ENSG00000093134
ACVRL1	Endothelial cells	94	ENSG00000139567
APLN	Endothelial cells	8862	ENSG00000171388
BCL6B	Endothelial cells	255877	ENSG00000161940
BMP6	Endothelial cells	654	ENSG00000153162
BMX	Endothelial cells	660	ENSG00000102010
CDH5	Endothelial cells	1003	ENSG00000179776
CLEC14A	Endothelial cells	161198	ENSG00000176435
DIPK2B	Endothelial cells	79742	ENSG00000147113
EDN1	Endothelial cells	1906	ENSG00000078401
ADGRL4	Endothelial cells	64123	ENSG00000162618
EMCN	Endothelial cells	51705	ENSG00000164035
ESAM	Endothelial cells	90952	ENSG00000149564

ESM1	Endothelial cells	11082	ENSG00000164283
FAM124B	Endothelial cells	79843	ENSG00000124019
HECW2	Endothelial cells	57520	ENSG00000138411
HHIP	Endothelial cells	64399	ENSG00000164161
KDR	Endothelial cells	3791	ENSG00000128052
MMRN1	Endothelial cells	22915	ENSG00000138722
MMRN2	Endothelial cells	79812	ENSG00000173269
MYCT1	Endothelial cells	80177	ENSG00000120279
PALMD	Endothelial cells	54873	ENSG00000099260
PEAR1	Endothelial cells	375033	ENSG00000187800
PGF	Endothelial cells	5228	ENSG00000119630
PLXNA2	Endothelial cells	5362	ENSG00000076356
PTPRB	Endothelial cells	5787	ENSG00000127329
ROBO4	Endothelial cells	54538	ENSG00000154133
C	Endothelial cells	8436	ENSG00000168497
SHANK3	Endothelial cells	85358	ENSG00000251322
SHE	Endothelial cells	126669	ENSG00000169291
TEK	Endothelial cells	7010	ENSG00000120156
TIE1	Endothelial cells	7075	ENSG00000066056
VEPH1	Endothelial cells	79674	ENSG00000197415
VWF	Endothelial cells	7450	ENSG00000110799
COL1A1	Fibroblasts	1277	ENSG00000108821
COL3A1	Fibroblasts	1281	ENSG00000168542
COL6A1	Fibroblasts	1291	ENSG00000142156
COL6A2	Fibroblasts	1292	ENSG00000142173
DCN	Fibroblasts	1634	ENSG00000011465
GREM1	Fibroblasts	26585	ENSG00000166923
PAMR1	Fibroblasts	25891	ENSG00000149090
TAGLN	Fibroblasts	6876	ENSG00000149591

Table 2: GO function enrichment analysis and related genes

ONTOLOGY	ID	Description	geneID
BP	GO:0030198	extracellular matrix organization	MMP2/AEBP1/COL5A2/COL5A1/COL1A2
BP	GO:0043062	extracellular structure organization	MMP2/AEBP1/COL5A2/COL5A1/COL1A2
BP	GO:0045229	external encapsulating structure organization	MMP2/AEBP1/COL5A2/COL5A1/COL1A2
BP	GO:0035987	endodermal cell differentiation	MMP2/INHBA/COL5A2/COL5A1
BP	GO:0001706	endoderm formation	MMP2/INHBA/COL5A2/COL5A1
BP	GO:0030199	collagen fibril organization	AEBP1/COL5A2/COL5A1/COL1A2
BP	GO:0007492	endoderm development	MMP2/INHBA/COL5A2/COL5A1

BP	GO:0032963	collagen metabolic process	MMP2/INHBA/COL5A1/COL1A2
BP	GO:0001704	formation of primary germ layer	MMP2/INHBA/COL5A2/COL5A1
BP	GO:0007369	gastrulation	MMP2/INHBA/COL5A2/COL5A1
CC	GO:0062023	collagen-containing extracellular matrix	SPARC/ANGPTL2/MMP2/COL6A3/AEBP1/COL5A2/COL5A1/COL1A2/MXRA5/THBS2
CC	GO:0005581	collagen trimer	COL6A3/COL5A2/COL5A1/COL1A2
CC	GO:0005788	endoplasmic reticulum lumen	COL6A3/COL5A2/COL5A1/COL1A2
CC	GO:0005583	fibrillar collagen trimer	COL5A2/COL5A1/COL1A2
CC	GO:0098643	banded collagen fibril	COL5A2/COL5A1/COL1A2
CC	GO:0098644	complex of collagen trimers	COL5A2/COL5A1/COL1A2
CC	GO:0005604	basement membrane	SPARC/COL5A1/THBS2
CC	GO:0031091	platelet alpha granule	SPARC/THBS2
CC	GO:0031092	platelet alpha granule membrane	SPARC
CC	GO:0071682	endocytic vesicle lumen	SPARC
MF	GO:0005201	extracellular matrix structural constituent	SPARC/COL6A3/AEBP1/COL5A2/COL5A1/COL1A2/MXRA5/THBS2
MF	GO:0030020	extracellular matrix structural constituent conferring tensile strength	COL6A3/COL5A2/COL5A1/COL1A2
MF	GO:0048407	platelet-derived growth factor binding	COL5A1/COL1A2
MF	GO:0005518	collagen binding	SPARC/AEBP1
MF	GO:0046332	SMAD binding	COL5A2/COL1A2
MF	GO:0019838	growth factor binding	COL5A1/COL1A2
MF	GO:0008201	heparin binding	COL5A1/THBS2
MF	GO:0008237	metallopeptidase activity	MMP2/AEBP1
MF	GO:0005539	glycosaminoglycan binding	COL5A1/THBS2
MF	GO:1901681	sulfur compound binding	COL5A1/THBS2
MF	GO:0070696	transmembrane receptor protein serine/threonine kinase binding	INHBA

Table 3: KEGG pathway enrichment and the related genes

ID	Description	geneID
hsa04974	Protein digestion and absorption	COL6A3/COL5A2/COL5A1/COL1A2
hsa04512	ECM-receptor interaction	COL6A3/COL1A2/THBS2
hsa04510	Focal adhesion	COL6A3/COL1A2/THBS2
hsa05165	Human papillomavirus infection	COL6A3/COL1A2/THBS2
hsa04151	PI3K-Akt signaling pathway	COL6A3/COL1A2/THBS2
hsa04933	AGE-RAGE signaling pathway in diabetic complications	MMP2/COL1A2
hsa04926	Relaxin signaling pathway	MMP2/COL1A2
hsa05415	Diabetic cardiomyopathy	MMP2/COL1A2
hsa05205	Proteoglycans in cancer	MMP2/COL1A2
hsa05219	Bladder cancer	MMP2
hsa05144	Malaria	THBS2

Table 4. Association of pathological response to clinical stage and histological features

		Clinical stage				P value	Uniform histological components		P value	COL5A2 expression (%)			Proportion of Necrosis $\bar{X} \pm SD$, (95% CI)	Proportion of Inflammation $\bar{X} \pm SD$, (95% CI)	Proportion of Fibrosis $\bar{X} \pm SD$, (95% CI)	Proportion of residual viable tumor cells $\bar{X} \pm SD$ (95% CI)
		IB (%)	IIA (%)	IIB (%)	IIIA (%)		Yes	No		low	high	P value				
Pathological response	LPR	0	0	4 (33.3)	8 (66.7)	<0.001	2 (16.7)	10 (83.3)	0.005	9 (75)	3 (25)	0.098	0.11±0.07, (0.07-0.15)	0.24±0.05, (0.21-0.27)	0.34±0.05, (0.31-0.37)	0.31±0.07 (0.27-0.35)
	MPR	0	4 (100)	0	0		3 (75)	1 (25)		1 (25)	3 (75)		0.18±0.05, (0.1-0.25)	0.28±0.17, (0.003-0.55)	0.43±0.19, (0.12-0.73)	0.075±0.07 (0.03-0.12)
	CPR	2 (50)	2 (50)	0	0		4 (100)	0		1(25)	3 (75)		0.3±0.22, (0.044-0.64)	0.28±0.13, (0.08-0.48)	0.43±0.13, (0.23-0.63)	0

LPR, less pathological response; MPR, major pathological response; CPR, complete pathological response.