

RESEARCH

Open Access



Functional gene signature offers a powerful tool for characterizing clinicopathological features and depicting tumor immune microenvironment of colorectal cancer

Ziyan Zhu¹, Jikun Li¹, Zhenzhong Fa^{2,3,4}, Xuezhong Xu^{2,3,4}, Yue Wang^{2,3,4}, Jie Zhou^{2,3,4} and Yixin Xu^{2,3,4*}

Abstract

Background Colorectal cancer, a prevalent malignancy worldwide, poses a significant challenge due to the lack of effective prognostic tools. In this study, we aimed to develop a functional gene signature to stratify colorectal cancer patients into different groups with distinct characteristics, which will greatly facilitate disease prediction.

Results Patients were stratified into high- and low-risk groups using a prediction model built based on the functional gene signature. This innovative approach not only predicts clinicopathological features but also reveals tumor immune microenvironment types and responses to immunotherapy. The study reveals that patients in the high-risk group exhibit poorer pathological features, including invasion depth, lymph node metastasis, and distant metastasis, as well as unfavorable survival outcomes in terms of overall survival and disease-free survival. The underlying mechanisms for these observations are attributed to upregulated tumor-related signaling pathways, increased infiltration of pro-tumor immune cells, decreased infiltration of anti-tumor immune cells, and a lower tumor mutation burden. Consequently, patients in the high-risk group exhibit a diminished response to immunotherapy. Furthermore, the high-risk group demonstrates enrichment in extracellular matrix-related functions and significant infiltration of cancer-associated fibroblasts (CAFs). Single-cell transcriptional data analysis identifies CAFs as the primary cellular type expressing hub genes, namely *ACTA2*, *TPM2*, *MYL9*, and *TAGLN*. This finding is further validated through multiple approaches, including multiplex immunohistochemistry (mIHC), polymerase chain reaction (PCR), and western blot analysis. Notably, *TPM2* emerges as a potential biomarker for identifying CAFs in colorectal cancer, distinguishing them from both colorectal cancer cell lines and normal colon epithelial cell lines. Co-culture of CAFs and colorectal cancer cells revealed that CAFs could enhance the tumorigenic biofunctions of cancer cells indirectly, which could be partially inhibited by knocking down CAF original *TPM2* expression.

Conclusions This study introduces a functional gene signature that effectively and reliably predicts clinicopathological features and the tumor immune microenvironment in colorectal cancer. Moreover, the identification of *TPM2* as a potential biomarker for CAFs holds promising implications for future research and clinical applications in the field of colorectal cancer.

Keywords Colorectal cancer, Functional gene signature, Clinicopathological features, Tumor immune microenvironment, Single-cell transcriptomics

*Correspondence:

Yixin Xu

xuyixin@wjrmmy.cn

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Colorectal cancer (CRC) stands as a formidable challenge in the field of medicine, given its high prevalence and mortality rates worldwide [1, 2]. A distressing 20% of patients receive a diagnosis of distant metastasis upon their initial visit, and half of the remaining patients progress to Stage IV during therapy, demonstrating the highly metastatic nature of this disease [3]. Unfortunately, the development of new therapeutic strategies for CRC has been met with limited success, as the tumor exhibits a high degree of insensitivity and resistance to treatment. In addition to the heterogeneous somatic mutations observed in CRC patients, extensive research has illuminated the significant role played by the stromal component, particularly cancer-associated fibroblasts (CAFs), in the progression of CRC [4].

The tumor microenvironment (TME) encompasses the cellular and non-cellular components surrounding the tumor. Numerous studies have underscored the contribution of non-malignant cells within the TME, particularly fibroblasts, in driving tumor progression through intricate interactions involving growth factors, cytokines, chemokines, enzymes, and the extracellular matrix [5–7]. These fibroblasts, known as CAFs, have been associated with poor prognosis not only in CRC but also in other malignancies such as breast cancer, highlighting their pro-tumor biological behavior [8–10]. Consequently, targeting CAFs has emerged as a focal point for therapeutic interventions in contemporary medical research. Although various prediction models have been developed to assess prognosis and therapeutic response in different tumor types, the clinical value of a functional gene signature (FGS) based on CAFs, capable of stratifying CRC patients into distinct groups with unique prognoses and therapeutic responses, remains significant.

In the present study, we have constructed an FGS comprising both anti-tumor and pro-tumor genes. Utilizing single-sample Gene Set Enrichment Analysis (ssGSEA) scores, we were able to classify CRC patients into high- and low-risk groups. To validate the prognostic efficacy of the FGS, we employed several independent Gene Expression Omnibus (GEO) datasets. Furthermore, we performed comprehensive analyses of the biological functions, tumor immune microenvironment (TIME), and mutation status using a multitude of algorithms. Then we identified several hub genes within the FGS and evaluated their expression patterns through single-cell data and some molecular and cellular experiments. Lastly, we verified the bio-function of *TPM2* in CAFs through some functional experiments.

In summary, this study introduces a novel FGS that encompasses anti-tumor and pro-tumor genes. Through its application, CRC patients can be stratified into

high- and low-risk groups, enabling tailored prognostic assessments and therapeutic approaches. Additionally, our investigation sheds light on the biological functions, tumor immune microenvironment, and mutation status associated with CRC. The identification and evaluation of hub genes, especially *TPM2*, further enhance our understanding of the molecular intricacies underlying CRC development and progression.

Material and methods

Construction of the multiple gene signatures list

Initially, we selected 10 gene signatures to build the multiple gene signatures (MGSs) list. All of them were listed as follows:

1. HALLMARK_HYPOXIA: This gene signature contained 200 genes and could be searched in GSEA database (http://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_HYPOXIA.html) [11]. It has been identified to be related to the clinicopathological features and immune status of colorectal cancer in our previous studies [12, 13].
2. Cell death-related gene signatures: Autophagy-related gene signatures including 232 genes, which were obtained from the Human Autophagy Database (<http://www.autophagy.lu/>); Senescence, ferroptosis, pyroptosis, and cuproptosis related gene signatures were obtained from previous studies [14–17].
3. Immune-related gene signature: It was obtained from the Tracking Tumor Immunophenotype Database (<http://biocc.hrbmu.edu.cn/TIP/index.jsp>) [18].
4. Cancer related gene signatures: HALLMARK_EPITHELIAL-MESENCHYMAL_TRANSITION and HALLMARK_ANGIOGENESIS included 200 genes and 36 genes, respectively (http://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_EPITHELIAL-MESENCHYMAL_TRANSITION.html and http://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_ANGIOGENESIS.html).
5. Epigenetics related gene signature: It contained 900 genes and was collected from three previous studies [19–21].

The list including all these gene signatures was shown in Supplementary Table 1.

Data source

The FPKM and counts data of the protein-coding RNA-seq and the related clinical information of CRC patients were obtained from the TCGA database (<https://portal.gdc.cancer.gov/>) on April 15, 2022. They contained 488 tumor and 42 adjacent normal samples. Among them, 468 tumor samples had complete clinicopathological

data, including survival, pathological stage, invasion depth, lymphatic metastasis, and distant metastasis.

The independent validation sets for the functional gene signature (FGS) included GSE39582 [22], GSE17538 [23–26], GSE38832 [27], GSE28722 [28], and GSE103479 [29]. GSE39582, GSE17538, and GSE38832 (GPL570 platform) contained 562, 232, and 122 CRC patients, respectively. GSE28722 (GPL13425 platform) contained 129 and GSE103479 (GPL23985 platform) contained 155 CRC patients. All GEO datasets mentioned above had clinicopathological and survival data. Additionally, GSE60331 (GPL15207 platform, 50 rectal tumor samples) was used to evaluate the ability of the FGS to classify different therapeutic responses to immunotherapy [30]. GSE132465 was applied to identify the levels and the specific cell type of the expression of the hub genes [31]. All the GEO sets were searched from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) on September 15, 2022. The details are shown in Supplementary Table 2.

The procedure of the construction of the functional gene signature

First, the count data were used to identify the differentially expressed genes (DEGs) between the tumor and adjacent normal samples. Supplementary Table 3 showed 5498 DEGs with $\log_2|FC| > 1$ & adjust. $p < 0.05$, which were screened out using *DESeq2* package (version 1.32.0) [32]. Second, we used univariate Cox regression analysis was used to evaluate the association between genes and overall survival (OS). The protective genes were defined as $HR < 1$ & $p < 0.05$, while the harmful genes as $HR > 1$ & $p < 0.05$. Supplementary Tables 4 and 5 showed the list of protective genes and harmful genes, respectively.

There were 24 genes in the intersection among DEGs, protective genes, and MGSs as the antitumor genes, and 22 genes in the intersection among DEGs, harmful genes, and MGSs as the protumor genes. The complete FGS consisted of the antitumor genes and protumor genes (Supplementary Table 6).

After the screening out procedure, the anti-tumor gene was defined as a gene having $HR < 1$ ($p < 0.05$) and higher expressed in the adjacent normal tissues ($\log_2|FC| > 1$ & adjusted $p < 0.05$); while the pro-tumor gene was defined as a gene having $HR > 1$ ($p < 0.05$) and higher expressed in the tumor tissues ($\log_2|FC| > 1$ & adjusted $p < 0.05$).

ssGSEA score and consensus clustering

The ssGSEA scores of the TCGA and GEO samples were computed with the GSVA algorithm using the *GSVA* package (version 1.40.1) [33]. After all samples were conferred with a specific ssGSEA score, which included anti-tumor score and pro-tumor score, they were classified

into low- and high-risk groups using the *Consensus-ClusterPlus* package (version 1.56.0) [34]. Because FGS had two gene lists, the optimal cluster number is always 2. Samples with relatively higher anti-tumor scores and lower pro-tumor scores were defined as the low-risk group, while the others were defined as the high-risk group.

Next, two groups would be compared in clinicopathological feature, gene expression, pathway activity, tumor immune microenvironment (TIME), etc. The complete pipeline of this study was shown in Fig. 1.

Analyses of cancer related signaling pathways and immune microenvironment

The activities of cancer related signaling pathways, including Androgen, EGFR, Estrogen, Hypoxia, JAK-STAT, MAPK, NF κ B, P53, PI3K, TGF β , TNF α , Trail, and WNT, were evaluated using *progeny* package (version 1.14.0) [35–37]. Every sample would be assigned a score according to the expressions of the marker genes of each pathway. Then, the scores of these pathways would be compared between the low-risk and high-risk groups using the Wilcox test.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were analyzed using the *msigdb* (version 7.5.1) [38], *GSVA*, *edgeR* (version 3.34.1) [39, 40], and *limma* (version 3.48.3) [41] packages. First, the FPKM data were converted into CPM form. Next, the scores of KEGG pathways were calculated using GSVA algorithm. Then, these fold changes of these scores would be compared using *limma* package.

Several packages, including *CIBERSORT* [42], *MCP-counter* (version 1.2.0) [43], *xCell* (version 1.1.0) [44], and *ESTIMATE* (version 1.0.13) [45], were used to assess the infiltration of immune cells and cancer associated fibroblast (CAF) in the tumor microenvironment (TME).

Differentially expressed genes and related analyses of gene function (kyoto encyclopedia of genes and genomes and gene ontology)

Due to controversies about the appropriate methods for screening out the DEGs using FPKM data, we calculated the fold change (FC) of the expression of genes using the *edgeR* package with counts data. While for the calculation of the microarray data, the *limma* package was used. The criterion was $\log_2|FC| > 1$ & adjust. $p < 0.05$.

After the DEGs were identified, they would be reordered according to the FC. Next, we used the *clusterProfiler* (version 4.0.5) package to analyze the gene function enrichment of the DEGs, including KEGG and Gene Ontology (GO) [46, 47]. To better describe the enrichment of gene function of different groups, we calculated the enrichment scores of hallmark-related pathways

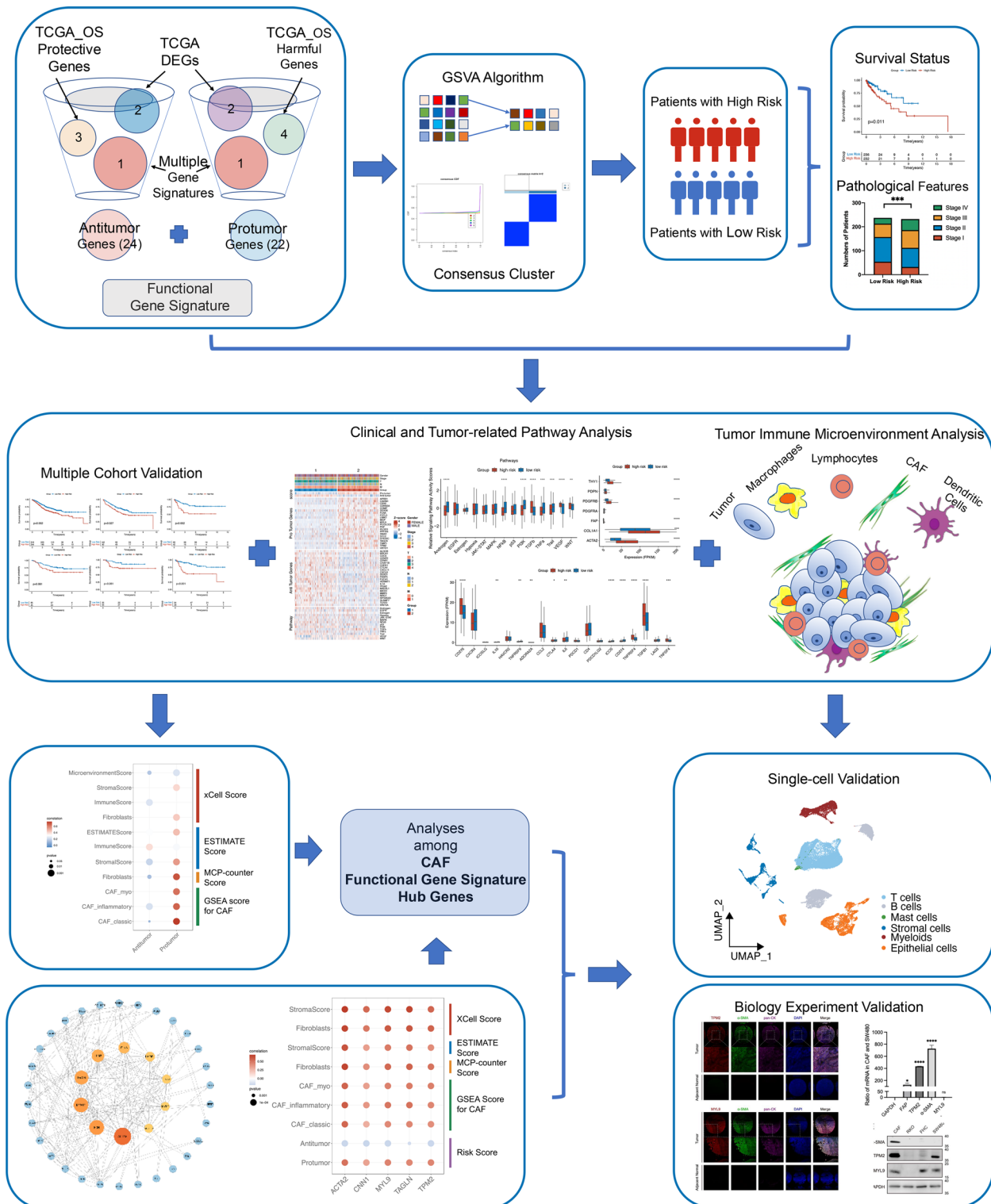


Fig. 1 The complete pipeline of the study

between two groups. The gene signatures of hallmark-related pathways were collected from the *msigdb* package. The enrichment score of each sample was calculated using the *GSEA* package. Then, the enrichments of hallmark-related pathways between the two groups were compared using the *limma* package. Pathways with $\log_2|FC| > 1$ & $\text{adjust.}p < 0.05$ were considered to be significantly different.

Somatic mutation and prediction of immunotherapy

The mutation ratio of specific genes, mutation type, and tumor mutation burden (TMB) were compared between the low- and high-risk groups using the *maftools* (version 2.8.05) [48].

After the TMB of every sample was calculated, the survival statuses between low- and high-TMB groups were compared. Meanwhile, the survival statuses of different groups with different risk scores plus TMBs were compared, too.

We also used Tumor Immune Dysfunction and Exclusion (TIDE), a website (<https://tide.dfci.harvard.edu/>) based on the ssGSEA algorithm, to evaluate the different infiltration of specific immune cells and the response to the immunotherapy [49, 50]. Higher dysfunction, exclusion, and TIDE score meant lower infiltration, migration, activity of the pro-immune cells, which were equal to the worse therapeutic effect of the immunotherapy. Next, GSE60331 was applied to assess whether FGSs could predict the response of immunotherapy in an independent cohort.

Single cell analysis for hub genes

GSE132465 was applied to confirm the specific cell type that expressed the hub genes. GSE132465 contained 33 samples of CRC patients in Korea, including 25 tumor and 10 adjacent normal samples. Ten paired tumor and adjacent normal samples from GSE132465.

To eliminate the batch effect, we used multiple CCA procedures in *Seurat* (version 4.1.1) [51, 52]. The thresholds of quality control (QC) were $200 < \text{genes} < 5000$ and $\text{percent.mt} < 20\%$. We annotated cells into 6 main types, including B cells, epithelial cells, mast cells, myeloid, stromal cells, and T cells, using annotations provided by Lee et al. For detailed cell subtype analysis, we used cell markers collected from CellMarker 2.0 and Panglao DB [53, 54]. We divided the stromal cells into 6 subtypes, including progenitor cells, CAF, fibroblasts, endothelial cells, enteric glial cells, and smooth muscle cells.

In GSE132465, there were 39,197 cells included in the analysis. After QC, 38,085 cells were left, which contained 5791 stromal cells.

Cell culture and culture conditions

Cell lines (SW480, RKO, SW620 and FHC) and CAFs were maintained at 37 °C in the humidified incubator with 5% CO₂. Dulbecco's Modified Eagle Medium (Thermo Fisher Scientific #11,965,092) was used as media with 10% FBS (Thermo Fisher Scientific #A5669401).

Cafs extraction and transfection

The tumor tissue samples were placed in Dulbecco's Modified Eagle Medium (Thermo Fisher Scientific #11,965,092) containing 1% FBS (Thermo Fisher Scientific #A5669401) after obtaining from patients. They were kept on ice and soaked in the medium during transference. Then the tissue was shifted into a sterile 10 cm culture dish, removed blood clots and adipose connective tissue with sterile forceps. Sterile scissors were used to cut the tissue into approximately 1mm³ fragments. Then the tissue was rinsed with sterile PBS (Thermo Fisher Scientific #10,010,023) and transferred to a new sterile 50 ml centrifuge tube. 20 mL of PBS was added and mixed and then discarded after 3–5 min. Repeat this 2–3 times until the PBS became clear. The tissue was then transferred to a 15ml centrifuge tube and centrifuged to remove the supernatant, and 8-10ml of 10% FBS culture medium containing collagenase I (1mg/ml, Millipore Sigma #1,148,089) was added. The tube was gently shaken on a 37°C shaker for 2–8 h until the tissue fragments appeared as a cloudy meat broth with no large fragments. The mixture was then filtered through a 40-micron nylon mesh and then through a 100-micron nylon mesh. Then the filtrate was centrifuged at 300 g for 10 min to discard the supernatant while the pellet was resuspended in fresh DMEM culture medium, and placed in a 10 cm culture dish together with the filtered tissue fragments. The dish was cultured in a humidified incubator with 5% CO₂ for approximately 3 days. If tissue fragments adhered to the dish and cells attached to the surface, the culture medium could be replaced. CAFs should start to emerge within approximately 1–2 weeks.

After fibroblast-like cells were gathered, tissue fragments were abandoned, and the remaining cells were collected to another cell dish for latter culturing and transfecting. pSLenti-U6-shRNA-CMV-EGFP-F2A-Puro-WPRE and shRNAs were purchased from OBiO Technology (Shanghai) Corp.,Ltd. Target sequences for shRNAs are included in Supplementary Table 10.

Co-culture

After transfected with different virus, CAFs were seeded in the upper Transwell inserts in 6-well plates (Corning® #3412) while Tumor cells and SW620s were seeded in the below well. The plates were cultured in a humidified

incubator with 5% CO₂ for 48 h, after which the inserts were removed and tumor cells were collected for the following functional experiments.

Ki-67 staining

Ki-67 Staining was conducted according to standard procedure. Briefly, tumor cells were seeded on the Corning BioCoat Cellware Poly-D-Lysine coverslips (Corning® #354,086) in 6-well plates overnight and fixed for the staining. 50–100 µl of permeabilization solution (Wuhan servicebio technology CO.,LTD #G1204) was used to permeabilize the cell membrane for 20 min at room temperature. 3% BSA (Wuhan servicebio technology CO.,LTD #GC305010) was used to block at room temperature for 30 min. The ki67 primary antibody (Wuhan servicebio technology CO.,LTD #GB151141) was diluted in PBS to 1:300 and then added to the cell culture wells. The culture plate was placed flat inside a humidified box and incubated at 4 °C overnight. CY3 marked second antibody was added and incubated at room temperature for 50 min. DAPI stain solution (Wuhan servicebio technology CO.,LTD #G1012) was used at room temperature for 10 min to stain the nucleus. An anti-fade mounting medium (Wuhan servicebio technology CO.,LTD #G1401) was used lastly. Images were acquired by NIKON ECLIPSE C1 for the following segments:

DAPI: excitation 330–380 nm, emission 420 nm;

CY3: excitation 510–560 nm, emission 590 nm;

Cell proliferation and cytotoxicity assay with cell counting kit-8

After co-cultured with different CAFs, Tumor cells were digested into cell suspension at the density of 4×10^4 and seeded in a 96-well plate (100 µL/well). The plate was placed in a humidified incubator (37 °C, 5% CO₂) and cultured for 24 h for the cells to adhere to the bottom of the plate. Then 10 µl of the Cell Counting Kit-8 solution (Abcam, #ab228554) was added to each well of the plate. The plate was incubated at 37 °C for 4 h in the incubator. Then absorbance at 450 nm of each well was measured using a Multiskan SkyHigh Microplate Spectrophotometer (Thermo Fisher Scientific #A51119500C). Data was analyzed by Prism GraphPad (version 8.4.0).

Wound healing

After co-cultured with different CAFs, tumor cells were digested and seeded into 6-well plates for 4×10^7 cells to make sure the cells reach full confluence to mimic a continuous tissue layer. After adhered to the well bottom, a sterile pipette tip was used to create a straight scratch across the cell monolayer, simulating a wound. The wells

were then washed with PBS and added DMEM without serum. Images of the wound were taken every 24 h until 96 h after the first picture. The medium was replaced everyday to wash out dead floating cells. The area of the wound was analyzed by ImageJ (version 2.14.0) and Prism GraphPad (version 8.4.0).

Clonogenic assay

CAFs Co-cultured tumor cells, *TPM2* shRNA CAFs Co-cultured tumor cells, and non-treated Tumor cells were digested by Trypsin–EDTA (0.25%) (Thermo Fisher Scientific #25,200,072) and resuspend multiple times by DMEM into a single-cell suspension. The density of the suspension was calculated by cell counting chamber. Then cells were seeded into 6-well plates for 800 cells per well. The plates were placed an incubator with 5% CO₂ for approximately 2 weeks (medium was replaced once every three days) until the cells have formed sufficiently large colonies, typically consisting of at least 50 cells. After the incubation period, the colonies were fixed with 4% glutaraldehyde and stained with crystal violet (Biosharp #BS941-5 g). Then the stereomicroscope was used to count the colonies with more than 50 cells. The plating efficiency and survival fraction was analyzed by Prism GraphPad (version 8.4.0).

Transwell assay

CAFs Co-cultured Tumor cells, *TPM2* shRNA CAFs Co-cultured Tumor cells, and non-treated Tumor cells were digested by Trypsin–EDTA (0.25%) and resuspend by serum-free DMEM. The density of the suspension was calculated by cell counting chamber. Then cells were seeded into the upper chamber of Transwell inserts in 24-well plates (Corning® #3422) coated with Corning® Matrigel® Matrix (Corning® #356,237) for 2.5×10^4 cells per well. Then DMEM medium with 20% fetal serum was placed in the lower chamber. The plates were placed an incubator with 5% CO₂ for 48 h. After incubation, the Transwell inserts were washed by PBS and then fixed by 4% glutaraldehyde, then stained with crystal violet. Cells that have migrated to the lower surface were counted under a microscope. Data was analyzed by Prism GraphPad (version 8.4.0).

RNA extraction, RT-qPCR, and multiple immunohistochemistry staining

RNA extraction, RT-qPCR, and multiple immunohistochemistry staining followed the routine procedures, which were reported in our previous study [13]. The primers used in RT-qPCR was shown in Supplementary Table 9.

Western blot

Upon reaching the appropriate cell culture density, the cells were lysed using RIPA Lysis Buffer (Beyotime Biotechnology #P0013B), supplemented with Protease and Phosphatase Inhibitor Cocktail (Beyotime Biotechnology #P1045). The concentration of total cellular protein was determined using the Enhanced BCA Protein Assay Kit (Beyotime Biotechnology #P0009). Subsequently, 25 µg of protein samples were combined with SDS-PAGE Loading Buffer (New cell & Molecular Biotech Co., Ltd #WB2001), and the mixture was boiled at 95 °C for 5 min. The protein samples were then loaded onto 12% Tris–Glycine-SDS PAGE gels (Nanjing Vazyme Biotech #E304-01) and subjected to electrophoresis in Tris SDS running buffer (Epizyme Biotech #PS105), focusing on the proteins GAPDH, MYL9, TPM2, and -SMA.

Following electrophoresis, the proteins were transferred to membranes using the Trans-Blot Turbo system (BioRad). To prevent nonspecific binding, the membranes were incubated with Blocking Buffer (New cell & Molecular Biotech Co., Ltd #P30500) for 15 min at room temperature. Afterward, the membranes were washed three times with 10% TBST buffer and subsequently incubated overnight at 4°C with primary antibodies. Following another round of washing with 10% TBST buffer, the membranes were incubated overnight at room temperature with secondary antibodies for 1 h. The membranes were then washed three times, and the protein bands were visualized using the Odyssey® DLx Imaging System (LI-COR, Inc). Finally, the acquired images were analyzed using Image J 9.0. All the antibodies used in this assay were shown in Supplementary Table 8.

Statistical analysis

Continuous variables, including the FPKM data, ssGSEA scores, TIDE scores, etc., were compared between different groups using the Wilcoxon test. Ranked variables, such as pathological stage, invasion depth, lymphatic metastasis, and distant metastasis, were analyzed using the Mann–Whitney U test. The prediction ability of hub genes was assessed using the receiving operation curve (ROC). Univariate Cox regression analysis was applied to identify genes that were significantly associated with the overall survival (OS) of the TCGA CRC cohort. The survival statuses, including OS, disease-free survival (DFS), and regression-free survival (RFS), were evaluated using Kaplan–Meier analysis. The association between continuous variables was compared using Spearman analysis. A variable with $p < 0.05$ was considered to be statistically significant. All statistical analyses were performed using R (version 4.2.1) and R studio (version 1.3.1093) software. Some histograms, which displayed different distributions of the clinicopathological features between

different groups, were drawn using Prism GraphPad (version 8.4.0).

Results

Construction of functional gene signature

First and foremost, we embarked upon our analysis by employing univariate Cox regression analysis to screen and identify genes that exhibited a significant association with the overall survival (OS) of the TCGA CRC cohort. Subsequently, a total of 1019 genes, characterized by a hazard ratio (HR) of less than 1, were deemed as protective genes, thereby signifying that patients with a higher expression of these genes would be associated with a more favorable OS status. On the other hand, a cohort of 816 genes were categorized as harmful genes. In the next phase of our investigation, we identified 5498 genes that were differentially expressed between CRC tumor samples and their adjacent normal tissue counterparts, utilizing counts data analyzed with the *DESeq2* package. The antitumor genes were defined as the intersection of protective genes, differentially expressed genes (DEGs), and genes within the gene set signatures (MGSs). Conversely, the protumor genes were delineated as the intersection of harmful genes, DEGs, and genes within the MGSs. Ultimately, the Functional Gene Signature (FGS) was composed of 24 antitumor genes and 22 protumor genes (Fig. 2A).

To further evaluate the intricate relationship between the antitumor and protumor genes, we conducted a comprehensive assessment of their expressions. These analyses revealed a positive correlation among the expressions of most antitumor genes, as well as a positive correlation among the expressions of most protumor genes (Supplementary Fig. 1A and B). In contrast, a negative correlation was observed between the expressions of antitumor and protumor genes (Supplementary Fig. 1C).

To assign each sample with two distinct scores, namely the antitumor score and the protumor score, we employed the *GSVA* package. Subsequently, we applied the consensus clustering method to classify the CRC patients into discrete groups based on their GSEA scores. Given the composition of the FGS, the optimal number of clusters consistently remained at 2.

In our pursuit of unraveling the association between the antitumor and protumor scores, we conducted a comprehensive Spearman analysis. The results demonstrated a negative correlation between the antitumor and protumor scores ($R = -0.17$, $p = 0.00025$) (Supplementary Fig. 1D). Consequently, we designated the group exhibiting higher antitumor scores as the low-risk group, while the remaining group was deemed the high-risk group. Subsequent assessments revealed that the low-risk group exhibited significantly higher antitumor scores and lower

expressed, while the protumor genes showed lower expression levels in the antitumor group (Fig. 2D). Finally, to ascertain the efficacy of the expression of the genes within the FGS in effectively distinguishing between the two groups, we employed dimension reduction techniques, including principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE). The results revealed a separation of samples along one dimension in both the PCA and t-SNE analyses (Fig. 2C). We also checked whether patients in two groups had distinct clinicopathological features. The results showed that patients in high-risk group has poorer OS and pathological characteristics, including tumor invasion depth, lymph node metastasis, and distant metastasis (Fig. 2E).

To fortify the robustness of our evidence, we subjected additional 5 independent CRC cohorts, namely GSE39582, GSE17538, GSE38832, GSE28722, and GSE103479, to comprehensive validation analyses. Beginning with GSE39582, our findings demonstrated that the low-risk group exhibited a lower number of deceased and relapsed patients (Supplementary Fig. 2A). Moreover, both the overall survival (OS) and the relapse-free survival (RFS) of the low-risk group were superior to those of the high-risk group (OS: $p=0.002$; RFS: $p<0.001$). Importantly, akin to the TCGA CRC cohort, the low-risk group in GSE39582 showed significantly better pathological features as well (Supplementary Fig. 2B). Similarly, in GSE17538, patients belonging to the low-risk group experienced improved overall survival (OS), and disease-free survival (DFS), and exhibited more favorable pathological stages (Supplementary Fig. 2C and E). The FGS was also effective in accurately classifying the DFS and pathological stage in GSE38832 (Supplementary Fig. 2D and Fig. 2F). In GSE28722, although the overall survival (OS) did not exhibit a statistically significant difference between the low-risk and high-risk groups, the alive patients in the low-risk group outnumbered those in the high-risk group. Besides, the low-risk group displayed significantly better disease-free survival (DFS) and more favorable pathological stages, including the TNM stage and Duke's stage (Supplementary Fig. 2G, I, and J). Furthermore, in GSE103479, where only disease-free survival (DFS) information was available, our analysis revealed a significantly superior DFS in the low-risk group compared to the high-risk group (Supplementary Fig. 2H).

Molecular features and immune microenvironment

In this study, we sought to evaluate the disparities in the immune microenvironment between the two groups by utilizing *CIBERSORT* (Fig. 3A). The analysis unveiled that the low-risk group displayed higher infiltration of plasma cells, CD8⁺ T cells, CD4⁺ T cells

(memory resting and memory activated), activated dendritic cells, eosinophils, and neutrophils. In contrast, the high-risk group exhibited a higher infiltration of M0 macrophages. These findings suggest that patients in the low-risk group tend to possess a "pro-inflammatory" tumor immune microenvironment (TIME), while the other group tends to have an "anti-inflammatory" TIME. Moreover, these results imply that patients in the low-risk group may exhibit a more favorable response to immunotherapy.

Furthermore, we employed the *progeny* package to assess cancer-related signaling pathways. Our analysis revealed that 5 pathways, namely androgen, NFκB, PI3K, TNFα, and trail, were significantly upregulated in the high-risk group. Conversely, 3 pathways, including TGFβ, VEGF, and WNT, exhibited upregulation in the low-risk group (Fig. 3B). Basically, the hypoxic condition would activate epithelial-mesenchymal transition (EMT) and angiogenesis through VEGF and WNT pathways [55, 56]. Meanwhile, TGFβ signaling also promoted tumor metastasis and proliferation through EMT [57]. These findings were different from ours. However, solid conclusions needed both bioinformatic analyses and biology validation, which would be done through additional experiments.

Moreover, Fig. 3C elucidated that several immune checkpoint molecules, including CD276, CTLA4, TNFRSF4, TGFβ1, and TNFSF4, displayed higher expression in the high-risk group. This observation further confirms the suppression of immune cell survival, infiltration, and migration, signifying a "colder" TME in the high-risk group. On the other hand, the secretion of ICOS occurs when T cells are activated during infections or cancer invasion. Increased infiltration of ICOS⁺ effector T cells indicates a favorable response to immunotherapy [56]. Additionally, CD274, also known as PD-L1, exhibited higher expression in the low-risk group, suggesting that the TME in this group may have a greater infiltration of tumor-infiltrating lymphocytes (TILs) and increased expression of interferon-γ [58, 59]. Interestingly, we discovered that most chemokines, including those belonging to the CXCL and CCL families, exhibited higher expression in the low-risk group (Supplementary Fig. 3A and B). These chemokines play a crucial role in shaping and regulating the immune system, and their heightened expression can induce the activation of both innate and adaptive immune responses. Furthermore, such elevated expression is associated with improved prognosis in various types of cancer, including colorectal, endometrial, and breast tumors [60, 61]. Additionally, a majority of interleukins (IL) were also found to be highly expressed in the low-risk group, which is characterized by smaller tumor

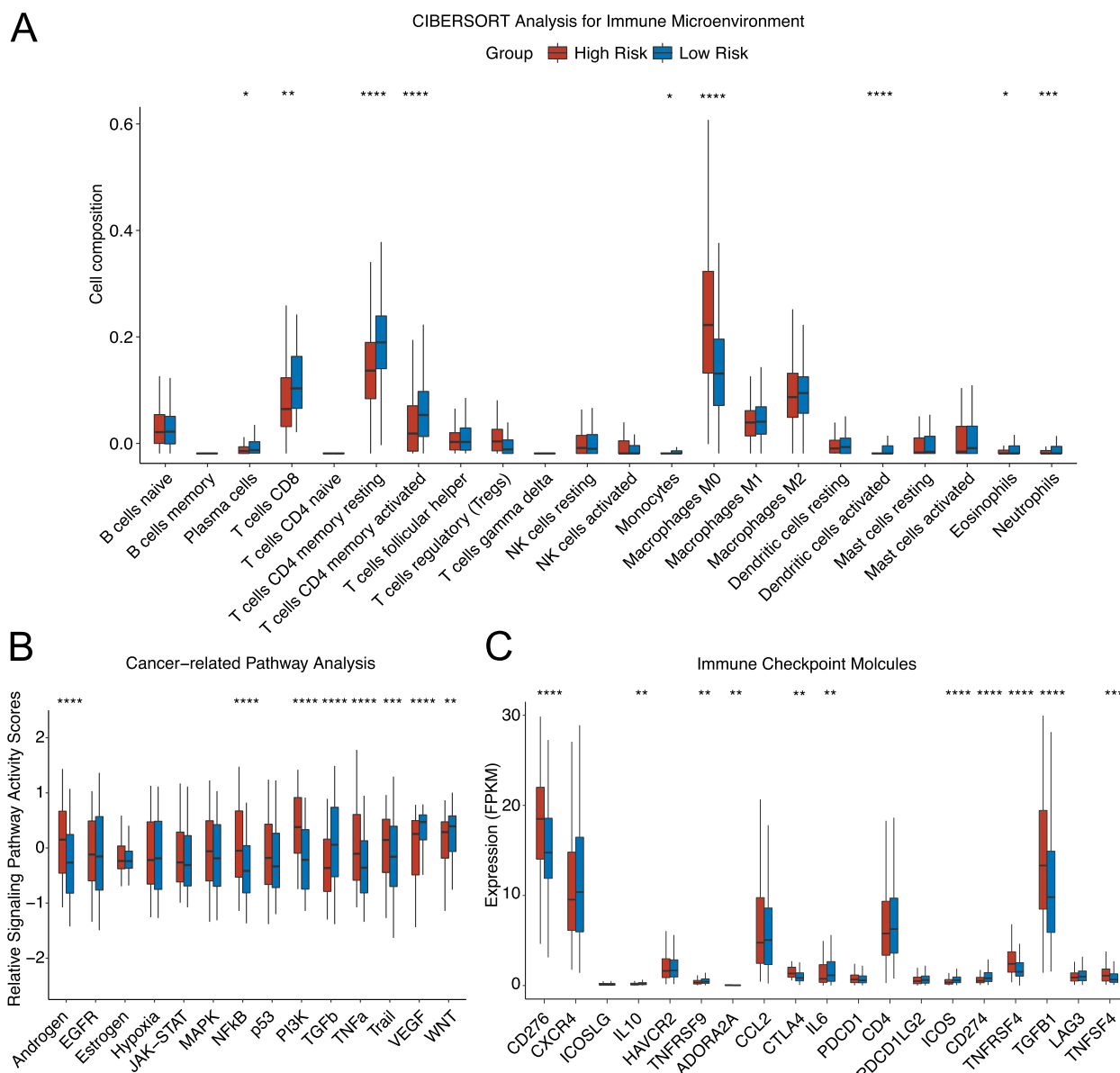


Fig. 3 The analyses of tumor immune microenvironment and molecular expression. **A** The results of CIBERSORT showed that plasma cells, CD8+T cells, CD4+memory resting T cells, CD4+memory activated T cells, monocytes, activated dendritic cells, eosinophils, and neutrophils were enriched in TIME of the low-risk group. While the TIME of the high-risk group had more infiltration of M0 macrophages. **B** The analysis of cancer-related pathways. It indicated that the Androgen, NFkB, PI3K, TNFa, and Trail pathways were upregulated in the high-risk group. While the TGFb, VEGF, and WNT pathways were upregulated in the low-risk group. **C** Most immune checkpoint molecules were enriched in the high-risk group, except IL10, TNFRSF9, IL6, ICOS, and CD274. * represents $p < 0.05$, ** represents $p < 0.01$, *** represents $p < 0.001$, **** represents $p < 0.0001$

size, fewer infiltrating anti-inflammatory cells, and a normoxic TME (Supplementary Fig. 3C) [62].

Somatic mutation and prediction of immunotherapy

Upon thorough examination, we have successfully confirmed the disparities in clinicopathological and molecular characteristics between the two groups. We proceeded with an assessment of the distinct patterns of

somatic mutations, which were visualized in the waterfall plot (Fig. 4A). It became apparent that the low-risk group exhibited a higher frequency of most types of mutations. Among the top 30 frequently mutated genes, such as *APC*, *TP53*, *TTN*, *KRAS*, and *FAT4*, the ranking remained consistent between the two groups. Notably, most of the shared genes showed a higher mutation frequency in the low-risk group, emphasizing the prevalence of these

mutations in this group. Furthermore, the forest plot demonstrated that genes with a p -value less than 0.01 displayed a higher mutation frequency in the low-risk group (Fig. 4B).

Moving forward, we delved into the assessment of tumor mutation burden (TMB). Our findings revealed a significantly higher TMB in the low-risk group compared to the high-risk group (Fig. 4C). By stratifying CRC patients based on different TMB levels, we observed a greater number of patients with high TMB in the low-risk group (Fig. 4H). Notably, we identified a significant positive correlation between antitumor scores and TMB (Fig. 4D). These results suggest that our FGS approach not only enables the classification of CRC patients based on clinicopathological and molecular features but also facilitates the distinction of somatic mutation patterns among these patients. Although the differences in TMB levels did not reach statistical significance in terms of overall survival (OS), the Kaplan–Meier curves delineated the separation between the two groups (Fig. 4E). This observation aligns with the previously observed higher frequency of somatic mutations and elevated TMB in the low-risk group. Moreover, when combining TMB information with risk grouping, we observed the following ranking in terms of OS for CRC patients: high TMB plus low risk, low TMB plus low risk, high TMB plus high risk, and low TMB plus high risk (Fig. 4F). The differences among these four groups were statistically significant ($p < 0.001$), underscoring the enhanced predictive capacity of FGS in OS for CRC patients when incorporating TMB information.

Furthermore, we employed the TIDE algorithm to predict the response to immunotherapy. According to TIDE, higher TIDE, dysfunction, and exclusion scores correspond to a worse response to immunotherapy [50]. Our results uncovered that the low-risk group exhibited less infiltration of anti-inflammatory components, including cancer-associated fibroblasts (CAF) and type 2

tumor-associated macrophages (TAM.M2). Additionally, the low-risk group displayed higher levels of interferon γ (IFNG) and CD274 (Fig. 4G). Moreover, it comprised a greater number of patients with microsatellite instability (MSI) and higher levels of the T cell-inflamed signature (merck18). In the TCGA CRC cohort, 178 out of 238 patients in the low-risk group, while 75 out of 234 in the high-risk group, responded favorably to immunotherapy (Fig. 4I). To substantiate this evidence, we searched the GEO database for CRC cohorts in which patients received immunotherapy. Among the 50 rectal cancer patients from the GSE60331 dataset who received bevacizumab and chemoradiation therapy, we stratified them into low- and high-risk groups using FGS. Notably, 83.3% of patients (10 out of 12) in the low-risk group responded to immunotherapy, while only 34.21% of patients (13 out of 38) in the high-risk group showed a positive response (Fig. 4I).

In summary, our FGS approach effectively stratifies CRC patients into distinct groups based on clinicopathological, molecular, and somatic mutation features. Furthermore, it demonstrates the ability to predict the response to immunotherapy, a validation that has been solidified through the analysis of both TCGA and GEO databases.

Differentially expressed genes and related bio-functional analysis

Next, we examined the differential gene expression and related pathways between the two groups. Utilizing the *edgeR* package, we calculated the differentially expressed genes (DEGs) based on the counts data. Our comprehensive analysis unveiled 958 genes that exhibited increased expression in the high-risk group, while 273 genes displayed decreased expression (Fig. 5A). We added the list of the significantly changed DEGs to Supplementary Table 7.

(See figure on next page.)

Fig. 4 The analyses of somatic mutation and the prediction of immunotherapy. **A** The waterfall plots showed that the low-risk group had larger quantity and more types of somatic mutation. Additionally, many genes in the top 30 mutation list of the low-risk group had lower mutation rate in the high-risk group. **B** In the analyses of the differences of the quantity of the mutated genes ($p < 0.01$) between two groups, we found that most genes were more frequently mutated in the low-risk group. **C** The low-risk group had significantly higher TMB. **D** The spearman analysis revealed that the antitumor score was positively related to the TMB. **E** Although the K-M curves between low-TMB and high-TMB groups were well separated, but the difference had no statistical significance. **F** The high-TMB plus low-risk group had best OS, followed by the low-TMB plus low-risk, high-TMB plus high-risk, and low-TMB plus high-risk group. **G** The high-risk group had more infiltration of CAF and M2 type TAM, higher TIDE scores, dysfunction scores, and exclusion scores, and fewer expression of interferon γ , MSI, merck18, and CD274. It meant that the high-risk group had relatively “cold” TIME and worse response to the immunotherapy. **H** The stacking histogram showed different distributions of TMB between two groups. The low-risk group had larger ratio of CRC patients with high levels of TMB. **I** According to the results of TIDE for TCGA CRC cohort, the low-risk group had more patients that were responded to the immunotherapy. Meanwhile, in GSE60331 (CRC immunotherapy cohort), the low-risk group had significantly larger quantity of patients who were responded to the immunotherapy. * represents $p < 0.05$, ** represents $p < 0.01$, *** represents $p < 0.001$, **** represents $p < 0.0001$

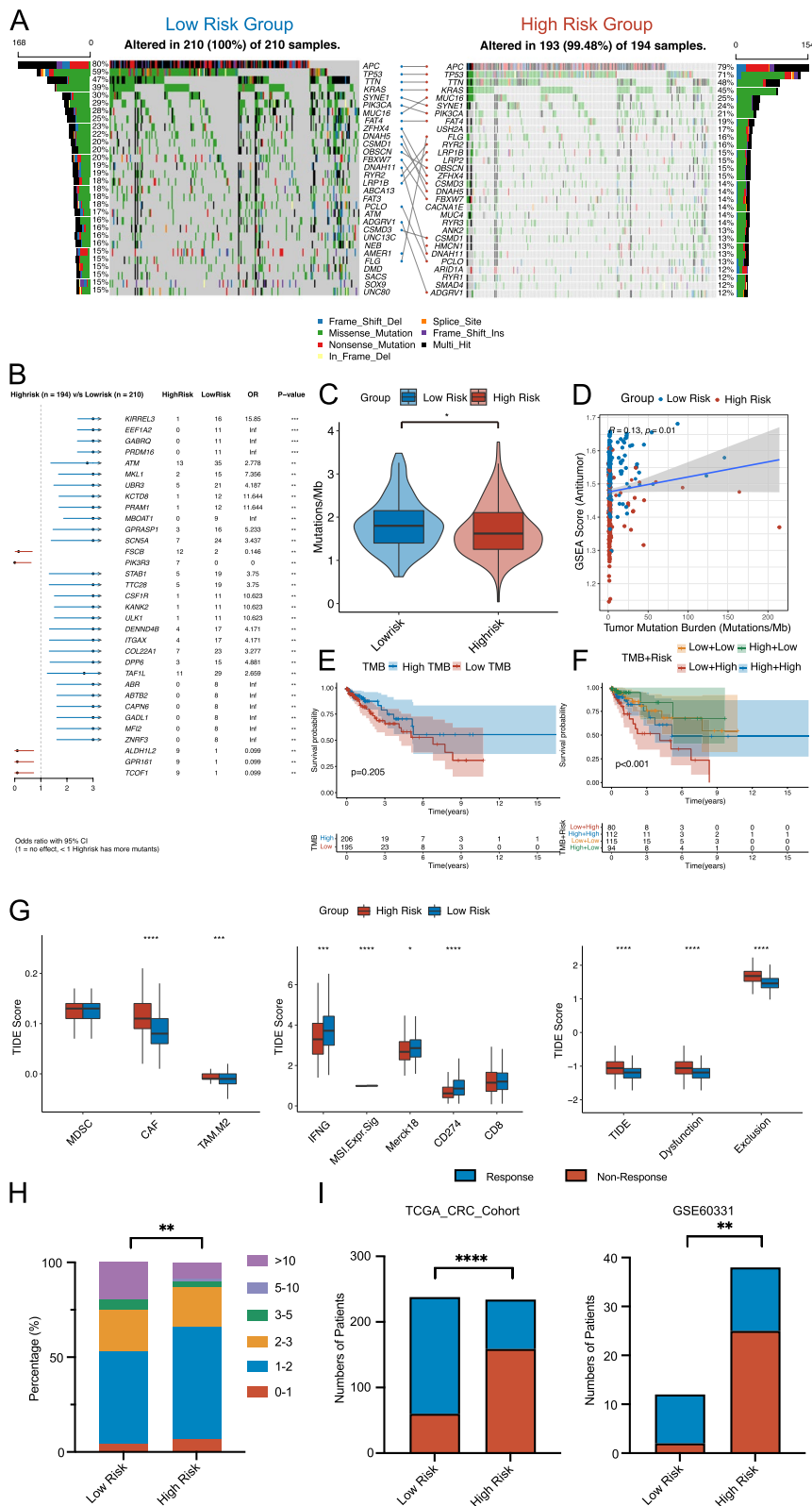


Fig. 4 (See legend on previous page.)

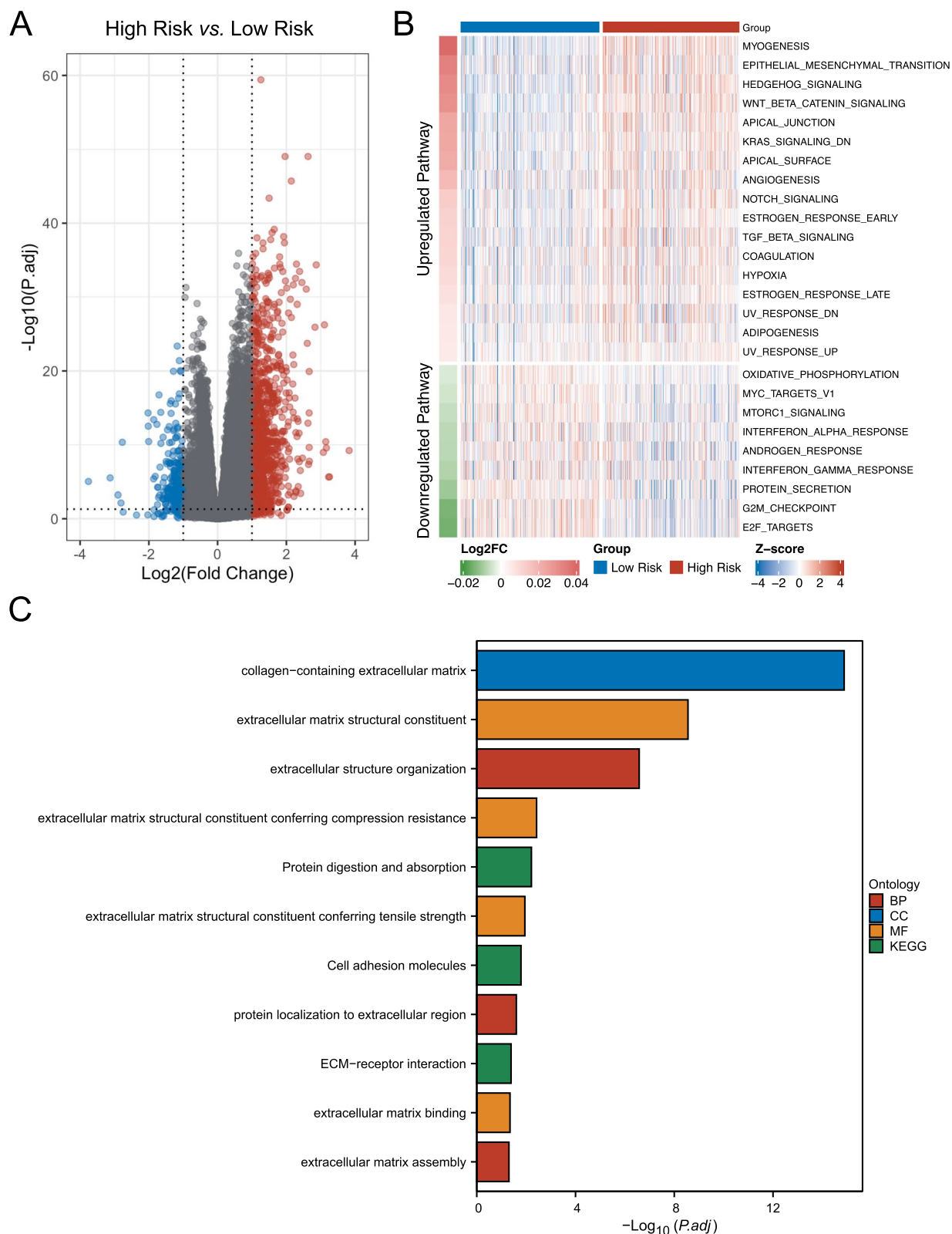


Fig. 5 The differentially expressed genes between two groups and related bio-functional analyses. **A** In TCGA CRC cohort, compared to the low-risk group, the high-risk group had 958 upregulated genes and 273 downregulated genes. **B** Based on the molecular signatures database, 17 KEGG signaling pathways were upregulated, and 9 KEGG signaling pathways were downregulated in the high-risk group. **C** The biofunction enrichment analyses of upregulated DEGs

To further delve into the mechanism of this clustering method, we conducted an analysis of the hallmark-related pathways using the Molecular Signature Database (MSigDB). We discovered 26 pathways that exhibited significant changes between the two groups (adjust. $p < 0.05$) (Fig. 5B). Notably, several cancer-related pathways, including EMT, Hedgehog, WNT-beta catenin, KRAS, angiogenesis, Notch, TGF β , and hypoxia-related pathways, were upregulated in the high-risk group. This observation partially elucidates why patients in the high-risk group demonstrated poorer clinicopathological features and overall survival.

Moreover, we observed that the mTORC1 signaling pathway was impeded due to DNA damage and the deprivation of essential nutrients and oxygen [63]. Additionally, the oxidative phosphorylation signaling pathway was impaired in the high-risk group, most likely due to the hypoxic tumor microenvironment (TME). These impairments in signaling pathways suggest that the TME in the high-risk group is hypoxic and lacks crucial nutrients, such as amino acids and glucose. Consequently, this hampers the migration, survival, and tumoricidal function of immune cells. In contrast, interferon α and interferon γ responses were upregulated in the low-risk group, indicating a TME with potentially greater tumoricidal activities.

Subsequently, we utilized the *clusterProfiler* package to analyze the biological functions based on the DEGs. Regrettably, due to the limited number of downregulated DEGs, we were unable to perform the cellular component (CC) analysis. Furthermore, the results of the KEGG, molecular function (MF), and biological processes (BP) analyses did not highlight any specific functions (Supplementary Fig. 4E-G). However, for the upregulated DEGs, our analyses revealed that extracellular matrix remodeling related biofunctions, including collagen-containing extracellular matrix, extracellular matrix organization, and extracellular structure organization, were all upregulated in the high-risk group (Fig. 5C and Supplementary Fig. 4A-D). This observation prompts us to consider that, in addition to cancer-related pathways, cancer-associated fibroblasts (CAFs) might also contribute to the worse clinicopathological

features, survival outcomes, and response to immunotherapy observed in the high-risk group.

Moving forward, our attention will be directed toward exploring the hub genes and their association with CAFs, providing further insight into the intricate mechanisms at play.

The procedure of searching for hub genes

Initially, we undertook an analysis of the 582 colorectal cancer (CRC) patients in the GSE39582 dataset. Employing the FGS method, we divided the patients into low- and high-risk groups. Subsequently, we utilized the *limma* package to calculate the differentially expressed genes (DEGs) between the two groups. Our analysis identified 74 upregulated genes and 12 downregulated genes, which are visually represented in Fig. 6A.

Furthermore, we conducted a comparative analysis between the upregulated DEGs in the TCGA dataset and the GSE39582 dataset. This analysis revealed an intersection of 53 genes (Fig. 6B). Additionally, we identified 3 genes that intersected between the downregulated DEGs in both datasets. These genes were then ranked based on their betweenness centrality, as illustrated in Fig. 6C.

To gain further insights, we assessed the differential expression of these genes between tumor and adjacent normal tissues, as well as their association with patient survival outcomes, including overall survival (OS) and disease-specific survival (DSS), in the TCGA CRC cohort. Intriguingly, we observed that five genes (*ACTA2*, *TAGLN*, *TPM2*, *MYL9*, and *CNN1*) exhibited consistent expression patterns across adjacent normal tissues, early-stage tumors (stage I), and advanced-stage tumors (stage II-IV) (Supplementary Fig. 5). Notably, the expression levels were highest in adjacent normal tissues, followed by early-stage and advanced-stage tumor tissues. In terms of survival outcomes, the higher expression group for all five genes demonstrated significantly worse OS and DSS compared to the lower expression group, except for *ACTA2* and OS ($p = 0.15$). The receiver operating characteristic (ROC) curves revealed that these genes held valuable predictive abilities, with all areas under the curves (AUCs) above 0.7 (Fig. 6E). Furthermore, we observed a strong positive correlation in the expression levels of these genes within the TCGA CRC cohort (Fig. 6D and

(See figure on next page.)

Fig. 6 The procedures of searching for hub genes. **A** There were 74 genes upregulated in the high-risk group and 12 genes upregulated in the low-risk group. **B** Fifty-three genes were the intersection of upregulated DEGs in TCGA CRC cohort and upregulated DEGs in GSE39582. Three genes were the intersection of downregulated DEGs in TCGA CRC cohort and downregulated DEGs in GSE39582. **C** Using String database and Cytoscape software, top 10 gene were screened out through these 56 genes according to the betweenness centrality. **D** The expressions of these genes were strongly positively related. **E** The area under curve (AUCs) of 5 hub genes were all above 0.7. It meant that these genes were valuable to predict the OS of CRC patients. **F** The spearman correlation coefficients of these hub genes were all above 0.85

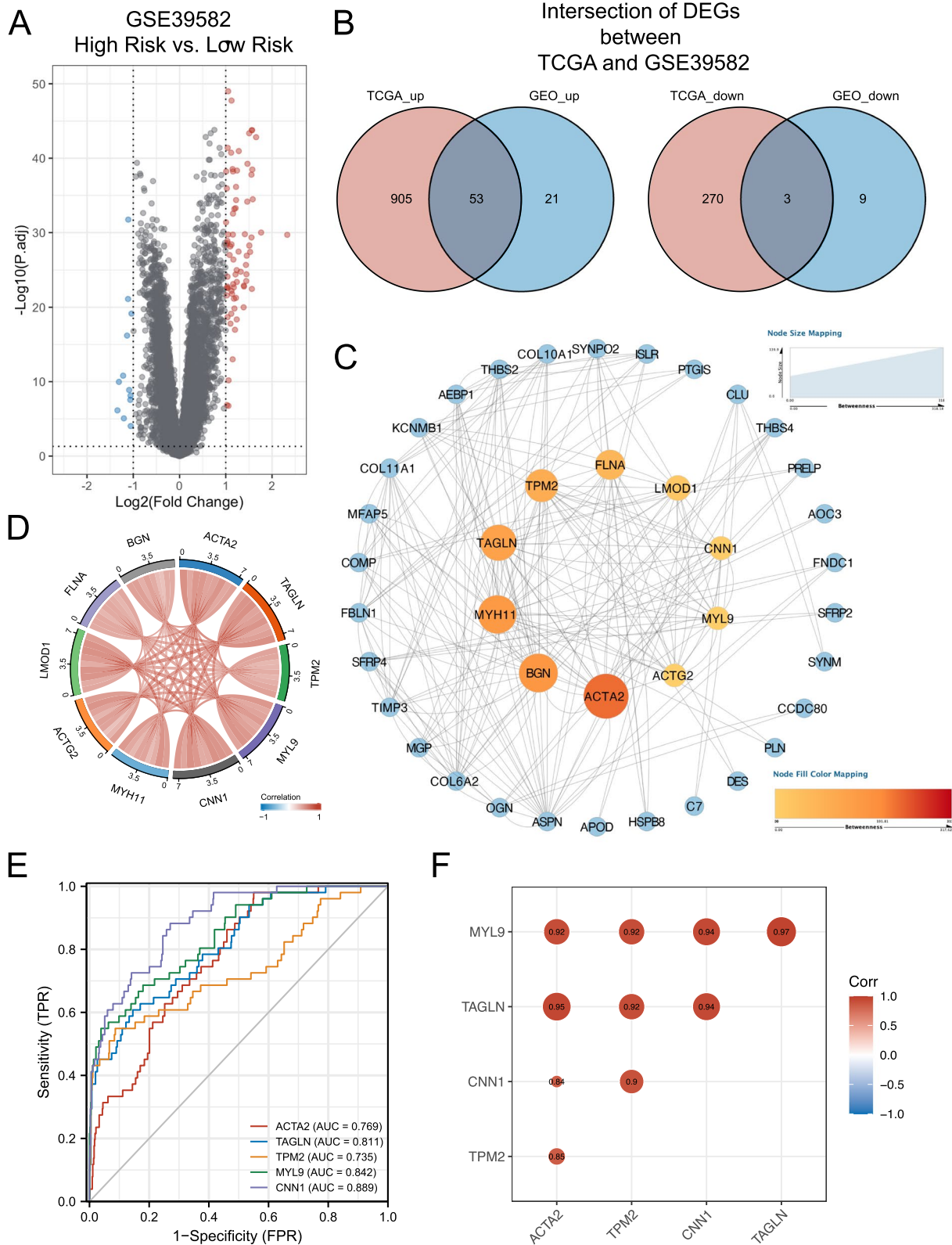


Fig. 6 (See legend on previous page.)

F). Consequently, these five genes were selected as the hub genes for further analyses.

Conversely, the remaining five genes (*ACTG2*, *MYH11*, *LMOD1*, *FLNA*, and *BGN*) were excluded from further analyses for the following reasons: 1) The expression levels in early-stage tissues did not differ significantly from those in advanced-stage tissues (*ACTG2*, *MYH11*, *LMOD1*, and *FLNA*); 2) Some genes did not possess significant survival prediction abilities (*ACTG2*, *MYH11*, *LMOD1*, and *BGN*); 3) *BGN* emerged as a typical oncogenic gene, with its expression increasing in tandem with the deterioration of the pathological tumor stage (Supplementary Fig. 6).

Association between hub genes and CAF

We further explored the potential association between the FGS and cancer-associated fibroblasts (CAFs) using multiple algorithms, including *xCell*, *ESTIMATE*, *MCP-counter*, and *GSEA*. In Fig. 7A, we observed that the protumor scores exhibited a higher positive correlation with CAF-related scores compared to antitumor scores. This finding suggests a potential relationship between the high-risk group and increased CAF infiltration.

To validate this relationship, we utilized *GSEA* analysis and confirmed that CAFs were indeed significantly enriched in the high-risk group (Fig. 7B). Furthermore, based on the results from *ESTIMATE*, the high-risk group displayed higher stromal scores and lower immune scores, indicating a relative 'colder' tumor immune microenvironment (TIME) that could potentially impair the tumoricidal function of immune cells (Fig. 7C). Similarly, in *MCP-counter* and *xCell* analyses, we observed higher fibroblast and stromal scores, as well as lower immune scores, in the high-risk group (Fig. 7E and F). The consistency of these results across all algorithms greatly enhances their reliability.

Subsequently, we investigated the relationship between the hub genes and CAFs. The findings revealed that all five hub genes exhibited a positive association with protumor scores and a negative association with anti-tumor scores (Fig. 7D). Moreover, their expressions were significantly positively correlated with CAFs. These findings further support the notion that these hub genes may play a role in the interaction between the tumor and the CAF-rich microenvironment.

We have successfully confirmed a highly positive association between the protumor scores and the hub genes with the infiltration of cancer-associated fibroblasts (CAF) through multiple algorithms. To further investigate the precise expression location of the hub genes and identify additional genes for subsequent biology experiments, we utilized an independent single-cell dataset (GSE132465).

In GSE132465, we selected 39,197 cells from 10 paired normal and tumor samples and divided them into the normal and tumor groups (Fig. 7G). Through careful annotation using the provided annotations by the authors, we identified six cell types, including T cells, B cells, mast cells, stromal cells, myeloid, and epithelial cells (Fig. 7H). Notably, the epithelial cells in the tumor samples exhibited a highly active proliferative state, which may induce antitumor activity of the self-immune system, consistent with previous research [64].

We further analyzed the expression of the hub genes in these datasets and found that they were highly expressed in stromal cells (Fig. 7I). Specifically, *CNN1* was expressed in approximately 10% of stromal cells, while *TPM2* and *MYL9* were expressed in around 60% of stromal cells and 20% of epithelial cells.

Based on these findings and several considerations, we have decided to screen out two genes, *MYL9* and *TPM2*, for molecular and cellular experiments. Firstly, these genes are expressed in both epithelial and stromal cells, allowing us to investigate their effects on the phenotype using CRC cell lines. Additionally, we can explore the impact of these genes on the communication between epithelial and stromal cells through knockdown or over-expression experiments. Lastly, *MYL9* and *TPM2* exhibit specific associations with tumor CAFs while maintaining comparable expressions between tumor and normal samples in other cell types.

Identification of the expression of hub genes

First, we used mIHC to identify the expression of two hub genes described previously, *MYL9* and *TPM2*, and the association between them and tumor cells at the tissue level. Because α -SMA was a classical biomarker for CAF in types of malignancies [65], we used it to locate the CAF in both adjacent normal and tumor tissues. In line with our expectation, *TPM2* and *MYL9* were expressed mostly in tumor tissues, conversely, they were barely expressed in the adjacent normal tissues (Fig. 8A, B, and Supplementary Fig. 7). Meanwhile, a high similarity of the localization was found between cellular components marked by *TPM2* and α -SMA. It meant that *TPM2* might be used as a potential CAF marker. Pan-CK was often used to localize epithelial cells. The results of mIHC showed that cellular components marked by α -SMA were different from those marked by Pan-CK, which also proved that α -SMA was a CAF marker.

To further evaluate the expression of α -SMA, *TPM2*, and *MYL9* at the mRNA and protein levels, we employed RT-qPCR and Western blotting techniques. Initially, we assessed the transcriptional levels of these genes in tumor cell lines (SW480 and RKO) and primary CAFs. Subsequently, we calculated the mRNA expression ratios of

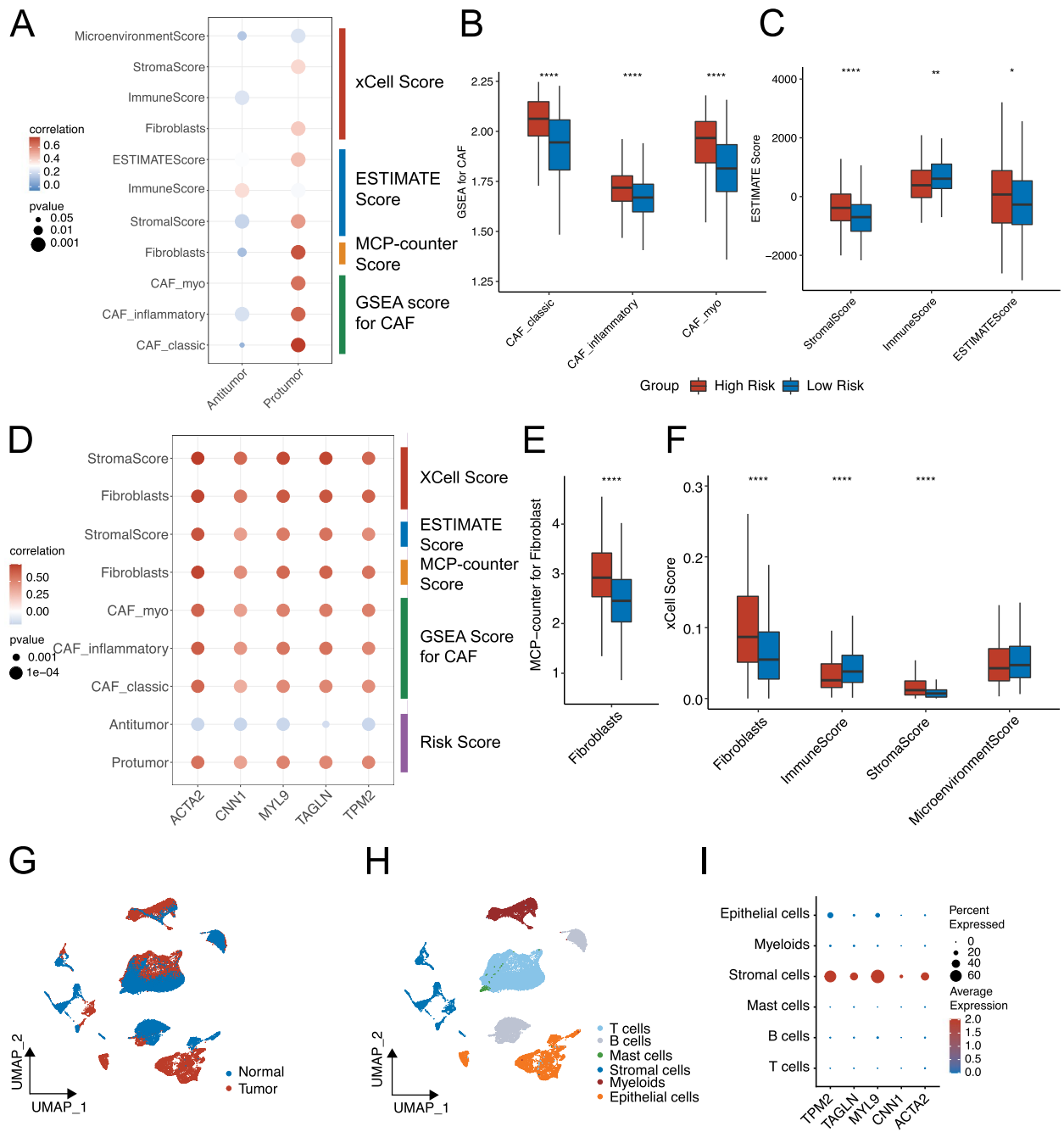


Fig. 7 The association between hub genes and CAF, and the analyses of the expression location of the hub genes. **A** The correlation between protumor scores and CAF was stronger than that between antitumor scores and CAF. **B** The GSEA analysis showed that the high-risk group had more infiltration of classic, inflammatory, and myo CAF. **C** The results of ESTIMATE indicated that the low-risk group had smaller stromal scores and larger immune scores, suggesting its TME was beneficial for immune cells to enhance the tumoricidal function. **D** The dotplot showed that the hub genes were positively related to the protumor and negatively related to the antitumor scores. Meanwhile, these genes were also positively associated with the CAF. **E** The MCP-counter analysis showed that the high-risk group had more infiltration of fibroblasts. **F** The xCell analysis revealed that the high-risk group had more infiltration of fibroblasts, larger stromal scores, and smaller immune scores, which was consistent with the previous results. **G** In single cell cohort (GSE132465), cells in normal and tumor samples could be separated using uMAP analysis. **H** Six types of cells were annotated using the annotations provided by the authors. **I** The hub genes were mostly expressed by the stromal cells. * represents $p < 0.05$, ** represents $p < 0.01$, *** represents $p < 0.001$, **** represents $p < 0.0001$

TPM2, two classical markers (*FAP* and α -*SMA*), between CAFs and SW480, as well as CAFs and RKO. The results revealed that the mRNA expression levels of *TPM2* and the two classical markers were approximately one hundred times higher in CAFs compared to tumor cells, including SW480 (Fig. 8C) and RKO (Fig. 8D). However, there was no statistically significant difference in the mRNA expression of *MYL9* between CAFs and tumor cells.

Interestingly, the Western blot results exhibited slight discrepancies compared to the RT-qPCR results (Fig. 8E). Specifically, α -*SMA* expression was only observed in primary CAFs, confirming its specificity as a CAF marker rather than being present in tumor cells or normal colon epithelial cell lines (FHC). *TPM2* displayed strong expression in CAFs, as well as in SW480. Similarly, *MYL9* exhibited comparable expression in CAFs, SW480, and FHC.

Although *TPM2* demonstrated potential as a CAF marker, in this study, we showed that it is also expressed in tumor cell lines based on mRNA and protein levels. This phenomenon was not found in a CAF-specific marker like α -*SMA*.

TPM2 function in the interaction between cafs and tumor cells

We silenced *TPM2* in CAFs using shRNA of 3 different clones focusing on different sites and another empty vector clone. Western blot and RT-qPCR were used to verify the knocking-down efficacy. The results of Western blot and RT-qPCR assays showed that all three shRNA clones could efficiently knock down the *TPM2* expression at the protein and mRNA levels without influencing the expression of the classic CAF marker, α -*SMA* (Fig. 9A and B). Among the clones, clone 3 displayed the superiority of the knocking-down ability. As a result, this clone was named shCAF3 and used for further functional experiments.

We chose the co-culture system described in the previous study (Fig. 9C) [66] to validate the biofunction of *TPM2*. Initially, we found that the co-culture with CAFs could greatly enhance the tumorigenic ability of colorectal cancer cells. However, knocking down *TPM2* in CAFs could partially reduce these enhanced malignant

phenotypes of tumor cells brought by CAFs. These results were proven by the following assays. First, knocking down *TPM2* in CAFs could reduce the proliferation of colorectal cancer cells indirectly, which was validated using Ki67 staining (Fig. 9D) and CCK-8 (Fig. 9E). Next, the invasion (Fig. 10A), capacity of forming colonies (Fig. 10B), and migration (Fig. 10C) were also validated by transwell, clonogenic assays, and wound healing, respectively.

In conclusion, CAFs could significantly increase colorectal cancer cells' tumorigenic biofunctions through co-culture, which could be partially inhibited by knocking down CAFs' *TPM2* expression in vitro.

Discussion

In this study, ssGSEA scores were employed to partition CRC patients into distinct high- and low-risk groups utilizing FGS. Notably, patients classified as high risk exhibited a more unfavorable survival status, accompanied by progression in the TMN stage. These findings were further substantiated across five independent datasets, lending credence to their robustness. Furthermore, an intriguing disparity in the TIME (Tumor Immune Micro-environment) between the high-risk and low-risk groups was observed. The high-risk group exhibited diminished infiltration of tumor-infiltrating lymphocytes (TIL) and elevated expression levels of immune checkpoint molecules, setting it apart from the other group. This discrepancy in TIME implies that the low-risk group possesses a more vibrant and dynamic TME, ultimately culminating in a more favorable prognosis. In addition, a comprehensive analysis of TCGA and GEO databases enabled the identification of differentially expressed genes (DEGs) between the two groups. Notably, five genes, namely *ACTA2*, *TAGLN*, *TPM2*, *MYL9*, and *CNN1*, displayed consistent expression patterns and exhibited a significant association with a poorer survival status. Furthermore, the correlation of these genes with cancer-associated fibroblasts (CAFs) was confirmed using single-cell datasets, adding another layer of evidence to their relevance in CRC. To validate these findings, the expression levels of two central hub genes, *TPM2* and *MYL9*, were assessed in CAFs through western blot and RT-qPCR techniques. The results were consistent with the hypothesis, further

(See figure on next page.)

Fig. 8 Biology experiments to verify the bioinformatic results. **A** Multicolor immunofluorescence tissue staining results of *TPM2*, α -*SMA*, pan-CK and DAPI on tissue chips of colorectal cancer tissues and adjacent normal tissues. **B** Multicolor immunofluorescence tissue staining results of *MYL9*, α -*SMA*, pan-CK and DAPI on tissue chips of colorectal cancer tissues. **C** RT-qPCR results of relative expression of *FAP*, *TPM2*, α -*SMA* and *MYL9* compared with *GAPDH* in SW480 cells. **D** RT-qPCR results of relative expression of *FAP*, *TPM2*, α -*SMA* and *MYL9* compared with *GAPDH* in SW620 cells. **E** Western blot analysis of α -*SMA*, *TPM2*, *MYL9* with *GAPDH* as the loading control (Full-length blots/gels are presented in Supplementary Fig. 9). Relative protein expression (measured by grayscale value) on the right

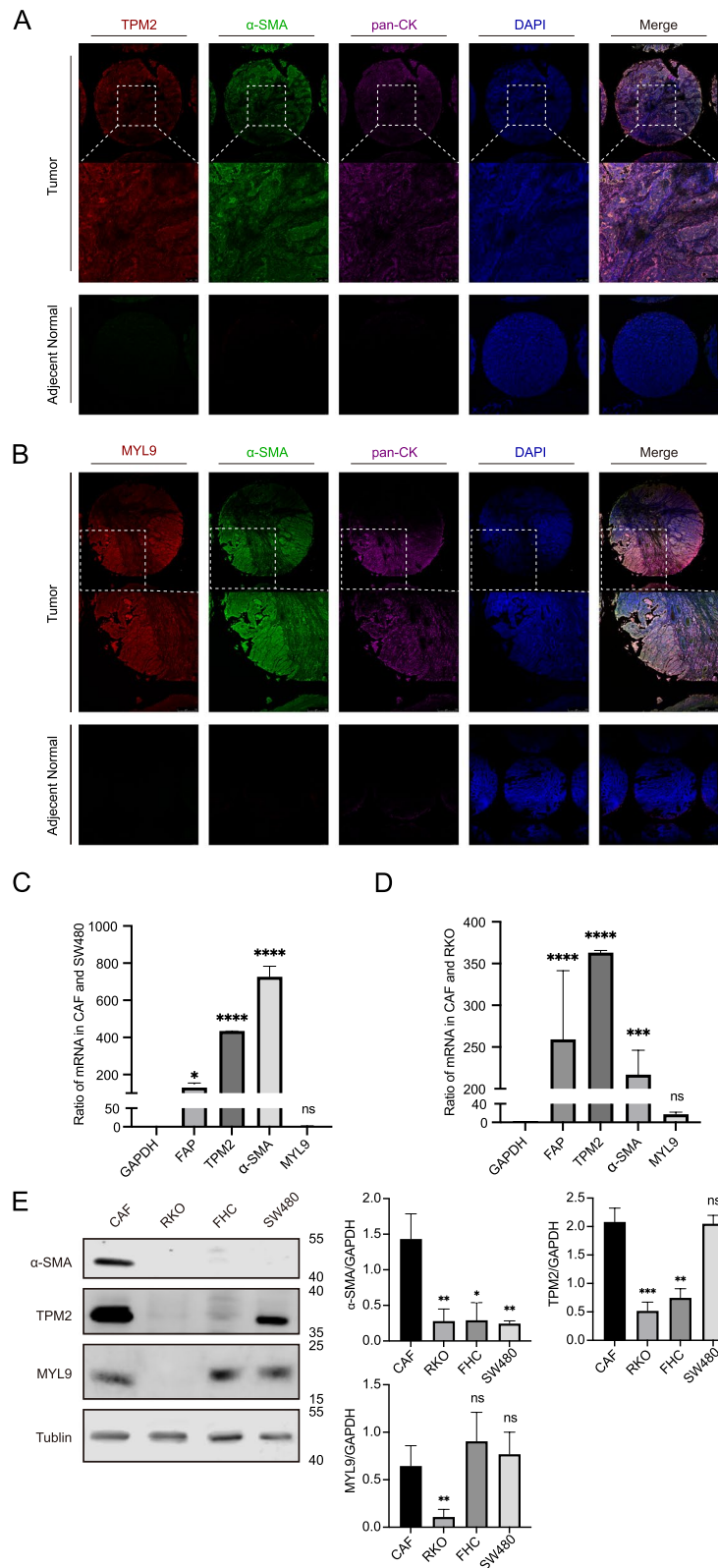


Fig. 8 (See legend on previous page.)

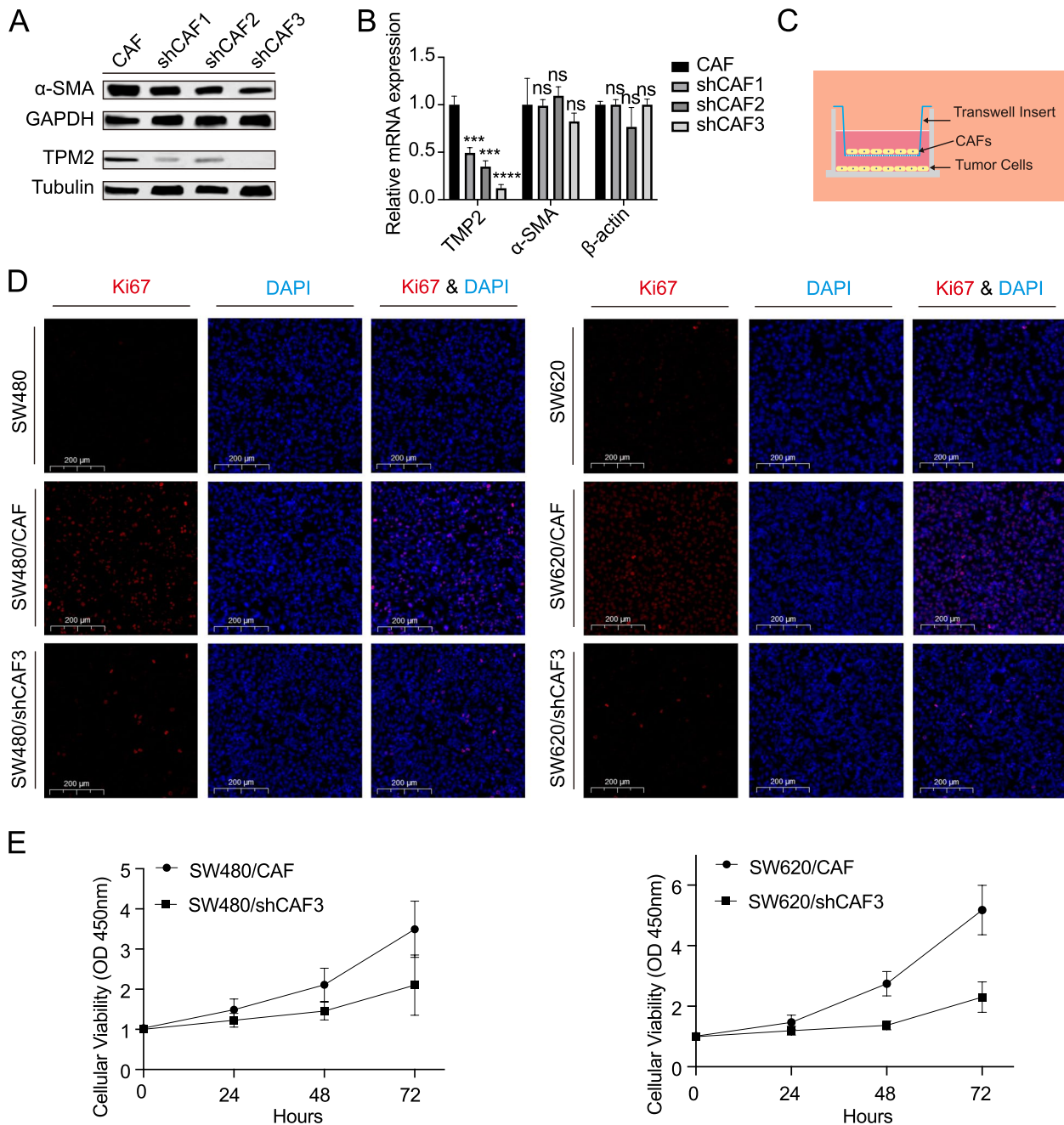


Fig. 9 Biology function of TPM2. TPM2 was knocked down using three different clones of shRNA against TPM2. Western blot (A) and RT-qPCR (B) results showed clone 3 had the highest silencing efficiency (Full-length blots/gels were presented in Supplementary Fig. 10). C Schematic of co-culture of CAFs and tumor cells. CAFs were seeded in the upper transwell inserts and Tumor cells were placed in the below well in a 6-well plate. The two cells were indirectly co-cultured for 48 h. TPM2 knockdown in CAFs influenced the proliferation of SW480 and SW620, which was validated by Ki67 staining (D) and CCK-8 (E). * represents $p < 0.05$, ** represents $p < 0.01$, *** represents $p < 0.001$, **** represents $p < 0.0001$

solidifying their role in the context of CAFs. This validation was also corroborated by mIHC analysis of tissue chips, providing additional support to the previously established findings. Overall, this study sheds light on the intricate relationship between ssGSEA scores, risk

stratification, TIME, DEGs, and CAFs in CRC. The comprehensive analysis of multiple datasets contributes to a deeper understanding of the molecular mechanisms underlying CRC progression and prognosis, potentially paving the way for more effective therapeutic strategies.

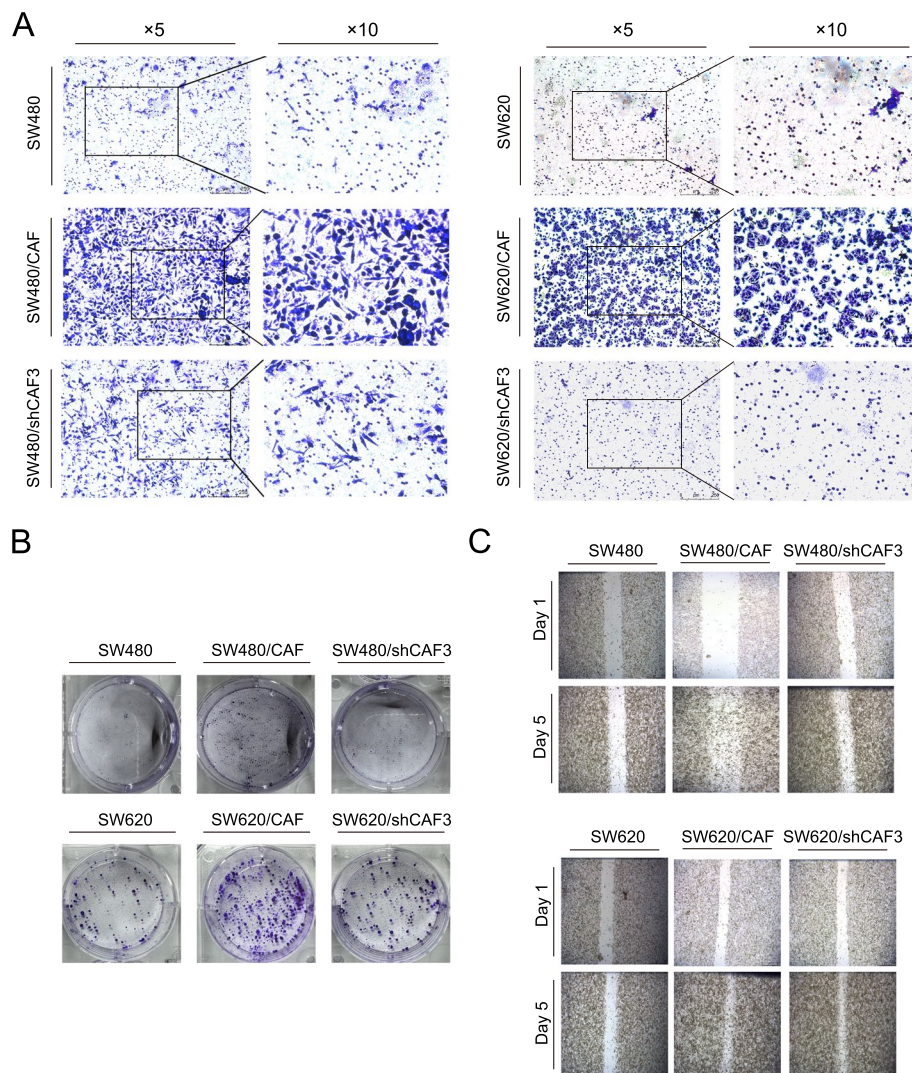


Fig. 10 **A** TPM2 knockdown in CAFs influenced the SW480 and SW620 tumor cell invasion ability. **B** TPM2 knockdown in CAFs influenced the tumorigenic capacity of SW480 and SW620. **C** TPM2 knockdown in CAFs influenced the SW480 and SW620 tumor cell migration ability

The focus of recent studies has shifted towards exploring the intricate components of the tumor microenvironment (TME) rather than solely concentrating on tumor cells themselves. The TME encompasses four major constituents, as elucidated by the study [67]: 1) immune cells, 2) vascular components such as blood and lymphatic endothelial cells, 3) the extracellular matrix (ECM), and 4) the stromal component comprised of cancer-associated fibroblasts (CAFs) and mesenchymal stem cells (MSCs). Among these components, CAFs play a pivotal role, exhibiting numerous functions that promote tumor growth and contribute to drug resistance [68, 69]. Additionally, CAFs can exert indirect effects by secreting ECM, which forms a barrier that hinders the infiltration of immune cells [70]. However, it is important to

note that CAFs are not a homogeneous cell population, as demonstrated by recent studies [71–73]. Instead, they consist of distinct subpopulations that can be identified using various markers. These subpopulations exhibit differences in their origin, phenotype, and functions across different types of cancer [71, 74]. This emerging understanding highlights the need for a more nuanced and comprehensive analysis of CAFs in order to fully comprehend their impact on tumor progression and therapeutic responses.

In our study, we have selected four hub genes, namely *ACTA2*, *MYL9*, *TAGLN*, and *TPM2*, which were expressed in fibroblasts of tumor samples. Consequently, they were deemed as CAF-related genes. Of particular interest is *ACTA2*, a protein-coding gene that encodes

ACTA2, an exquisite form of actin known as alpha-actin, alpha-actin-2, aortic smooth muscle, and alpha-smooth muscle actin [75]. ACTA2, a universal cytoskeletal protein, plays a pivotal role in cell differentiation and migration, as elucidated in a recent study. Inhibition of ACTA2 has been shown to significantly curtail cellular motility and Erk1/2 phosphorylation [76]. Patients exhibiting overexpression of ACTA2 have demonstrated heightened distant metastasis and an unfavorable prognosis [77]. Another gene of notable significance is MYL9, also referred to as MLC2, which encodes myosin light chain 9. This gene regulates myosin II activity through post-translational phosphorylation [78]. Recent research [68] has demonstrated the indispensable role of MYL9 in cytoskeletal dynamics and experimental metastasis [79]. Furthermore, its overexpression has been associated with enhanced matrix remodeling and the promotion of invasion in CAFs [80]. Moving on to TAGLN, it intricately governs the expression of an actin-binding protein that is crucial in cytoskeletal alterations and transformations [81]. Overexpression of TAGLN has been shown to induce cell invasion and facilitate tumor metastasis [82]. Notably, in lung cancer, its expression can be augmented under hypoxic conditions [83] whereas the loss of TAGLN inhibits tumor growth [84]. In the context of ovarian cancer, TAGLN has been found to promote disease progression through its influence on matrix stiffness [85]. Lastly, we turn our attention to tropomyosin beta chain isoform 2 (*TPM2*), which has been implicated in colorectal cancer recurrence and has been identified as a TASC marker closely associated with a poor prognosis, as per recent single-cell analysis [86, 87]. Another illuminating study [75], focusing on the single-cell analysis of CRC patients, has substantiated the presence of differentially expressed genes (DEGs) between CAFs and normal tissue fibroblasts, including MYL9 and TPM2, thereby affirming their correlation with an adverse prognosis in the TCGA cohort of CRC [86].

Among those hub genes, our discerning analysis has illuminated the distinct overexpression of MYL9 and TPM2 in CAFs, while exhibiting relatively low expression levels in both tumor and normal samples. These findings align with the results obtained from our biology experiments. On a histological level, both TPM2 and MYL9 manifested heightened expression in tumor tissue when compared to normal tissue, elegantly coinciding with the presence of α -SMA sites. On a cellular level, although displaying slight deviations from the histological results, TPM2 demonstrated significant upregulation at both the mRNA and protein levels in CAFs, in contrast to normal colorectal epithelial cells (FHC) and colorectal cancer cells (SW480, SW620, and RKO). Conversely, MYL9 did not exhibit notable specificity between CAFs and other

cell lines. It is noteworthy that Zhou et al.'s single-cell analysis of colorectal cancer patients corroborated the differential expression of MYL9 in normal tissue and primary tumors through immunohistochemistry [86].

CAF original MYL9 high expression has been proven to bring tumorigenic enhancement to tumor cells by regulating the secretion of CCL2 and TGF- β 1 [88] in colorectal cancer, which is consistent with our findings. However, another hub gene, TPM2, is also worthy of exploration. Tumor original TPM2 has been identified to play a pro-tumor role in the progression of several types of cancers, including prostate [89], hepatocellular [90], and colorectal cancer [91]. Besides, several previous studies have already found that the higher expression of CAF original TPM2 in CAFs might accelerate the disease progression in colorectal cancer [86, 92]. However, there is still some blank that should be filled in, because until now studies about CAF original TPM2 still stay at the bioinformatics level. In the present study, we identified that CAFs expressed higher TPM2 than MYL9 in both mRNA and protein levels. It might be more valuable to be used as a biomarker or therapeutic target. The co-culture system of CAFs and colorectal cancer cells in the present study revealed that CAFs could enhance malignant cells' tumorigenic phenotype indirectly without physical contact. Knocking down CAF original TPM2 expression could partially inhibit this kind of enhancement. Selectively inhibiting TPM2 expression in CAFs instead of tumor cells might bring an anti-tumor effect in treating colorectal cancer. However, in TME, it is more complicated, the cellular interactions and molecular expression patterns are dynamically changing [93–95]. Whether inhibition of CAF original TPM2 has a similar anti-tumor performance in vivo is valuable to be further explored.

There are several limitations in the present study. While our diligent exploration of single-cell databases has bolstered our findings, we have not yet extinguished the existence of various CAF subgroups. Furthermore, although we successfully isolated CAFs from tumor tissue and validated their identity through α -SMA, we have yet to fractionate them into distinct subpopulations to unravel the diverse characteristics of CAFs. In future studies, we intend to employ cell cytometry to sort these cells and delve into the disparities among these subgroups.

Conclusion

In summary, we built a functional gene signature (FGS) that could stratify CRC patients into low- and high-risk groups. Compared to the high-risk group, the low-risk group had significantly better clinicopathological features and prognosis. Meanwhile, the predicting ability

of FGS was assessed in several other independent CRC cohorts, which proved the FGS was valuable and stable. The further analyses into the diverse Tumor Immune Microenvironments (TIMEs) in colorectal cancer patients has unveiled that "hotter" TIMEs, characterized by intensified infiltration of tumor-infiltrating lymphocytes (TILs) and diminished expression of immune checkpoint molecules, portend a more favorable prognosis. Notably, CAFs have been intricately associated with varying prognoses, and TPM2 may serve as a valuable marker for CAFs, potentially serving as a prognostic factor of utmost import.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-024-12996-y>.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.
Supplementary Material 8.
Supplementary Material 9.
Supplementary Material 10.
Supplementary Material 11.
Supplementary Material 12.
Supplementary Material 13.
Supplementary Material 14.
Supplementary Material 15.
Supplementary Material 16.
Supplementary Material 17.
Supplementary Material 18.

Acknowledgements

We appreciated the TCGA and GEO databases for providing the original data.

Authors' contributions

Ziyan Zhu performed the biology experiment, wrote the related results, and polished the language. Jikun Li revised the manuscript. Xuezhong Xu prepared the figures. Zhenzhong Fa performed statistical analysis. Yue Wang edited the figures. Jiezhou performed bioinformatic analysis. Yixin Xu designed the research, wrote and revised the manuscript, and edited the figures. All authors agree to be accountable for the content of this work. All authors contribute to the article and approved the submitted version.

Funding

This study was funded by the Jiangsu University Medical Education Collaborative Innovation Fund Project (No. JDY2023018 and No. JDY2023017), the Medical Research Project of Jiangsu Health Commission (No. Z2021010), and Changzhou Sci & Tech Program (No. CJ20230007).

Availability of data and materials

Data supporting these research findings were collected from Wujin Hospital. The original contributions presented in the study are included in the article/

Supplementary Material, further inquiries can be directed to the corresponding author.

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Wujin Hospital (Protocol code: 2023-SR-114; Approval date: April 25, 2023). Informed consent to participate was obtained from all patients and the study has been performed in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of General Surgery, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. ²Department of General Surgery, Wujin Hospital Affiliated With Jiangsu University, Changzhou, Jiangsu Province, China. ³Department of General Surgery, the Wujin Clinical College of Xuzhou Medical University, Changzhou, Jiangsu Province, China. ⁴Changzhou Key Laboratory of Molecular Diagnostics and Precision Cancer Medicine, Changzhou, Jiangsu Province, China.

Received: 26 April 2024 Accepted: 26 September 2024

Published online: 28 September 2024

References

- Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209–49.
- Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023. *CA Cancer J Clin.* 2023;73(1):17–48.
- Siegel RL, Wagle NS, Cercek A, et al. Colorectal cancer statistics, 2023. *CA Cancer J Clin.* 2023;73(3):233–54.
- Kobayashi H, Gieniec KA, Lannagan TRM, et al. The Origin and Contribution of Cancer-Associated Fibroblasts in Colorectal Carcinogenesis. *Gastroenterology.* 2022;162(3):890–906.
- Winkler J, Abisoye-Ogunniyan A, Metcalf KJ, et al. Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat Commun.* 2020;11(1):5120.
- Darvishi B, Eisavand MR, Majidzadeh AK, et al. Matrix stiffening and acquired resistance to chemotherapy: concepts and clinical significance. *Br J Cancer.* 2022;126(9):1253–63.
- Anderson NM, Simon MC. The tumor microenvironment. *Curr Biol.* 2020;30(16):R921–5.
- Zhong B, Cheng B, Huang X, et al. Colorectal cancer-associated fibroblasts promote metastasis by up-regulating LRG1 through stromal IL-6/STAT3 signaling. *Cell Death Dis.* 2021;13(1):16.
- Costa A, Kieffer Y, Scholer-Dahirel A, et al. Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. *Cancer Cell.* 2018;33(3):463–79.e10.
- Hanley CJ, Waise S, Ellis MJ, et al. Single-cell analysis reveals prognostic fibroblast subpopulations linked to molecular and immunological subtypes of lung cancer. *Nat Commun.* 2023;14(1):387.
- Liberzon A, Birger C, Thorvaldsdóttir H, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1(6):417–25.
- Xu Y, Hu J, Cao C, et al. Distinct Hypoxia-Related Gene Profiling Characterizes Clinicopathological Features and Immune Status of Mismatch Repair-Deficient Colon Cancer. *J Oncol.* 2021;2021:2427427.
- Xu Y, Cao C, Zhu Z, et al. Novel Hypoxia-Associated Gene Signature Depicts Tumor Immune Microenvironment and Predicts Prognosis of Colon Cancer Patients. *Front Genet.* 2022;13: 901734.

14. Avelar RA, Ortega JG, Tacutu R, et al. A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biol.* 2020;21(1):91.
15. Lin R, Fogarty CE, Ma B, et al. Identification of ferroptosis genes in immune infiltration and prognosis in thyroid papillary carcinoma using network analysis. *BMC Genomics.* 2021;22(1):576.
16. Jiang A, Meng J, Bao Y, et al. Establishment of a prognosis Prediction Model Based on Pyroptosis-Related Signatures Associated With the Immune Microenvironment and Molecular Heterogeneity in Clear Cell Renal Cell Carcinoma. *Front Oncol.* 2021;11: 755212.
17. Bian Z, Fan R, Xie L. A Novel Cuproptosis-Related Prognostic Gene Signature and Validation of Differential Expression in Clear Cell Renal Cell Carcinoma. *Genes (Basel).* 2022;13(5):851.
18. Xu L, Deng C, Pang B, et al. TIP: A Web Server for Resolving Tumor Immunophenotype Profiling. *Cancer Res.* 2018;78(23):6575–80.
19. Huether R, Dong L, Chen X, et al. The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat Commun.* 2014;5:3630.
20. Zuber J, Shi J, Wang E, et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature.* 2011;478(7370):524–8.
21. Park SG, Lee D, Seo HR, et al. Cytotoxic activity of bromodomain inhibitor NVS-CECR2-1 on human cancer cells. *Sci Rep.* 2020;10(1):16330.
22. Marisa L, de Reyniès A, Duval A, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* 2013;10(5): e1001453.
23. Smith JJ, Deane NG, Wu F, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology.* 2010;138(3):958–68.
24. Freeman TJ, Smith JJ, Chen X, et al. Smad4-mediated signaling inhibits intestinal neoplasia by inhibiting expression of β -catenin. *Gastroenterology.* 2012;142(3):562–71.e2.
25. Williams CS, Bernard JK, Demory Beckler M, et al. ERBB4 is over-expressed in human colon cancer and enhances cellular transformation. *Carcinogenesis.* 2015;36(7):710–8.
26. Chen MS, Lo YH, Chen X, et al. Growth Factor-Independent 1 Is a Tumor Suppressor Gene in Colorectal Cancer. *Mol Cancer Res.* 2019;17(3):697–708.
27. Tripathi MK, Deane NG, Zhu J, et al. Nuclear factor of activated T-cell activity is associated with metastatic capacity in colon cancer. *Cancer Res.* 2014;74(23):6947–57.
28. Loboda A, Nebozhyn MV, Watters JW, et al. EMT is the dominant program in human colon cancer. *BMC Med Genomics.* 2011;4:9.
29. Allen WL, Dunne PD, McDade S, et al. Transcriptional subtyping and CD8 immunohistochemistry identifies poor prognosis stage II/III colorectal cancer patients who benefit from adjuvant chemotherapy. *JCO Precis Oncol.* 2018;2018:PO.17.00241.
30. Verstraete M, Debucquoy A, Dekervel J, et al. Combining bevacizumab and chemoradiation in rectal cancer. Translational results of the AXEBeam trial. *Br J Cancer.* 2015;112(8):1314–25.
31. Kaur S, Saldana AC, Elkahloun AG, et al. CD47 interactions with exportin-1 limit the targeting of m(7)G-modified RNAs to extracellular vesicles. *J Cell Commun Signal* 2022;16(3):397–419.
32. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
33. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.
34. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics.* 2010;26(12):1572–3.
35. Schubert M, Klinger B, Klünemann M, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun.* 2018;9(1):20.
36. Holland CH, Szalai B, Saez-Rodriguez J. Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochim Biophys Acta Gene Regul Mech.* 2020;1863(6): 194431.
37. Holland CH, Tanevski J, Perales-Patón J, et al. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol.* 2020;21(1):36.
38. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–50.
39. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–40.
40. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 2012;40(10):4288–97.
41. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7): e47.
42. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol.* 2019;37(7):773–82.
43. Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016;17(1):218.
44. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18(1):220.
45. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.
46. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb).* 2021;2(3).
47. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–7.
48. Mayakonda A, Lin DC, Assenov Y, et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018;28(11):1747–56.
49. Fu J, Li K, Zhang W, et al. Large-scale public data reuse to model immunotherapy response and resistance. *Genome Med.* 2020;12(1):21.
50. Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med.* 2018;24(10):1550–8.
51. Butler A, Hoffman P, Smibert P, et al. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* 2018;36(5):411–20.
52. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021;184(13):3573–87.e29.
53. Zhang X, Lan Y, Xu J, et al. Cell Marker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* 2019;47(D1):D721–8.
54. Franzén O, Gan LM, Björkregren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford).* 2019;2019:baz046.
55. Miao Z, Zhao X, Liu X. Hypoxia induced β -catenin lactylation promotes the cell proliferation and stemness of colorectal cancer through the wnt signaling pathway. *Exp Cell Res.* 2023;422(1): 113439.
56. Caporarello N, Lupo G, Olivieri M, et al. Classical VEGF, Notch and Ang signalling in cancer angiogenesis, alternative approaches and future directions (Review). *Mol Med Rep.* 2017;16(4):4393–402.
57. Chen Z, Li H, Li Z, et al. SHH/GLI2-TGF- β 1 feedback loop between cancer cells and tumor-associated macrophages maintains epithelial-mesenchymal transition and endoplasmic reticulum homeostasis in cholangiocarcinoma. *Pharmacol Res.* 2023;187: 106564.
58. Secintini IE, Ozgur T, Dede I. PD-L1 Expression in Colorectal Adenocarcinoma Is Associated With the Tumor Immune Microenvironment and Epithelial-Mesenchymal Transition. *Am J Clin Pathol.* 2022;158(4):506–15.
59. Padmanabhan S, Gaire B, Zou Y, et al. IFN γ -induced PD-L1 expression in ovarian cancer cells is regulated by JAK1, STAT1 and IRF1 signaling. *Cell Signal.* 2022;97: 110400.
60. Zlotnik A, Yoshie O. The chemokine superfamily revisited. *Immunity.* 2012;36(5):705–16.
61. Vilgelm AE, Richmond A. Chemokines Modulate Immune Surveillance in Tumorigenesis, Metastasis, and Response to Immunotherapy. *Front Immunol.* 2019;10:333.
62. Czajka-Francuz P, Cisoń-Jurek S, Czajka A, et al. Systemic Interleukins' Profile in Early and Advanced Colorectal Cancer. *Int J Mol Sci.* 2021;23(1):124.
63. Saxton RA, Sabatini DM. mTOR Signaling in Growth, Metabolism, and Disease. *Cell.* 2017;168(6):960–76.
64. Kumar V, Ramnarayanan K, Sundar R, et al. Single-Cell Atlas of Lineage States, Tumor Microenvironment, and Subtype-Specific Expression Programs in Gastric Cancer. *Cancer Discov.* 2022;12(3):670–91.

65. Nurmik M, Ullmann P, Rodriguez F, et al. In search of definitions: Cancer-associated fibroblasts and their markers. *Int J Cancer*. 2020;146(4):895–905.
66. Goers L, Freemont P, Polizzi KM. Co-culture systems and technologies: taking synthetic biology to the next level. *J R Soc Interface*. 2014;11(96):20140065.
67. Mhaidly R, Mechta-Grigoriou F. Role of cancer-associated fibroblast subpopulations in immune infiltration, as a new means of treatment in cancer. *Immunol Rev*. 2021;302(1):259–72.
68. Monteran L, Erez N. The Dark Side of Fibroblasts: Cancer-Associated Fibroblasts as Mediators of Immunosuppression in the Tumor Microenvironment. *Front Immunol*. 2019;10:1835.
69. Takahashi H, Sakakura K, Kawabata-Iwakawa R, et al. Immunosuppressive activity of cancer-associated fibroblasts in head and neck squamous cell carcinoma. *Cancer Immunol Immunother*. 2015;64(11):1407–17.
70. Barrett R, Puré E. Cancer-associated fibroblasts: key determinants of tumor immunity and immunotherapy. *Curr Opin Immunol*. 2020;64:80–7.
71. Mhaidly R, Mechta-Grigoriou F. Fibroblast heterogeneity in tumor microenvironment: Role in immunosuppression and new therapies. *Semin Immunol*. 2020;48: 101417.
72. Gentric G, Mechta-Grigoriou F. Tumor Cells and Cancer-Associated Fibroblasts: An Updated Metabolic Perspective. *Cancers (Basel)*. 2021;13(3):399.
73. Vickman RE, Faget DV, Beachy P, et al. Deconstructing tumor heterogeneity: the stromal perspective. *Oncotarget*. 2020;11(40):3621–32.
74. Cano A, Pérez-Moreno MA, Rodrigo I, et al. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat Cell Biol*. 2000;2(2):76–83.
75. Wang Y, Li Z, Zhang Z, et al. Identification ACTA2 and KDR as key proteins for prognosis of PD-1/PD-L1 blockade therapy in melanoma. *Animal Models and Experimental Medicine*. 2021;4(2):138–50.
76. Rockey DC, Weymouth N, Shi Z. Smooth Muscle α Actin (Acta 2) and Myofibroblast Function during Hepatic Wound Healing. *PLoS ONE*. 2013;8(10): e77166.
77. Lee HW, Park YM, Lee SJ, et al. Alpha-smooth muscle actin (ACTA2) is required for metastatic potential of human lung adenocarcinoma. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*. 2013;19(21):5879–89.
78. Sandquist JC, Swenson KI, Demali KA, et al. Rho kinase differentially regulates phosphorylation of nonmuscle myosin II isoforms A and B during cell rounding and migration. *J Biol Chem*. 2006;281(47):35873–83.
79. Medjkane S, Perez-Sanchez C, Gaggioli C, et al. Myocardin-related transcription factors and SRF are required for cytoskeletal dynamics and experimental metastasis. *Nat Cell Biol*. 2009;11(3):257–68.
80. Calvo F, Ege N, Grande-Garcia A, et al. Mechano-transduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer associated fibroblasts. *Nat Cell Biol*. 2013;15(6):637–46. <https://doi.org/10.1038/ncb2756>.
81. Liu Y, Wu J, Huang W, et al. Development and validation of a hypoxia-immune-based microenvironment gene signature for risk stratification in gastric cancer. *J Transl Med*. 2020;18:201.
82. Yu B, Chen X, Li J, et al. Stromal fibroblasts in the microenvironment of gastric carcinomas promote tumor metastasis via upregulating TAGLN expression. *BMC Cell Biol*. 2013;14:17.
83. Wu X, Dong L, Zhang R, et al. Transgelin overexpression in lung adenocarcinoma is associated with tumor progression. *Int J Mol Med*. 2014;34(2):585–91.
84. Fu J, Wang X, Yue Q. Functional loss of TAGLN inhibits tumor growth and increases chemosensitivity of non-small cell lung cancer. *Biochem Biophys Res Commun*. 2020;529(4):1086–93.
85. Wei X, Lou H, Zhou D, et al. TAGLN mediated stiffness-regulated ovarian cancer progression via RhoA/ROCK pathway. *Journal of Experimental & Clinical Cancer Research : CR*. 2021;40:292.
86. Zhou Y, Bian S, Zhou X, et al. Single-Cell Multiomics Sequencing Reveals Prevalent Genomic Alterations in Tumor Stromal Cells of Human Colorectal Cancer. *Cancer Cell*. 2020;38(6):818–28.e5.
87. Lee H-O, Hong Y, Etliglu HE, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat Genet*. 2020;52(6):594–603.
88. Deng S, Cheng D, Wang J, et al. MYL9 expressed in cancer-associated fibroblasts regulate the immune microenvironment of colorectal cancer and promotes tumor progression in an autocrine manner. *J Exp Clin Cancer Res*. 2023;42(1):294.
89. Wu Z, Ge L, Ma L, et al. TPM2 attenuates progression of prostate cancer by blocking PDLIM7-mediated nuclear translocation of YAP1. *Cell Biosci*. 2023;13(1):39.
90. Tian Z, Zhao J, Wang Y. The prognostic value of TPM1-4 in hepatocellular carcinoma. *Cancer Med*. 2022;11(2):433–46.
91. Huang Q, Li XM, Sun JP, et al. Tumor-derived endomucin promotes colorectal cancer proliferation and metastasis. *Cancer Med*. 2023;12(3):3222–36.
92. Mele V, Basso C, Governa V, et al. Identification of TPM2 and CNN1 as Novel Prognostic Markers in Functionally Characterized Human Colon Cancer-Associated Stromal Cells. *Cancers (Basel)*. 2022;14(8):2024.
93. Mempel TR, Lill JK, Altenburger LM. How chemokines organize the tumour microenvironment. *Nat Rev Cancer*. 2024;24(1):28–50.
94. Cox TR. The matrix in cancer. *Nat Rev Cancer*. 2021;21(4):217–38.
95. van Weverwijk A, de Visser KE. Mechanisms driving the immunoregulatory function of cancer cells. *Nat Rev Cancer*. 2023;23(4):193–215.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.