

Identification Of Gene Signature For Renal Cell Carcinoma-Associated Fibroblasts Mediating Cancer Progression And Affecting Prognosis

Bitian Liu

Shengjing Hospital of China Medical University <https://orcid.org/0000-0002-3747-9567>

Xiaonan Chen

Shengjing Hospital of China Medical University

Yunhong Zhan

Shengjing Hospital of China Medical University

Bin Wu

Shengjing Hospital of China Medical University

Shen Pan (✉ panshencmu@163.com)

<https://orcid.org/0000-0002-3431-7423>

Primary research

Keywords: renal cell carcinoma, cancer-associated fibroblast, weighted gene co-expression network analysis, single sample Gene Set Enrichment Analysis, gene signature

Posted Date: July 30th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-49601/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Cancer-associated fibroblasts (CAFs) are most abundant in stroma and are critically involved in cancer progression. However, the specific signature of CAFs and related clinicopathological parameters in renal cell carcinoma (RCC) remain unclear.

Methods: In this work, methods using recognized gene signatures were employed to roughly assess the infiltration level of the stroma and CAFs in RCC based on the data in The Cancer Genome Atlas. Weighted gene co-expression network analysis (WGCNA) was used to cluster transcriptomes and correlate with CAFs to identify specific markers. A comparison of fibroblast versus urothelial carcinoma cell lines and correlation with previously reported CAF markers were performed to demonstrate the specific expressed of the gene signature. The gene signature was used to compare fibroblast infiltration of each sample through single sample gene set enrichment analysis, and the clinical significance of fibroblasts was analyzed via Cox risk assessment and the chi-square test. Finally, we used validation data to verify the clinical significance of the fibroblast gene signature in RCC.

Results: Roughly calculated tumor matrix and CAF levels were significantly higher in kidney cancer than in normal tissues. More than 85% of fibroblast-specific markers identified by WGCNA were consistent with markers obtained via single-cell sequencing. These markers were more highly expressed in fibroblast cell lines and were significantly correlated with canonical CAFs makers. Data validation also showed that CAFs were significant correlation with survival and pathological grade.

Conclusions: In summary, our findings indicate that the gene signature potentially serves as a biomarker of CAFs in RCC and that infiltration of fibroblasts in RCC is an independent prognostic factor associated with pathological grade and stage of tumor. The ability to recognize specific CAF markers using WGCNA is comparable to single-cell sequencing.

Funding Sources

The present study was supported by 345 Talent Project from Shengjing Hospital of China Medical University (grant nos. M0312 and M0340), Doctoral Scientific Research Foundation of Liaoning Province (grant no. 2020-BS-120) and Joint plan of research and development program of Liaoning Province (grant no. 2020JH 2/10300137).

1. Background

Renal cell carcinoma (RCC) is one of the ten most deadly cancers in the world, causing more than 140,000 deaths each year[1]. Over the past decade, advanced RCC treatment has changed from a nonspecific immune approach to vascular endothelial growth factor targeted therapy, and now to immune checkpoint blockers[2]. Nevertheless, the overall prognosis for patients with advanced RCC is still poor[3]. Cancer-associated fibroblasts (CAFs) are an important part of the tumor microenvironment (TME), and they can function as tumor promoters and inhibitors[4]. Recently, CAFs have been gradually

recognized predominantly for their benefits for tumor progression[5], so targeting CAFs is a potential therapeutic strategy in the future[6]. However, there are few studies on CAFs in RCC.

According to previous reports, CAFs often carry recognized markers for identification, such as alpha smooth muscle actin (α SMA) and fibroblast activation protein alpha (FAP)[7]. There are specific CAF genes in different cancer tissues, so research using these recognized markers as gene signatures may cause deviations in the evaluation of CAFs in the TME. Single-cell transcriptomes can reveal the specificity of cells in specific tumor tissues, such as the cellular identity of human kidney tumors, which contain fibroblasts[8]. Although single-cell sequencing technology can classify cells and identify specific markers, the number of cells measured and the source of cases are limited, which may cause bias.

The R package weighted gene co-expression network analysis (WGCNA), which has been widely used recently[9], has the potential to recognize CAF-specific markers[10]. In the TME, different cells have varying specific gene expression, and the fluctuation of cell class proportions affects the expression of their specific genes. However, the expression of highly specific genes is less disturbed by the proportion of other cells. These genes show a strong correlation that cannot be offset as the fraction of cells changes. WGCNA can make good use of the special properties of the TME to identify cell-specific gene sets in a large number of heterogeneous samples.

In our study, we evaluated the infiltration status of the stroma and CAFs using previously identified gene signatures by the R packages Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) and Estimate the Proportion of Immune and Cancer cells (EPIC). By correlating the fraction of CAFs with the gene module calculated by WGCNA, the specific CAF gene set can be found. Comparison with the marker obtained by the algorithm of single-cell sequencing can verify the recognition ability of WGCNA. The gene signatures obtained by WGCNA were used to calculate the infiltration of fibroblasts in kidney renal clear cell carcinoma (KIRC), which is the main pathological type of RCC. Fibroblasts were found to be independent prognostic factors, related to pathological grade and stage.

2. Materials And Methods

2.1 Data download and processing

RNA-Seq and related clinical information for human KIRC samples were obtained from The Cancer Genome Atlas (TCGA) database (portal.gdc.cancer.gov), containing 611 tissues and 530 cases. These data were updated on April 10th, 2020. HTSeq data of 72 normal samples and 539 cancer samples were combined into matrix files.

For data verification, series GSE29609 containing various TNM-stage samples from 39 patients with survival information and series GSE53757 containing 101 normal tissue-tumor pair samples were downloaded from the Gene Expression Omnibus (GEO) database.

2.2 Stromal and immune components

Stromal and immune fractions were evaluated by ESTIMATE[11]. Using the ESTIMATE R package in R 3.6.2, tumor stromal and immune infiltrations of TCGA-KIRC samples were calculated from the profiles of 2 gene sets including 141 genes. The preliminary calculated stromal and immune scores are used to observe the differences between stage and grade.

2.3 Fractions of various cells in the TME

In order to continue to explore the changes in matrix components in tumor tissues, EPIC (gfellerlab.shinyapps.io/EPIC_1-1/) was used to estimate the fraction of CAFs. The EPIC application is designed to estimate the proportion of immune and cancer cells from bulk tumor gene expression data[12]. This is done by fitting gene expression reference profiles from the main non-malignant cell types, simultaneously accounting for an uncharacterized cell type without prior knowledge about it. EPIC establishes reference gene expression profiles for major tumor-invasive immune cell types (CD4⁺ T, CD8⁺ T, B, NK, and macrophages) and further deduces the reference spectra of CAFs and endothelial cells.

2.4 WGCNA

According to the principle of WGCNA calculation, highly co-expressed gene modules recognized by WGCNA can be considered as a set of specific expressed genes of a certain type of cells in tumor tissues. More tissue samples and greater fluctuations in the composition ratio make this method more suitable for identifying cell-specific expression gene sets. In a complex tumor environment, the detection of cell-specific expression genes is mainly determined by the proportion of one type of cell, while other cells will have less interference with such genes.

2.4.1 Highly co-expressed gene set - gene module

The WGCNA package in R was used to perform weighted correlation network analysis[13]. To exclude genes highly correlated with no significant changes, the differentially expressed genes with the highest variance of 25% were chosen[9, 14]. After filtering RNA-seq data to remove outliers, we constructed a Pearson correlation matrix and generated a weighted adjacency matrix emphasizing strong correlations and penalizing weak correlations. After selecting an appropriate β value via power calculation, a topological overlap matrix (TOM) was produced[15]. Based on TOM-dependent dissimilarity measurements, average-linkage hierarchical clustering and module dendrograms were used to construct modules with a minimum gene dendrogram size of 30.

2.4.2 Identification of interested modules

Gene significance (GS) was calculated to measure the correlation between genes and cell fractions and determine the significance of each module. The expression patterns of all module eigengenes were summarized as a single feature within a given module[14]. We selected a cutoff threshold of < 0.25 to merge some modules with similar heights and increase module capacity.[9]

2.4.3 Representative genes in a module

GS and module membership (MM, the correlation between the module's own genes and gene expression profiles) can assess gene-phenotype relationships and their importance in modules. Similar to most studies, we defined high MM and GS values (MM.cor and GS.cor, respectively) as the threshold to identify representative genes in a module[9].

2.5 Cellular identify of renal tumors

Single-cell transcriptomes revealed the cell-specific genes of different cells in RCC, and the study summarized the markers of canonical cell types in the existing literature[8]. To explore whether the representative genes screened by WGCNA are consistent with the cell-specific genes revealed by single-cell transcriptomes, we first looked for the specific genes of different cells in kidney cancer tissue revealed by single-cell sequencing. In particular, cell-specific gene expression of a cell type cluster was 20% higher than all other clusters, and no other cluster had more than 10% of cells expressing this gene[8]. Due to the variety and complexity of immune cells, we plan to compare only the specifically expressed genes of endothelial cells and fibroblasts in kidney tumor tissues.

2.6 Pathway and process functional enrichment analysis

In order to further verify that the gene module generated by WGCNA is a class of cell-specific expressed gene set, a Metascape (metascape.org) search was performed for gene enrichment analysis. The Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), Reactome gene sets, and CORUM provide ontology sources for pathway and process enrichment analysis[16]. Terms with a P-value < 0.01, a minimum count of 3, and an enrichment factor > 1.5 were collected and grouped into clusters based on their similarities. Subsets of enriched terms with a similarity score > 0.3 were connected by the edges to render a network plot and further capture the relationships between the terms.

2.7 Gene expression of different cell types

The gene expression profiles of different cell lines can prove that genes are specifically expressed in a class of cells. We download gene expression data of cell lines from the depmap portal (depmap.org/portal), which contains 1270 cell lines and 19144 genes. According to the tumor environment of kidney cancer, we selected cell lines labeled as blood, fibroblast, kidney, lymphocyte, and plasma cells as the comparison objects.

2.8 CAFs markers

The markers of CAFs have been collated and summarized[17, 18]. They comprise specific markers of CAFs as well as non-specific markers of CAFs. We will prove the reliability of the signature of CAFs in KIRC through correlation analysis using TCGA data.

2.9 Single sample gene set enrichment analysis

Single sample gene set enrichment analysis (ssGSEA) can identify and distinguish the changes in a class of cells in different samples based on gene signatures[14, 19]. We used the GSEA program to obtain the absolute enrichment scores from representative gene signatures previously identified by WGCNA. The infiltration level of one cell type was quantified by ssGSEA in the R package gsva, where ssGSEA utilized a deconvolution approach including myofibroblasts and fibroblasts.

2.10 Statistical analysis

Graphpad Prism 8.4 was used for statistical analysis. Statistical significance was determined using the Student's t-test (two-tailed) for 2 groups, one-way analysis of variance, and/or Tukey's test for more than 2 groups. Pearson's chi-square test was used to analyze the correlation between fibroblasts and clinicopathological parameters. The Cox proportional hazards regression model and the Kaplan-Meier curve were used to calculate the association of the clinicopathological characteristics and fractions of various cells with overall survival. Each group of data is presented as mean \pm SD. A p-value < 0.05 in statistical analysis was considered significantly different.

3. Results

3.1 Changing trends in the stroma and CAFs

First, we used ESTIMATE and EPIC to approximately calculate the infiltration of the stroma, CAFs, and immune cells in KIRC. The ESTIMATE result found that tumor tissues had a higher level of stromal and immune infiltration than normal tissues (Figure 1A). In EPIC, the proportion of CAFs showed a tendency to increase with tumor stage and grade progression, but the proportion of CD4⁺ and CD8⁺ T cells had no significant change in KIRC (Figure 1B). To visualize the scale changes in all cells in EPIC, we show the proportion of CAFs with the progression of stage and grade using the average fraction pie chart (Figure 1C). Based on the above results, the proportion of CAFs increases during tumor progression, or can be considered as an increase in the rate of proliferation of CAFs (Figure 2).

3.2 Rough CAFs and survival

Because the CAF fraction of EPIC is estimated based on gene signatures from other tissues, we suspected that this fraction may contain different cells and defined them as rough CAFs. In the prognosis significance of the Cox hazards model analysis of rough CAFs, it was found that the rough CAFs were associated with prognosis, but they were not an independent prognosis factor (Table 1). Age, grade, stage, and endothelial cells were all independent prognostic factors. Endothelial cells were associated with good prognosis (Table 1).

3.3 Identification of gene modules correlated to CAFs

TCGA KIRC sequencing samples provided WGCNA with 530 cases to effectively and objectively identify cell-specific gene sets. We chose $\beta = 6$ (no scale $R^2 = 0.814$) as a soft threshold to construct a scale-free

network. Three modules had a correlation with the CAF fraction of about 0.7. Among them, darkturquoise and lightgreen had a very close relationship in the cluster tree, which could originate from the identical type of cells, but the grey60 module was clearly not homologous to them (Figure 3A). According to the canonical cell types in normal human kidney tissue, with the exception of epithelial cells of different microanatomical regions and immune cells, the remaining cells are mainly vascular endothelial cells, fibroblasts, and myofibroblasts[8]. We compared the endothelium, fibroblast, and myofibroblasts markers[8], with MM.cor of different modules. The results show that the darkturquoise and lightgreen modules are closely related to fibroblasts, and that the grey60 and royalblue modules are closely related to myofibroblasts and endothelial cells, respectively (Figure 3B).

3.4 Markers recognized by single-cell sequencing technology and WGCNA

In order to confirm that WGCNA can identify cell-specific genes, we compared the differences in the identified genes. The markers identified by single-cell sequencing technology can be considered as the current standard for identifying cell-specific expressed genes. Genes specifically expressed by endothelial cells and fibroblasts could be found in the RCC single-cell sequencing results (Figure 3C)[8]. False discovery rate (FDR) < 0.05 is considered significant for the cell-specific expression of genes. Since there are too many endothelial markers, we selected the 500 most specific markers (smallest FDR) for comparison. According to the results shown in Figure 3A and Figure 3B, the royalblue module defines the specific endothelial gene set, and the darkturquoise and lightgreen modules define the specific fibroblast gene set. Although the intersection of the single-cell markers and module genes is not ideal, more than 85% of the representative genes (GS.cor > 0.6 and MM.cor > 0.6) in the module are single-cell markers (Figure 3D).

3.5 WGCNA can recognize cell-specific genes

Through the comparison of single-cell markers, the representative genes in the WGCNA recognition module can be regarded as cell-specific genes. However, not all cells have cell-specific gene sets, which can be illustrated by the WGCNA KIRC results. We only identified five specific gene modules for cells (Figure 3A). CAFs contain two kinds of cells, myofibroblasts and fibroblasts. According to the clustering relationship of modules shown in Figure 2A and the correlation between markers and modules shown in Figure 2B, the grey60 module is a representative gene set of myofibroblasts, and the darkturquoise and lightgreen modules are representative gene sets of fibroblasts. The WGCNA calculation principle is to identify highly co-expressed gene sets, while the expression of cell-specific genes changes according to the change in the proportion of cells. Although other cells may interfere with the expression of these genes, the impact on their highly co-expressed relationship is limited. We have drawn a schematic diagram to show this WGCNA function (Figure 4A). Obtaining cell-specific gene sets relies on a large number of samples with high tumor heterogeneity and cells with specific expression genes.

3.6 Function of cell-specific modules

After analyzing and identifying cell-specific genes on the principle of high co-expression, we used Metascape to further verify the function of these specific genes. The modules representing five types of cell-specific expressed genes were enriched and showed their respective functions. For example, fibroblasts and myofibroblasts, both of which belong to CAFs, are functionally similar, but the latter are more abundant in muscle-related functions (Figure 4B).

3.7 Fibroblast and myofibroblast markers - module representative genes

The representative genes in the module serve as specific markers for fibroblasts and myofibroblasts. In order to ensure the quantity and quality of the markers, we chose $GS.cor > 0.6$ and $MM.cor > 0.6$ as the screening conditions (Figure 5A). The lightgreen and darkturquoise fibroblast modules contain a total of 37 genes, including one non-coding protein gene. The grey60 myofibroblast module contains 24 genes. To verify the specificity of the 36 protein-coding fibroblast genes, we used Cancer Cell Line Encyclopedia data. Regarding the possible cell types in KIRC, it was found that the expression of 30 out of 36 genes was significantly higher in fibroblasts than in other cells (Figure 5B). We defined the 36 genes as signatures of CAFs in KIRC.

Among the CAF markers and CAF-specific markers, there were six and three intersections with the signature of CAFs in KIRC, respectively. The 36 key genes showed very high co-expression (Figure 5C). CAF-specific markers showed significantly higher correlation with the signature than non-specific markers did. In CAF-specific markers, FN1 and TNC were not correlated with multiple genes in the signature. Therefore, the CAF markers exhibit tissue specificity.

3.8 Fibroblasts in CAFs – a prognostic factor

Although rough CAFs were related to prognosis, they were not an independent prognostic factor. Now, we have identified two cellular components in rough CAFs estimated by EPIC, fibroblasts and myofibroblasts. Using representative genes calculated by WGCNA as gene signatures, ssGSEA was employed to score the infiltration level of fibroblasts and myofibroblasts in each sample (Figure 6A). According to the redefinition of CAFs as fibroblasts and myofibroblasts, the Cox risk regression analysis suggests that fibroblasts in KIRC are independent prognostic factors (Table 2).

3.9 Fibroblasts in CAFs and related clinicopathological parameters

Fibroblasts in CAFs are indeed related to tumor progression in KIRC. The ssGSEA score of fibroblasts was grouped according to the median of the corresponding cases of different clinicopathological parameters. In TCGA, increased fibroblast infiltration is related to tumor tissue, age < 60, male, stage III & IV, and Grade 3 & 4 parameters (Figure 6B). Moreover, in the data verification of series GSE29609 and GSE53757, we used the same gene signatures to score fibroblast infiltration (Figure 6C). In GSE29609, highly infiltrating fibroblasts were related to a worse prognosis (Figure 6D). Regarding clinicopathological parameters, a higher fibroblast count is associated with pathological grade 3 & 4 and tumor tissue (Figure 6E). Although the TNM staging and fibroblast infiltration level did not show significant differences, the Odds Ratio (OR)

value shows that their trend is consistent with the KIRC stage data in TCGA. However, gender and age do not show a clear relationship with fibroblasts.

4. Discussion

Fibroblast infiltration in KIRC is an independent prognostic factor related to the pathological grade and stage. The identification of fibroblast-specific markers in KIRC by WGCNA can be compared to the markers calculated by single-cell sequencing. WGCNA for identification of cell-specific gene sets in the complex TME may also be applicable to other cancers, and it can provide a basis for the identification and treatment of CAFs in different tumors in the future.

The standard treatment of many advanced cancers, including RCC, has changed from non-specific immunotherapy or chemotherapy to targeted therapy and immune checkpoint block therapy, from a single treatment to a combined treatment, and from a unified treatment to a personal treatment[20]. The combined precise treatment of cancer and stroma should be the future treatment strategy, which may greatly improve patient prognosis[21]. The TME and heterogeneity within the tumor provide invasive cells with advantageous conditions for cloning and growth[22]. In our research, we have shown that fibroblasts are potential invasive cells in the stroma of RCC.

CAFs affect the prognosis of cancer patients and are related to treatment resistance[23]. The interaction between tumor cells and CAFs may be the cause of treatment failure[24]. CAFs promote tumor matrix deposition and remodeling in the TME[25], help cancer cells evade immune surveillance[26], and achieve resistance to immunotherapy[27]. A study showed that the extracellular matrix can reduce the effectiveness of immune checkpoint blockers[28]. Due to the limited understanding of the origin and function of CAFs, it will be a challenge to target them in the future[25].

Rough CAFs in KIRC include fibroblasts and myofibroblasts. They are identified and distinguished in WGCNA clustering. Unlike fibroblasts, myofibroblasts have no demonstrated significance in prognostic analysis. Myofibroblasts play a major role in the fibrosis of chronic kidney disease, where compression of the blood vessels in the glomeruli leads to a decline in kidney function[29]. The grey60 module that represents myofibroblasts is closely related to the royalblue module of vascular endothelial cells in clustering. This can also explain the similarity of cell fractions between vascular endothelial cells and myofibroblasts.

It is very important to accurately define the specific markers of fibroblasts in various cancers, not only to identify fibroblasts, but also to accurately assess their infiltration level. Previous cancer studies have always confirmed fibroblasts based on the recognition of FAP and α SMA[30]. However, this only constitutes a basic distinction from cancer cells, and CAFs cannot be further classified. In the TME of different cancers, fibroblasts also have different specific expression markers from the other surrounding cells. Therefore, identifying cancer-specific CAF markers as targets is a goal for precise treatment or diagnosis.

WGCNA can even outperform single-cell sequencing in identifying cell-specific markers. First, the number of samples that WGCNA can process can come from hundreds or thousands of cases. The number of cells contained far exceeds the processing capacity of single-cell sequencing. Secondly, the module clustered by WGCNA is the specific gene set of different cells, which can be calculated only by high-throughput data sampling. However, no studies using WGCNA have reached such a conclusion so far. Most studies have used prognosis and tumor stage[31, 32] or immune cells[33] as the relevant phenotype to find the key immune genes related to cancer.

Regarding cancer gene prognosis models, kidney cancer research has been reported, but the genes used to build the models were different[34, 35]. This is highly related to the proportion of cells in the detected cancer tissue, so the results will also vary greatly. Relying on gene prognostic models presents very limited options for accurate prognosis. Conversely, assessing the degree of tumor fibroblast infiltration to determine prognosis is more widely applicable and stable. Prognosis by fibroblasts intervention in KIRC patients was confirmed in our study.

5. Conclusions

In summary, we demonstrated that the increased infiltration of fibroblasts in KIRC is significantly associated with tumor stage, pathological grade, and prognosis. WGCNA could identify the specifically expressed gene set of fibroblasts in the complex TME. WGCNA provides a new method for identifying specific markers of fibroblasts for specific tumors and lays the foundation for future tumor-targeted CAF treatment.

Abbreviations

CAFs: cancer-associated fibroblasts

RCC: renal cell carcinoma

WGCNA: weighted gene co-expression network analysis

TME: tumor microenvironment

ESTIMATE: Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data

EPIC: Estimate the Proportion of Immune and Cancer cells

KIRC: kidney renal clear cell carcinoma

TCGA: The Cancer Genome Atlas

GEO: Gene Expression Omnibus

TOM: topological overlap matrix

GS: gene significance

MM: module membership

KEGG: Kyoto Encyclopedia of Genes and Genomes

GO: Gene Ontology

ssGSEA: single sample gene set enrichment analysis

Declarations

Acknowledgments

Not applicable.

Funding

The present study was supported by 345 Talent Project from Shengjing Hospital of China Medical University (Grant Nos. M0312 and M0340), the Doctoral Scientific Research Foundation of Liaoning Province (Grant No. 2020-BS-120), and the Joint Plan of Research and Development Program of Liaoning Province (Grant No. 2020JH 2/10300137).

Conflicts of interests

The authors declare no conflict of interest.

Authors' contributions

BL proposed the study concept, designed, and drafted the manuscript. BL and SP collected, analysed, and interpreted the data. YZ, XC, and BW participated in revising the manuscript. All authors read and approved the final manuscript.

Data availability statement

The datasets used and/or analyzed during the current study are available from GEO (<https://www.ncbi.nlm.nih.gov/geo>) and TCGA (<https://portal.gdc.cancer.gov/repository>).

Ethics approval and consent to participate

Not applicable.

Reference

1. Capitanio U, Bensalah K, Bex A, Boorjian SA, Bray F, Coleman J, Gore JL, Sun M, Wood C, Russo P. Epidemiology of Renal Cell Carcinoma. *Eur Urol*. 2019;75(1):74-84. doi: 10.1016/j.eururo.2018.08.036.
2. Barata PC, Rini BI. Treatment of renal cell carcinoma: Current status and future directions. *CA Cancer J Clin*. 2017;67(6):507-524. doi: 10.3322/caac.21411.
3. Bosse D, Ong M. Evolution in upfront treatment strategies for metastatic RCC. *Nat Rev Urol*. 2020;17(2):73-74. doi: 10.1038/s41585-019-0271-6.
4. Park D, Sahai E, Rullan A. SnapShot: Cancer-Associated Fibroblasts. *Cell*. 2020;181(2):486-486.e481. doi: 10.1016/j.cell.2020.03.013.
5. Saini H, Rahmani Eliato K, Veldhuizen J, Zare A, Allam M, Silva C, Kratz A, Truong D, Mouneimne G, LaBaer J *et al*. The role of tumor-stroma interactions on desmoplasia and tumorigenicity within a microengineered 3D platform. *Biomaterials*. 2020;247:119975. doi: 10.1016/j.biomaterials.2020.119975.
6. Chen X, Song E. Turning foes to friends: targeting cancer-associated fibroblasts. *Nat Rev Drug Discov*. 2019;18(2):99-115. doi: 10.1038/s41573-018-0004-1.
7. Nurmik M, Ullmann P, Rodriguez F, Haan S, Letellier E. In search of definitions: Cancer-associated fibroblasts and their markers. *Int J Cancer*. 2020;146(4):895-905. doi: 10.1002/ijc.32193.
8. Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, Ferdinand JR, Collord G, Botting RA, Popescu DM, Loudon KW *et al*. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science*. 2018;361(6402):594-599. doi: 10.1126/science.aat1699.
9. Pan S, Zhan Y, Chen X, Wu B, Liu B. Identification of Biomarkers for Controlling Cancer Stem Cell Characteristics in Bladder Cancer by Network Analysis of Transcriptome Data Stemness Indices. *Front Oncol*. 2019;9:613. doi: 10.3389/fonc.2019.00613.
10. Liu B, Zhan Y, Chen X, Hu X, Wu B, Pan S. Weighted gene co-expression network analysis can sort cancer-associated fibroblast-specific markers promoting bladder cancer progression. *J Cell Physiol*. 2020. doi: 10.1002/jcp.29939.
11. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA *et al*. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612. doi: 10.1038/ncomms3612.
12. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife*. 2017;6. doi: 10.7554/eLife.26476.
13. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. doi: 10.1186/1471-2105-9-559.
14. Pan S, Zhan Y, Chen X, Wu B, Liu B. Bladder Cancer Exhibiting High Immune Infiltration Shows the Lowest Response Rate to Immune Checkpoint Inhibitors. *Front Oncol*. 2019;9:1101. doi: 10.3389/fonc.2019.01101.
15. Botia JA, Vandrovcova J, Forabosco P, Guelfi S, D'Sa K, Hardy J, Lewis CM, Ryten M, Weale ME. An additional k-means clustering step improves the biological features of WGCNA gene co-expression

- networks. *BMC Syst Biol*. 2017;11(1):47. doi: 10.1186/s12918-017-0420-6.
16. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10(1):1523. doi: 10.1038/s41467-019-09234-6.
 17. Gascard P, Tlsty TD. Carcinoma-associated fibroblasts: orchestrating the composition of malignancy. *Genes & development*. 2016;30(9):1002-1019. doi: 10.1101/gad.279737.116.
 18. Yu Z, Liao J, Chen Y, Zou C, Zhang H, Cheng J, Liu D, Li T, Zhang Q, Li J *et al*. Single-Cell Transcriptomic Map of the Human and Mouse Bladders. *J Am Soc Nephrol*. 2019;30(11):2159-2176. doi: 10.1681/asn.2019040335.
 19. Zhang L, Zhao Y, Dai Y, Cheng JN, Gong Z, Feng Y, Sun C, Jia Q, Zhu B. Immune Landscape of Colorectal Cancer Tumor Microenvironment from Different Primary Tumor Location. *Front Immunol*. 2018;9:1578. doi: 10.3389/fimmu.2018.01578.
 20. Gotwals P, Cameron S, Cipolletta D, Cremasco V, Crystal A, Hewes B, Mueller B, Quaratino S, Sabatos-Peyton C, Petruzzelli L *et al*. Prospects for combining targeted and conventional cancer therapy with immunotherapy. *Nat Rev Cancer*. 2017;17(5):286-301. doi: 10.1038/nrc.2017.17.
 21. Valkenburg KC, de Groot AE, Pienta KJ. Targeting the tumour stroma to improve cancer therapy. *Nat Rev Clin Oncol*. 2018;15(6):366-381. doi: 10.1038/s41571-018-0007-1.
 22. Parker TM, Henriques V, Beltran A, Nakshatri H, Gogna R. Cell competition and tumor heterogeneity. *Semin Cancer Biol*. 2020;63:1-10. doi: 10.1016/j.semcancer.2019.09.003.
 23. Paulsson J, Micke P. Prognostic relevance of cancer-associated fibroblasts in human cancer. *Semin Cancer Biol*. 2014;25:61-68. doi: 10.1016/j.semcancer.2014.02.006.
 24. Errarte P, Larrinaga G, Lopez JI. The role of cancer-associated fibroblasts in renal cell carcinoma. An example of tumor modulation through tumor/non-tumor cell interactions. *J Adv Res*. 2020;21:103-108. doi: 10.1016/j.jare.2019.09.004.
 25. Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, Fearon D, Greten FR, Hingorani SR, Hunter T *et al*. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer*. 2020;20(3):174-186. doi: 10.1038/s41568-019-0238-1.
 26. De Jaeghere EA, Denys HG, De Wever O. Fibroblasts Fuel Immune Escape in the Tumor Microenvironment. *Trends Cancer*. 2019;5(11):704-723. doi: 10.1016/j.trecan.2019.09.009.
 27. Galvani E, Mundra PA, Valpione S, Garcia-Martinez P, Smith M, Greenall J, Thakur R, Helmink B, Andrews MC, Boon L *et al*. Stroma remodeling and reduced cell division define durable response to PD-1 blockade in melanoma. *Nat Commun*. 2020;11(1):853. doi: 10.1038/s41467-020-14632-2.
 28. Wang L, Saci A, Szabo PM, Chasalow SD, Castillo-Martin M, Domingo-Domenech J, Siefker-Radtke A, Sharma P, Sfakianos JP, Gong Y *et al*. EMT- and stroma-related gene expression and resistance to PD-1 blockade in urothelial cancer. *Nat Commun*. 2018;9(1):3503. doi: 10.1038/s41467-018-05992-x.
 29. Ruiz-Ortega M, Rayego-Mateos S, Lamas S, Ortiz A, Rodrigues-Diez RR. Targeting the progression of chronic kidney disease. *Nat Rev Nephrol*. 2020;16(5):269-288. doi: 10.1038/s41581-019-0248-y.

30. Wu X, Zhou Z, Xu S, Liao C, Chen X, Li B, Peng J, Li D, Yang L. Extracellular vesicle packaged LMP1-activated fibroblasts promote tumor progression via autophagy and stroma-tumor metabolism coupling. *Cancer Lett.* 2020;478:93-106. doi: 10.1016/j.canlet.2020.03.004.
31. Jiang J, Bi Y, Liu XP, Yu D, Yan X, Yao J, Liu T, Li S. To construct a ceRNA regulatory network as prognostic biomarkers for bladder cancer. *J Cell Mol Med.* 2020;24(9):5375-5386. doi: 10.1111/jcmm.15193.
32. Giulietti M, Occhipinti G, Righetti A, Bracci M, Conti A, Ruzzo A, Cerigioni E, Cacciamani T, Principato G, Piva F. Emerging Biomarkers in Bladder Cancer Identified by Network Analysis of Transcriptomic Data. *Front Oncol.* 2018;8:450. doi: 10.3389/fonc.2018.00450.
33. Lin J, Yu M, Xu X, Wang Y, Xing H, An J, Yang J, Tang C, Sun D, Zhu Y. Identification of biomarkers related to CD8(+) T cell infiltration with gene co-expression network in clear cell renal cell carcinoma. *Aging (Albany NY).* 2020;12(4):3694-3712. doi: 10.18632/aging.102841.
34. Zhang Z, Lin E, Zhuang H, Xie L, Feng X, Liu J, Yu Y. Construction of a novel gene-based model for prognosis prediction of clear cell renal cell carcinoma. *Cancer Cell Int.* 2020;20:27. doi: 10.1186/s12935-020-1113-6.
35. Zhang C, Wang F, Guo F, Ye C, Yang Y, Huang Y, Hou J, Tian F, Yang B. A 13-gene risk score system and a nomogram survival model for predicting the prognosis of clear cell renal cell carcinoma. *Urol Oncol.* 2020;38(3):74.e71-74.e11. doi: 10.1016/j.urolonc.2019.12.022.

Tables

Table 1. Analysis of clinicopathological characteristics and infiltration of various cells calculated using EPIC with overall survival in the KIRC cohort

	Univariate analysis		Multivariate analysis	
	HR (95% CI)	P value	HR (95% CI)	P value
TCGA KIRC (n=519)				
Gender (male vs. female)	0.964(0.774-1.202)	0.748		
Age (>=60 vs. <60)	1.493(1.201-1.856)	< 0.001	1.419(1.137-1.771)	0.002
Grade (3 & 4 vs. 1 & 2)	1.987(1.559-2.533)	< 0.001	1.438(1.111-1.861)	0.006
Stage (iii & iv vs. I & ii)	2.522(2.014-3.158)	< 0.001	2.082(1.639-2.646)	< 0.001
Rough CAFs (>=median vs. <median)	1.322(1.066-1.639)	0.011	1.19(0.952-1.487)	0.127
B cells (>=median vs. <median)	0.888(0.718-1.098)	0.273		
CD4 T cells (>=median vs. <median)	1.042(0.842-1.288)	0.707		
CD8 T cells (>=median vs. <median)	0.931(0.753-1.15)	0.506		
Endothelial (>=median vs. <median)	0.685(0.551-0.852)	< 0.001	0.765(0.613-0.954)	0.017
Macrophages (>=median vs. <median)	1.201(0.969-1.489)	0.094		

Table 2. Analysis of clinicopathological characteristics and infiltration of various cells with overall survival in the KIRC cohort

	Univariate analysis		Multivariate analysis	
	HR (95% CI)	P value	HR (95% CI)	P value
TCGA KIRC (n=519)				
Gender (male vs. female)	0.964(0.774-1.202)	0.748		
Age (>=60 vs. <60)	1.493(1.201-1.856)	< 0.001	1.409(1.131-1.756)	0.002
Grade (3 & 4 vs. 1 & 2)	1.987(1.559-2.533)	< 0.001	1.437(1.11-1.86)	0.006
Stage (iii & iv vs. I & ii)	2.522(2.014-3.158)	< 0.001	2.054(1.614-2.614)	< 0.001
Fibroblasts (>=median vs. <median)	1.428(1.151-1.772)	0.001	1.282(1.026-1.6)	0.029
Myofibroblasts (>=median vs. <median)	1.021(0.756-1.379)	0.891		
B cells (>=median vs. <median)	0.888(0.718-1.098)	0.273		
CD4T cells (>=median vs. <median)	1.042(0.842-1.288)	0.707		
CD8 T cells (>=median vs. <median)	0.931(0.753-1.15)	0.506		
Endothelial (>=median vs. <median)	0.685(0.551-0.852)	< 0.001	0.77(0.617-0.961)	0.021
Macrophages (>=median vs. <median)	1.201(0.969-1.489)	0.094		

Figures

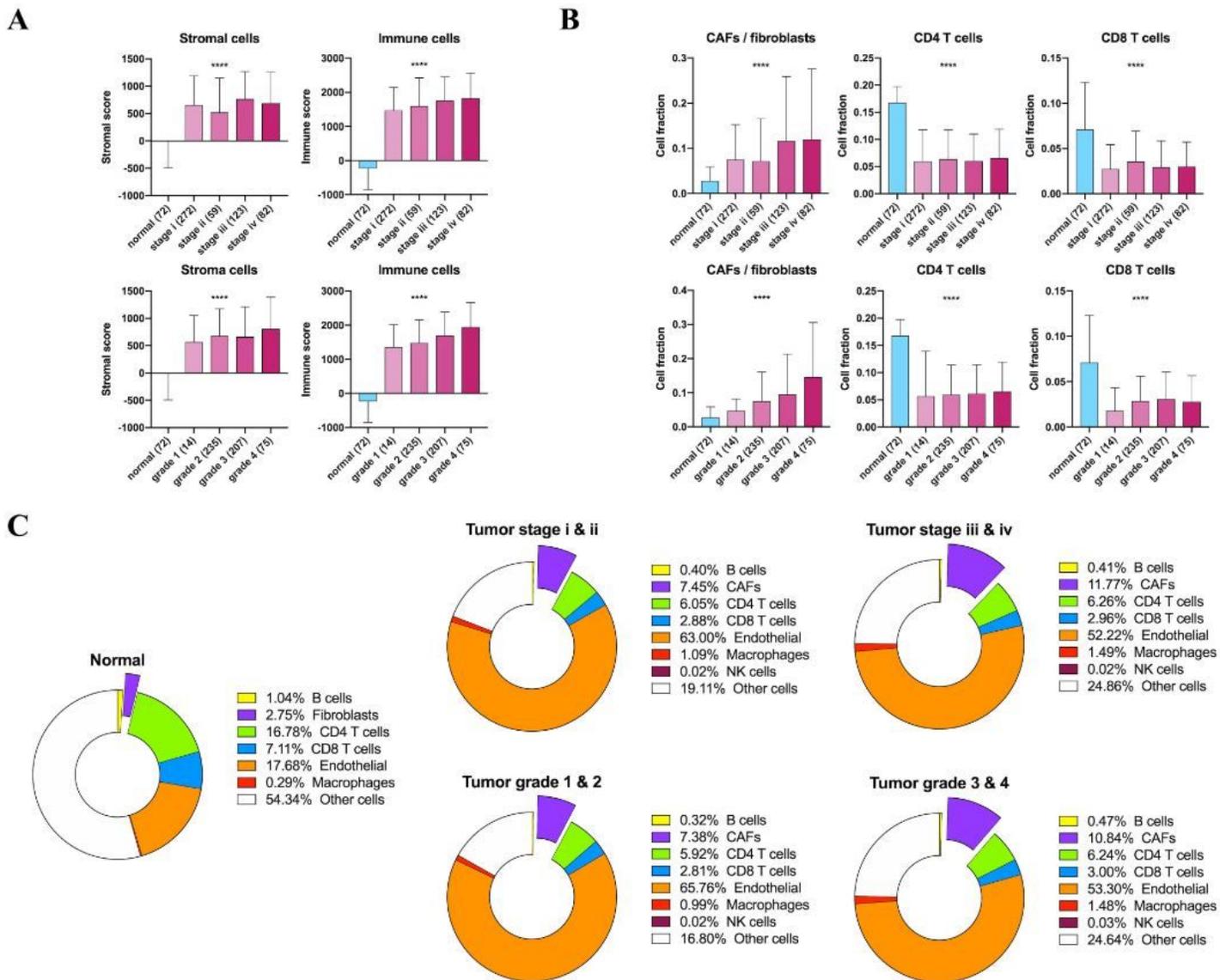


Figure 1

Stroma and CAFs in KIRC. (A) Stromal and immune scores in various stages and grades. (B) CAFs, CD4, and CD8 T cells in various stages and grades. (C) Cell fractions of fibroblasts or CAFs differ in tissue types, stages, and grades.

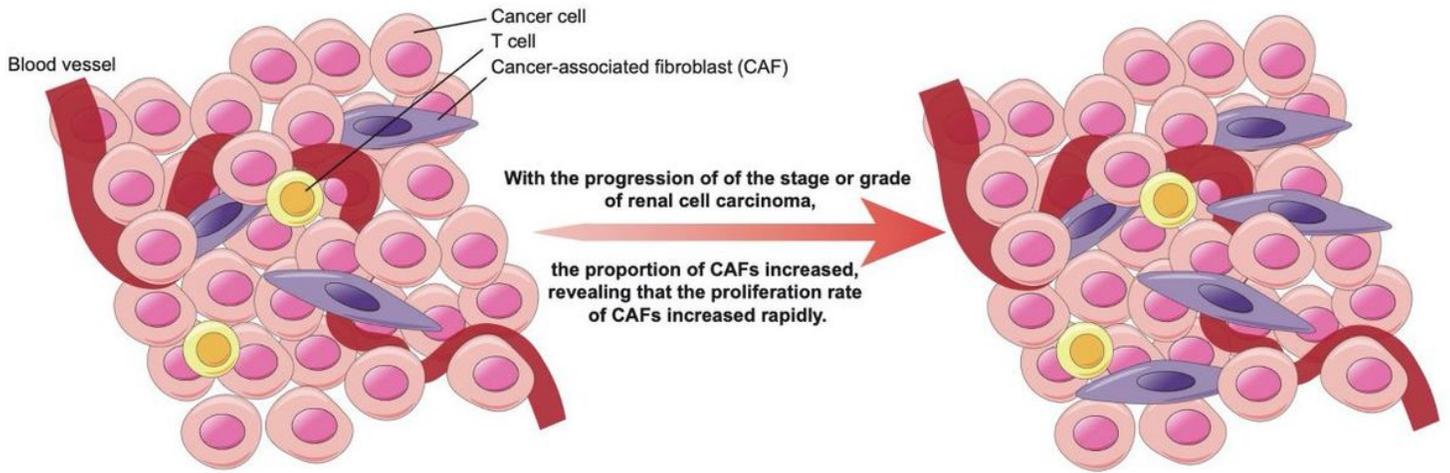


Figure 2

Increased proliferation rate of CAFs during tumor progression.

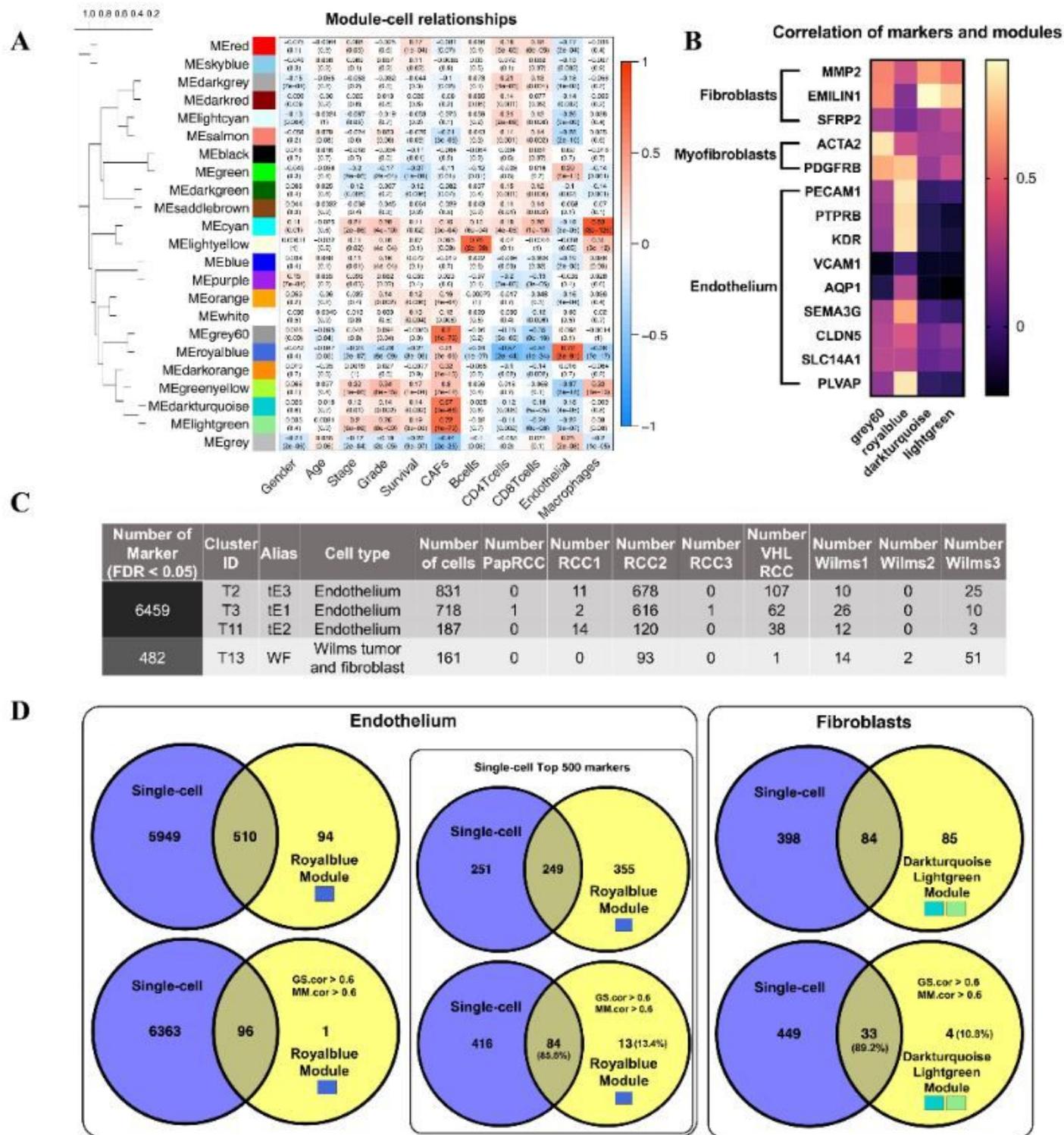


Figure 3

Modules identified by WGCNA contain cell-specific markers. (A) Correlation between the gene modules and traits, including clinical parameters and cell fractions of various cells estimated by EPIC. Correlation coefficients and p-values are presented in each cell. The dendrogram on the left shows the degree of difference between the modules. (B) MM.cor calculated by WGCNA of canonical markers and some

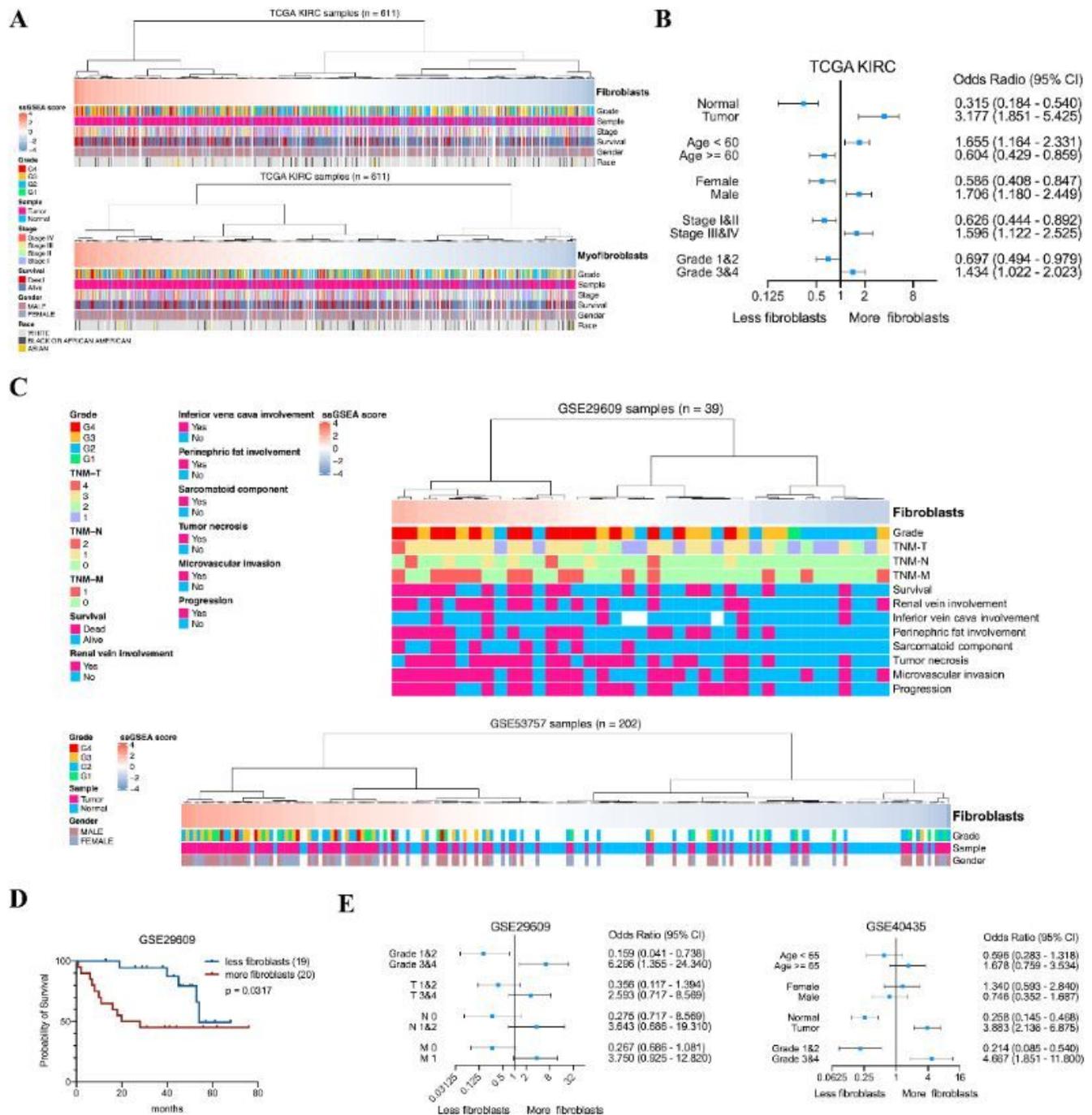


Figure 6

Verification of the clinical significance of fibroblasts. (A) Comparison of the infiltration level of fibroblasts and myofibroblasts in each sample in KIRC according to gene signature. (B) Relationship between fibroblasts and clinicopathological parameters in KIRC. (C) The infiltration level of fibroblasts of each sample in series GSE29609 and GSE53757 was calculated based on gene signature. (D) In GSE29609, patients with higher fibroblast infiltration had a worse prognosis. (E) Relationship between fibroblasts in KIRC and clinicopathological parameters in GSE29609 and GSE53757.