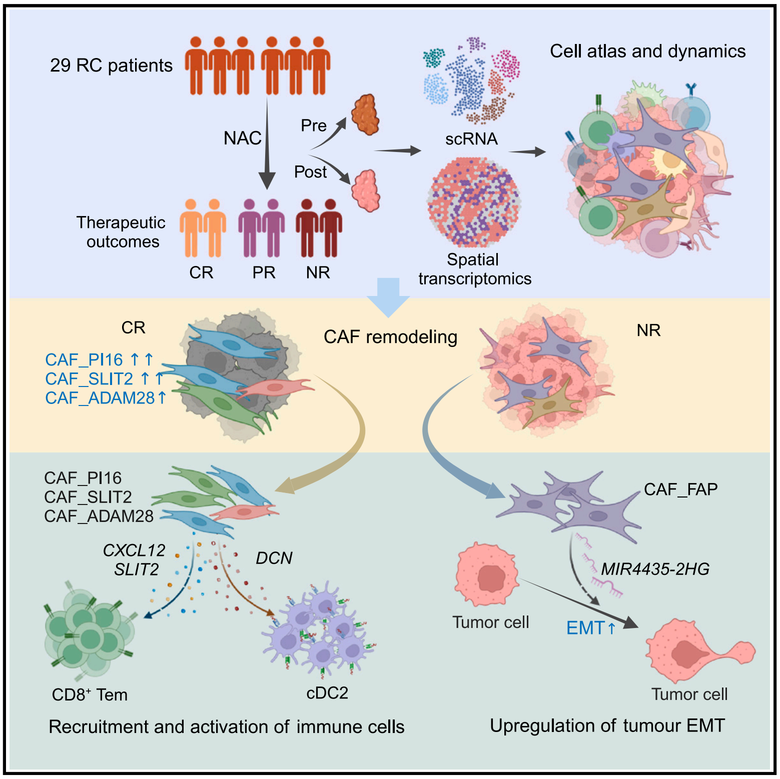
Cell Reports
Medicine

Article

Cancer-associated fibroblasts undergoing neoadjuvant chemotherapy suppress rectal cancer revealed by single-cell and spatial transcriptomics

Graphical abstract



Highlights

- NAC reshapes the CAF subsets in rectal cancer, associating with therapeutic outcomes
- Remodeled CAF subsets regulate immune cells by spatial recruitment and crosstalk
- CAF_FAP upregulates EMT of tumor cells through MIR4435-2HG induction

Authors

Pengfei Qin, Huaxian Chen,
Yuhang Wang, ..., Jufang Wang,
Shiping Liu, Hongcheng Lin

Correspondence

rendl@mail.sysu.edu.cn (D.R.),
wuliang@genomics.cn (L.W.),
jufwang@scut.edu.cn (J.W.),
liushiping@genomics.cn (S.L.),
lhcheng@mail.sysu.edu.cn (H.L.)

In brief

Qin et al. analyze cellular dynamics in 29 patients with rectal cancer who are sampled pairwise before and after neoadjuvant chemotherapy through single-cell and spatial transcriptome sequencing, revealing that the remodeled cancer-associated fibroblasts activate immunity and suppress tumor progression.

Qin et al., 2023, Cell Reports Medicine *4*, 101231
October 17, 2023 © 2023 The Author(s).
https://doi.org/10.1016/j.xcrm.2023.101231

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Cancer-associated fibroblasts undergoing neoadjuvant chemotherapy suppress rectal cancer revealed by single-cell and spatial transcriptomics

Pengfei Qin,[2,6,16] Huaxian Chen,[1,3,4,16] Yuhang Wang,[2,5,16] Liang Huang,[1,3,4,16] Ke Huang,[2,16] Guozhong Xiao,[1,3,4] Changpeng Han,[8] Jiancong Hu,[3,4,9] Dezheng Lin,[3,4,9] Xingyang Wan,[1,3,4] Yihui Zheng,[1,3,4] Yufeng Liu,[10] Guiming Li,[1,3,4] Haojie Yang,[8] Shubiao Ye,[1,3,4] Minyi Luo,[1,3,4] Yuanji Fu,[1,3,4] Hao Xu,[11] Luping Wen,[12] Zhiwei Guo,[13] Xunan Shen,[14] Zeyu Li,[2,15] Chunqing Wang,[2,15] Xi Chen,[2] Linying Wang,[2] Liuyong Sun,[2] Donglin Ren,[1,3,4,*] Liang Wu,[2,6,*] Jufang Wang,[5,*] Shiping Liu,[2,7,*] and Hongcheng Lin[1,3,4,17,*]

[1]Department of Coloproctology, Department of General Surgery, the Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou 510655, China
[2]BGI Research, Shenzhen 518083, China
[3]Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, the Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou 510655, China
[4]Guangdong Institute of Gastroenterology, Guangzhou 510655, China
[5]School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China
[6]BGI Research, Chongqing 401329, China
[7]BGI Research, Hangzhou 310030, China
[8]Department of Coloproctology, Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai 200437, China
[9]Department of Endoscopic Surgery, the Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong 510655, China
[10]Center for Medical Research on Innovation and Translation, The Second Affiliated Hospital, School of Medicine, South China University of Technology, Guangzhou 510320, China
[11]Department of Coloproctology, Shanghai Fourth People's Hospital, School of Medicine, Tongji University, Shanghai 200434, China
[12]Department of Coloproctology, Affiliated Hospital of Xuzhou Medical University, Xuzhou 221005, China
[13]Department of Coloproctology, The Eighth Hospital of Wuhan, Wuhan 430000, China
[14]BGI Research, Beijing 102601, China
[15]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[16]These authors contributed equally
[17]Lead contact
*Correspondence: rendl@mail.sysu.edu.cn (D.R.), wuliang@genomics.cn (L.W.), jufwang@scut.edu.cn (J.W.), liushiping@genomics.cn (S.L.), lhcheng@mail.sysu.edu.cn (H.L.)
https://doi.org/10.1016/j.xcrm.2023.101231

SUMMARY

Neoadjuvant chemotherapy (NAC) for rectal cancer (RC) shows promising clinical response. The modulation of the tumor microenvironment (TME) by NAC and its association with therapeutic response remain unclear. Here, we use single-cell RNA sequencing and spatial transcriptome sequencing to examine the cell dynamics in 29 patients with RC, who are sampled pairwise before and after treatment. We construct a high-resolution cellular dynamic landscape remodeled by NAC and their associations with therapeutic response. NAC markedly reshapes the populations of cancer-associated fibroblasts (CAFs), which is strongly associated with therapeutic response. The remodeled CAF subsets regulate the TME through spatial recruitment and cross-talk to activate immunity and suppress tumor progression through multiple cytokines, including *CXCL12*, *SLIT2*, and *DCN*. In contrast, the epithelial-mesenchymal transition of malignant cells is upregulated by CAF_FAP through *MIR4435-2HG* induction, resulting in worse outcomes. Our study demonstrates that NAC inhibits tumor progression and modulates the TME by remodeling CAFs.

INTRODUCTION

Colorectal cancers (CRCs) are the third most common malignancy worldwide,[1] approximately 30% of which are rectal cancer (RC).[2] In China, patients with RC account for a larger proportion of patients than that in Western countries.[3] Based on the clinical guidelines, neoadjuvant chemoradiotherapy followed by surgery is recommended for locally advanced RC (LARC).[4] However, researchers have found that chemoradiotherapy has no advantage in outcomes over neoadjuvant chemotherapy (NAC) without radiation for LARC, which may avoid severe side effects following radiotherapy. A neoadjuvant-modified FOLFOX6 study reported that FOLFOX6 without radiotherapy had no significant differences in outcomes compared to fluorouracil with radiotherapy in patients with LARC.[5] Some other recent studies demonstrated that in patients with LARC who were eligible for sphincter-sparing surgery, preoperative FOLFOX was not inferior to preoperative chemoradiotherapy with respect to disease-free survival (DFS).[6,7] In addition, NAC for LARC can achieve tumor downstaging and even a pathologically complete response, which increases the possibility of complete resection with negative pathological margins and the chance of performing a sphincter preservation procedure. Recently, immunotherapy has achieved exciting results in patients with mismatch repair deficiency (dMMR) or high microsatellite instability (MSI-H) CRC.[8,9] However, only 15% of patients with CRC harbor dMMR/MSI-H.[10,11] Strategies for the remaining patients with mismatch repair proficiency (pMMR) or who are microsatellite stable (MSS) demand further investigation. Considering that immune suppression and drug resistance of the tumor microenvironment (TME) lead to the unsatisfactory efficacy of NAC,[12] it is necessary to explore the connections between the TME and outcome in patients with pMMR/MSS CRC.

Most cytotoxic chemicals in NAC target the apparatus of mitosis to disrupt the core features of cell proliferation.[13] Despite its cytotoxic properties, there is mounting evidence that NAC promotes the antitumor immune response of the TME, motivating its use in tumor therapy.[13–15] However, the immunomodulatory mechanisms of action of NAC remain unclear. One common explanation is that chemotherapy initiates an antitumor immune response through the immunogenic cell death (ICD) process,[15] which induces endoplasmic reticulum stress and autophagy of damaged tumor cells and subsequently infiltrates immune cells into the TME. However, other mechanisms underlying the regulation of immunity remain unclear because TME heterogeneity and interactions are highly complex. In addition to immune cells, stromal cells, a major component of the tumor context, have rarely been reported regarding their functions and reactions to NAC. Cancer-associated fibroblasts (CAFs) and endothelial cells (ECs) release a series of extracellular factors to establish a cellular communication network and regulate the immune response or tumor progression.[16–19] Although the TME significantly affects tumor progression and clinical outcomes, the association between cellular diversity and NAC response in tumors has not been well investigated. Therefore, there is an urgent need to comprehensively study the TME modulation of NAC and determine how it predicts and reacts to therapy.

Considering these concerns, we enrolled 29 patients with RC of pMMR/MSS treated with NAC. Most of the samples were pairwise collected before and after treatment and categorized according to therapeutic responses to facilitate a comparative study across treatments or outcomes. We leveraged the single-cell RNA sequencing (SC) and spatial transcriptome sequencing (ST) technologies of Stereo-seq[20] to characterize the heterogeneity and spatial architecture of tumors. We investigated TME complexity and alterations modulated by NAC and the cellular and molecular composition associated with the therapeutic response. We uncovered the dynamics of CAF subsets remodeled by NAC and the roles of CAFs in modulating immune and malignant cells through multiple-cell interactions. Our data demonstrate the findings of NAC reshaping the TME and restraining tumors, which will facilitate future therapeutic targeting and outcome improvements.

RESULTS

Composition dynamics of tumor-infiltrating cells following NAC

To investigate the dynamics of the TME in RC following NAC treatment, we enrolled 29 patients with RC of MSS/pMMR, most of whom were pairwise sampled before (27 samples) and after (29 samples) the treatment (Figure 1A; Table S1). Platinum-based NAC was administered to all patients. Paired tumor endoscopic biopsies and surgical resections were collected pretherapy (PE) and post-therapy (PO), respectively. Surgical specimens were subjected to pathological examination to determine the outcome of NAC using tumor regression grade (TRG) systems, which included TRG0 (complete regression: no remaining viable cancer cells), TRG1 (nearly complete regression: only small clusters or single cancer cells remaining), TRG2 (partial regression: residual cancer remaining with predominant fibrosis), and TRG3 (poor or no regression: minimal or no tumor kill; extensive residual cancer).[21] Based on their TRGs, we further divided the patients into three groups: complete (or nearly complete) regression (CR; 7 patients) (TRG0 and TRG1), partial regression (PR; 15 patients) (TRG2), and no regression (NR; 5 patients) (TRG3). Two patients underwent endoscopic sampling rather than surgical resection after NAC, so the TRGs of these two patients could not be evaluated.

The composition of the TME was investigated using SC data. After quality control and filtering, we obtained SC data of 333,285 high-quality cells, which were clustered and identified as epithelial cells (24.4%), T cells (23.9%), plasma cells (18.2%), CAFs (10.0%), ECs (7.2%), B cells (5.4%), pericytes (3.8%), myeloid cells (2.6%), smooth-muscle cells (SMCs) (1.7%), mast cells (1.5%), and enteric glial cells (1.3%) (Figures 1B and S1A; Table S2). Malignant epithelial cells were distinguished from normal ones using copy-number variant (CNV) identification and cluster analysis (see STAR Methods). The cell composition and the abundance of the TME were dramatically remodeled by NAC and showed remarkable differences among the responsive groups (Figures 1C–1E and S1B). Significantly elevated by NAC, CAFs expanded to a larger population following therapy, especially in the responsive groups (CR and PR) (Figure 1E). T cells shrank in abundance, and ECs increased following therapy, suggesting their connection with treatment and outcomes (Figures 1E and S1B).

Remodeled CAF subsets associate with favorable outcomes

NAC altered the composition of CAF subsets in the responsive samples. We identified three CAF subsets, namely CAF_PI16, CAF_SLIT2, and CAF_ADAM28, which substantially expanded following therapy, especially in the responsive groups (CR and PR) (Figures 2A–2C and S2A). Three subsets, CAF_BMP4, CAF_FAP, and CAF_MMP1, decreased to knockdown levels after therapy in CR samples, whereas CAF_BMP4 maintained a higher abundance in NR samples (Figures 2B, 2C, and S2A).

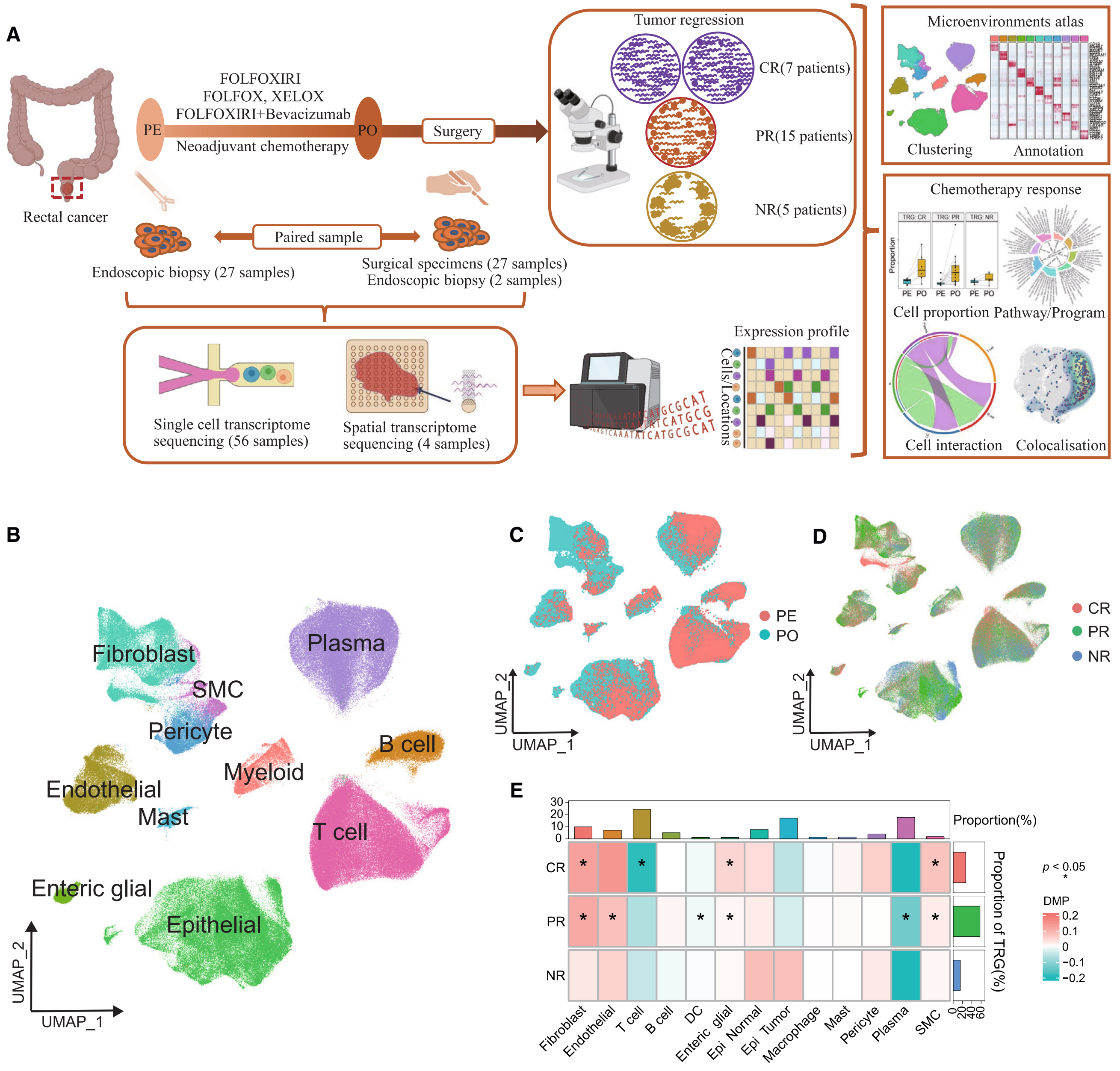


Figure 1. Dynamics of tumor-infiltrating cell composition following NAC

(A) Schematic overview of experimental design and analytical workflow.

(B) Uniform manifold approximation and projection (UMAP) plot of the major cell types of SC data. There are 333,285 cells after filtration from 56 samples. Cell types were annotated according to their differential genes.

(C) UMAP plot presents the dynamic changes of cell types between pretherapy (PE) and post-therapy (PO).

(D) UMAP plot presents the difference of cell types among outcome groups.

(E) Dynamic changes of the major cell types among the treatments and outcome groups. Difference of mean proportion (DMP) for each cell type between PE and PO were calculated. DMP >0 means an increase in the proportion of cell type following treatment, and DMP <0 means a decrease. Patients included CR: 6, PR: 14, and NR: 5. Statistical significance was determined using the Wilcoxon signed-rank test. *p <0.05 was considered statistically significant.

The evident alterations and differences in CAF subsets suggest their diverse roles and association with responses to therapy. CAF subsets identified in the SC data (Figures 2A and S2A) were validated in four ST samples of CR and NR, which showed more details of specific spatial distribution (Figures 2D and S2B). Consistent with the abundance observed in the SC data, CAF_PI16, CAF_SLIT2, and CAF_ADAM28 dominated the TME space of CR samples, whereas CAF_BMP4, CAF_FAP, and CAF_MMP1 were the major components in the TME space of NR samples. Trajectory inference recovered multiple lineages

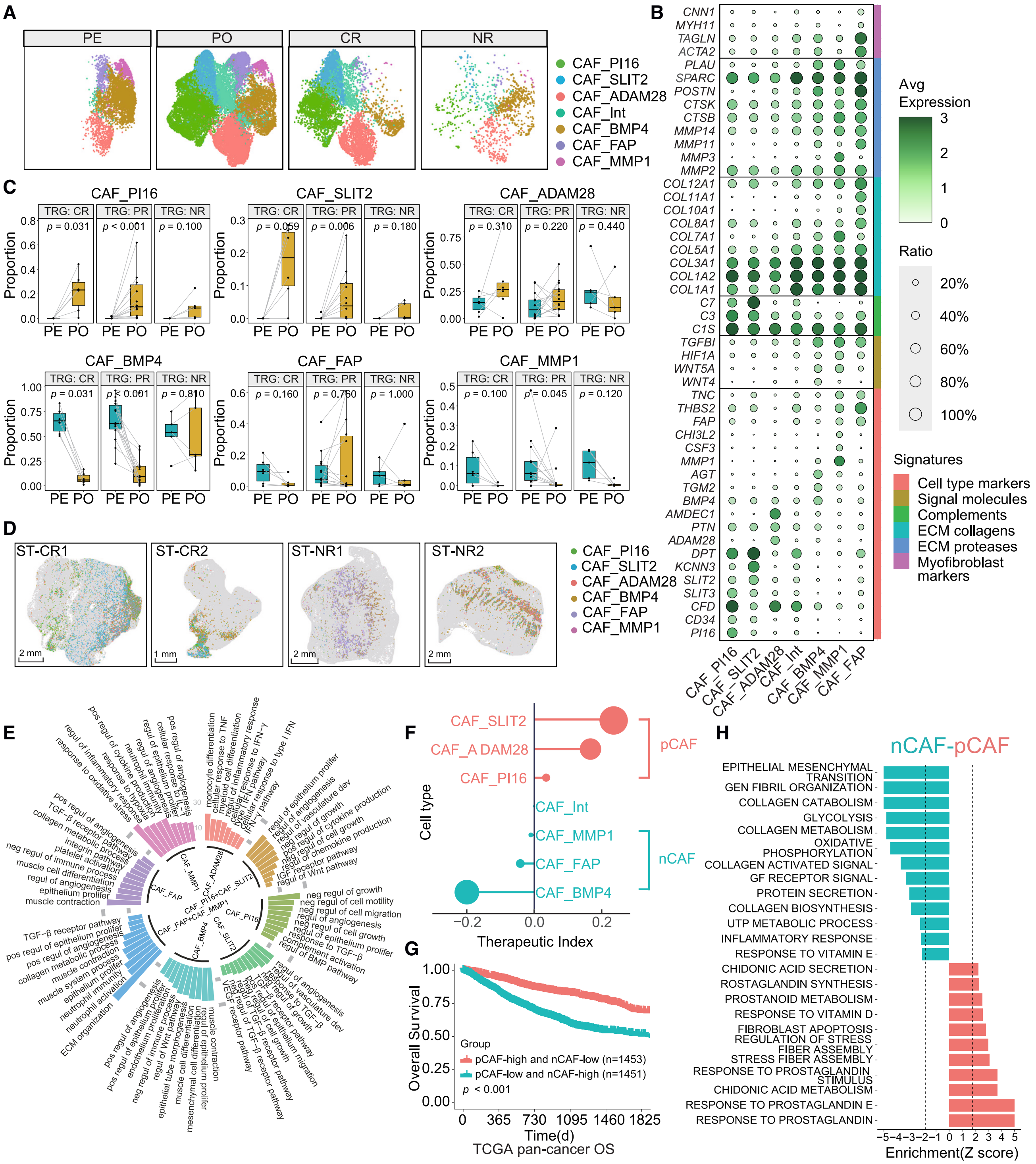


Figure 2. CAF subsets associate with therapeutic responses

(A) UMAP plots of the CAF subsets identified in SC data.

(B) The expression of the marker genes of the CAF subsets. “Ratio” reflects the percentage of cells expressing the gene at non-zero levels. “Avg Expression” reflects the averaged log-normalized expression.

(C) Proportion comparisons of the CAF subsets among treatments and outcome groups. Box middle lines: median; box limits: upper and lower quartiles; box whiskers: 1.5× the interquartile range; black dots: samples, connected with lines if they were collected from the sample patient. Patients included CR: 6, PR: 14, and NR: 5; p <0.05 of Wilcoxon signed-rank test was considered statistically significant.

(legend continued on next page)

of differentiation between CAF_PI16 and other cells, including CAF_PI16-CAF_SLIT2-CAF_FAP, CAF_PI16-CAF_BMP4, and CAF_PI16-CAF_ADAM28-CAF_BMP4 (Figures S2C and S2D).

The expressions of *PI16*, *CD34*, *SLIT2*, *CFD*, *C3*, and *C7*, which are associated with regulation of cell differentiation and immunoregulation,[22–24] were upregulated in CAF_PI16, CAF_SLIT2, and CAF_ADAM28, whereas pronounced collagen, matrix metalloproteinase (MMP), and WNT genes associated with ECM remodeling and tumor promotion[25,26] were upregulated in subsets of CAF_BMP4, CAF_FAP, and CAF_MMP1 (Figure 2B). Gene Ontology described the functional differences among the CAF subsets (Figure 2E). CAF_PI16 and CAF_SLIT2 enriched genes of negative regulation of cell growth and migration, as well as the positive regulation of cytokine production and complements. CAF_ADAM28 enriched genes of the interferon-gamma signaling pathway, regulation of monocyte differentiation, and regulation of the inflammatory response. Highly expressed genes of CAF_FAP and CAF_MMP1 were enriched in the pathways related to ECM organization, collagen metabolism, and epithelial proliferation. CAF_BMP4 enriched genes in the pathway related to mesenchymal cell differentiation (Figure 2E). We screened for substantially different cytokines genes expressed in CAF subsets (Figure S2E). CAF_PI16 and CAF_SLIT2 showed high levels of *DCN*, *CXCL12*, *SLIT2*, *C7*, and *C3*, which are related to cell differentiation, recruitment, and immunoregulation.[18,24,27] In contrast, CAF_MMP1 highly expressed chemokines (*CXCL1*, *CXCL3*, *CXCL5*, *CXCL6*, *CXCL8*, and *MIF*), CAF_BMP4 expressed tumor necrosis factors (*TNFSF11*), and CAF_MMP1, CAF_BMP4, and CAF_FAP expressed the growth factor *TGFB1*, which is associated with the promotion of cell chemotaxis and growth.[28–32]

We investigated the correlation between CAF subsets and NAC outcomes and tumor prognosis. We applied a therapeutic index (Ti) to measure whether the cellular proportion of CAF subsets correlated with therapeutic response (see STAR Methods). CAF_PI16, CAF_SLIT2, and CAF_ADAM28 cells were strongly correlated with favorable therapeutic outcomes, whereas worse outcomes were observed for CAF_FAP, CAF_MMP1, and CAF_BMP4 (Figure 2F). Therefore, we grouped CAF_PI16, CAF_SLIT2, and CAF_ADAM28 as positive-response-associated CAFs (pCAFs) and CAF_FAP, CAF_MMP1, and CAF_BMP4 as negative-response-associated CAFs (nCAFs). To further investigate the clinical relevance of CAF subsets, we conducted survival analysis on pan-cancer datasets from The Cancer Genome Atlas (TCGA). The overall survival rate of integrated pan-cancer samples significantly increased in the pCAF-high and nCAF-low score groups (Figure 2G). Furthermore, we investigated multiple indicators, including overall survival (OS), progression-free interval (PFI), disease-free interval (DFI), and disease-specific survival (DSS) in 21 cancer types by evaluating the signature scores of each CAF subset. The results showed that the pCAF subsets extended the values of survival rates and free intervals in numerous cancer types (Figure S2F).

Comparing the signaling pathways of gene set enrichment analysis (GSEA) between nCAFs and pCAFs (Figure 2H), we observed that the most significantly upregulated pathways of nCAFs were involved in epithelial-mesenchymal transition (EMT), collagen metabolism, glycolysis metabolism, and oxidative phosphorylation, indicating their tumor-supporting roles in tumor cell migration, ECM remodeling, and energy support. However, the most significantly upregulated pathway in pCAF was the prostaglandin (PG) response, suggesting that PG is involved in reshaping CAF populations.

Although myofibroblasts have traditionally been considered to be terminally differentiated, PG prevents or reverses myofibroblast differentiation, as characterized by its ability to diminish the expression of collagen I and smooth-muscle actin in myofibroblasts.[33,34] Consistent with previous studies, the PG response and prostanoid metabolism pathways were upregulated in pCAFs compared with in nCAFs (Figures 3A and S2G), and the gene score of the PG response pathway was positively correlated with the abundance of CAF_PI16 and CAF_SLIT2 across samples (Figure S2H), suggesting a connection between PG and CAF remodeling. Chemotherapy induces PG release from vascular ECs[35] and elevates PG levels in the TME.[36] We screened all possible sources of PG and identified a PG connection between CAF and EC subsets. EC_Vein, which expanded following therapy and was enriched in the CR group, expressed higher levels of the PG enzymes *PTGDS*, *PTGIS*, *PTGS2*, and *PTGES* than the other ECs (Figures 3B, S2I, and S2J). A positive correlation was observed between EC_Vein and pCAF subsets and a negative correlation was observed with nCAF subsets (Figure 3C), which indicated that ECs regulated the expansion of pCAF subsets following therapy. We investigated the correlation between the PG enzyme genes expressed in EC_Vein and the PG response pathway in pCAF subsets. Positive correlations suggested that the pCAF subsets were regulated by PG produced from EC_Vein (Figures 3D and S2K). Furthermore, we examined the spatial expression of PG synthase in EC_Vein, PG receptors in CAFs, and the colocalization of EC_Vein with PG synthase expression and CAFs with PG response (Figure 3E). They were spatially colocalized in CR samples but not in NR samples, indicating their interactions and outcome relevance. These findings suggest that PG from ECs extends the pCAF subsets associating with better outcomes.

(D) The cell distribution of the CAF subsets in ST data. CAF subsets were annotated using the marker genes identified in SC data. There are four ST samples after treatments from two patients with CR (ST-CR1 and ST-CR2) and two NR patients (ST-NR1 and ST-NR2).

(E) Circular bar plot of GO enrichment (biological processes) of the CAF subsets. The bar height presents the gene counts enriched in the pathway. Only gene sets with adjusted p value < 0.01 were plotted. pos, positive; neg, negative; regul, regulation; dev, development; prolifer, proliferation.

(F) Therapeutic index (Ti) of the CAF subsets. A positive Ti indicates the CAF subset is associated with a better response, whereas a negative Ti indicates worse response. pCAF, positive-response-associated CAF; nCAF, negative-response-associated CAF.

(G) pCAF subsets associate with the extended survivals in pan-cancer samples from TCGA. The gene scores of pCAF and nCAF subsets were calculated from the averaged expression level of their differential genes. Red line: group of pCAF-high and nCAF-low scores (n = 1,453); blue line: group of pCAF-low and nCAF-high scores (n = 1,451); the p value was calculated using a two-sided log-rank test.

(H) Comparison of pathway enrichment between pCAF subsets (red) and nCAF subsets (blue) using GSEA. The dashed lines of Z score = ±1.8 were plotted.

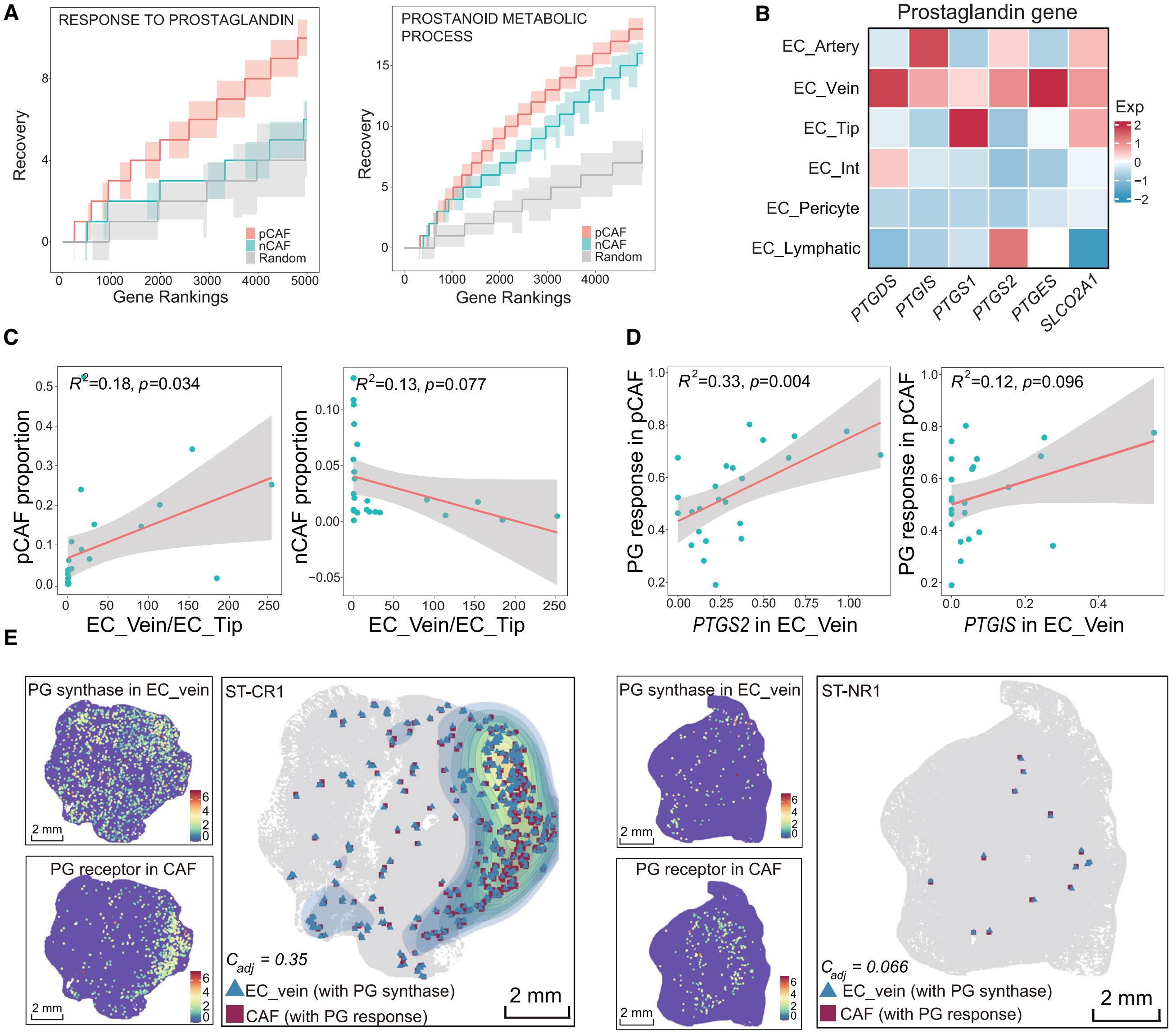


Figure 3. Prostaglandins from endothelial cells remodel CAF subsets

(A) Enrichment of prostaglandin (PG) pathways in the pCAF subsets (red) and the nCAF subsets (blue) compared with a background (gray) revealed by GSEA.

(B) Heatmap of expression of PG synthase in EC subsets.

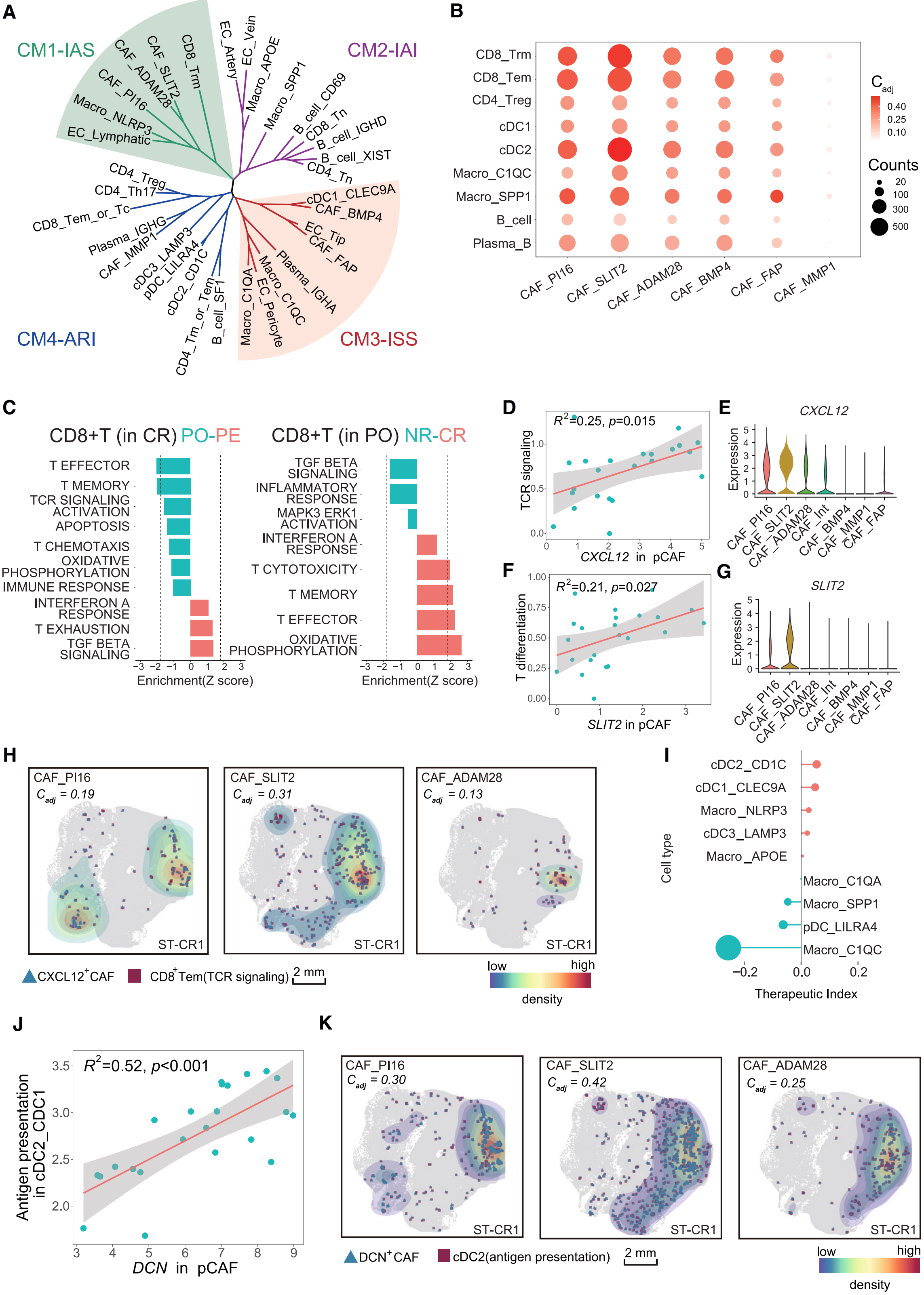
(C) Scatterplot presents the connection between EC_Vein and the pCAF subsets (left) and the nCAF subsets (right) after therapy. The proportion of EC_Vein was adjusted by the proportion of EC_Tip. R^2 value tells how much variation is explained by the model. The p value of F statistic indicates if the model fitting is significant.

(D) Scatterplot presents the correlation of the PG synthase genes expressed by EC_Vein with PG responses in the pCAF subsets after therapy.

(E) Spatial colocalization of EC_Vein expressing genes of PG synthase and CAFs responding to PG in ST-CR1 (left) and ST-NR1 (right) samples. For each panel, top left: the distribution of EC_Vein expressing genes of PG synthase (total expression of PTGDS, PTGIS, PTGS1, PTGS2, and PTGES); bottom left: the distribution of CAFs expressing PG receptors (total expression of PTGER3, PEGER4, PTGDR2, PTGER1, PTGER2, and PTGDR1); right: the colocalization of EC_Vein expressing genes of PG synthase and CAFs responding to PG; C_{adj} represents the intensity of colocalization between two cell sets. The color gradient indicates the density of the colocalized cells.

Immune microenvironment is modulated by CAFs following therapy

We investigated the heterogeneous complexity of immune cells, showing specific alterations in each cell type between PE and PO and differences among the responsive groups (Figures S3A and S3B). We examined the coenrichment patterns of TME cells after therapy to investigate how CAFs remodel the immune microenvironment associated with therapy. Clustering analysis performed using a neighbor-joining tree from the SC data identified four stable cellular modules (CM1–CM4) (Figure 4A). We



(legend on next page)

annotated these four modules as immune active stromal cells (CM1-IASs), inactive immune cells (CM2-IAIs), immune suppressive stromal cells (CM3-ISSs), and active-regulatory immune cells (CM4-ARIs), according to the correlated cell subtypes. All pCAF subsets were enriched in the CM1-IAS module with CD8$^+$ tissue-resident memory T (Trm) cells and Macro_NLRP3. High levels of CD8$^+$Trm in the CR group were related to better therapeutic outcomes (Figure S3B). CM2-IAIs were enriched with naive CD8$^+$ and CD4$^+$ T cells, with Macro_SPP1 and EC subsets showing an inactive immune state. CAF_FAP and CAF_BMP4, along with subtypes of ECs and macrophages, were composed of the CM3-ISSs, promoting immunosuppression and tumor progression. Conventional DCs (cDCs), plasmacytoid DCs (pDCs), plasma, and regulatory T cells (Tregs) correlated with CD8$^+$ effector memory T (Tem) cells were enriched in CM4-ARIs, indicating the immune-regulating and -activating functions of this community.

Consistent with the cellular modules of the SC data, we observed that the CAF subsets in ST data showed spatially diverse colocalization with immune cells (Figures 4B and S4A). CAF_SLIT2 and CAF_PI16 strongly enriched CD8$^+$Trm, CD8$^+$Tem, and cDC2s in CR samples, with higher levels than nCAF subsets. These immune cells are more enriched than Tregs, cDC1, Macro_SPP1, Macro_C1QC, and B cells (Figure 4B). In NR samples, CAF_FAP and CAF_BMP4 colocalized with more Macro_SPP1 than other immune cells and were stronger than pCAF subsets (Figure S4A). Therefore, both cellular modules and spatial colocalization suggested interactions and regulations between CAF subsets and other immune cells.

NAC significantly activated cell differentiation and immune response of CD8$^+$T in CR samples of PO, as we observed upregulated pathways of effector and memory cells and T cell receptor (TCR) signals after therapy (Figure 4C), which were not observed in NR samples (Figure S4B). Comparison of the CR and NR samples after therapy highlighted the differences in pathways of oxidative phosphorylation, cytotoxicity, and effector and memory cells, indicating a stronger immune response in the CR samples (Figure 4C). CD4$^+$T cells of NR samples after therapy, especially Tregs, showed stronger activation, as the interferon, interleukin-2 (IL-2), and transforming growth factor β (TGF-β) signaling pathways were upregulated, while these cells of CR samples were turned into an immune quiescent state (Figure S4B).

As T cells and CAF subsets were closely correlated in abundance and spatially colocalized (Figures 4A and 4B), we investigated how CAF subsets modulate T cells. We systematically calculated the correlation between CAF ligands and T cell signaling pathways across all samples following therapy. *CXCL12* and *SLIT2* from CAF_SLIT2 and CAF_PI16 cells upregulated T cell chemotaxis, differentiation, and activation, as suggested by significant correlations (Figures 4D–4G and S4C). Validated by ST data, we observed strong colocalization of CXCL12$^+$pCAF subsets and CD8$^+$Tem with active TCR signaling in the CR sample (ST-CR1) (Figure 4H), supporting their interaction at adjacent distance. Compared with the CR sample, this colocalization was not evident because fewer pCAFs and lower *CXCL12* expression were identified in the NR sample (ST-NR1) (Figure S4D). This spatial colocalization of nCAF subsets were not as strong as that of pCAF subsets (Figures S4E and S4F). The receptor *CXCR4* of *CXCL12* was expressed at a much higher level in CD8$^+$Tem cells than in Tregs after therapy, suggesting stronger recruitment of CD8$^+$Tem cells than Tregs (Figure S4G).

DCs are the major antigen-presenting cells (APCs) that regulate T cell activation in the TME. Although DCs shrank to a lower level after therapy, cDC2-CD1C and cDC1-CLEC9A levels were much higher in the CR samples than in the NR samples (Figures S3A and S3B). cDC2-CD1C showed active signaling pathways of antigen processing and presentation and inflammatory response in GSEA, indicating immune activation (Figure S4H). Population abundance of cDC2 correlated with therapeutic outcome, confirming the positive effect of tumor restraining (Figure 4I). cDC2 was the major activator of T cells, as antigen processing and presentation in cDC2 were significantly correlated with T cell activation; however, this correlation was not observed for other myeloid cells (e.g., Macro_NLRP3, Macro_C1QC) (Figure S4I), confirming immune activation and regulation of cDC2. Considering the enrichment of cDC2 around pCAF subsets (Figure 4B), we investigated the potential interactions between pCAF subsets and DCs. The pronounced cytokine *DCN*, with higher expression

Figure 4. The immune microenvironment is modulated by the pCAF subsets following therapy

(A) The neighbor-joining tree presents the cellular modules of TME of SC data in all samples after therapy. Pairwise distance between cell types were measured by the Spearman correlation. Four cellular modules were identified as immune-active stromal cells (CM1-IASs), inactive immune cells (CM2-IAIs), immune-suppressive stromal cells (CM3-ISSs), and active-regulatory immune cells (CM4-ARIs).

(B) Spatial colocalization of CAF subsets and immune cells in ST data of the CR sample. "Counts" indicate the number of immune cells surrounding CAFs in less than a 3-bin range. C_{adj} represents the intensity of colocalization (see STAR Methods).

(C) Comparison of pathway enrichment among treatments and outcomes of CD8$^+$ T cells. Left: comparison between PO (blue) and PE (red) in the CR samples. Right: comparison between NR (blue) and CR (red) in the PO samples.

(D) Scatterplot shows the correlation between *CXCL12* expressed by the pCAF subsets and the TCR signaling pathway in the T cells after therapy.

(E) Violin plot of *CXCL12* expression in the CAF subsets.

(F) Scatterplot shows the correlation between *SLIT2* expressed by the pCAF subsets and the differentiation pathway in the T cells after therapy.

(G) Violin plot of SLIT2 expression in the CAF subsets.

(H) Spatial colocalization of CAF subsets expressing *CXCL12* and CD8$^+$Tem with active TCR pathway in the ST-CR1 sample. C_{adj} represents the intensity of colocalization between two cell sets. The blue triangle represents CAFs expressing *CXCL12*, and the red square represents CD8$^+$Tem with active TCR pathway. The color gradient indicates density of colocalized cells.

(I) The Ti of myeloid subsets. A positive Ti indicates the myeloid subset is associated with a better response, whereas a negative Ti indicates a worse response.

(J) Scatterplot presents the correlation between *DCN* expressed by pCAF subsets and the antigen presentation pathway in cDC2_CDC1.

(K) Spatial colocalization of CAF subsets expressing *DCN* and cDC2s with active antigen presentation pathway in the ST-CR1 sample. C_{adj} represents the intensity of colocalization between two cell sets. The blue triangle represents CAFs expressing *DCN*, and the red square represents cDC2 with an active antigen presentation pathway. The color gradient indicates density of colocalized cells.

in CAF_PI16 and CAF_SLIT2, strongly correlated with the antigen processing and presentation pathway in DCs (Figures 4J and S4J). *DCN* is the main component of the ECM, acting as a tumor suppressor and preventing tissue fibrosis, and studies have indicated that *DCN* is an effective candidate for reducing TGF-β bioavailability.[27,37,38] *DCN* prevents the binding of TGF-β to its receptor to enhance the function of major histocompatibility complex (MHC) genes, consistent with the positive correlation between *DCN* and the antigen processing and presentation pathway observed in our data. Furthermore, we observed strong colocalization of DCN^+pCAF subsets and cDC2 with active antigen processing and presentation in the CR sample and not in the NR sample (Figures 4K and S4K), supporting the regulation between pCAF subsets and cDC2 through *DCN*. This spatial colocalization of nCAF subsets was not as strong as that of pCAF subsets (Figures S4L and S4M). Both the correlation in SC data and the spatial colocalization in ST data confirmed the immunological regulation of pCAF subsets in T cells and DCs.

CAF_FAP enhances the tumor EMT program through MIR4435-2HG

Malignant epithelial cells in solid tumors show high cell heterogeneity and diverse abnormalities.[17,39] Malignant cells were distinguished from healthy epithelial cells by cluster analysis and CNV detection (Figures S5A and S5B). The size of the malignant cell population was consistent with the TRG determined by pathological examination (Figure S1B). Comparing the malignant cells that resisted therapy with those before therapy, we observed that malignant cells became quiescent after chemo inhibition, whereas malignant cells before therapy showed universally pronounced pathways such as phosphorylation, glycolysis oxidative, and E2F targets (Figure S5C). Comparing surviving malignant cells in CR and NR samples, the most pronounced pathways in NR samples were EMT, inflammatory response, MTORC1 signals, MYC targets, TNFA signals, KRAS signals, TGF-β signals, and NOTCH signals, which promote tumor progression (Figure S5C). The more strongly enriched pathways in the CR samples after therapy were antigen processing and presentation, which induced the ICD process and were consistent with their better therapeutic outcomes. Furthermore, we screened for ligands expressed at higher levels in NR samples than in CR samples after therapy, including *MIF*, *IL32*, *AREG*, *HBEGF*, and *ADAM9* and the chemokines *CXCL1*, *CXCL3*, *CXCL8*, and *CCL20* (Figure S5D). MIF attracts macrophages and regulates IL-2 induction to increase Tregs and inhibit $CD8^+$T cytotoxicity.[29,30] IL-32 induces the production of IL-6 and IL-1β and activates nuclear factor κB (NF-κB) signaling.[40,41] AREG mediates EMT in cancer cells.[42] Their receptors, including *CXCR4*, *CD74*, *CD9*, *SDC1/2*, and *ITGA/B*, which are involved in immune resistance and tumor progression and are correlated with worse outcomes, were expressed in macrophages, $CD4^+$ T cells, nCAFs, and ECs (Figure S5E).

To robustly identify tumor programs associated with clusters, functions, and clinical outcomes, we adopted a non-negative matrix factorization (NMF) procedure[43] and identified the 53 programs accounting for the variance in both healthy and malignant epithelia (see STAR Methods). Programs were hierarchically clustered into four modules (M1–M4) (Figure 5A), which were annotated as EMT and epithelial migration (M1), hypoxia response (M1), metal-ion response (M2), mitosis (M3), and T cell activation (M4) according to their gene characters. Comparing the module activity among CR, NR, and healthy epithelia, we observed that M1 was upregulated in NR malignant cells, whereas M2–M4 were enriched in healthy epithelia (Figures 5B and S5F), indicating that M1 associated with resisting tumors.

Because CAFs play important roles in EMT and ECM remodeling, associating with M1 in tumors, we investigated the connections between CAF and malignant cells. nCAF subsets (CAF_FAP and CAF_MMP1) were positively correlated with M1 activity, whereas pCAF subsets were negatively correlated (Figure 5C). Further investigation showed that CAFs were significantly synergistical with malignant cells in the EMT pathway, which accounts for tumor migration and invasion (Figure S5G). Comparison of EMT between CAFs and malignant cells indicated their different roles in EMT, as CAFs highly expressed collagen genes and, in contrast, malignant cells highly expressed *AREG*, a member of the epidermal growth factor (EGF) family (Figure S5H). Comparing the EMT pathway between NR and CR malignant cells also showed that AREG was one of the most pronounced factors in tumor EMT (Figure 5D). A significantly positive correlation confirmed the important role of AREG in EMT pathways in malignant cells (Figure 5E).

Considering that microRNAs (miRNAs) can bind targeted genes to regulate various tumor processes, such as proliferation and metastasis,[44–46] and that miRNAs can be transported to neighboring cells through extracellular vesicles,[47,48] we screened for miRNAs and identified *MIR4435-2HG*, which was highly expressed in CAF_FAPs (Figure 5F) and showed a significant positive correlation with *AREG* and EMT levels in malignant cells across the PO samples (Figures 5G, 5H, andS5I). The enhancement was consistent with previous findings showing that *MIR4435-2HG* promotes CRC growth and metastasis and that knockdown of MIR4435-2HG downregulates *YAP1*, *CTGF*, and *AREG* expression.[49] In addition, spatial proximity highlighted the interaction between CAF_FAPs and malignant cells (Figure 5I), which was significantly stronger than that between other CAFs, consistent with the correlation between CAF_FAP and M1 (Figure 5C). Malignant cells in a close neighborhood of CAF_FAP showed significantly higher EMT activity than cells at increasing distances (Figure S5J). Malignant cells approximating other CAF subsets showed lower EMT gene scores than CAF_FAP (Figure S5K). The spatial colocalization of $MIR4435\text{-}2HG^+$CAF_FAP and malignant cells with active EMT confirmed their interactions in NR samples (Figure 5J). Finally, malignant cells around $MIR4435\text{-}2HG^+$CAF_FAP showed higher EMT scores than those around $MIR4435\text{-}2HG^-$CAF_FAP, confirming the EMT upregulation of CAF_FAP through *MIR4435-2HG* (Figure S5L). This interaction was not observed in CR samples, as the population of CAF_FAP decreased after NAC and consequently reduced the EMT promotion (Figure 5J).

DISCUSSION

The advantages of this study include the following: (1) large-scale pairwise samples of patients with RC before and after NAC were used to directly compare the cellular and molecular

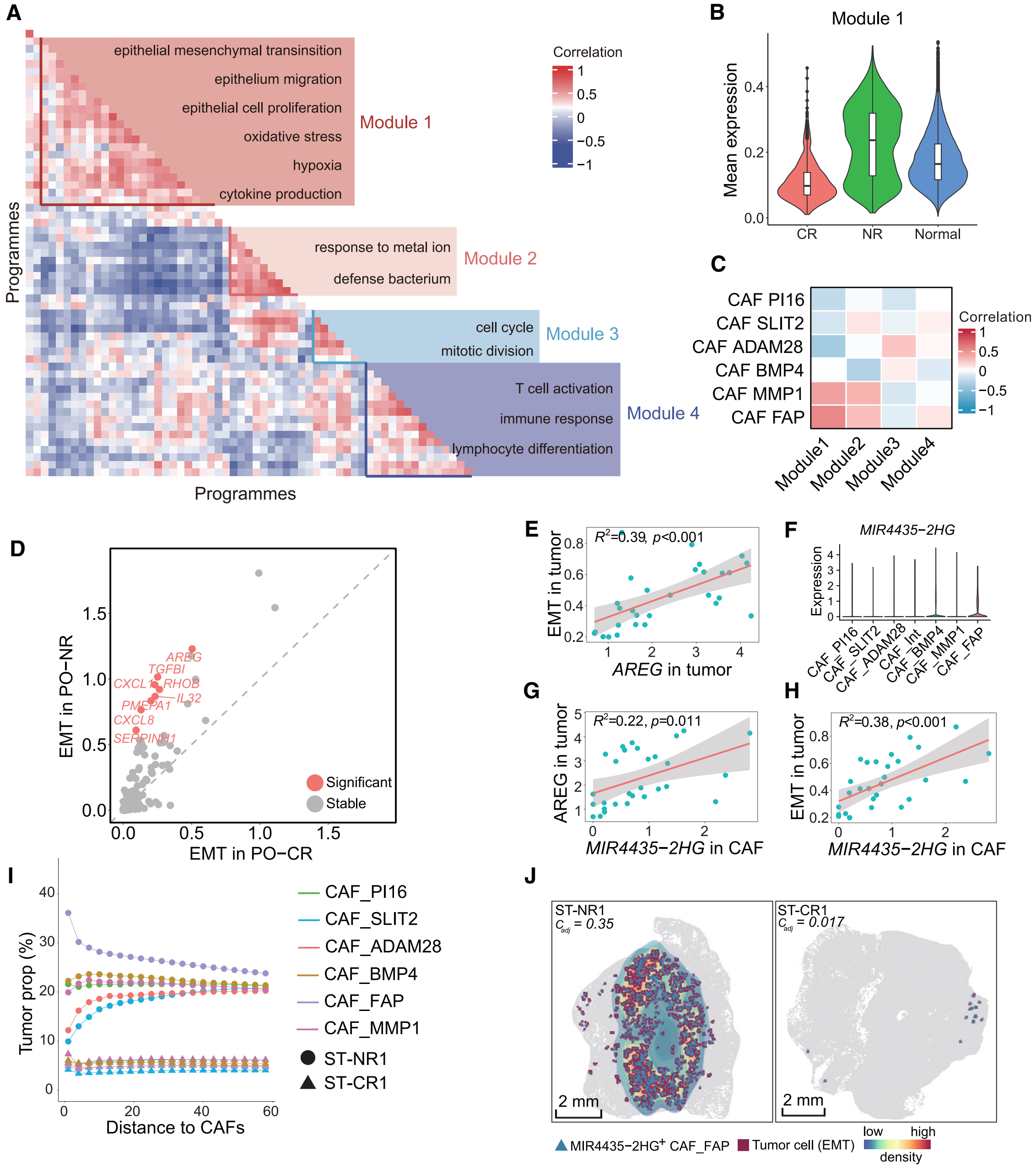


Figure 5. Malignant programs and the CAF regulation

(A) Program modules of both healthy and malignant epithelial cells. Modules were clustered from 53 expression programs identified by the NMF method. Correlations between programs were calculated following the Spearman method.

(B) Expression of module 1 in NR and CR malignant cells and healthy epithelial cells.

(C) Heatmap presents the correlation (Spearman method) between modules (averaged expression) of the malignant cells and the proportion of the CAF subsets.

(D) Differentially expressed genes of the EMT pathway between malignant cells of CR samples and NR samples after therapy. Red dot: differential genes with $p < 0.05$ from Wilcoxon rank-sum test and log2(fold change) >0.7.

(E) Scatterplot shows the correlation between *AREG* expression and EMT pathway in the tumor cells after therapy.

(F) Violin plot of *MIR4435-2HG* expression in the CAF subsets.

(G) Scatterplot shows the correlation between *MIR4435-2HG* expressed by the CAFs and AREG expressed by the tumor cells after therapy.

(H) Scatterplot shows the correlation between *MIR4435-2HG* expressed by the CAFs and EMT pathway in the tumor cells after therapy.

(I) Spatial proximity between the tumor cells and the CAF subsets. The x axis represents the different radius centered on the CAF cells. The y axis represents the probability of tumor cells presenting in the neighborhood of the CAF cells.

(J) Spatial colocalization of CAF_FAP expressing *MIR4435-2HG* and the tumor cells with active EMT pathway in ST samples. C_{adj} represents the intensity of colocalization between the two cell sets. The blue triangle represents CAF_FAP expressing *MIR4435-2HG*, and the red square represents tumor cells with active EMT pathway. The color gradient indicates density of colocalization cells.

differences between PE and PO samples and between CR and NR samples. (2) The comprehensive integration of SC and ST data with the fine resolution provided a clear viewing field and robust confirmation of cell composition, proximity, interaction, and signaling pathways. Considering these advantages, our study provides a valuable single-cell and spatial transcriptome landscape of RC following NAC. Systematically, we investigated the TME alterations modulated by therapy and associated cellular composition with the therapeutic response. We demonstrated that NAC restrained tumor progression and regulated the TME by remodeling CAFs. In addition, the findings on the tumor-promoting mode of CAF_FAP through miRNAs will have important implications for future investigations and tumor therapeutic strategies.

Our study has suggested two coinciding roles for CAFs in tumor promotion and restraint. CAFs are highly heterogeneous in types and functions affecting cancer progression and immunomodulation. Several studies have reported immunosuppression by CAFs; however, the mechanism underlying immune activation remains unclear. In this study, we revealed the complexity of CAF subsets and the dual roles of tumor promoters and immune activators. Several CAF subsets contained numerous chemokines, collagens, and MMPs related to tumor-supporting functions, such as angiogenesis, ECM remodeling, EMT, and tumor invasion. We observed that CAF_FAP enhanced tumor EMT through MIR4435-2HG associating with worse NAC outcomes. In contrast, pCAF subsets, including CAF_PI16 and CAF_SLIT2, were substantially expanded in the responsive group and exhibited immune-promoting properties by enhancing immune cell recruitment and activation through multiple cytokines. Several cell-cell interactions were confirmed through correlation analysis and spatial colocalization. CXCL12 helped recruit CD8$^+$Tem cells instead of Tregs, as stronger receptor expression was observed in CD8$^+$Tem (Figure S4G). DCN, as a TGF-β blocker, upregulated antigen processing and presentation of DCs, subsequently activating T cells spatially. These findings have several implications for tumor interference through CAF modulation, such as altering the composition of CAF subsets in the TME, blocking tumor-promoting cytokines or molecules in nCAF subsets, and enhancing the immunoactivation of pCAF subsets.

Analytical strategies were adapted to satisfy the complex requirements of large-scale SC data and their integration with ST data. One of them is to integrate the analysis of cell-cell interactions for both SC and ST data. False positives in cell-cell interactions overflowed in SC data analyses because of poor consideration of downstream pathways and a lack of sample repeats. In this study, cell-cell interactions were estimated between a pair of cell subsets by systematically calculating the correlation between ligands from one subset and pathways from another, with a background control to avoid systematic inflation due to cell population correlations (see STAR Methods). Correlation computations were performed for numerous samples. Moreover, the spatial neighborhood of interactive cells was examined, which has advantages over immunohistochemistry (IHC) and immunofluorescence (IF). IHC and IF use limiting antibodies to provide visual details of certain locations. In contrast, ST considers whole transcriptomes and signals that can be conveniently combined from any gene or pathway.

PI16$^+$ fibroblasts are considered progenitors of CAF differentiation,[22] consistent with the observation of highly expressed gene related to stemness (*CD34*) in our data (Figure 2B). Comparison of the trajectories in CR and NR samples revealed differences in CAF differentiation between the CR and NR samples, with the CAF_ADAM28-CAF_BMP4 lineage in NR samples and the major lineage of CAF_PI16-CAF_SLIT2-CAF_BMP4 in CR samples (Figure S2C). Using ST data, which provide the spatial distance between cells, we confirmed multiple lineages with more spatial regional details of differentiation (Figure S2D). The diverse trajectories of the CR and NR samples suggested that distinct CAF differentiation was associated with NAC responses and outcomes. Therefore, different strategies should be considered when interfering with differentiation.

Limitations of the study

There are several limitations of the data analysis and the findings in our study. First, the associations between the CAF cytokines and immune pathways will not conclude the direct immunoregulation of remodeled CAFs through these cytokines. The cause-effect relationships need further investigation by *in vitro* and *in vivo* experiments. Second, we observed the connection between PG and CAF dynamics. But statistical analysis and trajectory inference are incapable of uncovering the process of dedifferentiation of myofibroblasts as reported by previous study.[33,34] Therefore, cell experiment is recommended. Moreover, ST data were produced using Stereo-seq technology, which provides nanoscale resolution. However, it was difficult to segment cells of tumor tissue and assemble the exact transcripts of single cells. We therefore merged 50 × 50 spots into a single informative “bin” as the minimum unit (25 × 25 μm square), representing one cell (see STAR Methods). Future development of the technology will provide more accurate transcripts of single cells.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - ○ Human subjects
- METHOD DETAILS
  - ○ Single-cell isolation
  - ○ Library construction for single cell RNA sequencing
  - ○ Single-cell RNA sequencing using DNBelab C4
  - ○ Sample preparation for spatial transcriptomic sequencing
  - ○ Spatial transcriptomic sequencing using Stereo-seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Raw data processing for single-cell RNA sequencing
  - ○ Raw data processing of single-cell spatial transcriptome from Stereo-seq

- ○ Dimension reduction, clustering, and identification of differentially expressed genes of SC data
- ○ Identification of transcriptional programmes and gene modules using NMF
- ○ Cell type identification and deconvolution of ST data
- ○ Measurements of cell colocalisation and proximity in ST data
- ○ Trajectory inference of cell differentiation
- ○ Cell-cell interaction analysis
- ○ Gene set enrichment analysis
- ○ Definition of therapeutic index
- ○ Survival analysis
- ○ Correlation analysis
- ○ Identification of malignant cells

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2023.101231.

ACKNOWLEDGMENTS

This study was supported by fundings from the Natural Science Foundation of Guangdong Province for Distinguished Young Scholars (no. 2022B1515020003); the National Natural Science Foundation of China (nos. 82174369 and 81973847); the Natural Science Foundation of Guangdong Province (no. 2020A1515011254); the Guangzhou Clinical High-Tech Project (no. 2023P-GX09); the Bethune Charitable Foundation (no. zllcaxw-14); the Open Project of BGI-Shenzhen, Shenzhen 518000, China (no. BGIRSZ20210005); the National Key Research and Development Program of China (nos. 2021YFA0805100 and 2021YFC2501900); and the Guangdong Basic and Applied Basic Research Foundation (no. 2021A1515110832). This work was supported by the China National GeneBank (CNGB) and the National Key Clinical Discipline.

AUTHOR CONTRIBUTIONS

Conceptualization, H.L., S.L., J.W., L. Wu, and D.R.; methodology, P.Q., H.C., Y.W., L.H., and K.H.; investigation, P.Q., H.C., Y.W., L.H., K.H., G.X., C.H., J.H., D.L., X.W., Y.Z., Y.L., G.L., H.Y., S.Y., M.L., Y.F., H.X., L. Wen, Z.G., X.S., Z.L., C.W., X.C., L. Wang, and L.S.; resources, L.H., J.H., D.L., Y.L., and S.Y.; writing – original draft, P.Q., H.C., Y.W., L.H., and K.H.; writing – review & editing, H.L., S.L., J.W., L. Wu, and D.R.; funding acquisition, H.L., S.L., and L. Wu; supervision, H.L., S.L., J.W., L. Wu, and D.R.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 30, 2023
Revised: July 17, 2023
Accepted: September 14, 2023
Published: October 17, 2023

REFERENCES

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA. Cancer J. Clin. *71*, 209–249. https://doi.org/10.3322/caac.21660.
2. Siegel, R., Desantis, C., and Jemal, A. (2014). Colorectal cancer statistics, 2014. CA. Cancer J. Clin. *64*, 104–117. https://doi.org/10.3322/caac.21220.
3. Deng, Y. (2017). Rectal Cancer in Asian vs. Western Countries: Why the Variation in Incidence? Curr. Treat. Options Oncol. *18*, 64. https://doi.org/10.1007/s11864-017-0500-2.
4. Dou, R., He, S., Deng, Y., and Wang, J. (2021). Comparison of guidelines on rectal cancer: exception proves the rule? Gastroenterol. Rep. *9*, 290–298. https://doi.org/10.1093/gastro/goab034.
5. Deng, Y., Chi, P., Lan, P., Wang, L., Chen, W., Cui, L., Chen, D., Cao, J., Wei, H., Peng, X., et al. (2019). Neoadjuvant Modified FOLFOX6 With or Without Radiation Versus Fluorouracil Plus Radiation for Locally Advanced Rectal Cancer: Final Results of the Chinese FOWARC Trial. J. Clin. Oncol. *37*, 3223–3233. https://doi.org/10.1200/jco.18.02309.
6. Schrag, D., Shi, Q., Weiser, M.R., Gollub, M.J., Saltz, L.B., Musher, B.L., Goldberg, J., Al Baghdadi, T., Goodman, K.A., McWilliams, R.R., et al. (2023). Preoperative Treatment of Locally Advanced Rectal Cancer. N. Engl. J. Med. *389*, 322–334. https://doi.org/10.1056/NEJMoa2303269.
7. Zhang, J., Chi, P., Lan, P., Cui, L., Wei, H., Zhao, R., Huang, Z., Zhang, H., Cai, Y., Wang, J., and Deng, Y. (2023). Long-term outcome of neoadjuvant mFOLFOX6 with or without radiation versus fluorouracil plus radiation for locally advanced rectal cancer: A multicenter, randomized phase III trial. J. Clin. Oncol. *41*, 3505. https://doi.org/10.1200/JCO.2023.41.16_suppl.3505.
8. Hu, H., Kang, L., Zhang, J., Wu, Z., Wang, H., Huang, M., Lan, P., Wu, X., Wang, C., Cao, W., et al. (2022). Neoadjuvant PD-1 blockade with toripalimab, with or without celecoxib, in mismatch repair-deficient or microsatellite instability-high, locally advanced, colorectal cancer (PICC): a single-centre, parallel-group, non-comparative, randomised, phase 2 trial. Lancet. Gastroenterol. Hepatol. *7*, 38–48. https://doi.org/10.1016/s2468-1253(21)00348-4.
9. Cercek, A., Lumish, M., Sinopoli, J., Weiss, J., Shia, J., Lamendola-Essel, M., El Dika, I.H., Segal, N., Shcherba, M., Sugarman, R., et al. (2022). PD-1 Blockade in Mismatch Repair-Deficient, Locally Advanced Rectal Cancer. N. Engl. J. Med. *386*, 2363–2376. https://doi.org/10.1056/NEJMoa2201445.
10. Boland, C.R., and Goel, A. (2010). Microsatellite instability in colorectal cancer. Gastroenterology *138*, 2073–2087.e3. https://doi.org/10.1053/j.gastro.2009.12.064.
11. Vilar, E., and Gruber, S.B. (2010). Microsatellite instability in colorectal cancer-the stable evidence. Nat. Rev. Clin. Oncol. *7*, 153–162. https://doi.org/10.1038/nrclinonc.2009.237.
12. Guo, Y., Wang, M., Zou, Y., Jin, L., Zhao, Z., Liu, Q., Wang, S., and Li, J. (2022). Mechanisms of chemotherapeutic resistance and the application of targeted nanoparticles for enhanced chemotherapy in colorectal cancer. J. Nanobiotechnology *20*, 371. https://doi.org/10.1186/s12951-022-01586-4.
13. Galluzzi, L., Buqué, A., Kepp, O., Zitvogel, L., and Kroemer, G. (2015). Immunological Effects of Conventional Chemotherapy and Targeted Anticancer Agents. Cancer Cell *28*, 690–714. https://doi.org/10.1016/j.ccell.2015.10.012.
14. Bracci, L., Schiavoni, G., Sistigu, A., and Belardelli, F. (2014). Immune-based mechanisms of cytotoxic chemotherapy: implications for the design of novel and rationale-based combined treatments against cancer. Cell Death Differ. *21*, 15–25. https://doi.org/10.1038/cdd.2013.67.
15. Opzoomer, J.W., Sosnowska, D., Anstee, J.E., Spicer, J.F., and Arnold, J.N. (2019). Cytotoxic Chemotherapy as an Immune Stimulus: A Molecular Perspective on Turning Up the Immunological Heat on Cancer. Front. Immunol. *10*, 1654. https://doi.org/10.3389/fimmu.2019.01654.
16. Öhlund, D., Handly-Santana, A., Biffi, G., Elyada, E., Almeida, A.S., Ponz-Sarvise, M., Corbo, V., Oni, T.E., Hearn, S.A., Lee, E.J., et al. (2017). Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. J. Exp. Med. *214*, 579–596. https://doi.org/10.1084/jem.20162024.
17. Pelka, K., Hofree, M., Chen, J.H., Sarkizova, S., Pirl, J.D., Jorgji, V., Bejnood, A., Dionne, D., Ge, W.H., Xu, K.H., et al. (2021). Spatially organized multicellular immune hubs in human colorectal cancer. Cell *184*, 4734–4752.e20. https://doi.org/10.1016/j.cell.2021.08.003.

18. Righetti, A., Giulietti, M., Šabanović, B., Occhipinti, G., Principato, G., and Piva, F. (2019). CXCL12 and Its Isoforms: Different Roles in Pancreatic Cancer? J. Oncol. *2019*, 9681698. https://doi.org/10.1155/2019/9681698.
19. Zhang, Y., Guan, X.Y., and Jiang, P. (2020). Cytokine and Chemokine Signals of T-Cell Exclusion in Tumors. Front. Immunol. *11*, 594609. https://doi.org/10.3389/fimmu.2020.594609.
20. Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., et al. (2022). Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell *185*, 1777–1792.e21. https://doi.org/10.1016/j.cell.2022.04.003.
21. Amin, M.B., Edge, S.B., Greene, F.L., Byrd, D.R., Brookland, R.K., Washington, M.K., Gershenwald, J.E., Compton, C.C., Hess, K.R., and Sullivan, D.C. (2017). AJCC Cancer Staging Manual, Eight edition (Springer).
22. Buechler, M.B., Pradhan, R.N., Krishnamurty, A.T., Cox, C., Calviello, A.K., Wang, A.W., Yang, Y.A., Tam, L., Caothien, R., Roose-Girma, M., et al. (2021). Cross-tissue organization of the fibroblast lineage. Nature *593*, 575–579. https://doi.org/10.1038/s41586-021-03549-5.
23. Pilling, D., Zheng, Z., Vakil, V., and Gomer, R.H. (2014). Fibroblasts secrete Slit2 to inhibit fibrocyte differentiation and fibrosis. Proc. Natl. Acad. Sci. USA *111*, 18291–18296. https://doi.org/10.1073/pnas.1417426112.
24. Chen, K., Wang, Q., Li, M., Guo, H., Liu, W., Wang, F., Tian, X., and Yang, Y. (2021). Single-cell RNA-seq reveals dynamic change in tumor microenvironment during pancreatic ductal adenocarcinoma malignant progression. EBioMedicine *66*, 103315. https://doi.org/10.1016/j.ebiom.2021.103315.
25. Kessenbrock, K., Plaks, V., and Werb, Z. (2010). Matrix metalloproteinases: regulators of the tumor microenvironment. Cell *141*, 52–67. https://doi.org/10.1016/j.cell.2010.03.015.
26. Perugorria, M.J., Olaizola, P., Labiano, I., Esparza-Baquer, A., Marzioni, M., Marin, J.J.G., Bujanda, L., and Banales, J.M. (2019). Wnt-β-catenin signalling in liver development, health and disease. Nat. Rev. Gastroenterol. Hepatol. *16*, 121–136. https://doi.org/10.1038/s41575-018-0075-9.
27. Zhang, W., Ge, Y., Cheng, Q., Zhang, Q., Fang, L., and Zheng, J. (2018). Decorin is a pivotal effector in the extracellular matrix and tumour microenvironment. Oncotarget *9*, 5480–5491. https://doi.org/10.18632/oncotarget.23869.
28. Zhou, C., Gao, Y., Ding, P., Wu, T., and Ji, G. (2023). The role of CXCL family members in different diseases. Cell Death Discov. *9*, 212. https://doi.org/10.1038/s41420-023-01524-9.
29. Guda, M.R., Rashid, M.A., Asuthkar, S., Jalasutram, A., Caniglia, J.L., Tsung, A.J., and Velpula, K.K. (2019). Pleiotropic role of macrophage migration inhibitory factor in cancer. Am. J. Cancer Res. *9*, 2760–2773.
30. Noe, J.T., and Mitchell, R.A. (2020). MIF-Dependent Control of Tumor Immunity. Front. Immunol. *11*, 609948. https://doi.org/10.3389/fimmu.2020.609948.
31. Raje, N.S., Bhatta, S., and Terpos, E. (2019). Role of the RANK/RANKL Pathway in Multiple Myeloma. Clin. Cancer Res. *25*, 12–20. https://doi.org/10.1158/1078-0432.Ccr-18-1537.
32. Li, M.O., Wan, Y.Y., Sanjabi, S., Robertson, A.K.L., and Flavell, R.A. (2006). Transforming growth factor-beta regulation of immune responses. Annu. Rev. Immunol. *24*, 99–146. https://doi.org/10.1146/annurev.immunol.24.021605.090737.
33. Garrison, G., Huang, S.K., Okunishi, K., Scott, J.P., Kumar Penke, L.R., Scruggs, A.M., and Peters-Golden, M. (2013). Reversal of myofibroblast differentiation by prostaglandin E(2). Am. J. Respir. Cell Mol. Biol. *48*, 550–558. https://doi.org/10.1165/rcmb.2012-0262OC.
34. Wettlaufer, S.H., Scott, J.P., McEachin, R.C., Peters-Golden, M., and Huang, S.K. (2016). Reversal of the Transcriptome by Prostaglandin E2 during Myofibroblast Dedifferentiation. Am. J. Respir. Cell Mol. Biol. *54*, 114–127. https://doi.org/10.1165/rcmb.2014-0468OC.
35. Matsunuma, S., Handa, S., Kamei, D., Yamamoto, H., Okuyama, K., and Kato, Y. (2019). Oxaliplatin induces prostaglandin E2 release in vascular endothelial cells. Cancer Chemother. Pharmacol. *84*, 345–350. https://doi.org/10.1007/s00280-019-03901-7.
36. Altorki, N.K., Port, J.L., Zhang, F., Golijanin, D., Thaler, H.T., Duffield-Lillico, A.J., Subbaramaiah, K., and Dannenberg, A.J. (2005). Chemotherapy induces the expression of cyclooxygenase-2 in non-small cell lung cancer. Clin. Cancer Res. *11*, 4191–4197. https://doi.org/10.1158/1078-0432.Ccr-05-0108.
37. Järvinen, T.A.H., and Prince, S. (2015). Decorin: A Growth Factor Antagonist for Tumor Growth Inhibition. BioMed Res. Int. *2015*, 654765. https://doi.org/10.1155/2015/654765.
38. Hu, X., Villodre, E.S., Larson, R., Rahal, O.M., Wang, X., Gong, Y., Song, J., Krishnamurthy, S., Ueno, N.T., Tripathy, D., et al. (2021). Decorin-mediated suppression of tumorigenesis, invasion, and metastasis in inflammatory breast cancer. Commun. Biol. *4*, 72. https://doi.org/10.1038/s42003-020-01590-0.
39. Joanito, I., Wirapati, P., Zhao, N., Nawaz, Z., Yeo, G., Lee, F., Eng, C.L.P., Macalinao, D.C., Kahraman, M., Srinivasan, H., et al. (2022). Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. Nat. Genet. *54*, 963–975. https://doi.org/10.1038/s41588-022-01100-4.
40. Hong, J.T., Son, D.J., Lee, C.K., Yoon, D.Y., Lee, D.H., and Park, M.H. (2017). Interleukin 32, inflammation and cancer. Pharmacol. Ther. *174*, 127–137. https://doi.org/10.1016/j.pharmthera.2017.02.025.
41. Xin, T., Chen, M., Duan, L., Xu, Y., and Gao, P. (2018). Interleukin-32: its role in asthma and potential as a therapeutic agent. Respir. Res. *19*, 124. https://doi.org/10.1186/s12931-018-0832-x.
42. Wang, L., Wang, L., Zhang, H., Lu, J., Zhang, Z., Wu, H., and Liang, Z. (2020). AREG mediates the epithelial-mesenchymal transition in pancreatic cancer cells via the EGFR/ERK/NF-κB signalling pathway. Oncol. Rep. *43*, 1558–1568. https://doi.org/10.3892/or.2020.7523.
43. Kotliar, D., Veres, A., Nagy, M.A., Tabrizi, S., Hodis, E., Melton, D.A., and Sabeti, P.C. (2019). Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. Elife *8*, e43803. https://doi.org/10.7554/eLife.43803.
44. Han, C., Li, H., Ma, Z., Dong, G., Wang, Q., Wang, S., Fang, P., Li, X., Chen, H., Liu, T., et al. (2021). MIR99AHG is a noncoding tumor suppressor gene in lung adenocarcinoma. Cell Death Dis. *12*, 424. https://doi.org/10.1038/s41419-021-03715-7.
45. Massillo, C., Dalton, G.N., Farré, P.L., De Luca, P., and De Siervi, A. (2017). Implications of microRNA dysregulation in the development of prostate cancer. Reproduction *154*, R81–R97. https://doi.org/10.1530/rep-17-0322.
46. Zhang, L., Liao, Y., and Tang, L. (2019). MicroRNA-34 family: a potential tumor suppressor and therapeutic candidate in cancer. J. Exp. Clin. Cancer Res. *38*, 53. https://doi.org/10.1186/s13046-019-1059-5.
47. Garcia-Martin, R., Wang, G., Brandão, B.B., Zanotto, T.M., Shah, S., Kumar Patel, S., Schilling, B., and Kahn, C.R. (2022). MicroRNA sequence codes for small extracellular vesicle release and cellular retention. Nature *601*, 446–451. https://doi.org/10.1038/s41586-021-04234-3.
48. Sun, I.O., Bae, Y.U., Lee, H., Kim, H., Jeon, J.S., Noh, H., Choi, J.S., Doh, K.O., and Kwon, S.H. (2022). Circulating miRNAs in extracellular vesicles related to treatment response in patients with idiopathic membranous nephropathy. J. Transl. Med. *20*, 224. https://doi.org/10.1186/s12967-022-03430-7.
49. Dong, X., Yang, Z., Yang, H., Li, D., and Qiu, X. (2020). Long Non-coding RNA MIR4435-2HG Promotes Colorectal Cancer Proliferation and Metastasis Through miR-206/YAP1 Axis. Front. Oncol. *10*, 160. https://doi.org/10.3389/fonc.2020.00160.
50. Van Rossum, G., and Drake, F.L. (2009). Python 3 Reference Manual (Valley, CA: CreateSpace: Scotts).
51. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019).

Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

52. Tickle, T., Tirosh, I., Georgescu, C., Brown, M., and Haas, B. (2019). inferCNV of the Trinity CTAT Project (Klarman Cell Observatory, Broad Institute of MIT and Harvard).
53. Kolde, R. (2019). pheatmap: Pretty Heatmaps (R package version 1.0.12).
54. Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2020). dplyr: A Grammar of Data Manipulation.
55. Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinf. *11*, 367. https://doi.org/10.1186/1471-2105-11-367.
56. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). lusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.
57. Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., and Dunnington, D. (2023). ggplot2: Create elegant data visualisations using the grammar of graphics.
58. Kassambara, A. (2022). Ggpubr: ‘Ggplot2’ Based Publication Ready Plots. https://cran.r-project.org/web/packages/ggpubr.
59. Wei, T., Simko, V. R., Levy, M., Xie, Y., Jin, Y., and Zemla, J. (2021). package “corrplot”: Visualization of a Correlation Matrix. 2017. Version 0.84. https://cran.r-project.org/package=corrplot.
60. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.
61. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Systems *8*, 329–337.e4. https://doi.org/10.1016/j.cels.2019.03.003.
62. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using. CellChat. Nat Commun *12*, 1088. https://doi.org/10.1038/s41467-021-21246-9.
63. Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc *15*, 1484–1506. https://doi.org/10.1038/s41596-020-0292-x.
64. Bonnardel, J., T’Jonck, W., Gaublomme, D., Browaeys, R., Scott, C.L., Martens, L., Vanneste, B., De Prijck, S., Nedospasov, S.A., Kremer, A., et al. (2019). Stellate Cells, Hepatocytes, and Endothelial Cells Imprint the Kupffer Cell Identity on Monocytes Colonizing the Liver Macrophage Niche. Immunity *51*, 638–654.e9. https://doi.org/10.1016/j.immuni.2019.08.017.
65. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I., and Heyn, H. (2021). SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. Nucleic Acids Res. *49*, e50. https://doi.org/10.1093/nar/gkab043.
66. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.
67. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740. https://doi.org/10.1093/bioinformatics/btr260.
68. Cang, Z., Zhao, Y., Almet, A.A., Stabell, A., Ramos, R., Plikus, M.V., Atwood, S.X., and Nie, Q. (2023). Screening cell-cell communication in spatial transcriptomics via collective optimal transport. Nat. Methods *20*, 218–228. https://doi.org/10.1038/s41592-022-01728-4.
69. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982. https://doi.org/10.1038/nmeth.4402.
70. Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genom. *19*, 477. https://doi.org/10.1186/s12864-018-4772-0.
71. Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. *20*, 59. https://doi.org/10.1186/s13059-019-1663-x.
72. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414. https://doi.org/10.1038/s41587-020-0591-3.
73. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498. https://doi.org/10.1038/s41586-018-0414-6.
74. Qiu, X., Zhang, Y., Martin-Rufino, J.D., Weng, C., Hosseinzadeh, S., Yang, D., Pogson, A.N., Hein, M.Y., Hoi Joseph Min, K., Wang, L., et al. (2022). Mapping transcriptomic vector fields of single cells. Cell *185*, 690–711.e45. https://doi.org/10.1016/j.cell.2021.12.045.
75. Zhang, Y., Chen, H., Mo, H., Hu, X., Gao, R., Zhao, Y., Liu, B., Niu, L., Sun, X., Yu, X., et al. (2021). Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. Cancer Cell *39*, 1578–1593.e8. https://doi.org/10.1016/j.ccell.2021.09.010.
76. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401. https://doi.org/10.1126/science.1254257.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
APC-Cyanine7 Anti-Human CD45	Tonbo Biosciences	Cat# 25-0459-T100; RRID:AB_2621631
Chemicals, peptides, and recombinant proteins		
7-AAD	Abam	Cat# ab228563
RPMI 1640	Thermo Fisher	Cat# C11875500BT
Red Blood Cell Lysis Buffer	Beyotime	Cat# C3702-500mL
FICOLL PAQUE PLUS	GE hyclone	Cat# 17-1440-03
DPBS	HyClone	Cat# SH30028.02
FBS	Gibco	Cat# 10270-106
Nuclease-Free Water	Monad	Cat# MR80101S
Tumor Dissociation Kit, human	Miltenyi	Cat# 130-095-929
Trypan Blue	Sigma	Cat# T8154-100ML
20×SSC	Ginbio	Cat# GB20212-500
Cryo Embedding Medium	biosharp	Cat# BL557A
Dead Cell Removal Kit	Miltenyi	Cat# 130-090-101
Bovine serum albumin, fraction V, heat shock isolation	Sangon biotech	Cat# A600332-0025
20× Saline sodium citrate (SSC)	Thermo Fisher	Cat# AM9770
RNase inhibitor	NEB	Cat# M0314L
Qubit^{TM} dsDNA kit	Invitrogen	Cat#Q32854
Qubit^{TM} ssDNA kit	Thermo Fisher	Cat#Q10212
Deposited data		
Transcriptome expression	This paper	CNGBdb: CNP0004138
Bulk RNA-seq datasets from TCGA (Pan-Cancer Atlas Hub of UCSC Xena database)	https://xenabrowser.net/datapages/?cohort=TCGA%20Pan-Cancer%20(PANCAN)&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443	Batch effects normalized mRNA data (ID: EB++AdjustPANCAN_IlluminaHiSeq_RNASeqV2.geneExp.xena); Curated clinical data (ID: Survival_SupplementalTable_S1_20171025_xena_sp)
Software and algorithms		
R-4.1.2	R Core Team (2021)	https://www.r-project.org/
Python 3.9.12	Van Rossum et al.[50]	https://www.python.org/
spaTrack	Github	https://github.com/yzf072/spaTrack
RStudio	RStudio (2015)	https://rstudio.com/
Seurat v4.1.1	Stuart et al.[51]	https://github.com/satijalab/seurat
inferCNV	Tickle et al.[52]	https://github.com/broadinstitute/inferCNV
pheatmap	Kolde et al.[53]	https://cran.r-project.org/web/packages/pheatmap
dyplr	Wickham et al.[54]	https://cran.r-project.org/web/packages/dplyr
NMF	Gaujoux and Seoighe[55]	https://cran.r-project.org/web/packages/NMF
clusterProfiler	Wu et al.[56]	https://github.com/YuLab-SMU/clusterProfiler
ggplot2	Wickham et al.[57]	https://cran.r-project.org/web/packages/ggplot2
ggpubr	Kassambara et al.[58]	https://cran.r-project.org/web/packages/ggpubr
corrplot	Wei et al.[59]	https://cran.r-project.org/web/packages/corrplot
harmony	Korsunsky et al.[60]	https://cran.r-project.org/web/packages/harmony
DoubletFinder	McGinnis et al.[61]	https://github.com/chris-mcginnis-ucsf/DoubletFinder
CellChat	Jin et al.[62]	https://github.com/sqjin/CellChat
CellphoneDB	Efremova et al.[63]	https://github.com/ventolab/CellphoneDB
Nichenet	Bonnardel et al.[64]	https://github.com/saeyslab/nichenetr

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SPOTlight	Elosua-Bayes et al.[65]	https://github.com/MarcElosua/SPOTlight
MSigDB	Subramanian et al.[66] and Liberzon et al.[67]	https://www.gsea-msigdb.org/gsea/msigdb/index.jsp
Biorender	Biorender	https://www.biorender.com/
Other		
70-μm cell strainer	Biologix	Cat# 15-1070
FACS tube	Corning	Cat# 352054-125EA
Cryostat	Leica	Cat# CM1950
Stereo-seq chips and apparatus	BGI Shenzhen	–
KAPA Hotstart Ready Mix	Roche	Cat# KK2602
SuperScript reverse transcription (RT) mix	Invitrogen	Cat# 18064-014
VAHTS DNA clean beads	Vazyme	Cat# N411-03

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and fulfilled by the lead contact, Hongcheng Lin (lhcheng@mail.sysu.edu.cn).

Materials availability

This study did not generate new or unique reagents.

Data and code availability

The processed expression data of SC and ST reported in this study can be obtained from the China National GeneBank Database (CNGBdb) with accession number CNP0004138. Bulk RNA-seq datasets of TCGA were obtained from Pan-Cancer Atlas Hub of UCSC Xena database (https://xenabrowser.net/datapages/?cohort=TCGA%20Pan-Cancer%20(PANCAN)&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443), from which the "Batch effects normalized mRNA data" (dataset ID: EB++AdjustPANCAN_IlluminaHiSeq_RNASeqV2.geneExp.xena) was used as the gene expression data, and "Curated clinical data" (dataset ID: Survival_SupplementalTable_S1_20171025_xena_sp) was used the clinical phenotype data. The raw FASTQ files of single cell transcriptomes in this study will be provided for scientific research upon reasonable request, in compliance with the law, due to human patient privacy concerns. This paper does not report custom code. Any additional information required to reanalyse the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human subjects

Twenty-nine RC patients with pMMR/MSS were enrolled in this study, who were pathologically diagnosed with locally advanced rectal adenocarcinoma at the Sixth Affiliated Hospital of Sun Yat-sen University. The clinical and pathological characteristics of the patients are listed in Table S1. These patients routinely underwent colonoscopy and biopsy. Fresh tumor tissues were collected through a colonoscopic biopsy <2 weeks before NAC in 27 patients (two samples were excluded due to low quality). After NAC, the paired freshly tumor tissues were obtained through surgical resection (27 patients) and endoscopic biopsy (2 patients). In addition, five adjacent normal mucosa tissue sample (located >3 cm from the tumor) were freshly collected from five patients. Informed consent was obtained from all patients enrolled in the study. All sampling and experimental procedures were approved by the Ethics Committee of the Sixth Affiliated Hospital of Sun Yat-sen University (license no. 2021ZSLYEC-033 and 2022ZSLYEC-165).

METHOD DETAILS

Single-cell isolation

Fresh tumor and adjacent normal tissue samples were cut into approximately 1 mm^3 pieces in the Roswell Park Memorial Institute (RPMI-1640 medium (Invitrogen) and then enzymatically digested using the MACS human tumor dissociation kit (Miltenyi Biotec) for 60 min on the gentle MACS Octo dissociator at 37°C. The sample was filtered using a 70 mm Cell-Strainer (Biologix) in RPMI-1640 medium (Invitrogen), and the suspended cells were centrifuged at 300 × g for 7 min. After removing the supernatant, the pelleted cells were suspended in red blood cell lysis buffer (Beyotime) and incubated at 4°C for 10 min to lyse red blood cells. The cell pellets were re-suspended in washing buffer (0.04% bovine serum albumin [BSA] in Dulbecco's phosphate-buffered saline [DPBS]) after washing two times with washing buffer. Viability was confirmed to be >80% in all samples by trypan blue staining (Sigma). Finally, for the samples obtained before NAC, the single-cell suspension was centrifuged at 300 × g for 10 min, re-suspended in a cryoprotectant (90% fetal bovine serum [FBS] with 10% dimethyl sulfoxide [DMSO]), cryopreserved in cryotubes (Corning), and stored in liquid nitrogen tanks. For samples obtained after NAC, the cell suspensions were maintained on ice to prepare for sequencing.

Library construction for single cell RNA sequencing

The SC libraries were constructed according to the instructions of the DNA Nanoball (DNB) elab C4 scRNA Preparation Kit. Briefly, the cells were diluted to a concentration of 1,000 cells/μL and loaded into the cell reservoir of the microfluidic chip. Barcoded beads and droplet-generation oil were successively added to the beads and oil reservoirs. Encapsulated droplets were generated and collected using a DNBelab C4 system. Beads that captured the mRNA were recovered for reverse transcription (RT). Following polymerase chain reaction (PCR) amplification by polymerase chain reaction, complementary DNA (cDNA) was purified and quantified using a Qubit dsDNA kit (Invitrogen, #Q32854). Libraries of 3′-end transcripts were subsequently constructed through cDNA fragmentation, size selection, end repair and A-tailing, adapter ligation, PCR for indexing libraries, and cyclisation of sequencing libraries, according to the manufacturer's instructions. Sequencing libraries were purified and quantified using the Qubit ssDNA kit (Thermo Fisher Scientific, #Q10212).

Single-cell RNA sequencing using DNBelab C4

The DNBelab C4 Series Single-Cell Library Prep Set (MGI) was used for sequencing. DNBs were loaded into the patterned nanoarrays and sequenced on an ultra-high-throughput DIPSEQ T1 sequencer using the following read lengths: 30 base pairs (bp) for read 1, including 10 bp cell barcode 1, 10 bp cell barcode 2, and 10 bp unique molecular identifier (UMI), 100 bp of transcript sequence for read 2, and 10 bp for the sample index.

Sample preparation for spatial transcriptomic sequencing

Fresh tumors were embedded in tissue-neck OCT within 30 min of resection, and the extra fluid was removed. After freezing using dry ice, embedded samples were transferred to an −80°C freezer for storage until use. In a Leica CM1950 cryostat, fresh tumor tissue was cut into sample sections (10 μm) and promptly attached to Stereo-seq chips.

Spatial transcriptomic sequencing using Stereo-seq

Spatial transcriptomic data were obtained using the Stereo-seq Transcriptomics Set following the Chip-on-a-slide protocol (https://www.stomics.tech/). Briefly, tissue sections were adhered to the Stereo-seq chip and incubated for 3 min at 37°C. The chip was then incubated at −20°C for 40 min, stained with a nucleic acid dye (Thermo Fisher, Q10212), and imaged using a Ti-7 Nikon Eclipse microscope. Tissue sections were washed with 0.1× saline sodium citrate (SSC, Thermo, AM9770) buffer supplemented with 0.05 U/μL RNAase Inhibitor (NEB, M0314L). Tissue sections were permeabilized using 0.1% pepsin (Sigma, P7000) in 0.01 M HCl buffer (37°C, 12 min). RNAs captured by DNBs on the chip were reverse transcribed using the SuperScript II RT mix (Invitrogen, 18064-014, 10 U/μL reverse transcriptase, 1 mM deoxynucleoside triphosphates (dNTPs), 1 M betaine solution PCR reagent, 7.5 mM $MgCl_2$, 5 mM DTT, 2 U/μL RNase inhibitor, 2.5 μM Stereo-seq template switch oligo, and 1× first-strand buffer). The cDNA products were released from the Stereo-seq chips and amplified with KAPA HiFi Hotstart Ready Mix (Roche, KK2602) with 0.8 mM cDNA-PCR primer. Sequentially, the cDNA libraries were constructed following the manufacturer's protocol. The major steps include fragmentation (using in-house Tn5 transposase), amplification (KAPA HiFi Hotstart Ready Mix), and purification (AMPure XP Beads). The prepared cDNA libraries were sequenced using the MGI DNBSEQ-Tx sequencer.

QUANTIFICATION AND STATISTICAL ANALYSIS

Raw data processing for single-cell RNA sequencing

Raw sequencing reads from DIPSEQ-T1 were filtered and demultiplexed using PISA (v.0.2) (https://github.com/shiquan/PISA). The reads were aligned to the human genome using STAR (v.2.7.4a) and sorted using Sambamba (v.0.7.0). Doublets and contaminant cells were strictly filtered based on clustering in the UMAP embedding space, disordered marker expression, and estimates using DoubletFinder. The gene-cell matrix was filtered based on the number of genes detected per cell (any cells with <200 or >6000 genes per cell were filtered out) and the percentage of mitochondrial UMI counts (any cells with >20% mitochondrial UMI counts were filtered out). Before filtering, the median number of genes per cell for all cells was 688, and the median UMI counts was 1537. After filtering, the median number of genes per cell for all cells was 865 [IQR 527–1503], and the median UMI counts was 2057 [IQR 1080–4546]. UMI normalisation was performed by dividing the UMI counts by the total UMI counts in each cell, followed by multiplying by the universal number of total UMI counts across cells. Finally, the UMI counts were transformed into a natural log scale.

Raw data processing of single-cell spatial transcriptome from Stereo-seq

Stereo-seq FASTQ files containing the CID, MID, and cDNA sequences were generated using the MGI DNBSEQ-Tx sequencer. To ensure high-quality reads, CID sequences were mapped to the designed coordinates on the chip with a 1 base mismatch tolerance, allowing for the conversion of CID sequences into spatial coordinates. Reads with MID quality score <10 were removed. The cDNA sequences were then aligned to the reference human genome (GRCH38) and the SAW pipeline (https://github.com/BGIResearch/SAW) was used to quantify gene expression levels. Finally, GEM files containing information on the DNB coordinates and gene UMI counts in each DNB were generated after data pre-processing.

To obtain available expression abundance from sequence data of the sub-cellular resolution of 220 nm (one DNB), we merged 50 × 50 DNBs into a single highly informative ‘bin’ as the minimum unit (25 μm × 25 μm) for downstream analysis. The gene expression matrix and spatial coordinates of each bin were used as inputs to construct the Seurat object using Seurat (v.4.1.1). For two high-quality samples (ST-CR1 and ST-NR1), bins with less than 200 expressed genes, and a proportion of counts of mitochondrial genes greater than 20% were removed from downstream analysis. For the two low-quality samples (ST-CR2 and ST-NR2), the filtering criteria were adjusted to retain more spatial transcriptomic information. The filtering criteria were set to remove bins with fewer than 100 expressed genes, and a proportion of counts of mitochondrial genes greater than 20%. Additional quality details of the ST data are shown in Table S3.

Dimension reduction, clustering, and identification of differentially expressed genes of SC data

Dimension reductions were performed using PCA and UMAP implemented in Seurat (version 4.1.1). A total of 5000 highly variable genes were selected based on normalised dispersion. The top 30 principal components were used for UMAP projection and clustering analysis according to the elbow plot of standard deviation explained by PCs. Clustering of single cells was performed using KNN graph construction and the Louvain algorithm. Resolution parameter resolution = 2.0 was applied during clustering. Raw clusters with few differentially expressed genes were merged to avoid excessive classification. We used FindAllMarkers function of Seurat for the identification of differentially expressed genes between clusters, with parameters as pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25, test.use = “wilcox”. Genes with adjusted $p < 0.05$ were considered as differentially expressed genes.

Identification of transcriptional programmes and gene modules using NMF

We employed an NMF procedure to identify robust transcriptional programmes in all malignant and normal epithelial cells.[55] Expression matrix (genes × cells) could be factorized into two matrixes by NMF, which are gene × program and program × cell, therefore inferring identity and activity of expression programmes, including their relative contributions in each cell. We randomly sampled 50,000 epithelial cells and formed a 5,000 (genes) × 50,000 (cells) normalised expression matrix. The rank parameter of NMF defines the number of basis effects used to approximate the target matrix. For single-cell data, it can be set to the expected number of expression patterns or cell types. We set this to a large number (rank = 60), and will get plenty of programmes allowing some redundance. Next, we performed NMF using the “snmf/r” method and “nndsvd” random seed. We sorted gene weight on the transcriptional programmes obtained using NMF, and extracted top 50 genes with top high weights, which generated a gene × program matrix (50 × 60). Redundant programmes were merged if they shared most genes (80%). For the above matrix, we computed the Spearman correlation coefficients across all samples for the programmes and obtained a correlation matrix to define modules. We performed a hierarchical clustering with the “complete” method on the correlation matrix, which formed program modules. We de-duplicated all genes in each module. The gene ontology (GO) analysis was conducted based on GO and KEGG database. The GSEA was performed on hallmark gene sets of Molecular Signatures Database (MSigDB). Finally, we obtained epithelial expression programmes and annotated gene modules.

Cell type identification and deconvolution of ST data

We considered both expression of signatures and proportion of cell types to determine the cell type for an ST bin. We calculate a gene score (sum expression of the signatures for a cell type) for each of ST bins. Bins with top score will be defined as the cell type. The threshold was set according to the proportion of the cell type in the corresponding single cell data. The signatures were collected from two aspects (Table S2): (1) identified by single-cell data of RC and (2) widely adopted in the literatures. Regarding to CAFs, the CAF subtypes of ST data were determined with the similar approach. We defined ten markers for each CAF subset, and calculated the gene scores for each bin. Bins passing the threshold will be defined as the CAF subset. If a bin was assigned to various CAF subsets, we annotated the bin as the subset with the highest score after normalization across all subsets.

In addition, the cell types of ST data were also identified by SPOTlight,[65] and showed similar results with our approach for the major cell types. ST data were deconvoluted using the seeded NMF method implemented in SPOTlight v0.17. SC data were used as references to infer the composition of each ST bin. A threshold of 0.08 was applied to filter the composition of cell type. The distribution of cell types manually annotated by marker genes, exhibited a similar pattern to that of the high-density region deconvoluted using SPOTlight. However, for cells of small size, which are commonly diffused/mixed with tumor cells, SPOTlight shows underestimates compared with single cell data.

Measurements of cell colocalisation and proximity in ST data

We defined two cells were colocalized if their Euclidean distance (Equation 1) in spatial coordinate was $r < 3$ (approximately 75μm). For cell sets A and B, we count the A cells with any B cell is around within the cutoff ($r < 3$) as the colocalisation of A relative to B

(Equation 2). To compare the colocalisation of two cell sets A and B between samples, we need to adjust the value of colocalization by their population size (C_{adj}) because of the batch difference of samples (Equation 3), with A_{coloc} is the A cells colocalised with any B cells, B_{coloc} is the B cells colocalised with any A cells.

$$dist(a,b) = \sqrt{\left((x_a - x_b)^2 + (y_a - y_b)^2\right)} \quad \text{(Equation 1)}$$

$$Counts\ (\mathrm{A}, \mathrm{B}, \mathrm{r}) = \text{A with B around} \quad \text{(Equation 2)}$$

$$C_{adj}(\mathrm{A}, \mathrm{B}, \mathrm{r}) = \frac{|A_{coloc} \cup B_{coloc}|}{|\mathrm{A} \cup \mathrm{B}|} \quad \text{(Equation 3)}$$

We quantitively measured spatial proximity (Equation 3) of two cell sets A and B as the proportion of cells around set A are from set B, given a distance range r. We searched B cells around A, using each A cell as a center with a given radius r, to calculate the fraction of B within the total cells of the radius region. The proximity of A and B was averaged across all A cells. The final percentage represents the probability of cells around A space within r distance is B cells.

$$proximity(\mathrm{A}, \mathrm{B}, \mathrm{r}) = \frac{1}{n} \sum_{i}^{n} \frac{\text{B around } A_i}{all\ cells\ around\ A_i} \quad \text{(Equation 4)}$$

Trajectory inference of cell differentiation

We developed a new method by applying the optimal transport (OT) strategy to trajectory inference based on SC or ST data. OT is a strategy for finding least-cost schemes of coupling distributions and intuitively quantifying the distance between multiple datasets or samples represented as distributions, and it has recently been used for transcriptomic data analysis.[68] With the natural advantage of OT, both gene expression similarity and spatial distance from cell to cell can be conveniently introduced into the cost matrix and entropy regularisation to solve the OT problem. A fully captured cell-to-cell transition matrix facilitates the prediction of local developmental routes without limiting global lineage topology. In this process, we implanted the OT to calculate the transmission probability of each pair of cells and calculated the velocity for each data point in the embedding space of dimension reduction for the SC data or the morphometric space of tissue section for the ST data. Finally, we organised the streamlines into developmental trajectories. This method takes advantage of single-cell resolution, which depicts the details of local differentiation instead of population-averaged lineages.[69–71] In addition, this method has substantial advantages in computing efficiency and accuracy over methods based on the mRNA splicing ratio.[72–74] Firstly, the reads captured in ST data are often sparser than those in SC data, which can result in less accuracy of RNA-splice estimation and lead to worse trajectory inference. Secondly, most RNA-velocity methods were developed for SC data, neglecting the spatial locations of cells. Although they could be directly applied to ST data, the resulting trajectory may be discontinuous in space and usually lost the ability to recover local details of cell differentiations. Thirdly, estimating RNA splicing rate involves processing BAM file which can be computationally intensive and requires heavy memory loads. Python packages and tutorial pages were developed and deposited at https://spatrack-tutorials.readthedocs.io/en/latest/index.html.

Cell-cell interaction analysis

In our study, we implemented a new strategy to study the cell-cell interaction (CCI), taking advantage of our large cohorts. In short, CCIs were estimated between a pair of cell subsets by systematically screening the correlation between ligands from one subset and pathways from another with a background control, across large samples. The process is as follows.

(1) To screen the initial candidates, including ligands in subset A, receptors and the signaling pathways in subset B, we applied the following criteria:

$$candidates = \begin{cases} expression(L_A) \geq 1 \\ expression(R_B) \geq 1 \\ expression(G_B) \geq 1 \\ Zscore(regulation(L_A, G_B)) \geq 1.8 \end{cases}$$

L_A is the ligand from subset A, R_B is the receptor from subset B, G_B is the pathway from subset B. LR pairs were integrated from multiple datasets, including the LR candidates in Cellchat, CellphoneDB and Nichenet. Pathways were selected from Molecular Signatures Database (MSigDB). The initial thresholds are relaxing. The expressions are averaged from the cell type of all samples.

We selected pathways with targeted genes regulated by the ligand with a high potential score, as measured by the Nichenet. Nichenet has generated a regulatory network (a regulatory matrix) presenting the regulatory connection between a ligand and all targeted genes. We only consider the pathways that with a high regulatory score (Z score ≥ 1.8) compared to randoms.

(2) A real CCI should trigger a functional response. The regulation between a ligand and targeted signaling pathways is the most important factor of a real CCI. A Spearman correlation was calculated between the ligands in subset A, signaling pathways in subset B.
(3) The correlation between ligands and pathways may be inflated owing to coordination of cell populations. To reduce the false positives, we simulated 1000 correlations of ligand-pathway to generate a background distribution for each CCI candidate. Only significant correlations that deviated from the background distribution were considered (Z score ≥ 1.8). For a CCI candidate, we randomly sampled 1000 artificial ligands from subset A and pathways from subset B. The artificial pathways should keep the same number of genes as the real pathways.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) determines whether a priori-defined set of genes shows statistically significant enrichment in a biological state. Considering the inflated dropout rate of SC data and to reduce false positives owing to noise, we designed a GSEA approach as follows: (1) Gene expression was averaged from ten random cells from each cluster due to the high dropout rates. The genes were ranked according to their expression levels in each meta-cell. (2) A recovery curve was created by walking down the gene list, and the number of steps increased when a gene in the gene set was encountered. The area under the curve (AUC) was computed as an indicator of enrichment of a gene set. Only the AUC of the top 5000 ranked genes was considered. (3) To quantify the significance, we randomly sampled genes from the entire transcriptome gene list to generate an artificial gene set equal in size to the investigated gene set. Sampling was repeated 1000 times to produce an AUC distribution for the targeted gene set. By comparing the difference between a target gene set and its background distribution, we calculated the Z score of the recovery AUC of a gene set relative to the background in the cell population. Only significantly deviated gene sets (Z score>=1.8) were considered as being enriched. (4) To compare the different enrichments of the two cell populations for a gene set, we calculated the Z score of the AUC in one cell population relative to the AUC distribution in another cell population (Z score>=1.8 as significant). (5) Gene sets were selected from MSigDB.

Definition of therapeutic index

The therapeutic index (Ti), which was modified from the original proposal,[75] aims to determine whether the cellular proportion after NAC correlates with the chemotherapy response, as it measures the correlation between the change in the cellular fraction after NAC and TRG. Specifically, Ti defined as

$$Ti = -\frac{slope}{|slope|} \times R^2 \qquad \text{(Equation 5)}$$

slope and R^2 are obtained from the linear regression model lm (y $\sim$ x), where y is the alteration of cellular fraction of CAF subset following therapy and x is the tumor regression grade of the sample. A positive Ti indicates the CAF subset is positively associated with a favorable response, whereas a negative Ti indicates a worse response.

Survival analysis

Gene expression data and clinical information of TCGA cases were downloaded from Pan-Cancer Atlas Hub of UCSC Xena database (https://xenabrowser.net/datapages/?cohort=TCGA%20Pan-Cancer%20(PANCAN)&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443). We collected the “Batch effects normalized mRNA data” (dataset ID: EB++AdjustPANCAN_IlluminaHiSeq_RNASeqV2.geneExp.xena) as the gene expression data, and “Curated clinical data” (dataset ID: Survival_SupplementalTable_S1_20171025_xena_sp) as the clinical phenotype data.

Signatures for the individual CAF subset are chosen from the differentially expressed genes of one subset comparing with all others. All signatures are ranked according to their fold change (FC) values and we used top 10 genes for each subset in survival analysis of TCGA data. For combined CAF subsets, we similarly choose highly differentially expressed genes compared with the rest CAFs as their signatures. Gene score of each group was calculated in TCGA data by averaging the expressions of their signatures. The gene score of each subset was adjusted by calculated the ratio over one same CAF subset (i.e., CAF_FAP).

To assess the prognosis of pCAF and nCAF subsets in the pan-cancer patients of TCGA, we calculated the gene score of pCAF subsets and nCAF subsets in integrated pan-cancer samples. Subsequently, we divided the patients into two groups based on the median gene score, pCAF-high and nCAF-low, pCAF-low and nCAF-high, to conduct the overall survival (OS) analysis. We used the survfit and survdiff functions in R to generate Kaplan-Meier survival curves and calculated the p value of the log rank test.

We also evaluated the prognosis of each CAF subset. The gene score of each CAF subset were estimated from the averaged expression level of their differential expression genes. An adjusted ratio was calculated for each subset. All patients were divided into high and low groups based on the median ratio. OS, progression-free interval (PFI), disease-free interval (DFI), and disease-specific survival (DSS) were calculated and compared between groups. The survfit and survdiff functions were used to generate

Kaplan-Meier survival curves and calculate the p values of the log rank tests. The extended rate was calculated by subtracting the 3-year survival rate between high and low groups.

Correlation analysis

We investigated the correlations between cell populations, gene and pathways, gene/pathway and populations. Scatterplots and linear regression analysis were conducted using geom_points in ggplot2 package and lm function in R were implemented. These correlations were examined across 29 samples after therapy. Expression of gene or pathway is averaged from the cells in each sample. Considering the inflated missing values of gene expression, samples with fewer than 10 cells were excluded from the analysis.

Identification of malignant cells

We applied an integrative strategy of both copy number variants (CNVs) detection and expression clustering to identify malignant cells. We utilized inferCNV (https://github.com/broadinstitute/inferCNV)[76] to identify somatic alterations of large-scale chromosomal CNVs, either gains or losses for each cell. Para-carcinoma tissue of normal rectal mucosa was used as the reference for determining somatic variants. For the inferCNV analysis, the following parameters were used: ‘‘denoise’’, default hidden markov model (HMM) setting, and gene cutoff = 0.1. Cells were clustered according to their CNV pattern using hierarchical clustering method. We assigned each cell as CNV-malignant or CNV-normal at this step (CNV heavy or not) according to their clustering. We next identified sub-clusters of all epithelial cells (both normal and tumor tissue) through dimension reduction and clustering (by Seurat). We mapped the results of CNVs onto expression clusters. The clusters with heavy CNV-malignant cells but not from normal tissue were determined as malignant epithelial cells.