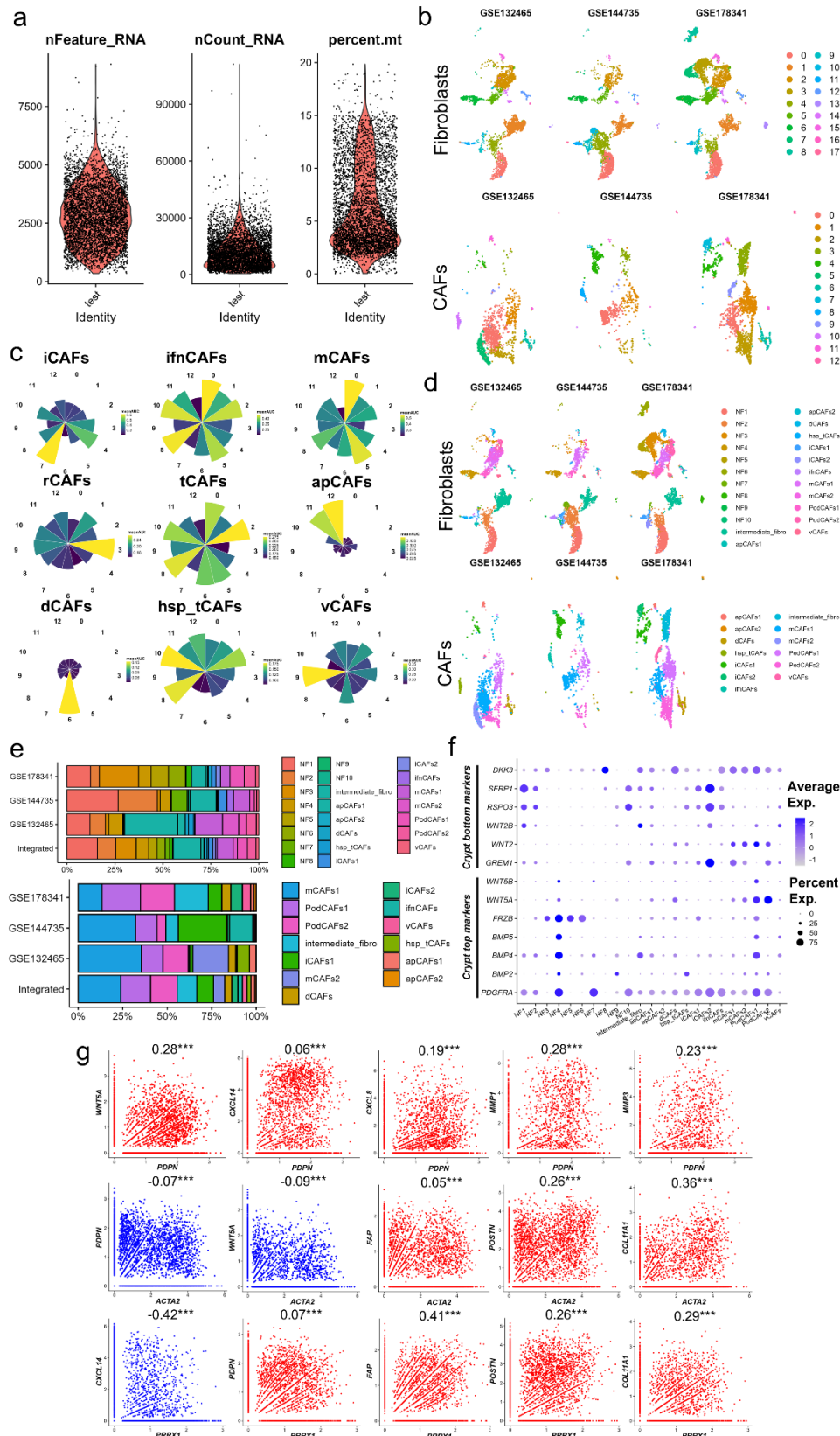


Supplementary information

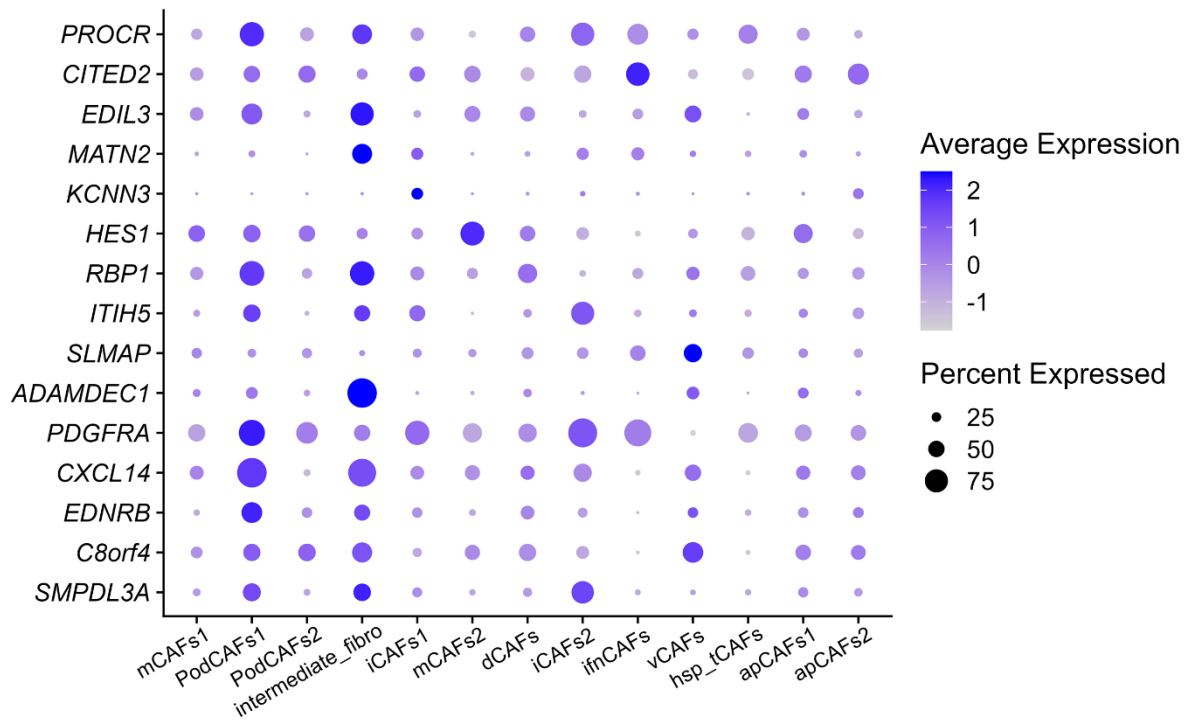
Title page	-p1
Supplementary Figure S1	-p2-3
Supplementary Figure S2	-p4
Supplementary Figure S3	-p5-6
Supplementary Figure S4	-p7
Supplementary Figure S5	-p8-9
Supplementary Figure S6	-p10
Supplementary Figure S7	-p11-12
Supplementary Figure S8	-p13
Supplementary Figure S9	-p14
Title and legend of Supplementary Data S1-S7	-p15-16
(Supplementary Data S1-S7 are provided as a separated Excel file)	



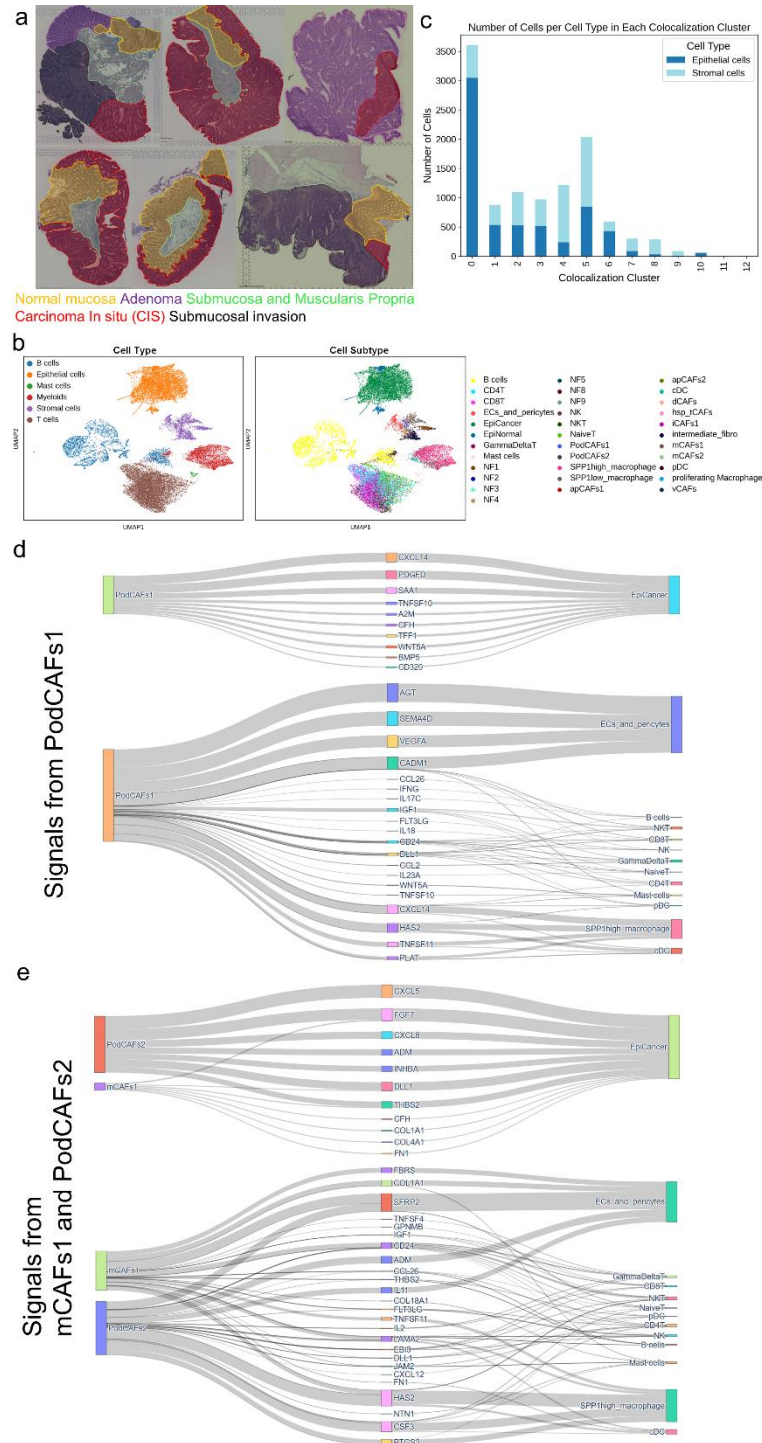
Supplementary Figure S1. Quality control, dataset integration, and annotation of fibroblasts in scRNA-seq datasets.

a: Violin plots showing the number of detected genes (nFeature_RNA), total RNA counts (nCount_RNA), and

percentage of mitochondrial gene counts (percent.mt) per cell in the integrated fibroblast dataset. Individual cells are shown as dots. b: UMAP plots of all fibroblasts (upper panels) and CAFs (lower panels), separated according to the original dataset and colored by unsupervised cluster assignment. c: Rose plots showing the mean AUCell enrichment scores for the indicated CAF gene signatures across CAF clusters. d: UMAP plots of all fibroblasts (upper panels) and CAFs (lower panels), separated according to the original dataset and colored by cell annotation based on *PDPN* expression and AUCell classification. e: Stacked bar plots showing the proportions of all fibroblast clusters (upper) and CAF clusters (lower) in each original scRNA-seq dataset and the integrated dataset. f: Dot plot showing the expression of marker genes associated with crypt-top and crypt-bottom fibroblasts across fibroblast clusters. g: Scatter plots showing pairwise associations between *PDPN*, *ACTA2*, or *PRRX1* expression and PodCAFs1/2- or mCAFs1-associated marker genes. Numbers above the plots indicate Pearson correlation coefficients; plots are colored red for positive correlations and blue for negative correlations. ***: $p < 0.001$



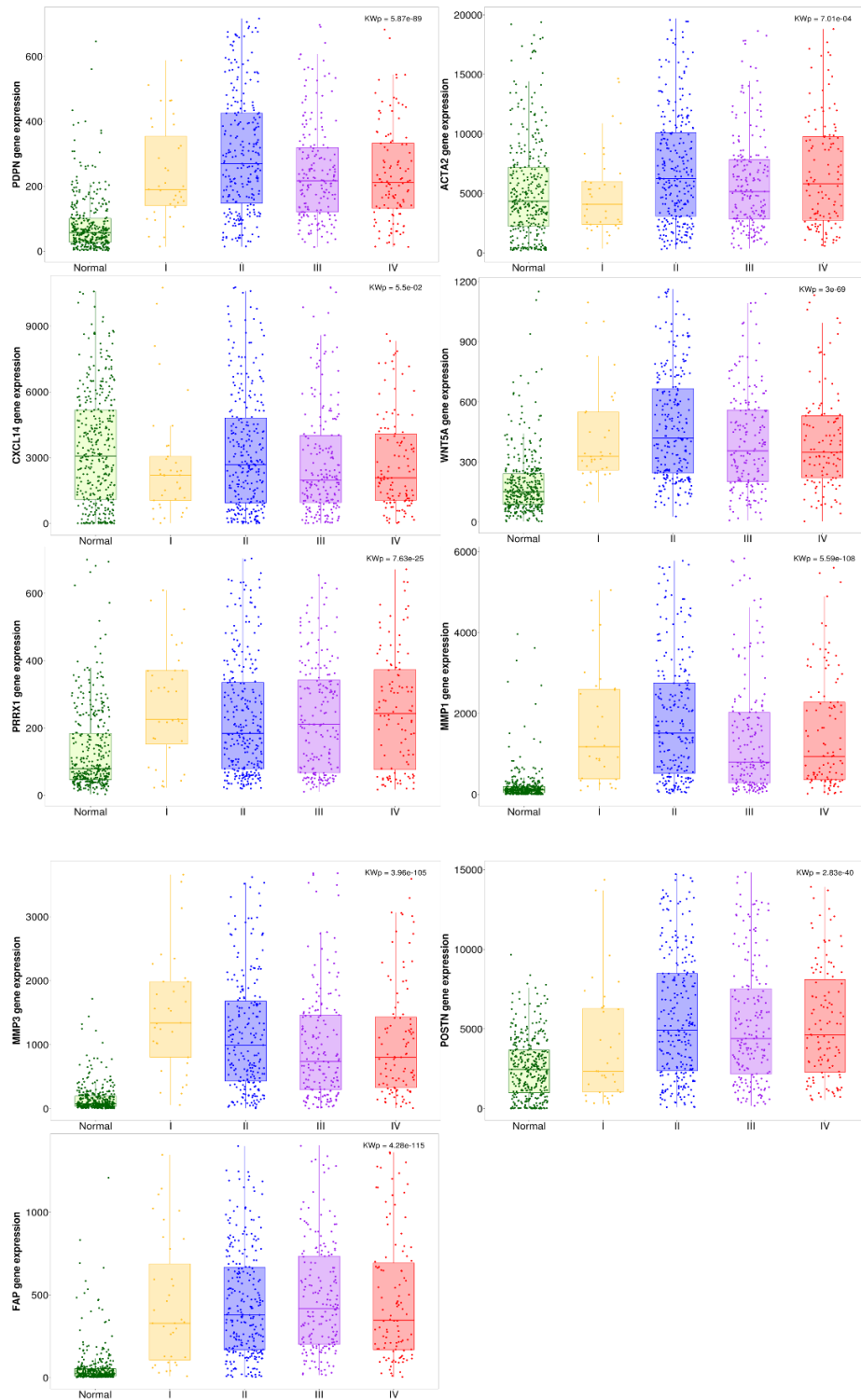
Supplementary Figure S2. Expression of individual genes in the low-risk-associated gene set across CAF clusters.



Supplementary Figure S3. Tissue segmentation, co-localization cluster composition, and predicted cell-cell interactions in cCRC environment.

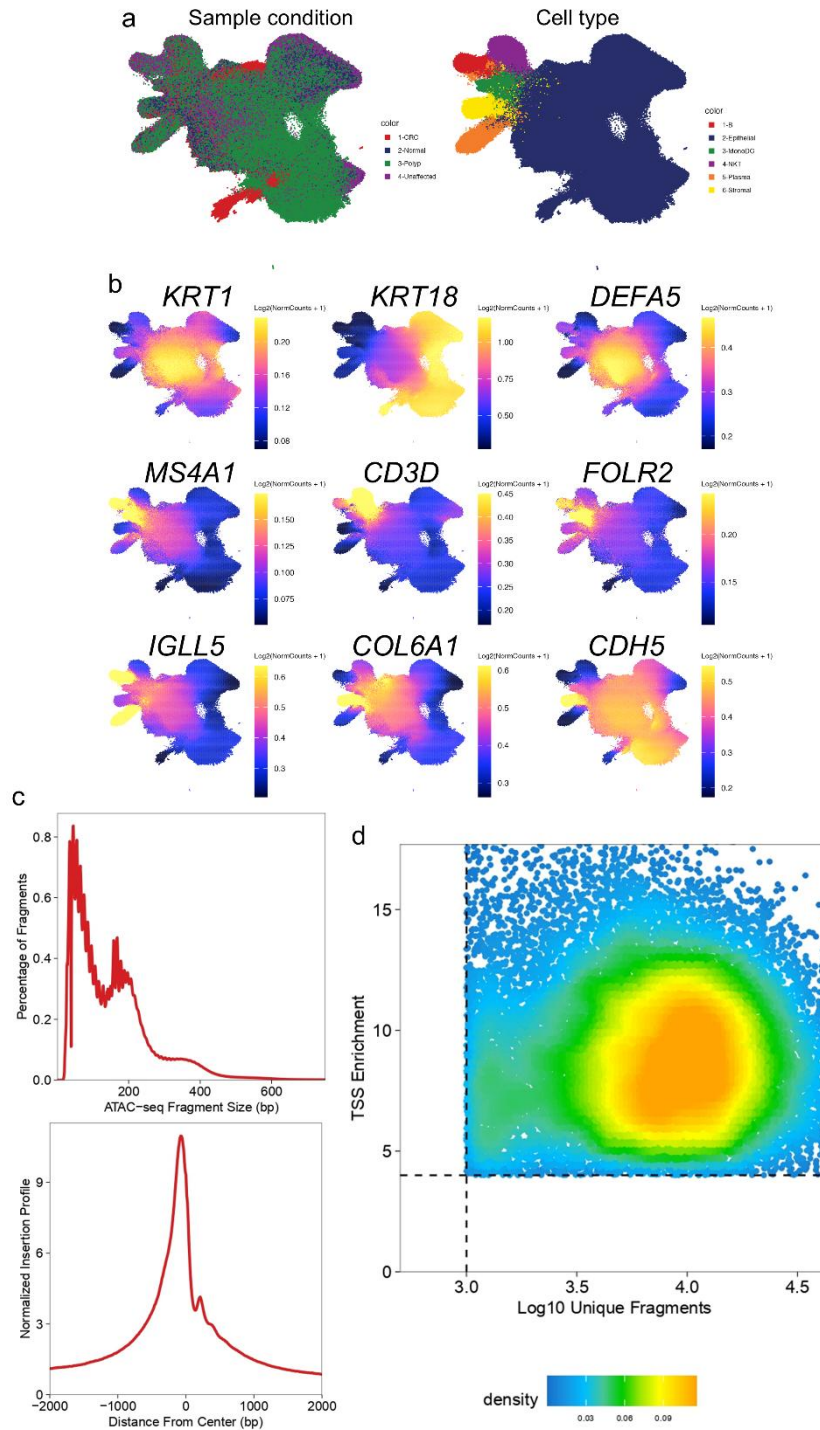
a: Tissue segmentation of all six colorectal cancer specimens included in the Visium spatial transcriptomic analysis. Representative regions are annotated as normal mucosa, submucosa and muscularis propria, adenoma, carcinoma in situ (CIS), and submucosal invasion. b: UMAP plots of the DeepCOLOR-defined “top10% cells” of the GSE132465 scRNA-seq dataset, colored by major cell type (left) or annotated cell subtype (right). c: Numbers of epithelial and stromal cells in each co-localization cluster. d: Predicted ligand-mediated cell-cell interactions originating from PodCAFs1 and involving the indicated recipient cell types. e: Predicted ligand-mediated cell-cell

interactions originating from PodCAFs2 or mCAFs1 and involving the indicated recipient cell types.



Supplementary Figure S4. Expression of PodCAFs1-, PodCAFs2-, and mCAFs1-associated genes across CRC stages in bulk RNA-sequencing data.

Box plots show the expression of *PDPN*, *ACTA2*, *CXCL14*, *WNT5A*, *PRRX1*, *MMP1*, *MMP3*, *POSTN*, and *FAP* in normal colorectal tissues and colorectal cancers of stages I–IV, obtained from the TNMplot website. Individual samples are shown as dots. Boxes indicate the interquartile range, and center lines indicate the median. p-values shown in the upper-right corner of each plot were calculated using the Kruskal–Wallis test.

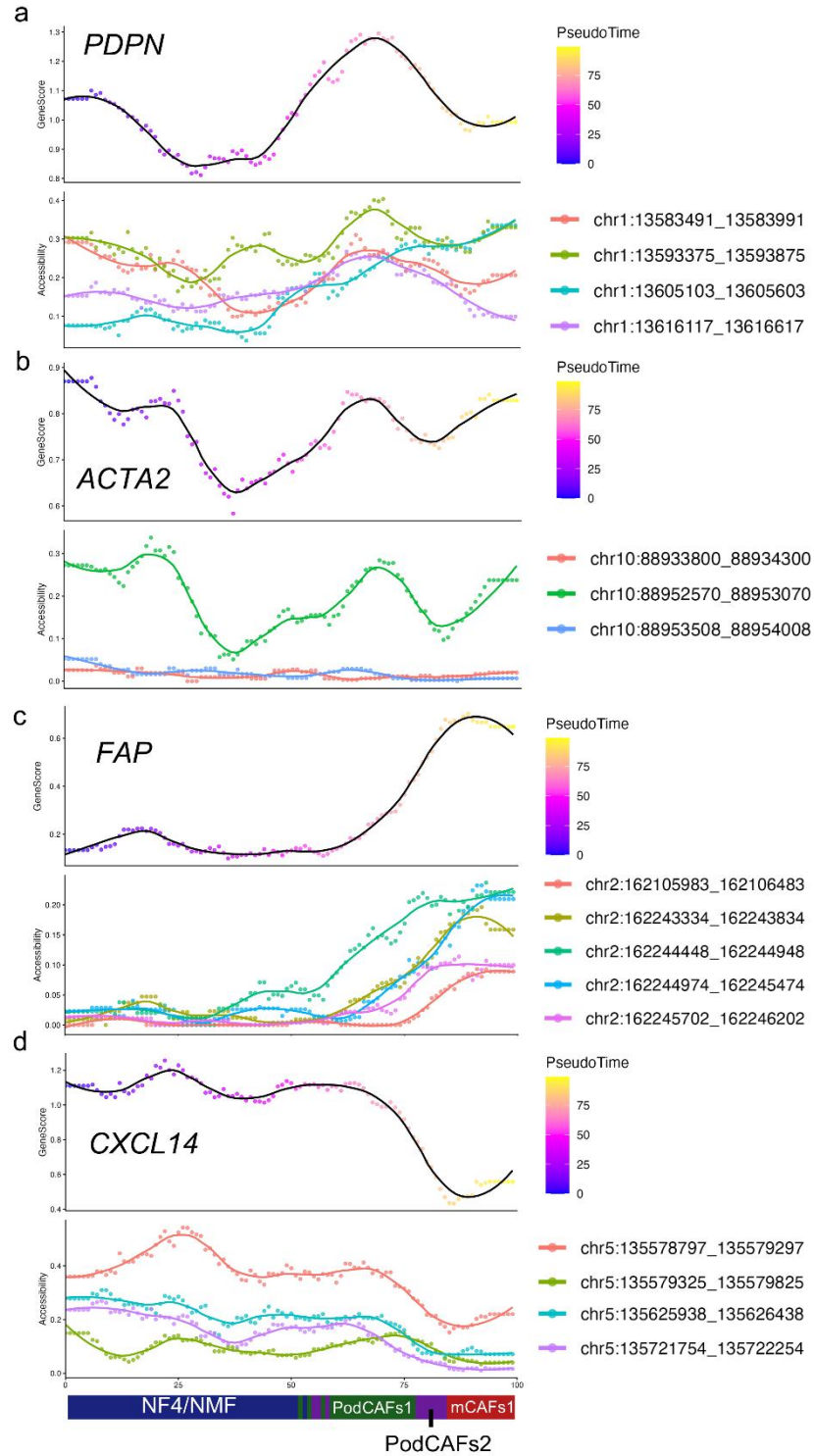


Supplementary Figure S5. Identification and quality control of fibroblasts in the single-cell chromatin accessibility dataset.

a: UMAP plots of the original single-cell assay for transposase-accessible chromatin sequencing dataset, colored by sample condition (left) or major cell type (right). Sample conditions comprised colorectal cancer, unaffected mucosa from patients with familial adenomatous polyposis, adenomatous polyps, and normal mucosa. b: UMAP feature plots showing gene activity scores for representative cell-type markers: *KRT1*, *KRT18*, and *DEFA5* for epithelial cells; *MS4A1* for B cells; *CD3D* for T and natural killer cells; *FOLR2* for monocytes and dendritic cells; *IGLL5* for plasma cells; *COL6A1* for stromal cells; and *CDH5* for endothelial cells. c: Fragment-size distribution

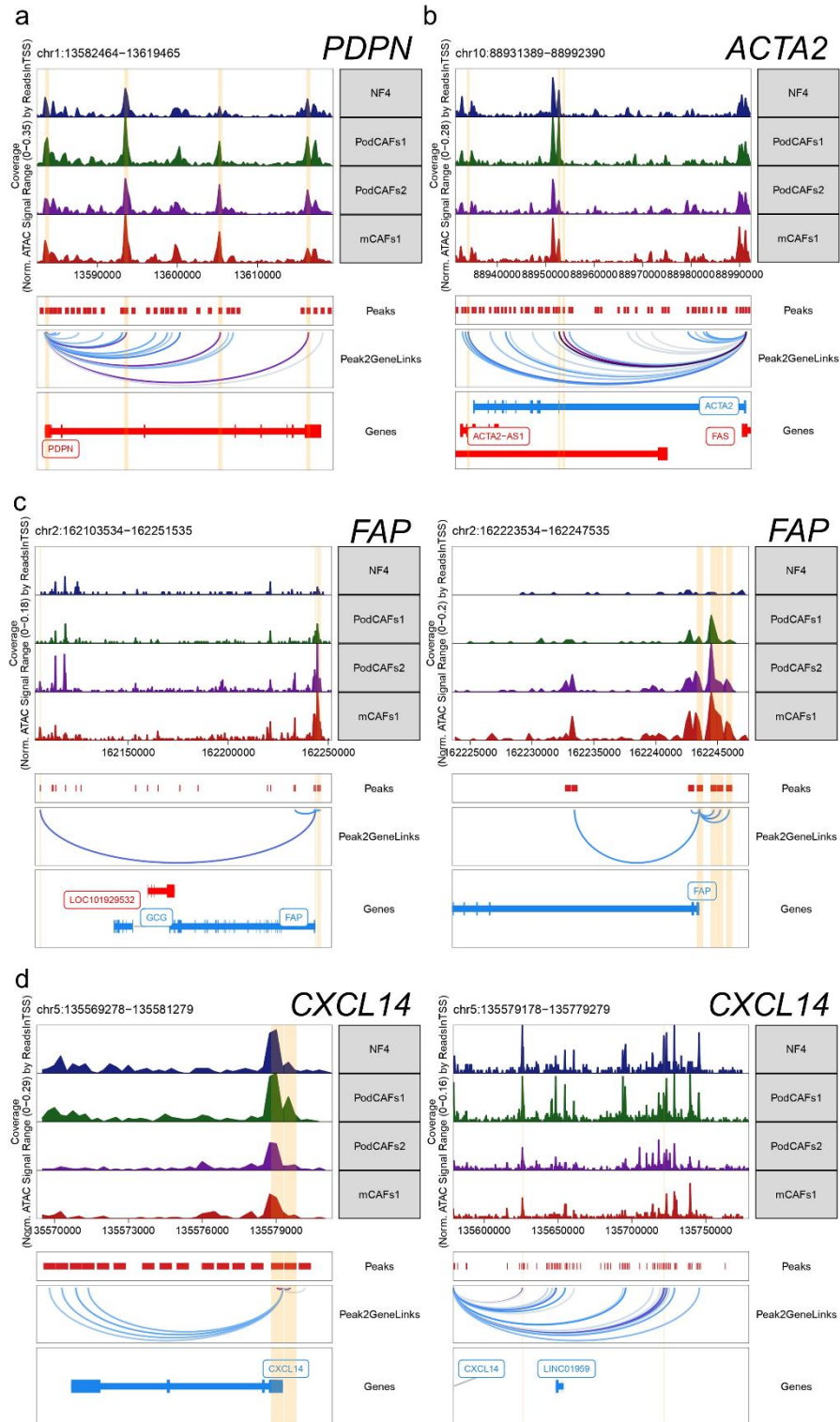
(upper) and transcription start site enrichment profile (lower) of the fibroblasts selected for subsequent analyses.

d: Density plot showing transcription start site enrichment scores versus the log10-transformed number of unique nuclear fragments in individual fibroblasts. Dashed lines indicate the quality-control thresholds.



Supplementary Figure S6. Pseudotime-associated changes in gene activity and linked-peak accessibility from NF4/NMF to mCAFs1.

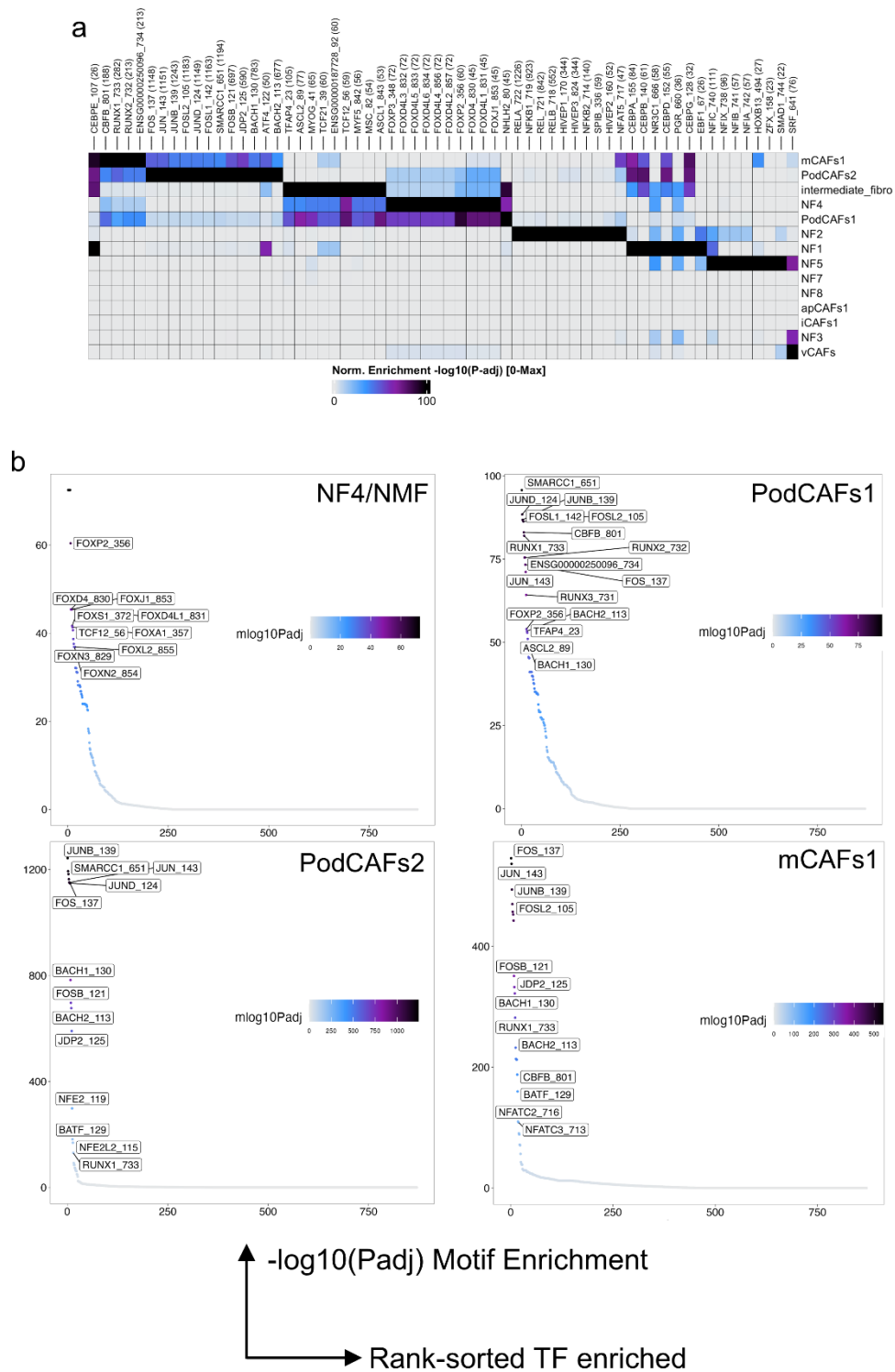
a-d: Gene activity and the accessibility of the indicated linked peaks are shown across the inferred pseudotime trajectory for *PDPN* (a), *ACTA2* (b), *FAP* (c), and *CXCL14* (d). Points represent individual cells, and lines indicate fitted trends. Colors in the gene-activity plots indicate pseudotime, whereas colors in the peak-accessibility plots distinguish individual genomic regions. The fibroblast-state annotations along pseudotime are shown at the bottom.



Supplementary Figure S7. Chromatin accessibility and peak-to-gene links across selected genomic regions associated with fibroblast marker genes.

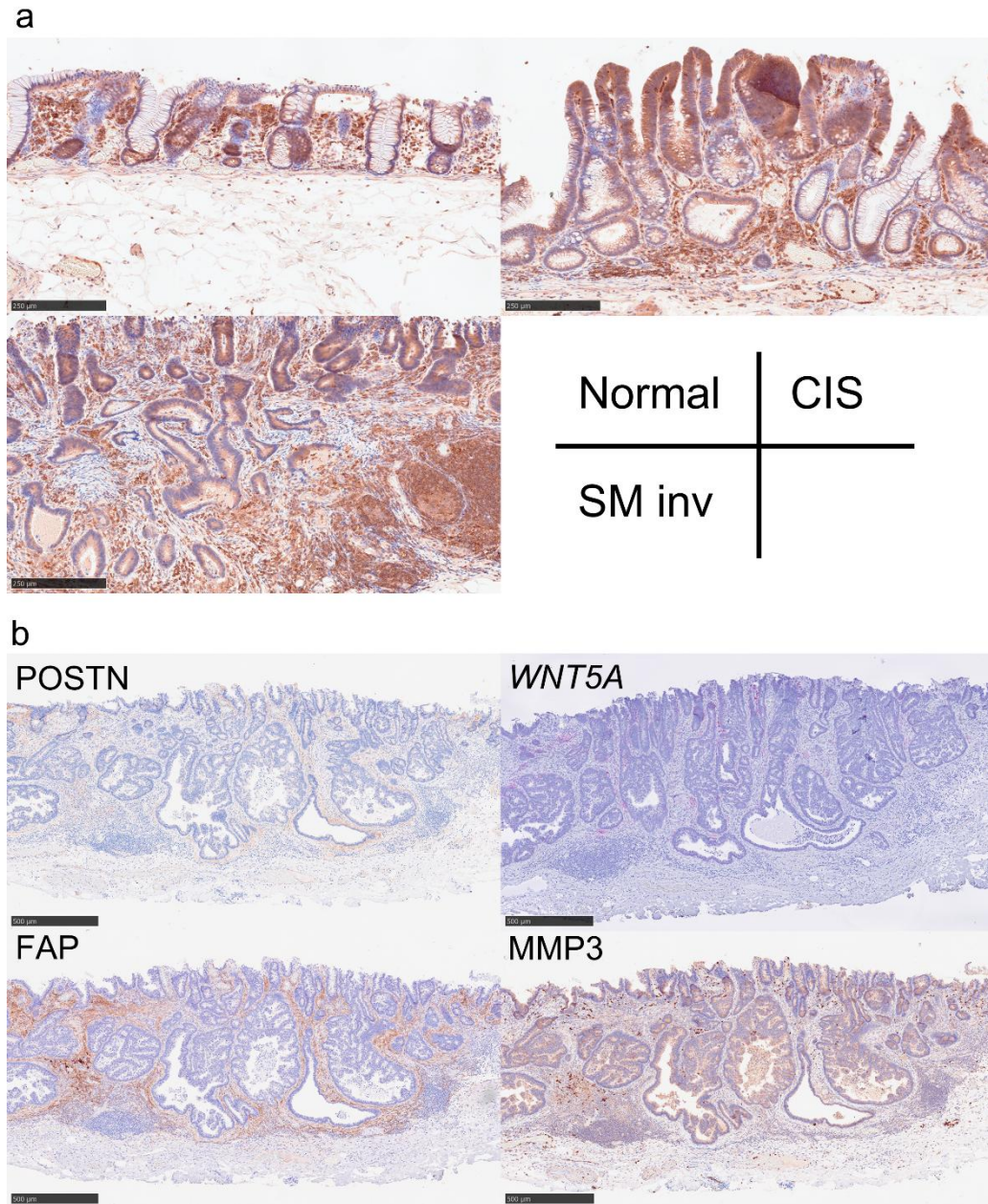
a–d: Coverage plots showing chromatin accessibility across NF4/NMF, PodCAFs1, PodCAFs2, and mCAFs1 at selected genomic regions associated with *PDPN* (a), *ACTA2* (b), *FAP* (c), and *CXCL14* (d). For *FAP* and *CXCL14*, two distinct genomic regions are shown in the left and right plots. Called peaks, peak-to-gene links, and gene

annotations are shown below each coverage plot. Orange shading highlights genomic regions with marked accessibility differences between NF4/NMF and PodCAFs1 or between PodCAFs1 and mCAFs1.



Supplementary Figure S8. Transcription factor motif enrichment in NF4/NMF, PodCAFs1, PodCAFs2, and mCAFs1.

a: Heatmap showing chromVAR motif deviation enrichment across fibroblast clusters. Fibroblast clusters were hierarchically clustered according to the similarity of their motif enrichment profiles. b: Rank-ordered transcription factors (TFs) identified in NF4/NMF, PodCAFs1, PodCAFs2, and mCAFs1. TFs are ranked according to activity score, and selected top-ranked TFs are labeled. Color intensity indicates statistical significance ($-\log_{10}$ adjusted p-value).



Supplementary Figure S9. Distribution of CXCL14 and MMP3 immunoreactivity and *WNT5A* RNA expression.

a: Representative images of CXCL14 immunohistochemistry in normal mucosa, carcinoma in situ (CIS), and submucosal invasion (SM inv). Small round cells, including mucosal fibroblasts and inflammatory cells, show CXCL14 immunoreactivity, whereas spindle-shaped stromal cells in the SM inv region show no detectable CXCL14 immunoreactivity. Scale bars, 250 µm. b: Representative serial sections from an SM inv-containing region, showing immunohistochemical staining for POSTN, FAP, and MMP3 and RNAscope detection of *WNT5A* RNA. POSTN and FAP immunoreactivity is broadly distributed throughout the stroma, whereas MMP3 and *WNT5A* signals are predominantly localised around neoplastic glands near the luminal surface and become less prominent toward the SM inv stroma. Scale bars, 500 µm.

Title and legend of Supplementary Data S1-S7

Supplementary Table S1: Logistic regression analysis results of Pedunculated vs Sessile or Flat category for Figure 2h.

ORs compare Sessile or Flat with Pedunculated eCRCs at the indicated tumor sizes. Size-specific ORs and 95% CIs were estimated from the logistic regression model including macroscopic morphology, tumor size as a continuous variable, and a morphology-by-tumor-size interaction term. p-values are two-sided. Data correspond to Figure 2h. CI, confidence interval; OR, odds ratio.

Supplementary Table S2: Tumor size-adjusted odds ratios for submucosal invasion according to three macroscopic classification systems

Adjusted ORs and 95% CIs were estimated using logistic regression models that included macroscopic classification and tumor size as a continuous covariate. The first category in each comparison was compared with the second category as the reference group; therefore, an OR greater than 1 indicates higher odds of submucosal invasion in the first category. p-values are two-sided. Data correspond to Figure 2i. CI, confidence interval; OR, odds ratio.

Supplementary Table S3: Significantly upregulated genes across scRNA-seq-defined CAF clusters

Differentially expressed genes were identified by comparing each CAF cluster with all other CAF clusters using the Wilcoxon rank-sum test implemented in Seurat FindAllMarkers. Only upregulated genes are shown.

Supplementary Table S4: Significantly up- and down-regulated genes across scRNA-seq-defined CAF clusters

Differentially expressed genes were identified by comparing each CAF cluster with all other CAF clusters using the Wilcoxon rank-sum test implemented in Seurat FindAllMarkers. Both up- and down-regulated genes are listed for gene set enrichment analysis.

Supplementary Table S5: Significantly enriched Hallmark gene sets in mCAFs1 and PodCAFs2

Hallmark gene sets with an adjusted p-value below 0.05 are shown. Gene set enrichment analysis was performed using genes ranked by average log2 fold change for each CAF cluster versus all other CAF clusters. Positive and negative NES values indicate enrichment toward genes upregulated and downregulated, respectively, in the indicated cluster. Leading-edge genes are the genes contributing to the enrichment signal. Adjusted p-values were calculated using the Benjamini–Hochberg method. NES, normalized enrichment score.

Supplementary Table S6: Significantly enriched GO Biological Process terms in mCAFs1, PodCAFs1, and PodCAFs2

GO terms with an adjusted p-value below 0.05 are shown. Gene set enrichment analysis was performed using genes ranked by average log2 fold change for each CAF cluster versus all other CAF clusters. Positive and negative NES values indicate enrichment toward genes upregulated and downregulated, respectively, in the

indicated cluster. Leading-edge genes are the genes contributing to the enrichment signal. Adjusted p-values were calculated using the Benjamini–Hochberg method. GO, gene ontology; NES, normalized enrichment score.

Supplementary Table S7: Numbers of cells of each annotated subtype in DeepCOLOR co-localization clusters

(No legend)