High-resolution integrative analysis allows characterization and spatial annotation of normal and cancer-associated colon fibroblasts

Emma W Viitala[1,2*], Ella Salminen[1,3], Mitro Miihkinen[4,5], Niki Chalkidi[6], Athanasia Stavropoulou[6], Toni T Lemmetyinen[1,2], Pekka Päivinen[7,8], Tuomas Kaprio[2,9], Jaana Hagström[2,9], Caj Haglund[2,9], Tomi P. Mäkelä[7,8], Tero Aittokallio[4,5,10,11], Pekka Katajisto[1,3,12], Vasiliki Koliaraki[6] & Saara Ollila[1,2*]

1 Molecular and Integrative Biosciences Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

2 Translational Cancer Medicine Program, University of Helsinki, Helsinki, Finland

3 Institute of Biotechnology, Helsinki Institute of Life Science, HiLIFE, University of Helsinki, Helsinki, Finland

4 Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland

5 iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

6 Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center Alexander Fleming, Vari, Greece

7 HiLIFE-Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland

8 iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki, Helsinki, Finland

9 Department of Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

10 Institute for Cancer Research, Department of Cancer Genetics, Oslo University Hospital, Oslo, Norway

11 Oslo Centre for Biostatistics and Epidemiology (OCBE), Faculty of Medicine, University of Oslo, Oslo, Norway

12 Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden

*Corresponding authors: Emma Viitala, emma.viitala@helsinki.fi; Saara Ollila, saara.ollila@helsinki.fi

ABSTRACT

Fibroblasts represent key regulators of colon homeostasis, and cancer-associated fibroblasts (CAFs) play pivotal roles in colorectal cancer (CRC). Despite their critical influence, a consensus view of adult human colon fibroblast and CRC CAF heterogeneity, spatial organization, and developmental trajectories is currently lacking. Here, we address this gap by performing a comprehensive characterization of colonic fibroblasts and CRC CAFs. Using large-scale integration of single-cell RNA-sequencing datasets from normal colon and CRC we mapped the fibroblast subpopulations. Spatial transcriptomics, immunohistochemistry, and *in situ* hybridization were used for validation and computational analyses to predict developmental trajectories and transcription factors underlying CAF activation. Subepithelial myofibroblasts (SEMFs), mucosa-associated fibroblasts (MAFs), and submucosa-associated fibroblasts (SAFs) were identified as the main colon fibroblast subtypes, and in mice, further divided into location-based subclusters. We also identified a novel colon fibroblast subset: muscle-embedded interstitial fibroblasts (MIFs). CRCs contained normal fibroblasts as well as four cancer-specific CAF populations: inflammatory CAFs (iCAFs), matrix CAFs (mCAFs), and two precursor CAF (preCAF) subtypes. Our data suggested that iCAFs originate from SEMFs through a preCAF1 intermediate phenotype, while mCAFs derive from SAFs/MIFs via preCAF2s. Transcription factors PRRX1, MAFB, and TWIST1 were uncovered as potential regulators of CAF identity and CTHRC1 was identified as a specific and sensitive pan-CAF marker in CRC. Our study presents a detailed framework for understanding colon fibroblast and CRC CAF heterogeneity. This work lays the groundwork for future research into the roles and potential therapeutic relevance of different CRC CAF subsets.

INTRODUCTION

Recent advances in single-cell technologies have revealed a vast transcriptional and functional heterogeneity of fibroblasts across various tissues [1-3], placing fibroblasts as a subject of intense research. In the colon, fibroblasts represent key regulators of epithelial turnover, and three main fibroblast subtypes have been characterized mostly in studies investigating mouse tissues: (1) *Pdgfra*^{high} fibroblasts, promoting epithelial cell differentiation e.g. through expression of BMP ligands, and *Pdgfra*^{low} fibroblasts, which further divide into (2) *Pi16- Cd81-* and (3) *Pi16+ Cd81+* subsets, supporting stem cell function by secreting BMP inhibitors and WNT agonists [4-11]. Variable nomenclature has been proposed for the identified colon fibroblast subsets, based on factors such as cluster numbers, markers, topology, morphology, or functional properties (S.Table 3). For example, *Pdgfra*^{high} cells have been referred to as subepithelial myofibroblasts (SEMFs) or telocytes; *Pdgfra*^{low} *Pi16- Cd81-* fibroblasts as mucosa-associated fibroblasts (MAFs) or crypt bottom fibroblast 1 (CBF1); and *Pdgfra*^{low} *Pi16+ Cd81+* cells as trophocytes or crypt bottom fibroblast 2 (CBF2) [5, 6, 12]. Human colon fibroblasts appear to exhibit transcriptional similarities to mouse cells [4, 13, 14], but a systematic analysis and annotation of adult human colon fibroblast subsets aligned with the current understanding of the mouse fibroblast subsets is lacking.

In addition to maintaining homeostatic tissues, fibroblasts play key roles in carcinogenesis, thus being increasingly recognized as potential drug targets in cancer [15, 16]. Cancer-associated fibroblasts (CAFs) are known to regulate tumor growth, immune responses, and drug resistance [17-19], but are heterogeneous and exhibit both tumor-promoting and tumor-restraining functions [20-22], emphasizing the acute need for careful characterization and annotation of different CAF subtypes. In colorectal cancer (CRC), the third most common and second most fatal cancer type worldwide (Cancer today 2022), the CAF-rich, mesenchymal CRC consensus molecular subtype 4 (CMS4) [23] is associated with poor survival and resistance to chemotherapy [24, 25], highlighting the importance of understanding the functional heterogeneity of CRC CAFs.

Several studies addressing the CAF heterogeneity in CRC have been conducted, reporting several CRC CAF subtypes, including inflammatory (iCAF), myofibroblastic/matrix (mCAFs), antigen-presenting (apCAFs), vascular (vCAFs), chemo-CAFs, tumor-like (tCAFs), and proliferative (pCAFs), along with tens of other, marker-based names (S.Table 4). However, a consensus view of CRC CAF subsets and their prognostic significance is currently lacking.

Furthermore, how normal tissue-resident colon fibroblast subsets relate to the developmental trajectories of CRC CAFs is currently unknown. One major bottleneck preventing reproducible characterization of CRC CAF subsets is the lack of consistent annotation of tissue-resident colon fibroblasts and other mesenchymal cells, such as pericytes, smooth muscle cells (SMCs), and myofibroblasts, leading to all mesenchymal cells in resected cancers often being labelled as CAFs [2, 3, 26-35] . This, together with limitations such as small sample size and/or low resolution [36-42], has made it difficult to create a coherent view on CRC CAF subsets. In spatial analyses, low resolution and lack of specific markers have challenged the formation of a clear view on CAF subtype localization [3, 32, 42, 43]. Finally, despite providing important information about the general properties of CAFs, cross-tissue atlases fail to capture the tissue-specific characteristics of CAF subtypes [2, 3, 37, 40, 42]. Therefore, a benchmarking analysis of CRC CAF subsets is urgently needed.

In this study, we adopt an integrative approach to fill the gaps in the normal colon fibroblast and CRC CAF characterization. First, we generate a high-resolution map of mouse colon fibroblast subclusters. Second, using the mouse analysis as a benchmark, we annotate human colon fibroblasts and validate their localization with spatial transcriptomics. Finally, we annotate CRC CAF subsets, establish their potential differentiation trajectories, and identify potential regulators underlying CAF development. Our work provides a solid foundation for future research into colon fibroblast heterogeneity and CAF subset-specific investigations in CRC.

RESULTS

Analysis and comparison of mouse and human colon fibroblast subsets. To investigate the heterogeneity of colonic fibroblasts, we first performed single-cell RNA-sequencing (scRNA-seq) of mouse colonic mesenchyme and analyzed the fibroblasts and smooth muscle cells (SMCs). Consistent with previous analyses [4, 5, 7, 8], the fibroblasts clustered into three main subsets: (1) *Pdgfra*^{high}, (2) *Pdgfra*^{low} *Pi16- Cd81-*, and (3) *Pdgfra*^{low} *Pi16+ Cd81+* populations (S1A and S1B Fig). Interestingly, however, the *Pdgfra*^{low} *Pi16- Cd81-* cluster was further divided into two subsets (S1A and S1C Fig), indicating heterogeneity within the main colon fibroblasts subtypes, as proposed previously [4, 11]. To investigate this with higher resolution, we integrated our data with four previously published datasets containing mouse colon mesenchymal cells [4, 5, 7, 8] (S1D Fig). The integration resulted in 36.085 cells including fibroblasts (n = 25.937), SMCs, pericytes, endothelial cells, glial cells, and epithelial cells (Fig 1A and 1B, S2A and S2B Fig). Using this approach, mouse colonic fibroblasts clustered into seven sub-populations, including three *Pdgfra*^{high} and four *Pdgfra*^{low} subsets (Fig 1B, S2A Fig). Among the *Pdgfra*^{high} fibroblasts, we observed two clusters, a *Lama5+* and a *Wif1+* cluster, both expressing *Foxl1, F3, Wnt5a, Bmp5,* and *Bmp7*, suggesting that these clusters represent the SEMF/telocyte population [5, 6, 12]. The third *Pdgfra*^{high} cluster expressed *Kcnn3+* but did not match the expression profile of previously identified colon fibroblasts or other mesenchymal cells (Fig 1B, S2B Fig). The *Pdgfra*^{low} cells consisted of three *Pi16- Cd81-* clusters, representing the previously identified mucosa-associated fibroblasts (MAFs) [8], and one *Pi16+ Cd81+* population (Fig 1B). The mucosal WNT-BMP gradient (Fig 1C) confirmed that the observed fibroblast populations (except for the *Pdgfra*^{high} *Kcnn3+* -cluster) represented the *Pdgfra*^{low}, stemness-supporting fibroblasts, and the *Pdgfra*^{high}, differentiation-promoting fibroblasts, previously shown to be distributed along the basal-luminal axis of colonic mucosa [5, 6, 12].

Fig 1. Annotation of mouse and human colon fibroblast subtypes using high-throughput analysis of scRNA-seq data. (A) Current model of mouse colon fibroblast subsets. (B) Uniform manifold approximation and projection (UMAP) plot of integrated mouse colon scRNA-seq data and (left) dot plot depicting markers of identified colonic fibroblast subpopulations in mouse (right). See also S2B Fig. (C) Signature feature plots of indicated WNT and BMP genes. (D) UMAP of integrated human colon scRNA-seq data (left) and dot plot of the corresponding human

markers shown for mice in B (right). See also S2D Fig. (E) Signature feature plots of indicated WNT and BMP genes. (F) Dot plot depicting top marker genes for each identified human colon fibroblast subpopulation. Common markers between mouse and human are marked in red. Asterisk (*) indicates genes included in the signatures used in G. See also S2 Table. (G) Signatures of human fibroblast subpopulations (marked by asterisk in F) projected onto spatial transcriptomics data of normal human colon. M, mucosa; SM, submucosa: MP, muscularis propria. (H) Updated model of mouse and human colon fibroblast subsets. SEMF, subepithelial myofibroblast; MAF, mucosa-associated fibroblasts; SAF, submucosa-associated fibroblasts; MIF, muscle-embedded interstitial fibroblasts; SMC, smooth muscle cells; peri, pericytes; EC, endothelial cells; VEC, vascular endothelial cells; LEC, lymphatic endothelial cells; epiStemProg, epithelial stem and progenitor cells.

To address how human colon fibroblasts correspond to the mouse counterparts, we analyzed scRNA-seq data of human colonic normal mucosal cells from four datasets (CRC16, KUL3, KUL5, SMC) sampled during surgery from CRC patients (n = 36) (S1E Fig) [26, 44]. Notably, the normal mesenchymal cells from these datasets have not been previously examined in detail while some studies have classified all (including normal colon-derived) fibroblasts from these datasets as CAFs [33, 38, 39], calling for clarification in annotation. After filtering out the immune cells (to match the mouse analysis), the integration resulted in 22.215 cells, including fibroblasts (n = 11.702), SMCs, pericytes, endothelial cells, glial cells, and epithelial cells, with similar contributions across datasets (Fig 1D, S2C and S2D Fig). Of note, also *KIT*+ intestinal cells of Cajal (ICCs) [45] and *CCL19*+ reticular fibroblasts [46] were identified, but not in sufficiently high numbers to form separate clusters (S2E Fig).

The human fibroblasts divided into two *PDGFRA*^{high} clusters, one expressing *FOXL1*+ and the other *KCNN3*+, and two *PDGFRA*^{low} populations that differed by *PI16* expression (Fig 1D). Except for the *KCNN3*+ cells, the BMP-WNT gradient suggested a basal-luminal distribution, as in the mouse analysis (Fig 1E). Despite some species-specific differences for individual marker genes (S2F and S2G Fig), the overall expression patterns allowed correlation of mouse and human fibroblast subsets (S2H Fig, Fig 1F). The human *PDGFRA*^{high} *FOXL1*+ cluster expressed *BMP5* and *BMP7*, consistently with SEMF/telocyte identity shared with mouse cells. The human *PDGFRA*^{low} *PI16*- cells also expressed several marker genes shared with their mouse counterparts,

including e.g., *CCL8,* a cytokine previously reported as a colon fibroblast subset-specific marker [4]. Unlike in mice, the human *PDGFRA*^{low} cells did not show differential expression of *CD81*, but expression of markers such as *ACKR4, C3,* and *PCOLCE2* suggested shared identity of *PDGFRA*^{low} *PI16*+ cells with the mouse *Pdgfra*^{low} *Pi16*+ *Cd81*+ cells (Fig 1F, S2H Fig). Interestingly, in mice, expression of *Grem1* and *Rspo3* was noted both in the *Pdgfra*^{low} *Pi16*- *Cd81*- and the *Pdgfra*^{low} *Pi16*+ *Cd81*+ cells (Fig 1B), while in humans *GREM1* and *RSPO3* were primarily derived from *PDGFRA*^{low} *PI16*+ cells, and the *PDGFRA*^{low} *PI16*- cells instead expressed the canonical Wnt ligand *WNT2B* (Fig 1d). Thus, the exact molecular mechanisms promoting canonical Wnt signaling in the colonic stem cell niche may slightly differ in mice and humans. Altogether, the gene signatures and growth factor expression patterns of human fibroblast subclusters suggested a similar organization to the mouse colon.

Next, we set out to annotate the identified human colon fibroblast clusters according to their spatial distribution by mapping the cluster-specific expression signatures (Fig 1F and 1G) in a high-resolution Visium HD spatial transcriptomics dataset of human colon [47]. SEMFs have been described as fibroblasts with expression of myofibroblast markers and a subepithelial location throughout the gastrointestinal tract, including the small intestine and colon [48]. Consistently, the *PDGFRA*^{high} *FOXL1*+ cells were enriched directly beneath the epithelial cells, especially in the luminal side of colonic mucosa, and expressed the myofibroblast marker *ACTG2* (S2D Fig), confirming the SEMF identity (Fig 1G). The *PDGFRA*^{low} *PI16*- cells localized within the mucosal lamina propria, and we named these cells mucosa-associated fibroblasts (MAFs). The *PDGFRA*^{low} *PI16*+ cluster genes were highly enriched in the submucosa, consistent with previous analysis of the mouse *Pdgfra*^{low} *Pi16*+ *Cd81*+ cells [8], and we named these cells submucosa-associated fibroblasts (SAFs). Interestingly, the *PDGFRA*^{high}*KCNN3*+ fibroblasts were embedded among the SMCs in the muscularis propria, and this cluster was named muscle-embedded interstitial fibroblasts (MIFs). Thus, we propose a coherent, location-based nomenclature for human and mouse fibroblast subsets, validated by spatial transcriptomics in human tissue and supported by previous literature and shared markers in mice (S3 Table, Fig 1H).

Sub-clustering of MAFs and SEMFs along the crypt-apex and proximal-distal axes in mouse colon. Next, we investigated the sub-clusters of MAFs identified in the mouse colon. Three separate clusters, MAF-1, MAF-2 and MAF-3, were identified (Fig 1B, Fig 2A). *Wnt4*, expressed

by fibroblasts residing in the top (apex) part of the colonic lamina propria [9], was seen in the MAF-1 cluster, while MAF-2 and MAF-3 expressed *Grem1*, a known niche factor for stem cells (Fig 2B) [6, 10, 49-52]. Therefore, the MAF-1 cluster likely represents MAFs localized at the crypt top (Top MAF), whereas MAF-2 and MAF-3 reside at the bottom part of the mucosa, closer to the stem cell zone (Bottom MAF). Interestingly, one of the analyzed datasets contained annotated cells deriving from either proximal or distal parts of the colon [8]. In this data, the MAF-2 cluster was primarily derived from proximal colon samples and expressed *Tnc,* while MAF-3 cells originated from the distal colon samples and expressed high levels of *Cd55* (Fig 2C and 2D). Tissue staining confirmed Tenascin C (encoded by *Tnc*) expression throughout the top part of the mucosal stroma as previously shown [53], while the crypt bottom stroma only expressed Tenascin C in the proximal colon (Fig 2E). To investigate the similarity between colonic and small intestinal (SI) mesenchymal cells, we also performed scRNA-seq on mouse SI mesenchyme and analyzed the fibroblasts and SMCs (S3A and S3B Fig). We identified *Pdgfra*^{high} *Foxl1*+ SEMFs (S3C Fig), two *Pdgfra*^{low} *Pi16*- MAF clusters (S3D Fig), and *Pdgfra*^{low} *Pi16*+ *Cd81*+ SAFs (S3E Fig). Consistent with the colon analysis, *Wnt4* expression separated Top MAFs from Bottom MAFs (S3D Fig). However, Bottom MAFs did not separate further, and *Tnc* and *Cd55* expressions were uniform across MAFs.

Fig 2. Sub-clustering mouse MAFs and SEMFs along the crypt-apex and proximal-distal axes. (A) UMAP of integrated mouse colon mesenchyme scRNA-seq data. SEMFs (subepithelial myofibroblasts) and MAFs (mucosa-associated fibroblasts) highlighted by dashed lines. (B) Violin plot of *Wnt4* and *Grem1* expression across the mesenchymal cell clusters. (C) UMAP of mouse colon mesenchymal cells from proximal and distal colons, MAFs highlighted by red circles. (D) Violin plots depicting differential expression of *Tnc* and *Cd55* in proximal (MAF-2) and distal (MAF-3) bottom MAFs. (E) Immunofluorescent staining of TNC in proximal and distal mouse colon. HOECHST: nuclei. (F) Feature plots of *Lama5, Lgr5* and *Adamts18* expression in mouse colon mesenchyme. Arrowheads point to Top SEMFs. (G) mRNA *in situ* hybridization of *Lama5* in mouse colon. (H) Feature plots of *Wif1*, *Bmp3,* and *Acta2* expression in mouse colon mesenchyme. Arrowheads point to Bottom SEMFs. (I) mRNA *in situ* hybridization of *Wif1* in mouse colon. (G and I), ECAD: E-cadherin, epithelial cells, DAPI: nuclei. (J) Updated model of mouse colon fibroblast subsets including MAF and SEMF subclusters and their markers.

Next, we addressed the heterogeneity of mouse colon SEMFs. SEMF-1 and SEMF-2 populations were identified, corresponding to *Lama5*+ and *Wif1*+ SEMFs, respectively (Fig 1B; Fig 2A, 2F and 2H). The *Lama5*+ SEMFs expressed *Wnt4,* indicating crypt top identity (Fig 2B) [9]. Consistently, the *Lama5*+ SEMF markers *Lgr5* and *Adamts18* (Fig 2F) were previously shown to be expressed in villus tip fibroblasts in the mouse small intestine [54]. Indeed, *Lama5* mRNA was localized at the top part of the colonic lamina propria (Fig 2G), and the *Lama5*+ SEMFs were annotated as "Top SEMFs". The *Wif1*+ SEMF population, on the other hand, expressed *Bmp3* and high levels of *Acta2* (Fig 2H). Since *Bmp3* encodes for a BMP inhibitor [55] and *Pdgfra*^{high} *Acta2*^{high} cells have previously been shown to be located in the bottom colonic stroma [11], we hypothesized that these cells could localize at the stem cell zone. Indeed, *Wif1* mRNA was localized to stromal cells adjacent to the crypt bottom (Fig 2I), and we named these cells "Bottom SEMFs". In SI, we only observed one SEMF population (S3C Fig), and *Wnt4* was expressed uniformly throughout the cluster (S3D Fig). However, the Top SEMF markers *Lama5* and *Adamts18* were enriched on one edge of the SEMF cluster, while *Wif1* was enriched on the other edge (S3C Fig), suggesting a similar bottom-to-top gradient in the SI SEMFs that could be more clearly separated in a larger dataset.

In summary, integration of multiple datasets allowed characterization of fibroblast subtypes of the mouse colon with high resolution, revealing differences along the basal-luminal and proximal-distal axes (Fig 2J). Interestingly, however, these subclusters were less clear in the human analysis, indicating subtle species-specific differences or lower resolution due to a smaller sample size (S4 Fig).

SMC subtypes in mouse and human colon. A thin layer of smooth muscle, the muscularis mucosae (MM), separates colonic mucosa from the submucosa, surrounded by the circular and longitudinal smooth muscle layers constituting the muscularis propria (MP). In literature, the colonic stromal cells expressing SMC markers such as *ACTA2, MYH11, ACTG2,* and *DES* are variably referred to as SMCs, myofibroblasts, or myocytes [4, 5, 9, 10, 13, 56, 57], calling for clarity in annotation. We first subsetted and re-clustered the SMCs from the integrated mouse dataset, resulting in two clusters expressing SMC markers but differing in expression of *Hhip* (n = 3.636) (Fig 3A and 3B). The *Hhip* negative (*Hhip*-) cluster expressed *Grem1*, *Grem2*, *Rspo3,* and *Chrdl1*, stem cell niche factors previously shown to be expressed in the intestinal MM [10] (Fig

3B), and we annotated the *Hhip*- cells as MM-SMCs. The *Hhip*+ cells, on the other hand, lacked the stem cell niche genes and instead expressed e.g., *Sncaip* and *Pdgfc* (Fig 3B). Interestingly, *Hhip* mRNA expression was limited to alpha smooth muscle actin (encoded by *Acta2*)-expressing tree-like structures within the mucosa of the colon and small intestine but was absent from the MM and MP (Fig 3C), confirming the *Hhip*-expressing cells being specific to mucosal stroma. Since they did not express fibroblast markers, we called them myocytes [9, 10]. In the integrated data, despite not clustering separately, the myocyte and MM-SMC signatures were observed in the opposite sides of the SMC cluster, highlighting their transcriptional differences (Fig 3D). In summary, the mouse SMCs clustered into two subpopulations – *Hhip*+ myocytes located in the mucosa and MM-SMCs expressing stem cell niche factors. Of note, cells from MP were not identified, likely due to their exclusion in the tissue dissociation.

Fig 3. SMC subtypes in mouse and human colon. (A) UMAP of re-clustered mouse SMCs (smooth-muscle cells) and violin plots depicting indicated SMC/myofibroblast /myocyte and Interstitial cells of Cajal (ICC) markers. (B) Dot plot depicting marker genes for myocytes and MM-SMC in mice. These displayed genes are used as signatures in D. (C) mRNA *in situ* hybridization of *Hhip* and immunofluorescent staining of αSMA in mouse colon and small intestine. ECAD: E-cadherin, epithelial cells; DAPI: nuclei. (D) Feature plots of myocyte and MM-SMC signatures (from B) in mouse colon mesenchyme. (E) UMAP of re-clustered human SMCs and violin plots of indicated SMC/myofibroblast/myocyte and ICC markers. (F) Dot plot depicting marker genes for indicated cell types in human. The genes shown are used in signatures in G and H. (G) Indicated signature (from F) feature plots in integrated human colon mesenchyme data. (H) Signatures (from F) of human myocytes, MM-SMCs and MP-SMCs projected onto spatial transcriptomics data of normal human colon. M, mucosa; SM, submucosa: MP, muscularis propria. (I) Updated model of mouse and human myocytes and SMCs.

Next, we investigated the human SMC cluster by subsetting and re-clustering (Fig 3E). Consistent with the mouse analysis, the dataset included *HHIP*+ cells, interpreted as mucosal myocytes, and *GREM1*+ *GREM2*+ *RSPO2*+ cells representing MM-SMCs (Fig 3E and 3F). In addition, a third population expressing the SMCs markers, but lacking *HHIP* and the stem cell niche genes, was noted, suggesting MP origin (Fig 3E and 3F). ICCs also clustered among the

SMCs but were distinguishable based on expression of *KIT* and lack of SMC markers (Fig 3E and 3F). In the whole integrated data, MM-SMCs and MP-SMCs clustered together, separately from the myocytes (Fig 3G), indicating close transcriptional similarity and distinction from myocytes. For example, *DES*, encoding for Desmin, was highly expressed in MM-SMCs and MP-SMCs but not in myocytes, indicating *DES* as a marker for non-mucosal SMCs in humans (Fig 3E).

To validate the suggested location of myocytes and SMCs in the human colon, we projected the signatures of each cell type onto spatial transcriptomics data of normal human colon [47]. As expected, myocyte signatures localized within the mucosa, MM SMCs in the muscularis mucosae, and MP SMCs in the muscularis propria (Fig 3H). Thus, our analysis provides a detailed annotation of colonic SMCs: mucosal myocytes, MM-SMCs, and MP-SMCs, together with their transcriptional and spatial characteristics (Fig 3I).

Characterization of MIFs. The *Pdgfra*^{high} *Kcnn3*+ MIFs were identified in both mouse and human integration analyses (Fig 1B, 1D and 1F; S2H Fig). Interestingly, MIFs did not resemble any of the previously identified colon fibroblast subclusters and, despite clustering with fibroblasts and expressing *PDGFRA,* also expressed a set of unique genes absent from all other fibroblast populations (Fig 4A and 4E). Interestingly, gene set enrichment analysis (GSEA) indicated neuron-related biological functions, including voltage-gated channel activity and ion transmembrane transporter activity for MIFs (Fig 4B and 4F). Tissue staining confirmed specific KCNN3 expression in the PDGFRA+ cells exclusively in the MP in the mouse (Fig 4C) and human colon (Fig 4H), as well as in the mouse small intestine (Fig 4D), consistent with the spatial transcriptomics data (Fig 1G).

Fig 4. Characterization of MIFs. (A) Dot plot depicting marker genes for mouse MIFs (muscle-embedded interstitial fibroblasts). (B) GSEA of the mouse MIF signature against indicated Gene Ontology Molecular Function (GOMF) gene sets. (C) Immunofluorescent staining of PDGFRA and KCNN3 in the muscularis propria in mouse colon and (D) small intestine. Dashed line indicates crypts. (E) Dot plot depicting marker genes for indicated human cell types. (F) GSEA of human MIF signature against indicated GOMF gene sets. (G) Normalized enrichment scores (NES) of GSEA of human MIF signature against gene sets from human colon-derived cells: PDGFRA+ cells of the smooth muscle (PaC), intestinal cells of Cajal (ICC), PDGFRA+ cells of the

colon mucosa (FIB), smooth muscle cells of the colon muscle (SMC), enteric neurons from colon myenteric ganglia (EN). (H) Immunofluorescent staining of PDGFRA and KCNN3 in the muscularis propria in human colon. (I) Graphical representation of MIFs in the mouse and human colon.

The neuron-related gene expression in MIFs and their clustering in the vicinity of ICCs (S2E Fig) prompted us to compare their transcriptional similarities with previously published datasets of human colon-derived cells, including 1) *PDGFRa*+ cells isolated from the smooth muscle layers, 2) ICCs, 3) *PDGFRa*+ fibroblasts isolated from the mucosa, 4) SMCs, and 5) enteric neurons [58, 59]. Notably, the MIF signature was highly enriched with the smooth muscle-derived *PDGFRa*+ cell gene set (NES >13) (Fig 4G). Thus, our analysis indicates that MIFs correspond to the previously described smooth muscle embedded PDGFRA+ fibroblasts (Fig 4I) [60-64] that have been largely missed in previous scRNA-seq analyses, presumably due to their small yield especially using mouse colon tissue dissection protocols.

Identification and characterization of CRC CAF subsets. Having annotated the normal fibroblast heterogeneity in human colon, we next set out to characterize the cancer-associated fibroblasts (CAF) in CRC. We integrated scRNA-seq data from 62 CRC patients (both cancer and normal adjacent tissue samples) across five cohorts (CRC16, JSC, KUL3, KUL5, SMC; S1E Fig) [26, 44], resulting in a total of 327.859 cells including epithelial cells, immune cells, and mesenchymal cells (Fig 5A; S5A, S5B and S5C Fig). To focus on the mesenchymal compartment, we re-clustered the mesenchymal cells (n = 56.484), including fibroblasts, SMCs, pericytes, endothelial cells, and glial cells (Fig 5B, S5D-F Fig). *KIT*+ ICCs and *CCL19*+ reticular fibroblasts were also present in the data but did not for separate clusters (S5G Fig). Interestingly, comparison of the mesenchymal cells derived from CRC and normal adjacent colon revealed major changes within fibroblasts, where four cancer-specific clusters emerged (Fig 5C). The endothelial tip cells were also primarily derived from cancer samples, reflecting their role in vessel sprouting during tumor vascularization. However, only minor changes were noted in other mesenchymal clusters, indicating that cancer context induces major transcriptional changes specifically in fibroblasts. The cancer-specific fibroblast clusters were named CAFs, consistent with their expression of the CAF marker *FAP* [65] (Fig 5D). Notably, however, the CRC samples also included all normal fibroblast

clusters, suggesting the presence of transcriptionally normal colon fibroblasts or representation of normal tissue in the resected CRC samples (Fig 5C and 5D, S5F Fig).

Fig 5. Identification of fibroblast subsets, their spatial distribution and transcriptional regulators in CRC. (A) UMAP of integrated scRNA-seq data from CRCs and normal adjacent tissue from 62 patients (n = 327.859 cells). (B) UMAP of re-clustered mesenchymal cells from A (n = 56.484). (C) UMAP shown in B split for CRC and normal adjacent tissue derived colon mesenchymal cells. (D) Dot plot depicting markers for the clusters shown in B and C. (E) Dot plot depicting CAF subtype signature genes across the whole dataset (clusters shown in A). (F) Spatial transcriptomics showing the stromal marker *VIMENTIN* (VIM) and CAF subtype signature (from E) expression in CRC tissue. (G) Zoom-ins of areas 2 and 3 (from F) with overlayed preCAF1 and preCAF2 signatures (left panels) and iCAF and mCAF signatures (right panels). (H) Fibroblast subtypes used for computational transcription factor knockout simulations. The pseudotime was calculated by allowing cells to take two simultaneous differentiation trajectories (from MIF to mCAF; from SEMF to iCAF). (I) Vector fields describing the development of the cellular fibroblast system without perturbation (control, gradient of pseudotime), or after simulated transcription factor knockouts. contrPeri, contractile pericytes; angPeri, angiogenic pericytes; venEC, venular endothelial cells; art&capEC, arteriolar and capillary endothelial cells; tipEC, endothelial tip cells, iCMS intrinsic consensus molecular subtype of CRC; macr, macrophages; tissueMacr, tissue-resident macrophages; neutr, neutrophils; DC, dendritic cells; mast, mast cells; KI67+ immune, proliferative immune cells; proT, proliferative T cells; Treg, regulatory T cells; Bcells, B cells; IgA+, IgA+ plasma cells; IgG+, IgG+ plasma cells; IgA+ IGKC+, IgA+ IGKC+ plasma cells; IgG+ IGKC+, IgG+ IGKC+ plasma cells.

Next, we investigated the CRC CAF subsets in more detail. One of the CAF populations specifically expressed *MMP1, CXCL1,* and *IL6* as well as other inflammatory genes, closely aligning with the previously described inflammatory CAFs (iCAFs) [66-69] (Fig 5D, S6A Fig). We also identified a cluster expressing the myofibroblast/matrix CAFs (mCAFs) genes, such as *MMP11, HOPX, COL10A1,* and *COL11A1* (Fig 5D, S6B Fig) [67, 70]. The two remaining CAF populations lacked clear specific markers and shared transcriptional features with both normal-like fibroblasts and iCAFs/mCAFs, suggesting transitional states, and were named precursor (pre)

CAFs. Namely, the preCAF1 cluster shared gene expression with SEMFs and iCAFs, (Fig 5D, S6C Fig), while the preCAF2 cells shared features with SAFs/MIFs and mCAFs (Fig 5D, S6D Fig). These observations prompted a hypothesis that preCAF1 and 2 may represent transitional states towards iCAFs or mCAFs, respectively, with different cells of origin: SEMFs for preCAF1 and SAFs/MIFs for preCAF2.

Next, we addressed the correlation of the CAF subset signatures (Fig 5E) to CRC patient survival probability. Interestingly, preCAF2 and mCAF signatures were correlated with reduced probability of overall survival, while a trend towards improved survival was observed for preCAF1 and iCAF signatures (S6E Fig). To explore the spatial distribution of the identified CAF subtypes, we projected their signatures (Fig 5E) onto CRC spatial transcriptomics data [47]. The preCAF1 and iCAF gene expression was found in the vicinity of cancer cells around the tumor edges, consistent with their suggested origin from SEMFs (Fig 5F and 5G). In contrast, preCAF2s and mCAFs were mainly located deeper in the tumor stroma, within the invasive front, consistent with SAF/MIF origin (Fig 5F and 5G). Importantly, the CAF signature genes were barely detectable in normal colon tissue, confirming their cancer-specificity (S7A Fig). Normal fibroblast signatures were, however, also observed in CRC tissue, suggesting that not all fibroblasts in CRC represent altered fibroblast states, consistently with our scRNA-seq analysis (S7B and S7C Fig).

In accordance with the transcriptional and spatial differences, trajectory and pseudo-time analyses using the Monocle3 algorithm [71] suggested that iCAFs and mCAFs develop via separate trajectories from distinct tissue-resident fibroblast subsets: SEMFs being the cells of origin for iCAFs through a preCAF1 state, and SAFs/MIFs polarizing into mCAF phenotype via preCAF2s (S7D-G Fig). To further explore gene regulatory changes during the CAF differentiation, we constructed gene regulatory networks (GRNs) for each identified fibroblast subtype and performed a gene ranking in which genes positioned across control points within the GRNs received a high score based on betweenness centrality (S7D Fig). Among the top-ranked genes, 22 were transcription factors (TFs). Next, we employed computational knockout simulations to assess how perturbations of these TFs would affect CAF differentiation. These simulations included the four CAF populations and normal fibroblast subtypes predicted to give rise to CAFs based on the Monocle3 analysis (SEMFs, SAFs, and MIFs) (S7E-H Fig), allowing them to undergo simultaneous differentiation along both identified trajectories towards iCAF and mCAF (Fig 5H). Using this approach, we identified *PRRX1*, *MAFB,* and *TWIST1* as potential TFs

regulating CRC CAF differentiation (Fig 5I). Consistently, the expression levels of the candidate TFs were highest in CAFs (S7H Fig).

Next, we investigated whether the relative abundances of the CAF subsets correlated with molecular and clinical features across CRCs (n = 40, S8A and S8B Fig). On average, 34.7 % of CAFs across these samples represented preCAF1; 14.1 % were iCAFs; 24.4 % preCAF2; and 26,8 % mCAFs (S9A Fig). Despite the relatively low numbers of samples falling to each category (n = 6 - 11 per group), we correlated the CAF subtype abundance across the CRC CMS 1-4 classification [23] (S9B and S9C Fig). Interestingly, the CMS2, the canonical subtype, presented with a significantly higher abundance of preCAF2s as compared to other subtypes (S9D Fig). Similarly, the CMS3, the metabolic subtype, showed a significantly ($p < 0.05$) higher abundance of preCAF1s (S9E Fig). Additionally, a tendency towards increased proportion of mCAFs in CMS4 was observed (S9F Fig). No consistent differences in CAF proportions between the iCMS2 and iCMS3 subtypes [44] or right-left sidedness were noted (S9G-J Fig), but rectal tumors presented with a higher abundance of preCAF1s as compared to colonic tumors (S9K and S9L Fig). MSI-H status showed a trend towards increased proportion of preCAF2, while iCAFs were significantly more abundant in early (1-2) versus late stage (3-4) CRC (S9M-R Fig).

We then examined the correlation of CAF proportions with the mutational status for *APC*, *KRAS*, *TP53, BRAF,* and *PIK3CA* (S10 Fig). Significant ($p < 0.05$) or borderline significant ($p = 0.055 - 0.066$) correlations were found with increased abundance of mCAFs in KRAS mutant cancers, and preCAF1s in *TP53* mutant and *PIK3CA* wild-type cancers (S9B and S9C Fig, S9f and S9G Fig, S9I and S9J Fig). Moreover, mCAFs were significantly enriched in late-stage *KRAS*-mutated cancers (S10D and S10E Fig).

CTHRC1 represents a pan-CAF marker in CRC. Multiple genes have been suggested as specific pan-CAF markers in CRC and other cancers, including *ACTA2, PDGFRA, PDGFRB, FAP, S100A4, PDPN, LRRC15, POSTN* or *TNC* [72, 73]. Of these, only *FAP* faithfully marked CAF populations in our analysis (Fig 6A). In addition to *FAP*, we identified *CTHRC1*, *INHBA*, *THBS2* and *WNT2* as the most specific and sensitive markers for CRC CAFs. Projecting the pan-CAF signature (*CTHRC1*, *INHBA*, *THBS2, WNT2, FAP)* (marked by asterisk in Fig 6A) onto CRC spatial transcriptomics data [47] showed clear signal in the tumor stroma but was barely detected in normal tissue as predicted (Fig 6B, S11A Fig). Following our data suggesting *CTHRC1* as a

pan-CAF marker in CRC, CTHRC1 protein staining was noted in mesenchymal cells in 96% of CRCs (n = 279/292) and was not detected in normal colon mesenchyme (n = 0/4) (Fig 6C and 6D, S11B). The extent of CTHRC1 staining did not, however, affect overall survival in this cohort (S11C Fig).

Fig 6. Identification of CTHRC1 as a pan-CAF marker in CRC. (A) Dot plot depicting previously proposed CAF markers (left, black box) and novel CRC pan-CAF markers identified in this study (right, red box). (B) CRC pan-CAF signature (*CTHRC1*, *INHBA*, *THBS2* and *WNT2*, from A) projected onto spatial transcriptomics data of CRC tissue. (C) Representative images of CTHRC1 staining in normal adjacent colon and CRC. (D) Quantification of CTHRC1 staining in normal adjacent colon and CRC mesenchyme. (E-G) Kaplan-Meier plots depicting the correlation between mRNA expression of the panCAF signature (*CTHRC1, INHBA, THBS2, WNT2, FAP*), as well as *CTHRC1* and *FAP* alone with overall survival probability in CRC. mRNA-sequencing data derived from: GSE12945, GSE12294, GSE14333, GSE123985, GSE17538, GSE10800, GSE26682, GSE20540, GSE21595, GSE33114, GSE34489, GSE37892, GSE38832, GSE39582, GSE41258, GSE92921. (H) Schematic representation and specific markers for the identified mouse and human colon fibroblast and SMC subpopulations (left) and CRC CAF subsets with suggested trajectories (right).

Since the abundance of CAFs appears to be increased in late-stage CRC (S11D Fig) and we found both *CTHRC1* and *FAP* highly expressed in CMS4 subtype of CRCs (S11E and S11F Fig) [74], we set out to investigate the effects of these markers on survival between early and late-stage CRC and across CMS statuses in multiple published RNAseq datasets. While the pan-CAF signature, as well as *CTHRC1* and *FAP* expression alone, correlated with poor overall survival in CRC (Fig 6E-G), the effects were specific to late-stage CRC (S11G-I Fig) and CMS4 patients (S11J-L Fig). Moreover, *CTHRC1* mRNA expression alone reflected the survival predictions obtained using the pan-CAF signature and *FAP,* indicating that *CTHRC1* mRNA expression alone is a predictor of overall survival in late-stage and CMS4 CRC. In summary, these results suggest that both the CAF signature expression and the pan-CAF marker *CTHRC1* expression alone are associated with reduced survival probability in late-stage and CMS4 CRC subtypes. Thus, CAFs indeed represent intriguing therapeutic targets.

Refined high-resolution model for colon fibroblast heterogeneity in homeostasis and CRC. Finally, we used our detailed analysis to construct a systematic and transparent summary of the fibroblast populations in both healthy colon and CRC (Fig 6H). We also compared our results with previously published data by matching the clusters using overlapping markers. Despite the very heterogeneous nomenclature, the reported gene expression patterns allowed us to match most of the previously proposed normal colon fibroblast (S3 Table) and CRC CAF subtypes (S4 Table) with the clusters identified in our analysis. Of note, our analysis indicates that the previously suggested CRC CAF subsets represented in addition to CAFs also cell types such as normal fibroblasts, pericytes, and SMCs. For example, we did not find a CAF cluster matching the proposed apCAFs (*CD74*+) [33, 35, 70, 75, 76], which may instead represent reticular fibroblasts (S5G Fig). We also failed to identify a separate pCAF (*MKI67*+) cluster [33, 42, 68]. Finally, to support easy comparison of our data and other studies, we provide an easy-to-use comparison of our annotation with previously suggested nomenclature, along with key marker genes and spatial information (Table 1).

Table 1. Colon fibroblast and CRC CAF subsets.

Organism	Annotation this study	Marker genes	Location	Annotation literature
Mouse	Top SEMF	Pdgfra-high, Foxl1, Wnt5a, Lama5	Crypt apex subepithelium	Subepithelial FB, Pdgfra-hi FB, Pdgfra-hi telocytes, telocytes, S2 FB, myofibroblasts, Pdgfra-hi SEMFs, iSEMFs, villus top FB, villus tip telocytes, crypt-apex FB, telocytes, epithelial proximal FB, WNT5B FB, Crypt-top FB,
Mouse	Bottom SEMF	Pdgfra-high, Foxl1, Wnt5a, Wif1	Crypt bottom subepithelium	Subepithelial FB, Pdgfra-hi FB, Pdgfra-hi telocytes, telocytes, S2 FB, myofibroblasts, Pdgfra-hi SEMFs, iSEMFs
Mouse	Top MAF	Pdgfra-low, Vcam1, Ccl8, Cd34, Wnt4	Crypt apex mucosal lamina propria	Crypt stromal cells, pericryptal myoFB, S1 FB, crypt-bottom/base FB, crypt telocytes, Cd81- stromal cells, F2 FB, parenchymal FB, villus FB, PDGFRA-lo interstitial FB, Mucosal associated FB, Pdgfra-lo Cd81- FB, Pdgfra-lo FB, crypt mesenchymal cells, Fgfr2+ FB
Mouse	Proximal bottom MAF	Pdgfra-low, Vcam1, Ccl8, Cd34, Tnc	Crypt bottom mucosal lamina propria, proximal colon	Crypt stromal cells, pericryptal myoFB, S1 FB, crypt-bottom/base FB, crypt telocytes, Cd81- stromal cells, F2 FB, parenchymal FB, villus FB, PDGFRA-lo interstitial FB, Mucosal associated FB, Pdgfra-lo Cd81- FB / stroma, Pdgfra-lo FB, crypt mesenchymal cells, RSPO3+GREM1+ FB 1&2, crypt FB, Igfbp5+/CD90+ FB
Mouse	Distal bottom MAF	Pdgfra-low, Vcam1, Ccl8, Cd34, Cd55	Crypt bottom mucosal lamina propria, distal colon	Crypt stromal cells, pericryptal myoFB, S1 FB, crypt-bottom/base FB, crypt telocytes, Cd81- stromal cells, F2 FB, parenchymal FB, villus FB, PDGFRA-lo interstitial FB, Mucosal associated FB, Pdgfra-lo Cd81- FB / stroma, Pdgfra-lo FB, crypt mesenchymal cells, RSPO3+GREM1+ FB 1&2, crypt FB, Igfbp5+/CD90+ FB
Mouse	SAF	Pdgfra-low, Pi16, Pcolce2, Cd34, Cd81	Submucosa	Crypt-bottom FB, crypt-base FB, crypt FB, CD81+ FB, Cd81+ trophocytes, trophocytes, adventitial fibroblasts, Interstitial FB, F1 FB, S3 FB, Pdgfra-lo FB, crypt mesenchymal cells
Mouse	MIF	Pdgfra-high, Kcnn3	Muscularis propria	PDGFRa+ KIT– cells of the musculature, FB-like cells, PDGFRα+ cells, telocytes, PDGFRa+ intramuscular cells, enteric mesothelial FB, Fibin+ FB, PDGFRA+ interstitial cells, Pdgfra+ stromal cells of thel intestinal muscle layers, PRAICs
Human	SEMF	PDGFRA-high, FOXL1, WNT5A, CXCL14	Subepithelium	Subepithelial FB, epithelial proximal FB, WNT5B FB, S2 FB, iSEMF
Human	MAF	PDGFRA-low, VCAM1, CCL8, ADAMDEC1, CXCL14	Mucosal lamina propria	Mesothelial serosa cells, WNT2B FB, S1 FB, lamina propria FB, WNT2-hi FB, TACI FB, Pdgfra-lo Cd81- FB
Human	SAF	PDGFRA-low, PI16, PCOLCE2, C3	Submucosa	S3 FB, trophocytes, mesothelial serosa cells
Human	MIF	PDGFRA-high, KCNN3, C3	Muscularis propria	Telocytes, PDGFRA+ interstitial cells, PDGFRα-expressing cells
Human	preCAF1	CTHRC1, WNT5A, CXCL14, F2RL2, TSPAN33	Tumor edge	CAFs5: endoplasmic reticulum-related CAF, HGF+ inflammatory FB, cluster 27: CXCL14+ CAF, CXCL14+INHBA+ FB, OGN+IGF+ FB
Human	preCAF2	CTHRC1, IGF1, GAS1, CRABP2, C3, COMP	Tumor core	CAFs5: endoplasmic reticulum-related CAF, HGF+ inflammatory FB, myoFB1, CAFs3: vascular-related CAF, PRELP FB, THBS2+ CAFs, C11_SERPINE1, cluster 31, CAF_Int, SFRP2+ myoFB, tCAF
Human	iCAF	CTHRC1, WNT5A, MMP1, MMP3, CXCL5	Tumor edge	C10_COMP, cluster 28: GREM1+ CAF, CTHRC1+ FB, myoFB 2, cluster 29: MMP3+ CAF, CAFs1: entoderm-related CAF, meCAF, CAF_MMP1, MMP1+ myoFB, IL6+ inflammatory FB, iCAF
Human	mCAF	CTHRC1, COL10A1, COL11A1, PLPP4, COMP	Tumor core	MyoFB1, CAFs3: vascular-related CAF, PRELP FB, THBS2+ CAFs, C10_COMP, cluster 28: GREM1+ CAF, CTHRC1+ FB, CAFinfla, mCAF, CAF_FAP, LRRC15+ myoFB, HOPX+ myoFB, MMP11+ myoFB, myCAF

Nomenclature, marker genes, localization, and previously suggested names for all mouse and human normal colon fibroblast and CRC CAF subtypes annotated in this study. For details and references, see S3 Table and S4 Table.

DISCUSSION

Recent discoveries highlighting fibroblasts as major regulators of tissue maintenance and pathological processes, including cancer, have fueled the need to understand fibroblast heterogeneity in healthy and pathogenic tissues [12, 77-81]. This study aimed to (1) generate a high-resolution map and annotation of mouse colon fibroblast subclusters; (2) characterize human colon fibroblasts using benchmarking data from mice and spatial transcriptomics; and (3) identify and annotate CRC CAF subsets and investigate their potential differentiation trajectories and association with clinical parameters.

Several studies have addressed fibroblast heterogeneity in mice, resulting in a consensus view of the three major subclasses [4-11] while the nomenclature remains variable (S3 Table). The heterogeneity of human colon fibroblasts is less well established. Our side-by-side integrated analyses of mouse and human colon fibroblasts, followed by spatial validation, allowed us to establish a coherent, location-based nomenclature using previously proposed (SEMF, MAF) [8] and novel terms (SAF, MIF). Importantly, we found remarkable similarities between mouse and human colonic fibroblast organization and provide both shared and species-specific markers for the identified fibroblast subsets, allowing easy comparison with other datasets (Fig 6H, Table 1). It is likely that the fibroblast identity represents a cell state, rather than cell type, and is largely regulated by the tissue context. Indeed, PI16-expressing fibroblasts have been proposed to be associated with vasculature and in colon, PI16 expression localizes close to the endothelium, suggesting that the proximity of endothelial cells may be essential for the SAF phenotype [82]. Moreover, fibroblast subtypes have been shown to be re-established at correct locations after tissue damage [83]. Interestingly, though, cell culture experiments addressing the functional properties of isolated primary fibroblast subsets suggest at least some degree of stability in short-time cell culture models [4, 6, 9, 10, 51]. Establishing conditions that maintain the fibroblast state for extended periods *ex vivo* would benefit the functional characterization of the identified fibroblast subsets.

Our study complements the existing knowledge of mouse colon fibroblasts subsets by showing that SEMFs and MAFs further divide into distinct subpopulations, reflecting gradients along the crypt-apex (basal-luminal) and proximo-distal axis. Notably, similar observations have been reported before in the small intestine and colon [4, 11]. The crypt-apex gradient appears to take place also in the mouse small intestine, where the low representation of top SEMF and MAF

subtypes and a lack of MIFs likely stems from the commonly used protocols, where both the intestinal villi and muscle layers are excluded in preparation of single cell suspensions. Our analysis suggests that human fibroblasts do not form similar subpopulations, but it remains possible that our dataset lacked sufficient power to reveal further subclusters.

SMCs are characterized by the expression of genes such as *ACTA2*, *ACTG2*, *MYH11,* and *DES*, while the cells co-expressing SMC markers and *HHIP* are often referred to as myofibroblasts [2, 4, 5, 13, 43, 56, 57, 84]. In our analysis of the normal colon, SEMFs were the only cell type with a clear fibroblast identity (clustering with other fibroblasts, expressing *PDGFRA*), together with expression of the SMC markers *Acta2* (in mice) and *ACTG2* (in humans), and could therefore be classified as myofibroblasts. Since the *Acta2*^{high} *Hhip*+ cells lacked fibroblast markers but were transcriptionally and spatially distinct from MM-SMCs and MP-SMCs, we propose the term myocyte for these cells, consistent with suggestions in the small intestine and stomach [9, 10]. The myocytes are located in the mucosal core of the colonic and intestinal lamina propria, and *HHIP* encodes for a Hedgehog signaling inhibiting factor. Since the intestinal and colonic epithelium express Hedgehog ligands [85], we hypothesize that the surface-core gradient of Hedgehog signaling could contribute to mucosal SEMF and MAF identity. Interestingly, myocytes also express *Pdgfc,* a PDGFRA ligand and a factor regulating endothelial growth. Thus, myocyte-derived PDGFC could play a role in programming the MAF and SEMF identities either directly or indirectly by regulating vascular growth, which affects oxygen abundance. Indeed, it was recently reported that hypoxia induces expression of WNT5A (SEMF marker) and an inflammatory fibroblast phenotype in the edges of colonic tumors [86], indicating that the SEMF identity might be driven, at least partly, by hypoxia.

Interestingly, we observed a *PDGFRA*^{high} *KCNN3*+ population clustering with fibroblasts and localizing within the muscularis propria and therefore named them MIFs. We propose that MIFs represent the fourth major colon fibroblast subset but were missed in previous mouse scRNA-seq experiments due to the removal of the muscle layers in the cell dissociation protocols. The corresponding human cluster was more prominent, and the human datasets also contained cells from muscularis propria, reflecting the inclusion of the muscle wall in CRC resections. Thus, the sampling method (e.g., biopsy vs. resection) is an important aspect to consider when comparing fibroblast compositions across different samples. MIFs likely correspond to the recently reported PDGFRA^{high} interstitial cells (PRAICs) of the mouse small intestine, although the marker genes

for PRAICs were not reported [10]. Interestingly, other methods have revealed PDGFRA and KCNN3-expressing cells in the vicinity of ICCs throughout the gastrointestinal tract smooth muscle layers [60-64]. Those cells have been suggested to participate in the regulation of SMCs in concert with ICCs and neurons [87-91]. Thus, we link the scRNA-seq findings with a previously characterized cell type and suggest the *PDGFRA*^{high} *KCNN3*+ cells represent fibroblasts that display neuronal properties and likely act as regulators of the smooth muscle layer functions, such as peristalsis.

In CRC, CAFs have emerged as key players and drug targets, mandating a detailed understanding of their heterogeneity. Tens of different names have been proposed for suggested CAF subsets (S.Table 4), but a coherent understanding has not been established. This is in part due to the incomplete understanding of the normal human colon fibroblasts and other mesenchymal cell subtypes. Our results indicate that the scRNA-seq analyses performed to date identify rather reproducible cell types/states, but only a subset of the proposed populations represents CAFs (S.Table 4). Importantly, our analysis revealed only four cancer-specific (CAF) clusters but showed that all the normal fibroblast subtypes persist within CRC. Interestingly, we identified iCAFs and mCAFs – the most well-characterized CAF subtypes – which expressed similar marker genes observed in corresponding CAF populations in other tissues [66-70]. However, in contrast to pancreatic cancer, where iCAFs were shown to localize at a distance from cancer cells and be associated with poor survival [92], in CRC, iCAFs are seem to be situated closer to the cancer cells, while mCAFs reside closer to the invasive front, and the mCAF signature expression predicts reduced survival probability. This is consistent with the recent report indicating expression of WNT5A in the tumor edge in iCAF-like cells that derive from BMP-expressing normal fibroblasts [86]. The other two CAF populations, preCAF1 and preCAF2, shared markers with SEMFs and SAFs/MIFs, respectively, and were interpreted as transitional states. Indeed, our trajectory and pseudotime analyses indicated that iCAFs develop from SEMFs via preCAF1s, and mCAFs arise from SAFs/MIFs through a preCAF2 state. Thus, only two functionally different CAF subtypes/trajectories may exist in CRC. Additionally, our *in silico* knock-out simulations identified three transcription factors, *PRRX1, MAFB,* and *TWIST1*, which showed potential regulatory functions in CAF differentiation. While the roles of MAFB and TWIST1 in CRC CAF differentiation are not yet characterized, PRRX1 has previously been shown to regulate CRC CAFs, supporting the validity of our data-driven approach for identifying transcriptional regulators

critical for fibroblast regulation during colon cancer [93]. Thus, these genes represent an intriguing avenue of research for the future.

The cellular origin of CAFs in CRC remains under debate, with different studies proposing tissue-resident fibroblasts and other cells such as bone marrow-derived mesenchymal stem cells [94], SMCs [95], pericytes [96], endothelial cells [97], epithelial cells [98], and adipocytes [99, 100] as sources of CAFs. Indeed, pericytes are commonly annotated as fibroblasts or CAFs in the CRC field despite their distinct transcriptional signature (*RGS5, NOTCH3, PDGFRB, NDUFA4L2*) [101-104]. In our analysis, pericytes were transcriptionally distinct from the CAF populations. Moreover, CAFs clustered with fibroblasts, and the trajectory analysis suggested a fibroblast origin. Thus, we propose that the majority of CAFs originate from tissue-resident fibroblasts, and that some of the previously identified subsets such as vascular CAFs [39, 68, 75, 105] may represent pericytes and CRC apCAFs [33, 35] represent reticular fibroblasts that also express high levels of *CD74* (S5G Fig). Additionally, we did not identify a distinct *MKI67*+ proliferative pCAF [33, 42] population (S5G) and hypothesize that while apCAFs and pCAFs may be represented in some cancer types, they do not form a specific entity in CRC. Of note, contributions from circulating mesenchymal stem cells cannot be excluded by our analysis.

The different CAF subset proportions between CRC CMS subtypes and genetic features suggest that fibroblast signatures may underlie some of the CMS classifications, and that different mutations could influence the CAF subtype composition, placing fibroblast phenotypes as important determinants of patient survival and treatment efficacy. Interestingly, our results suggest an increased abundance of mCAFs in CMS4 patients, commonly known to display a fibroblast-rich tumor phenotype and poor patient survival. Additionally, *KRAS* mutations were associated with high mCAF proportion in late-stage CRC. Consistently, a previous study in pancreatic cancer suggested that targeting mutant *KRAS* in cancer cells reprograms CAFs towards normal-like fibroblasts, raising the possibility that targeted therapies might also alter CAF subtype composition in CRC [106]. Thus, we hypothesize that targeting CAFs, and especially mCAFs, could benefit CRC patients belonging to the CMS4 subgroup and/or carrying the epithelial *KRAS* mutation, especially in late-stage CRC. It will be interesting to address in larger datasets if CAF composition may drive transcriptional differences between the CMS subtypes, and whether specific composition of mutations could play a role in the CAF reprogramming in CRC.

Specific and sensitive markers that differentiate CRC CAFs from normal fibroblasts could serve as valuable biomarkers or drug targets. Our analysis revealed that among previously suggested CAF markers, only *FAP* was expressed across CAFs and absent from other cell types. Instead, our analysis identified *CTHRC1*, *INHBA*, *THBS2,* and *WNT2* as CAF-specific markers. Interestingly, tissue staining confirmed that CTHRC1 protein expression was specific to the cancer stroma. CTHRC1 plays a crucial role in ECM remodeling by limiting collagen deposition and enhancing ECM plasticity, which are reported cancer-promoting functions driven by CAFs [107-111]. Therefore, CTHRC1 might be functionally linked to tumor progression by facilitating cancer cell migration and angiogenesis. Given its secreted nature and expression already in the preCAFs, CTHRC1 could potentially serve as a biomarker for early-stage CRC and thus represents a compelling target for further investigation. Interestingly, *CTHRC1* was previously suggested to mark a specific CAF subset in CRC [32, 112]. Our analysis suggests that *CTHRC1*+ fibroblasts represent all CAFs, while the other clusters represent normal tissue-resident fibroblasts.

In summary, our integrative analysis enabled the construction of a high-resolution model for fibroblast heterogeneity with consistent location-based nomenclature for both mouse and human colon fibroblasts, incorporating both previously validated and newly discovered marker genes. Additionally, our study provides a systematic characterization of CRC CAF subsets, offering a standardized framework for future studies into the colon fibroblast heterogeneity in both health and disease.

MATERIALS AND METHODS

Mice. Two C57Bl/6 mice (one male and one female) were used for the isolation of stromal cells for single-cell RNA sequencing (scRNA-seq) from the colon and small intestine. Mice were bred and maintained in the animal house facilities of BSRC Alexander Fleming under specific pathogen-free conditions, and experiments were performed in accordance with current European and national legislation (approved by BSRC Fleming's Institutional Committee of Protocol Evaluation in conjunction with the Veterinary Service Management of the Hellenic Republic Prefecture of Attika). Male and female C57BL/6JRccHsd mice (Mus musculus) were used for tissue collection for the validation experiments. These animals were housed at the University of Helsinki Laboratory Animal Center according to national and international legislation and guidelines (license: KEK23-006, approved by the Research Ethics Committee on Animal Research). Mice had free access to food and water.

Isolation of intestinal stromal cells by fluorescence-activated cell sorting (FACS). Isolation of intestinal stromal cells was performed as previously described [83]. Briefly, intestinal tissues were dissected, opened longitudinally, and washed with HBSS (Gibco) supplemented with antibiotic-antimycotic solution (Gibco). Epithelial cells were removed by incubating the tissue with 5 mM EDTA and 1 mM DTT in HBSS for 25 minutes at 37°C. The remaining tissues was pooled and digested in 300 U/mL Collagenase XI (Sigma-Aldrich) and 1 mg/mL Dispase II (Roche) in DMEM for 40-60 min at 37°C. The cell suspension was centrifuged, washed three times in FACS buffer (5% FBS (Biowest) in PBS), and the cells were counted. 1-2 million cells/100 μL were incubated with the APC-Cy7 conjugated anti-CD45 (clone 30-F11, BioLegend) and the APC-efluor780 conjugated anti-CD326 (clone G8.8, eBioscience) antibodies. Propidium Iodide (Sigma) was used for live-dead cell discrimination. Cell sorting was performed using the FACSAria III cell sorter (BD) and the FACSDiva (BD) software.

Single-cell RNA library preparation and sequencing. 100.000 FACS-sorted stromal cells from the colon and the small intestine were used fresh for the preparation of two single-cell RNA-sequencing libraries according to the 10X Chromium Single Cell 3' v3.1 protocol and sequenced using the DNBSEQ-G400 sequencer (PE100) (BGI Genomics). FASTQ files containing raw reads were initially processed, including read alignment, unique molecular identifier (UMI) counting,

filtering of empty droplets, and the generation of sample-specific count matrices, using the 10X Genomics CellRanger pipeline (version cellranger-7.1.0) function 'cellranger count', at default settings. Read pairs were aligned to the mouse reference genome (mm10, 10X-reference genomes, v2020). The number of read pairs assigned to each library and the estimated cell numbers were 237,166,873 bp and 6,990 cells for the colon and 210,226,057 bp and 11,576 cells for the small intestine, respectively. Original raw and processed scRNAseq data has been deposited at the Gene Expression Omnibus under the accession number GSE296873.

ScRNA-seq data analysis. All scRNA-seq data in this study was analyzed using the Seurat algorithm V4 [113] in R (version 4.2.3). All Seurat objects were created by setting the minimum number of cells per sample to 3 and the minimum number of genes expressed per cell to 200. All objects were normalized using SCTransform.

Mouse colonic scRNA-seq data produced for this publication was filtered using nFeature RNA 1000<n<4000, nCountRNA 1000<n<13.000, percent.mt<6, percent.rb 3<n<25 and the small intestinal data was filtered using nFeature RNA 500<n<5000, nCountRNA 750<n<15.000, percent.mt<5.5, percent.rb 2.5<n<20. Variable regression included %mt and %rb. The data was clustered (graph-based Louvain clustering) using 20 and 25 principal components (PCs) and resolution 0.5 and 0.4 for colon and small intestine, respectively. For further analysis, *Pdgfra*+ fibroblasts and *Myh11*+ SMCs were subsetted and re-clustered using 15 and 25 PCs and a resolution of 0.15 and 0.3 for colon and small intestine, respectively.

Published scRNA-seq data from healthy mouse colon was obtained from the GEO database, quality controlled (QC), and integrated with the scRNA-seq data produced for this study (S1 Table) [4, 5, 7, 8]. The data was filtered according to S1 Table. Variable regression included %mt, %rb, and cell cycle scores. The data was integrated using the Seurat anchor-based integration workflow: top 3.000 variable features, RPCA, and k.anchor 7. The integrated data was clustered (graph-based Louvain clustering) using 30 principal components (PCs) and resolution 0.7. Based on low overall gene expression (nFeatureRNA), low RNA counts (nCountRNA), high ribosomal RNA expression (%rb), and high expression of *Ptprc* (CD45), two small populations were filtered out, and the data were re-clustered using 15 PCs and resolution 0.6. Cell type annotation was done using validated marker genes (Fig 1B, S2B Fig).

Human scRNA-seq data were obtained from synapse.org under the project ID syn26844071 [44]. For analyzing homeostatic human colon tissue, only samples from normal adjacent tissue were included, and lymph node data was filtered out. The data had already been quality controlled by the original authors. Further QC and filtering were done on epithelial data using a cutoff for nFeatureRNA at 6.000 and nCountRNA at 25.00,0 and on stromal data using a cutoff for nFeatureRNA at 4.500 and nCountRNA at 35.000. The data was integrated using the Seurat anchor-based integration workflow: top 3.000 variable features, RPCA, and k.anchor 10. The integrated data was clustered (graph-based Louvain clustering) using 25 PCs and resolution 0.5. One sample (patient CRC2786) displayed features of tumor samples and was thus filtered out, resulting in a total of 36 patients. For further analysis, epithelial and immune cells were filtered out, and the data was re-clustered using 20 PCs and resolution 0.4. Cell type annotation was done using validated marker genes (Fig 1D, S2D Fiig).

For the integrative analysis of human normal adjacent colon tissue and CRC tumor tissue, data were obtained from synapse.org under the project ID syn26844071 [44]. Lymph node data was filtered out. The data had already been quality controlled by the original authors. Further QC and filtering were done on epithelial data using a cutoff for nCountRNA at 50.000 and on stromal data using a cutoff for nFeatureRNA at 5.500 and nCountRNA at 70.000. The sample from patient CRC278 was filtered out to match the normal tissue analysis. This resulted in samples from 62 patients. The data was integrated using the Seurat anchor-based integration workflow: top 3.000 variable features, RPCA, and k.anchor 10. The integrated data was clustered (graph-based Louvain clustering) using 30 PCs and resolution 0.5. Cell type annotation was done using validated marker genes (Fig 5D, S5B Fig). For the mesenchymal cell analysis (Fig 5B-D), only samples with a minimum of 50 cells in a minimum of 5 different stromal cell clusters were included and the small PECAM1 and RGS5 double-positive cluster likely representing doublets formed of endothelial cells and pericytes was filtered out. This resulted in 53 patients (S5F and S5G Fig). Mesenchymal cell type annotation was done using validated marker genes (Fig 5D). For CAF abundance analyses, only the four CAF populations were subsetted from the mesenchymal analysis, and samples with a minimum of 50 CAFs in total were included. This resulted in 40 patients and a total of 15.022 cells (S8A and S8B Fig). CAF abundances were compared using the Wilcoxon rank-sum test. Variability is represented as 1.5 times the interquartile range (IQR), as shown by the whiskers in box plots.

Spatial transcriptomics data analysis. Spatial transcriptomics data (Visium HD, 10x Genomics) from normal human colon and CRC tumors were obtained from https://www.10xgenomics.com/products/visium-hd-spatial-gene-expression/dataset-human-crc [47]. Analysis was executed in the Loupe Browser 8 app. Marker gene signatures (lists of genes, S.Table 2) were projected onto the spatial data and visualized using the following settings: Combination: FeatureSum; Scale value: LogNorm; Saturation: -100; Opacity: 30%, and images exported as tiffs.

Immunofluorescent staining of mouse tissue. Small intestinal and colon tissues were collected, washed with PBS, and fixed in 4% paraformaldehyde at 4°C for 24 h. The tissues were embedded in paraffin as Swiss rolls with a Sakura Tissue-Tek VIP 5 Jr tissue processing system. Then, 5 µm sections were cut using a microtome (Shandon AS325, Marshall Scientific). Paraffin-embedded sections were deparaffinized in a xylene and ethanol series. Antigen retrieval was performed using 1× Dako antigen retrieval solution at pH 9 (S2367, Agilent Technologies) at 95°C for 25 min, followed by a 30 min incubation at room temperature (RT). Samples were washed 2 x 5 min with PBS, and permeabilization was performed with 0.1% Triton X-100 in TBS (TBS-T) for 10 min at RT, followed by blocking for 30 min using TNB blocking buffer. The samples were incubated with anti-PDGFRa (AF1062, R&D systems, 1:100) and anti-KCNN3/KCa2.3 (APC-025, Alomone Labs, 1:400) antibodies overnight at 4°C in blocking buffer. The next day, samples were washed 5 min with TBS containing 0.1% Tween-20 and 2 x 10 min with PBS. Then, samples were incubated with secondary antibodies [AlexaFluor 546 donkey anti-goat IgG (A11056, Thermo Fisher Scientific) and AlexaFluor 647 donkey anti-rabbit IgG (A31573, Thermo Fisher Scientific)] were incubated in blocking buffer (1:500 dilution) for one hour at RT and washed 3 x 5 min in TBS containing 0.1% Tween-20. Nuclei were stained with 10 µg/ml Hoechst 33324 (62249, Thermo Fisher Scientific) in PBS for 3 min. Slides were rinsed with TBS and water before mounting with Immu-Mount (9990402, Thermo Fisher Scientific).

Immunofluorescent staining of vibratome sections from mouse tissue. The mice were perfused with 15 ml of PBS, followed by 15 ml 4% paraformaldehyde. Small intestinal and colon tissues were collected as a whole, washed with PBS, and flushed and fixed with 4% paraformaldehyde

overnight at 4°C. Tissues were cut into 1 cm pieces and embedded in 3 % TopVision Low Melting Point Agarose in PBS. Then, 70-100 µm sections were cut using a vibratome (Vibratome Series 1000 – Tissue Sectioning System) and stored in 5 % Na-Azide in PBS. Permeabilization was performed with 0.1% Triton X-100 in PBS (PBS-T) for 3 x 5 min at RT, followed by blocking for 1 h using 0.3% PBS-T containing 3 % goat serum and 3 % donkey serum. The samples were incubated with anti-PDGFRa (#3174, Cell Signaling, 1:100) and anti-TNC (MAB2138, R&D Systems, 1:200) antibodies overnight at 4°C in 0.3% PBS-T containing 1 % goat serum and 1 % donkey serum. Samples were then washed 3 x 15 min with 0.1 % PBS-T and incubated with secondary antibodies [AlexaFluor 594 goat anti-rat IgG (A11007 Thermo Fisher Scientific, 1:500) and AlexaFluor 647 donkey anti-rabbit IgG (A31573, Thermo Fisher Scientific, 1:500)] in 0.3% PBS-T containing 1 % goat serum and 1 % donkey serum for 1 h at RT and washed 3 x 10 min in 0.1% PBS-T. Nuclei were stained with 10 µg/ml Hoechst 33324 (62249, Thermo Fisher Scientific) in PBS for 10 min at RT. Samples were washed 3 x 10 min with PBS and mounted with Immu-Mount (9990402, Thermo Fisher Scientific).

mRNA *in situ* hybridization. RNA in situ hybridization was performed using RNAScope Multiplex Fluorescent Reagent Kit v2 (323110, Advanced Cell Diagnostics) according to the manufacturer's instructions. Small intestinal and colon tissues were collected, washed with PBS, and fixed in 4% paraformaldehyde for 24 h. The tissues were embedded in paraffin as Swiss rolls with a Leica Histocore Pearl tissue processing system. Then, 5 µm sections were cut using a microtome (Microm HM355, Thermo Fisher Scientific). Paraffin-embedded sections were deparaffinized in a xylene and ethanol series, and pre-treated with RNAScope Hydrogen Peroxide, Protease Plus, and Target Retrieval Reagents. Sections were hybridized with Lama5 (494911), Wif1 (412361), or Hhip (448441) probes (all from Advanced Cell Diagnostics). Signal amplification hybridization was performed, and signal detected with Opal 570 (FP1488001KT, Akoya Biosciences, 1:1000). Following hybridization steps, the tissue sections were immunostained. Samples were blocked with 5 % Normal Goat Serum in PBS for 45 minutes and incubated with primary antibodies [anti-E-Cadherin (610182, BD, 1:1000), anti-alpha smooth muscle Actin (ab5694, Abcam, 1:1000)] overnight at +4°C in blocking buffer. Samples were washed 3 x 10 minutes with PBS and incubated with secondary antibodies [Alexa Fluor 488 Goat anti-mouse IgG2a (115-547-186, Jackson ImmunoResearch, 1:500), Alexa Fluor 647 Goat anti-

Rabbit IgG (A21245, Thermo Fisher Scientific, 1:1000)] and 1 µg/ml DAPI (D9542, Sigma-Aldrich). Samples were washed 3 x 10 minutes with PBS and mounted with Immu-Mount (9990402, Thermo Fisher Scientific).

Gene regulatory networks, gene importance ranking and computational knockout simulations. To investigate the regulatory mechanisms underlying CAF differentiation, we employed the CellOracle pipeline [114] (version 0.20.0) to construct gene regulatory networks (GRNs) and perform *in silico* transcription factor (TF) knockout simulations. ScRNA-seq data was integrated with prior regulatory information to infer context-specific GRNs, which were then used to build linearized dynamical models approximating how gene expression responds to regulatory input. GRNs were constructed using the CellOracle Python package to model transcription factor (TF)–target gene interactions across fibroblast subtypes. We used previously described preprocessed scRNA-seq data, which included clustering and dimensionality reduction (UMAP) information, as well as raw expression counts, which were used for the modelling. For each fibroblast subtype, a context-specific GRN was inferred based on gene expression and a pre-defined base GRN. The context-specificity was achieved by trimming and refining the base GRN to fit it for the transcriptomic context for specific cell types or states. The trimming consisted of: (1) filtering the base GRN to retain TF-target pairs where both genes were expressed in the given cluster, (2) fitting a ridge-regression model to infer regulatory coefficients for each TF-target pair and (3) removing weak or statistically non-significant ($p<0.001$) edges based on statistical significance derived from the ridge-regression step. The gene importance ranking was directly performed from the trimmed, context-specific GRNs for each fibroblast subtype by calculating betweenness centrality scores to prioritize genes acting as major control points within the networks.

To allow computational knockout simulations, pseudotime was precomputed using scanpy's diffusion pseudotime algorithm [115]. After pseudotime calculation, for each TF of interest, the expression value was individually set to 0.0 (knocked out), and the downstream effects on gene expression were estimated by approximating signal passing within the gene regulatory network. The model then calculates updated cell states (post-perturbation) by simulating the immediate transcriptional response (detailed in the original publication for CellOracle), and the

updated cell states are then projected back onto the pre-existing low-dimensional cell manifold (e.g., UMAP used in this study), preserving biological interpretability.

Human colorectal cancer study population. This study comprised a cohort of 315 patients with colorectal cancer (CRC) who underwent surgical treatment between 1998 and 2003 at the Department of Surgery, Helsinki University Hospital. Clinical data were retrospectively collected from patient records. Survival data were obtained through linkage with the Finnish Population Registration Centre and Statistics Finland. The median age at diagnosis was 67.8 years (interquartile range, 57.8–67.8 years). The median disease-specific survival was 5.9 years, with a range of 0.0–18.8 years. Ethical approval for the study was granted by the Surgical Ethics Committee of Helsinki University Hospital (Dnro HUS/1223/2021 and permission to use the tissue samples was granted by The Finnish Medicines Agency (Dnro FIMEA/2021/006901).

Preparation of tumor tissue microarrays. Formalin-fixed, paraffin-embedded (FFPE) tumor samples from surgical specimens were retrieved from the archives of the Department of Pathology, University of Helsinki. All tumor samples were re-evaluated by an experienced pathologist using hematoxylin and eosin (H&E) -stained sections to confirm diagnoses and select representative tumor areas. Two punches, each 1.0 mm in diameter, were extracted from each tumor block using a semi-automated tissue microarray instrument (TMA) (Beecher Instruments, Silver Spring, MD, USA). Each TMA block was sectioned, resulting in two tissue cores per tumor sample.

Tumor microarray immunohistochemistry. TMA blocks were sectioned at 4 µm thickness, mounted onto slides, and dried at 37°C for 12–24 hours. Antigen retrieval was performed using the Envision Flex target retrieval solution (pH 9, DM828, Agilent Dako, CA, USA) in a PreTreatment module (Agilent Dako) at 98°C for 15 minutes. Immunohistochemical staining was conducted with an Autostainer 480S (LabVision Corp., Fremont, CA, USA) utilizing the Dako REAL EnVision Detection System, Peroxidase/DAB+, Rabbit/Mouse. Sections were initially treated with Envision Flex peroxidase-blocking reagent (SM801, Dako) for 5 minutes to block endogenous peroxidase activity. Subsequently, the slides were incubated with anti-CTCHR1 antibody (ab256458, Abcam, Cambridge, United Kingdom), diluted 1:400 in Dako REAL Antibody Diluent (S2022, Dako), for 60 minutes. Afterward, slides underwent a 30-minute

incubation with peroxidase-conjugated EnVision Flex/HRP rabbit/mouse reagent (SM802, Dako). Visualization was achieved using DAB chromogen (EnVision Flex DAB, DM827, Dako) for 10 minutes. Counterstaining was performed using Mayer's hematoxylin (S3309, Dako). Normal colon served as the positive control.

Interpretation of tumor microarray immunohistochemical staining. CTCHR1 immunoreactivity was independently assessed by two investigators (E.W.V. and J.H.), who were blinded to clinical outcomes. Immunostaining intensity was evaluated separately in tumor/epithelial cells and stromal cells, using a scoring system: negative (0), low (1), moderate (2), and strong immunoexpression (3). The highest score for each sample across all assessed compartments was considered representative of CTCHR1 expression. In cases of discrepancy between the two evaluators, consensus scoring was applied.

Statistical analysis of tumor microarray staining. Survival analysis was done by the Kaplan-Meier method and the log-rank test. Disease-specific overall survival was counted from the day of surgery to the date of death from CRC, or until the end of follow-up. A p-value of 0.05 was considered significant. All tests were two-sided. All statistical analyses were done with SPSS version 29.0 (IBM SPSS Statistics version 29.0 for Mac; SPSS Inc., Chicago, IL, USA, an IBM Company).

Immunofluorescent staining of human normal adjacent colon tissue. Normal adjacent colon tissue sections from the CRC study cohort described above were used for immunofluorescent staining. 5 µm-thick sections were deparaffinized in xylene and ethanol series. Antigen retrieval was performed using 1X Dako antigen retrieval solution pH 6 (S1699, Agilent) at +95°C for 20 min, followed by 30 min at room temperature (RT). Blocking was performed using TNB blocking buffer (SKU FP1012, Akoya Biosciences) for 30 min at RT. Primary antibodies anti-KCNN3 (APC-025, Alomone Labs, 1:100 dilution) and anti-PDGFRa (AF-307-NA, R&D Systems, 1:50 dilution) were incubated overnight at +4°C. Secondary antibodies Alexa Fluor 647 donkey anti-rabbit (A-31573, Thermo Fisher Scientific) and Alexa Fluor 594 donkey anti-goat (A11058, Thermo Fisher Scientific) were incubated in blocking buffer (1:500 dilution) for one hour at RT and washed 3x5 min in 0.1% TBS-Tween. The nuclei were stained with 10 µg/ml Hoechst 33324

(62249, Thermo Fisher Scientific) in PBS for 3 min. Slides were rinsed with TBS and water before mounting with Immu-Mount (9990402, Fisher Scientific).

Imaging. Imaging of immunofluorescent human paraffin and mouse vibratome sections was done using the Zeiss Axio Imager M2 microscope using 20x magnification. Image analysis was conducted using Fiji ImageJ [116]. Images of immunofluorescent mouse paraffin sections, CRC TMA and RNA *in situ* hybridization sections were generated using 3DHISTECH Pannoramic 250 FLASH III digital slide scanner (20x and 40x magnification, respectively) at Finnish Genome Editing Center and analyzed with QuPath [117], CaseViewer (3DHISTECH) and CaseViewer, respectively.

Gene set enrichment analysis. The GSEA was conducted using clusterProfiler [118] in R (version 4.2.3). Ranked lists for mouse MIF gene expression were created with the FindMarkers function in Seurat, comparing MIFs to all other clusters of the data using logfc.threshold = 0.1 and min.pct = 0.05. For humans, min pct was set to 0.25. For mouse GSEA, the mouse MSigDB ontology geneset M5 subcategory "molecular function (MF)" was used, and for human, the human MSigDB ontology geneset C5 subcategory "molecular function (MF)" was used [119-122]. Additionally, signatures of 250 top expressed genes were extracted from PDGFRA+ cells isolated from human colon muscle (PaC), ICCs isolated from human colon muscle (ICC), PDGFRA+ fibroblasts isolated from human colon mucosa (FIB), smooth muscle cells isolated from human colon muscle (SMC) [59] and enteric neurons (myenteric ganglia) isolated from human colon (EN) [58] and used for GSEA with a ranked list of human MIFs constructed with the FindMarkers function in Seurat comparing MIFs to all other clusters of the data using the default settings of the function.

Trajectory and pseudotime analysis. Trajectory analysis was performed using the Monocle 3 algorithm [71]. Cells were clustered with resolution = 0.00012, which resulted in the best match with the Seurat clustering. Based on our analysis of normal human colon fibroblasts and the partitioning, where CAFs were in the same group as SAFs, MIFs, and SEMFs, we set SAF, MIF, and SEMF populations as originating clusters for the trajectory and pseudotime analyses performed using Monocle 3.

CRC patient survival analysis. Survival analysis was performed using the Kaplan-Meier plotter website for colorectal cancer data (https://www.kmplot.com/analysis/index.php?p=service&cancer=colon) using the CAF subpopulation gene signatures (mean expression of selected genes). Autoselection for cutoff was used (percentile) for overall survival analysis. The cohorts included a total of 2.135 patients from the following datasets: GSE12945 (n = 62), GSE12294 (n = 155), GSE14333 (n = 123), GSE123985 (n = 91), GSE17538 (n = 232), GSE10800 (n = 53), GSE26682 (n = 331), GSE20540 (n = 35), GSE21595 (n = 37), GSE33114 (n = 90), GSE34489 (n = 33), GSE37892 (n = 65), GSE38832 (n = 70), GSE39582 (n = 514), GSE41258 (n = 185), GSE92921 (n = 59). Not all patients had the required clinical information for each analysis (e.g., CMS status, stage, etc.). In these cases, all patients with the required information were used in the analyses.

Ethics statement. For scRNA-seq, mice were bred and maintained in the animal house facilities of BSRC Alexander Fleming under specific pathogen-free conditions, and experiments were performed in accordance with current European and national legislation (approved by BSRC Fleming's Institutional Committee of Protocol Evaluation in conjunction with the Veterinary Service Management of the Hellenic Republic Prefecture of Attika). For validation experiments, mice were housed at the University of Helsinki Laboratory Animal Center according to national and international legislation and guidelines (license: KEK23-006, approved by the Research Ethics Committee on Animal Research). For the human CRC study cohort, ethical approval was granted by the Surgical Ethics Committee of Helsinki University Hospital (Dnro HUS/1223/2021 and permission to use the tissue samples was granted by The Finnish Medicines Agency (Dnro FIMEA/2021/006901).

ACKNOWLEDGEMENTS

We thank Reeta Huhtala, Melissa Montrose, and Pia Saarinen for technical assistance, Finnish Genome Editing Center (supported by HiLIFE and the Faculty of Medicine, University of Helsinki, and Biocenter Finland) for scanning services, and Emilia Kuuluvainen for critical comments on the manuscript.

REFERENCES

1. Buechler MB, Pradhan RN, Krishnamurty AT, Cox C, Calviello AK, Wang AW, et al. Cross-tissue organization of the fibroblast lineage - PubMed. Nature. 2021;593(7860). doi: 10.1038/s41586-021-03549-5.
2. Gao Y, Li J, Cheng W, Diao T, Liu H, Bo Y, et al. Cross-tissue human fibroblast atlas reveals myofibroblast subtypes with distinct roles in immune modulation - PubMed. Cancer Cell. 2024;42(10). doi: 10.1016/j.ccell.2024.08.020.
3. Liu K, Cui Y, Han H, Guo E, Shi X, Xiong K, et al. Fibroblast atlas: Shared and specific cell types across tissues - PubMed. Science advances. 2025;11(14). doi: 10.1126/sciadv.ado0173.
4. Kinchen J, Chen HH, Parikh K, Antanaviciute A, Jagielowicz M, Fawkner-Corbett D, et al. Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease - PubMed. Cell. 2018;175(2). doi: 10.1016/j.cell.2018.08.067.
5. Brügger MD, Valenta T, Fazilaty H, Hausmann G, Basler K. Distinct populations of crypt-associated fibroblasts act as signaling hubs to control colon homeostasis - PubMed. PLoS Biol. 2020;18(12). doi: 10.1371/journal.pbio.3001032.
6. McCarthy N, Manieri E, Storm EE, Saadatpour A, Luoma AM, Kapoor VN, et al. Distinct Mesenchymal Cell Populations Generate the Essential Intestinal BMP Signaling Gradient - PubMed. Cell Stem Cell. 2020;26(3). doi: 10.1016/j.stem.2020.01.008.
7. Roulis M, Kaklamanos A, Schernthanner M, Bielecki P, Zhao J, Kaffe E, et al. Paracrine orchestration of intestinal tumorigenesis by a mesenchymal niche - PubMed. Nature. 2020;580(7804). doi: 10.1038/s41586-020-2166-3.
8. Jasso GJ, Jaiswal A, Varma M, Laszewski T, Grauel AL, Omar A, et al. Colon stroma mediates an inflammation-driven fibroblastic response controlling matrix remodeling and healing - PubMed. PLoS Biol. 2022;20(1). doi: 10.1371/journal.pbio.3001532.
9. Kraiczy J, McCarthy N, Malagola E, Tie G, Madha S, Boffelli D, et al. Graded BMP signaling within intestinal crypt architecture directs self-organization of the Wnt-secreting stem cell niche - PubMed. Cell Stem Cell. 2023;30(4). doi: 10.1016/j.stem.2023.03.004.
10. McCarthy N, Tie G, Madha S, He R, Kraiczy J, Maglieri A, et al. Smooth muscle contributes to the development and function of a layered intestinal stem cell niche - PubMed. Dev Cell. 2023;58(7). doi: 10.1016/j.devcel.2023.02.012.
11. Pærregaard SI, Wulff L, Schussek S, Niss K, Mörbe U, Jendholm J, et al. The small and large intestine contain related mesenchymal subsets that derive from embryonic Gli1+ precursors. Nature Communications 2023 14:1. 2023;14(1). doi: 10.1038/s41467-023-37952-5.
12. Chalkidi N, Paraskeva C, Koliaraki V. Fibroblasts in intestinal homeostasis, damage, and repair - PubMed. Front Immunol. 2022;13. doi: 10.3389/fimmu.2022.924866.
13. Huang B, Chen Z, Geng L, Wang J, Liang H, Cao Y, et al. Mucosal Profiling of Pediatric-Onset Colitis and IBD Reveals Common Pathogenics and Therapeutic Pathways - PubMed. Cell. 2019;179(5). doi: 10.1016/j.cell.2019.10.027.
14. Elmentaite R, Ross ADB, Roberts K, James KR, Ortmann D, Gomes T, et al. Single-Cell Sequencing of Developing Human Gut Reveals Transcriptional Links to Childhood Crohn's Disease - PubMed. Dev Cell. 2020;55(6). doi: 10.1016/j.devcel.2020.11.010.
15. Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, et al. A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer. 2020;20(3). doi: 10.1038/s41568-019-0238-1.

16. Wu C, Von Schalscha T, Sansanwal D, Qian C, Jiang Q, Shepard RM, et al. Targeting Pancreatic Cancer Cell Stemness by Blocking Fibronectin-Binding Integrins on Cancer-Associated Fibroblasts. Cancer Res Commun. 2025;5(1). doi: 10.1158/2767-9764.CRC-24-0491.
17. Gao D, Fang L, Liu C, Yang M, Yu X, Wang L, et al. Microenvironmental regulation in tumor progression: Interactions between cancer-associated fibroblasts and immune cells. Biomed Pharmacother. 2023;167. doi: 10.1016/j.biopha.2023.115622.
18. Yang D, Liu J, Qian H, Zhuang Q. Cancer-associated fibroblasts: from basic science to anticancer therapy. Exp Mol Med. 2023;55(7). doi: 10.1038/s12276-023-01013-0.
19. Zhang C, Fei Y, Wang H, Hu S, Liu C, Hu R, et al. CAFs orchestrates tumor immune microenvironment—A new target in cancer therapy? Front Pharmacol. 2023;14. doi: 10.3389/fphar.2023.1113378.
20. Rhim AD, Oberstein PE, Thomas DH, Mirek ET, Palermo CF, Sastra SA, et al. Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma - PubMed. Cancer Cell. 2014;25(6). doi: 10.1016/j.ccr.2014.04.021.
21. Özdemir BC, Pentcheva-Hoang T, Carstens JL, Zheng X, Wu CC, Simpson TR, et al. Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival - PubMed. Cancer Cell. 2014;25(6). doi: 10.1016/j.ccr.2014.04.005.
22. Bhattacharjee S, Hamberger F, Ravichandra A, Miller M, Nair A, Affo S, et al. Tumor restriction by type I collagen opposes tumor-promoting effects of cancer-associated fibroblasts - PubMed. J Clin Invest. 2021;131(11). doi: 10.1172/JCI146987.
23. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer - PubMed. Nat Med. 2015;21(11). doi: 10.1038/nm.3967.
24. Mooi JK, Wirapati P, Asher R, Lee CK, Savas P, Price TJ, et al. The prognostic impact of consensus molecular subtypes (CMS) and its predictive effects for bevacizumab benefit in metastatic colorectal cancer: molecular analysis of the AGITG MAX clinical trial. Ann Oncol. 2018;29(11). doi: 10.1093/annonc/mdy410.
25. Källberg J, Harrison A, March V, Bērziņa S, Nemazanyy I, Kepp O, et al. Intratumor heterogeneity and cell secretome promote chemotherapy resistance and progression of colorectal cancer - PubMed. Cell Death Dis. 2023;14(5). doi: 10.1038/s41419-023-05806-z.
26. Lee HO, Hong Y, Etlioglu HE, Cho YB, Pomella V, Van den Bosch B, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer - PubMed. Nat Genet. 2020;52(6). doi: 10.1038/s41588-020-0636-z.
27. Becker WR, Nevins SA, Chen DC, Chiu R, Horning AM, Guha TK, et al. Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer - PubMed. Nat Genet. 2022;54(7). doi: 10.1038/s41588-022-01088-x.
28. Zhao J, Chen Y. Systematic identification of cancer-associated-fibroblast-derived genes in patients with colorectal cancer based on single-cell sequencing and transcriptomics - PubMed. Front Immunol. 2022;13. doi: 10.3389/fimmu.2022.988246.
29. Wang F, Long J, Li L, Wu ZX, Da TT, Wang XQ, et al. Single-cell and spatial transcriptome analysis reveals the cellular heterogeneity of liver metastatic colorectal cancer - PubMed. Sci Adv. 2023;9(24). doi: 10.1126/sciadv.adf5464.
30. Yu K, Wang J, Wang Y, He J, Hu S, Kuai S. Consensus clustering and development of a risk signature based on trajectory differential genes of cancer-associated fibroblast subpopulations

in colorectal cancer - PubMed. J Cancer Res Clin Oncol. 2024;150(8). doi: 10.1007/s00432-024-05906-z.
31. Zhou X, Han J, Zuo A, Ba Y, Liu S, Xu H, et al. THBS2 + cancer-associated fibroblasts promote EMT leading to oxaliplatin resistance via COL8A1-mediated PI3K/AKT activation in colorectal cancer - PubMed. Mol Cancer. 2024;23(1). doi: 10.1186/s12943-024-02180-y.
32. Lu Y, Chen Y, Wang Z, Shen H, Xu L, Huang C, et al. Single-cell and spatial transcriptome profiling reveal CTHRC1+ fibroblasts promote EMT through WNT5A signaling in colorectal cancer - PubMed. J Transl Med. 2025;23(1). doi: 10.1186/s12967-025-06236-5.
33. Ma M, Chu J, Zhuo C, Xiong X, Gu W, Li H, et al. Prognostic implications and therapeutic opportunities related to CAF subtypes in CMS4 colorectal cancer: insights from single-cell and bulk transcriptomics - PubMed. Apoptosis. 2025;30(3-4). doi: 10.1007/s10495-024-02063-z.
34. Wang Y, Qiu X, Li Q, Qin J, Ye L, Zhang X, et al. Single-cell and spatial-resolved profiling reveals cancer-associated fibroblast heterogeneity in colorectal cancer metabolic subtypes - PubMed. J Transl Med. 2025;23(1). doi: 10.1186/s12967-025-06103-3.
35. Zhan Y, Sun D, Gao J, Gao Q, Lv Y, Du T, et al. Single-cell transcriptomics reveals intratumor heterogeneity and the potential roles of cancer stem cells and myCAFs in colorectal cancer liver metastasis and recurrence - PubMed. Cancer Lett. 2025;612. doi: 10.1016/j.canlet.2025.217452.
36. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJL, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors - PubMed. Nat Genet. 2017;49(5). doi: 10.1038/ng.3818.
37. Qian J, Olbrecht S, Boeckx B, Vos H, Laoui D, Etlioglu E, et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling - PubMed. Cell Res. 2020;30(9). doi: 10.1038/s41422-020-0355-0.
38. Zheng H, Liu H, Ge Y, Wang XQ. Integrated single-cell and bulk RNA sequencing analysis identifies a cancer associated fibroblast-related signature for predicting prognosis and therapeutic responses in colorectal cancer - PubMed. Cancer Cell Int. 2021;21(1). doi: 10.1186/s12935-021-02252-9.
39. Zhao Z, Li W, Zhu L, Xu B, Jiang Y, Ma N, et al. Construction and Verification of a Fibroblast-Related Prognostic Signature Model for Colon Cancer - PubMed. Front Genet. 2022;13. doi: 10.3389/fgene.2022.908957.
40. Luo H, Xia X, Huang LB, An H, Cao M, Kim GD, et al. Pan-cancer single-cell analysis reveals the heterogeneity and plasticity of cancer-associated fibroblasts in the tumor microenvironment - PubMed. Nat Commun. 2022;13(1). doi: 10.1038/s41467-022-34395-2.
41. Liang C, Ji D, Qin F, Chen G. CAF signature predicts the prognosis of colorectal cancer patients: A retrospective study based on bulk RNA sequencing and single-cell RNA sequencing data - PubMed. Medicine. 2023;102(10). doi: 10.1097/MD.0000000000033149.
42. Ma C, Yang C, Peng A, Sun T, Ji X, Mi J, et al. Pan-cancer spatially resolved single-cell analysis reveals the crosstalk between cancer-associated fibroblasts and tumor microenvironment - PubMed. Mol Cancer. 2023;22(1). doi: 10.1186/s12943-023-01876-x.
43. Pelka K, Hofree M, Chen JH, Sarkizova S, Pirl JD, Jorgji V, et al. Spatially organized multicellular immune hubs in human colorectal cancer - PubMed. Cell. 2021;184(18). doi: 10.1016/j.cell.2021.08.003.
44. Joanito I, Wirapati P, Zhao N, Nawaz Z, Yeo G, Lee F, et al. Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus

molecular classification of colorectal cancer - PubMed. Nat Genet. 2022;54(7). doi: 10.1038/s41588-022-01100-4.
45. Hulzinga JD, Thuneberg L, Klüppel M, Malysz J, Mikkelsen HB, Bernstein A. W/kit gene required for interstitial cells of Cajal and for intestinal pacemaker activity. Nature 1995 373:6512. 1995;373(6512). doi: 10.1038/373347a0.
46. Link A, Vogt TK, Favre S, Britschgi MR, Acha-Orbea H, Hinz B, et al. Fibroblastic reticular cells in lymph nodes regulate the homeostasis of naive T cells. Nat Immunol. 2007;8(11). doi: 10.1038/ni1513.
47. Oliveira MF, Romero JP, Chung M, Williams SR, Gottscho AD, Gupta A, et al. High-definition spatial transcriptomic profiling of immune cell populations in colorectal cancer. Nat Genet. 2025;57(6). doi: 10.1038/s41588-025-02193-3.
48. Powell DW, Mifflin RC, Valentich JD, Crowe SE, Saada JI, West AB. Myofibroblasts. II. Intestinal subepithelial myofibroblasts - PubMed. Am J Physiol. 1999;277(2). doi: 10.1152/ajpcell.1999.277.2.C183.
49. Kosinski C, Li VS, Chan AS, Zhang J, Ho C, Tsui WY, et al. Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors - PubMed. Proc Natl Acad Sci U S A. 2007;104(39). doi: 10.1073/pnas.0707210104.
50. Stzepourginski I, Nigro G, Jacob JM, Dulauroy S, Sansonetti PJ, Eberl G, et al. CD34+ mesenchymal cells are a major component of the intestinal stem cells niche at homeostasis and after injury - PubMed. Proc Natl Acad Sci U S A. 2017;114(4). doi: 10.1073/pnas.1620059114.
51. Karpus ON, Westendorp BF, Vermeulen JLM, Meisner S, Koster J, Muncan V, et al. Colonic CD90+ Crypt Fibroblasts Secrete Semaphorins to Support Epithelial Growth - PubMed. Cell Rep. 2019;26(13). doi: 10.1016/j.celrep.2019.02.101.
52. Goto N, Goto S, Imada S, Hosseini S, Deshpande V, Yilmaz ÖH. Lymphatics and fibroblasts support intestinal stem cells in homeostasis and injury. Cell Stem Cell. 2022;29(8). doi: 10.1016/j.stem.2022.06.013.
53. Islam MS, Kusakabe M, Horiguchi K, Iino S, Nakamura T, Iwanaga K, et al. PDGF and TGF-β promote tenascin-C expression in subepithelial myofibroblasts and contribute to intestinal mucosal protection in mice. Br J Pharmacol. 2013;171(2). doi: 10.1111/bph.12452.
54. Bernier-Latmani J, Mauri C, Marcone R, Renevey F, Durot S, He L, et al. ADAMTS18+ villus tip telocytes maintain a polarized VEGFA signaling domain and fenestrations in nutrient-absorbing intestinal blood vessels - PubMed. Nat Commun. 2022;13(1). doi: 10.1038/s41467-022-31571-2.
55. Bahamonde ME, Lyons KM. BMP3: To Be or Not To Be a BMP. J Bone Joint Surg. 2001;83(1):S56-S62.
56. Halpern BK, Massalha H, Zwick RK, Moor AE, Castillo-Azofeifa D, Rozenberg M, et al. Lgr5+ telocytes are a signaling source at the intestinal villus tip - PubMed. Nat Commun. 2020;11(1). doi: 10.1038/s41467-020-15714-x.
57. Muhl L, Genové G, Leptidis S, Liu J, He L, Mocci G, et al. Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination - PubMed. Nat Commun. 2020;11(1). doi: 10.1038/s41467-020-17740-1.
58. May-Zhang AA, Tycksen E, Southard-Smith AN, Deal KK, Benthal JT, Buehler DP, et al. Combinatorial Transcriptional Profiling of Mouse and Human Enteric Neurons Identifies Shared and Disparate Subtypes In Situ - PubMed. Gastroenterology. 2021;160(3). doi: 10.1053/j.gastro.2020.09.032.

59. Schneider S, Hashmi SK, Thrasher AJ, Kothakapa DR, Wright CM, Heuckeroth RO. Single Nucleus Sequencing of Human Colon Myenteric Plexus-Associated Visceral Smooth Muscle Cells, Platelet Derived Growth Factor Receptor Alpha Cells, and Interstitial Cells of Cajal - PubMed. Gastro Hep Adv. 2023;2(3). doi: 10.1016/j.gastha.2022.12.004.
60. Iino S, Horiguchi K, Horiguchi S, Nojyo Y. c-Kit-negative fibroblast-like cells express platelet-derived growth factor receptor alpha in the murine gastrointestinal musculature - PubMed. Histochem Cell Biol. 2009;131(6). doi: 10.1007/s00418-009-0580-6.
61. Kurahashi M, Nakano Y, Hennig GW, Ward SM, Sanders KM. Platelet-derived growth factor receptor α-positive cells in the tunica muscularis of human colon. J Cell Mol Med. 2012;16(7). doi: 10.1111/j.1582-4934.2011.01510.x.
62. Baker SA, Hennig GW, Salter AK, Kurahashi M, Ward SM, Sanders KM. Distribution and Ca(2+) signalling of fibroblast-like (PDGFR(+)) cells in the murine gastric fundus - PubMed. J Physiol. 2013;591(24). doi: 10.1113/jphysiol.2013.264747.
63. Vannucchi MG, Traini C, Manetti M, Ibba-Mannesch L, Faussone-Pellegrini MS. Telocytes express PDGFRα in the human gastrointestinal tract - PubMed. J Cell Mol Med. 2013;17(9). doi: 10.1111/jcmm.12134.
64. Sanders KM, Ward SM, Koh SD. Interstitial cells: regulators of smooth muscle function - PubMed. Physiol Rev. 2014;94(3). doi: 10.1152/physrev.00037.2013.
65. Peltier A, Seban R-D, Buvat I, Bidard F-C, Mechta-Grigoriou F. Fibroblast heterogeneity in solid tumors: From single cell analysis to whole-body imaging. Semin Cancer Biol. 2022;86. doi: 10.1016/j.semcancer.2022.04.008.
66. Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M, et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer - PubMed. J Exp Med. 2017;214(3). doi: 10.1084/jem.20162024.
67. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, et al. Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts - PubMed. Cancer Discov. 2019;9(8). doi: 10.1158/2159-8290.CD-19-0094.
68. Grauel AL, Nguyen B, Ruddy D, Laszewski T, Schwartz S, Chang J, et al. TGFβ-blockade uncovers stromal plasticity in tumors by revealing the existence of a subset of interferon-licensed fibroblasts - PubMed. Nat Commun. 2020;11(1). doi: 10.1038/s41467-020-19920-5.
69. Forsthuber A, Aschenbrenner B, Korosec A, Jacob T, Annusver K, Krajic N, et al. Cancer-associated fibroblast subtypes modulate the tumor-immune microenvironment and are associated with skin cancer malignancy. Nat Commun. 2024;15(1). doi: 10.1038/s41467-024-53908-9.
70. Cords L, Tietscher S, Anzeneder T, Langwieder C, Rees M, de Souza N, et al. Cancer-associated fibroblast classification in single-cell and spatial proteomics data - PubMed. Nat Commun. 2023;14(1). doi: 10.1038/s41467-023-39762-1.
71. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature. 2019;566(7745). doi: 10.1038/s41586-019-0969-x.
72. Purcell JW, Tanlimco SG, Hickson J, Fox M, Sho M, Durkin L, et al. LRRC15 Is a Novel Mesenchymal Protein and Stromal Target for Antibody-Drug Conjugates - PubMed. Cancer Res. 2018;78(14). doi: 10.1158/0008-5472.CAN-18-0327.
73. Shintani Y, Kimura T, Funaki S, Ose N, Kanou T, Fukui E. Therapeutic Targeting of Cancer-Associated Fibroblasts in the Non-Small Cell Lung Cancer Tumor Microenvironment. Cancers (Basel). 2023;15(2). doi: 10.3390/cancers15020335.

74. Roelands J, Kuppen PJK, Ahmed EI, Mall R, Masoodi T, Singh P, et al. An integrated tumor, immune and microbiome atlas of colon cancer. Nat Med. 2023;29(5). doi: 10.1038/s41591-023-02324-5.
75. Zhang M, Yang H, Wan L, Wang Z, Wang H, Ge C, et al. Single-cell transcriptomic architecture and intercellular crosstalk of human intrahepatic cholangiocarcinoma - PubMed. J Hepatol. 2020;73(5). doi: 10.1016/j.jhep.2020.05.039.
76. Kieffer Y, Hocine HR, Gentric G, Pelon F, Bernard C, Bourachot B, et al. Single-Cell Analysis Reveals Fibroblast Clusters Linked to Immunotherapy Resistance in Cancer - PubMed. Cancer Discov. 2020;10(9). doi: 10.1158/2159-8290.CD-19-1384.
77. Roulis M, Flavell RA. Fibroblasts and myofibroblasts of the intestinal lamina propria in physiology and disease - PubMed. Differentiation. 2016;92(3). doi: 10.1016/j.diff.2016.05.002.
78. Hageman JH, Heinz MC, Kretzschmar K, van der Vaart J, Clevers H, Snippert HJG. Intestinal Regeneration: Regulation by the Microenvironment - PubMed. Dev Cell. 2020;54(4). doi: 10.1016/j.devcel.2020.07.009.
79. Felsenthal N, Vignjevic DM. Stand by me: Fibroblasts regulation of the intestinal epithelium during development and homeostasis - PubMed. Curr Opin Cell Biol. 2022;78. doi: 10.1016/j.ceb.2022.102116.
80. Brügger MD, Basler K. The diverse nature of intestinal fibroblasts in development, homeostasis, and disease - PubMed. Trends Cell Biol. 2023;33(10). doi: 10.1016/j.tcb.2023.03.007.
81. Cheng PSW, Zaccaria M, Biffi G. Functional heterogeneity of fibroblasts in primary tumors and metastases - PubMed. Trends in Cancer. 2025;11(2). doi: 10.1016/j.trecan.2024.11.005.
82. McCartney EE, Chung Y, Buechler MB. Life of Pi: Exploring functions of Pi16+ fibroblasts. F1000Research. 2024;13. doi: 10.12688/f1000research.143511.2.
83. Melissari MT, Henriques A, Tzaferis C, Prados A, Sarris ME, Chalkidi N, et al. Col6a1+/CD201+ mesenchymal cells regulate intestinal morphogenesis and homeostasis - PubMed. Cell Mol Life Sci. 2021;79(1). doi: 10.1007/s00018-021-04071-7.
84. Li S, Lu R, Shu L, Chen Y, Zhao J, Dai J, et al. An integrated map of fibroblastic populations in human colon mucosa and cancer tissues - PubMed. Commun Biol. 2022;5(1). doi: 10.1038/s42003-022-04298-5.
85. Xie Z, Zhang M, Zhou G, Lin L, Han J, Wang Y, et al. Emerging roles of the Hedgehog signalling pathway in inflammatory bowel disease. Cell Death Discov. 2021;7(1). doi: 10.1038/s41420-021-00679-7.
86. Harada A, Yasumizu Y, Harada T, Fumoto K, Sato A, Maehara N, et al. Hypoxia-induced Wnt5a-secreting fibroblasts promote colon cancer progression. Nat Commun. 2025;16(1). doi: 10.1038/s41467-025-58748-9.
87. Kurahashi M, Zheng H, Dwyer L, Ward SM, Koh SD, Sanders KM. A functional role for the 'fibroblast-like cells' in gastrointestinal smooth muscles - PubMed. J Physiol. 2011;589(Pt 3). doi: 10.1113/jphysiol.2010.201129.
88. Kurahashi M, Mutafova-Yambolieva V, Koh SD, Sanders KM. Platelet-derived growth factor receptor-α-positive cells and not smooth muscle cells mediate purinergic hyperpolarization in murine colonic muscles - PubMed. Am J Physiol. 2014;307(6). doi: 10.1152/ajpcell.00080.2014.

89. Lee H, Koh BH, Peri LE, Sanders KM, Koh SD. Purinergic inhibitory regulation of murine detrusor muscles mediated by PDGFRα+ interstitial cells - PubMed. J Physiol. 2014;592(6). doi: 10.1113/jphysiol.2013.267989.
90. Kurahashi M, Baker SA, Kito Y, Bartlett A, Hara M, Takeyama H, et al. PDGFRα+ Interstitial Cells are Effector Cells of PACAP Signaling in Mouse and Human Colon - PubMed. Cell Mol Gastroenterol Hepatol. 2022;14(2). doi: 10.1016/j.jcmgh.2022.05.004.
91. Hannigan KI, Ni Bhraonain EP, Gould TW, Keef KD, Cobine CA. Modulation of intracellular calcium activity in interstitial cells of Cajal by inhibitory neural pathways within the internal anal sphincter - PubMed. Am J Physiol. 2024;327(3). doi: 10.1152/ajpgi.00309.2023.
92. Biffi G, Oni TE, Spielman B, Hao Y, Elyada E, Park Y, et al. IL-1-induced JAK/STAT signaling is antagonized by TGF-β to shape CAF heterogeneity in pancreatic ductal adenocarcinoma. Cancer Discov. 2018;9(2). doi: 10.1158/2159-8290.CD-18-0710.
93. Lee K-W, Yeo S-Y, Gong J-R, Koo O-J, Sohn I, Lee WY, et al. PRRX1 is a master transcription factor of stromal fibroblasts for myofibroblastic lineage progression. Nat Commun. 2022;13(1). doi: 10.1038/s41467-022-30484-4.
94. Rynne-Vidal A, Jiménez-Heffernan JA, Fernández-Chacón C, López-Cabrera M, Sandoval P. The Mesothelial Origin of Carcinoma Associated-Fibroblasts in Peritoneal Metastasis. Cancers (Basel). 2015;7(4). doi: 10.3390/cancers7040872.
95. Wikström P, Marusic J, Stattin P, Bergh A. Low stroma androgen receptor level in normal and tumor prostate tissue is related to poor outcome in prostate cancer patients. The Prostate. 2009;69(8). doi: 10.1002/pros.20927.
96. Hosaka K, Yang Y, Seki T, Fischer C, Dubey O, Fredlund E, et al. Pericyte-fibroblast transition promotes tumor growth and metastasis - PubMed. Proc Natl Acad Sci U S A. 2016;113(38). doi: 10.1073/pnas.1608384113.
97. Zeisberg EM, Potenta S, Xie L, Zeisberg M, Kalluri R. Discovery of Endothelial to Mesenchymal Transition as a Source for Carcinoma-Associated Fibroblasts. Cancer Res. 2007;67(21). doi: 10.1158/0008-5472.CAN-07-3127.
98. Iwano M, Plieth D, Danoff TM, Xue C, Okada H, Neilson EG. Evidence that fibroblasts derive from epithelium during tissue fibrosis. J Clin Invest. 2002;110(3). doi: 10.1172/JCI15518.
99. Jotzu C, Alt E, Welte G, Li J, Hennessy BT, Devarajan E, et al. Adipose tissue derived stem cells differentiate into carcinoma-associated fibroblast-like cells under the influence of tumor derived factors. Cell Oncol. 2011;34(1). doi: 10.1007/s13402-011-0012-1.
100. Bochet L, Lehuédé C, Dauvillier S, Wang YY, Dirat B, Laurent V, et al. Adipocyte-Derived Fibroblasts Promote Tumor Progression and Contribute to the Desmoplastic Reaction in Breast Cancer. Cancer Res. 2013;73(18). doi: 10.1158/0008-5472.CAN-13-0530.
101. Lindahl P, Johansson BR, Levéen P, Betsholtz C. Pericyte Loss and Microaneurysm Formation in PDGF-B-Deficient Mice. Science. 1997;277(5323). doi: 10.1126/science.277.5323.242.
102. Cho H, Kozasa T, Bondjers C, Betsholtz C, Kehrl JH. Pericyte-specific expression of Rgs5: implications for PDGF and EDG receptor signaling during vascular maturation. The FASEB Journal. 2003;17(3). doi: 10.1096/fj.02-0340fje.
103. Mesa-Ciller C, Turiel G, Guajardo-Grence A, Lopez-Rodriguez AB, Egea J, De Bock K, et al. Unique expression of the atypical mitochondrial subunit NDUFA4L2 in cerebral pericytes fine tunes HIF activity in response to hypoxia - PubMed. J Cereb Blood Flow Metab. 2023;43(1). doi: 10.1177/0271678X221118236.

104. Miao Y, Li W, Jeansson M, Andaloussi Mäe M, Muhl L, He L. Different gene expression patterns between mouse and human brain pericytes revealed by single-cell/nucleus RNA sequencing. Vascul Pharmacol. 2024;157. doi: 10.1016/j.vph.2024.107434.
105. Bartoschek M, Oskolkov N, Bocci M, Lövrot J, Larsson C, Sommarin M, et al. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing - PubMed. Nat Commun. 2018;9(1). doi: 10.1038/s41467-018-07582-3.
106. Mahadevan KK, McAndrews KM, LeBleu VS, Yang S, Lyu H, Li B, et al. KRASG12D inhibition reprograms the tumor microenvironment of early and advanced pancreatic cancer to promote FAS-mediated killing by CD8+ T cells. Cancer Cell. 2023;41(9). doi: 10.1016/j.ccell.2023.07.002.
107. Attieh Y, Clark AG, Grass C, Richon S, Pocard M, Mariani P, et al. Cancer-associated fibroblasts lead tumor invasion through integrin-β3–dependent fibronectin assembly. J Cell Biol. 2017;216(11). doi: 10.1083/jcb.201702033.
108. Erdogan B, Ao M, White LM, Means AL, Brewer bM, Yang l, et al. Cancer-associated fibroblasts promote directional cancer cell migration by aligning fibronectin. J Cell Biol. 2017;216(11). doi: 10.1083/jcb.201704053.
109. Wang FT, Wei S, Zhang JT, Fan Y-Z. Cancer-associated fibroblast regulation of tumor neo-angiogenesis as a therapeutic target in cancer (Review). Oncol Lett. 2019;17(3). doi: 10.3892/ol.2019.9973.
110. Asif PJ, Longobardi C, Hahne M, Medema JP. The Role of Cancer-Associated Fibroblasts in Cancer Invasion and Metastasis. Cancers (Basel). 2021;13(18). doi: 10.3390/cancers13184720.
111. Szabo PM, Vajdi A, Kumar N, Tolstorukov MY, Chen BJ, Edwards R, et al. Cancer-associated fibroblasts are the main contributors to epithelial-to-mesenchymal signatures in the tumor microenvironment. Sci Rep. 2023;13(1). doi: 10.1038/s41598-023-28480-9.
112. Hu S, Qin J, Gao R, Xiao Q, Liu X, Pan Y, et al. Integrated analysis of single cell and bulk RNA sequencing identifies CTHRC1+ INHBA+ CAF as drivers of colorectal cancer progression - PubMed. Mol Carcinog. 2023;62(12). doi: 10.1002/mc.23615.
113. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data - PubMed. Cell. 2021;184(13). doi: 10.1016/j.cell.2021.04.048.
114. Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA. Dissecting cell identity via network inference and in silico gene perturbation. Nature. 2023;614(7949). doi: 10.1038/s41586-022-05688-9.
115. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1). doi: 10.1186/s13059-017-1382-0.
116. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7). doi: 10.1038/nmeth.2019.
117. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep. 2017;7(1). doi: 10.1038/s41598-017-17204-5.
118. Yu G, Wang L-G, Yanyan Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. https://homeliebertpubcom/omi. 2012;16(5). doi: 10.1089/omi.2011.0118.

119. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43). doi: 10.1073/pnas.0506580102.
120. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27(12). doi: 10.1093/bioinformatics/btr260.
121. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6). doi: 10.1016/j.cels.2015.12.004.
122. Castanza AS, Recla JM, Eby D, Thorvaldsdóttir H, Bult CJ, Mesirov JP. Extending support for mouse data in the Molecular Signatures Database (MSigDB). Nat Methods. 2023;20(11). doi: 10.1038/s41592-023-02014-7.

S1 Fig. ScRNA-seq analysis of mouse colon fibroblasts. (A) UMAP of the mouse colonic mesenchyme scRNA-seq performed in this study. (B) Feature plots depicting indicated marker genes for colon fibroblast subsets. (C) Violin plot showing unbiased marker genes for each identified fibroblast population. (d) Schematic representation of the datasets used for the integrated analysis of mouse mesenchymal cells. (E) Schematic representation of the datasets used for the integrated analysis of normal colon and CRC cells.

S2 Fig. Characterization of identified mouse and human colon fibroblast subtypes. (A) UMAPs split by individual datasets used for mouse integration analysis. (B) Dot plot depicting cell type marker genes in mouse colon. For fibroblast subsets, see Fig 1B. (C) UMAPs split by individual datasets used for human integration analysis. (D) Dot plot depicting cell type marker genes in human colon. For fibroblast subsets, see Fig 1D. (E) Feature plots showing a small *KIT*+ cluster representing ICCs (arrowhead) and a *CCL19*+ cluster representing reticular fibroblasts. (F) Dot plot of species-specific genes marking human but not mouse colon fibroblast subtypes (black genes in Fig 1D) in the mouse dataset. (G) Dot plot of species-specific genes marking mouse but not human colon fibroblast subtypes (black genes in S2H Fig) in the human dataset. (H) Dot plot depicting unbiased marker genes for each mouse fibroblast subpopulation (see Fig 1B for human genes). The markers shared between mouse and human are marked in red.

S3 Fig. ScRNA-seq analysis of mouse small intestinal fibroblasts. (A) UMAP of the mouse small intestinal fibroblasts and myocytes from scRNA-seq performed in this study. (B) Feature plot of the fibroblast marker *Pdgfra* and smooth-muscle marker *Myh11*. (C-E) Feature plots of

indicated genes marking colon fibroblast subsets in the small intestinal mesenchyme dataset. C, markers for Top and Bottom SEMFs; D, markers for MAF subclusters; E, markers for SAFs.

S4 Fig. Expression of mouse MAF and SEMF subset genes in human colon mesenchyme. (A) Feature plot showing expression of mouse bottom MAF genes *TNC* and *CD55* in human data. (B) UPAM showing the difference in fibroblast subpopulations in right-sided (proximal) and left-sided (distal) colon stroma in human. (C) Feature plot of *WNT4* expression in human data. (D) Feature plots of mouse top SEMF marker expression in human. (E) Feature plots of mouse bottom SEMF marker expression in human.

S5 Fig. Identification of CRC CAF subsets. (A) UMAP showing contribution of individual datasets used for integration of normal adjacent and tumor tissue from CRC patients. (B) Dot plot depicting genes used for identifying non-mesenchymal cells in the integration of normal adjacent and tumor tissue. See also Fig 5D. (C) UMAP of the whole integrated CRC patient dataset (see Fig 5A) separated by sample origin normal adjacent tissue and CRC. (D) UMAPs visualizing the contribution of individual datasets used for the mesenchymal cell analysis of normal adjacent and tumor tissue (Fig 5B). (E) UMAPs of the individual samples included in the colon mesenchyme analysis (n = 53, > 50 cells and ≥ 5 clusters identified per sample) (F) Stacked bar plots depicting the stromal cell proportions per patient in normal adjacent colon and CRC. See also S11C Fig. (G) Feature plots showing a small *KIT*+ cluster representing ICCs, a *CCL19*+ cluster representing reticular fibroblasts, the antigen-presenting CAF marker CD74 enriching in the CCL19+ cluster, and proliferation marker MKI67 (arrowheads).

S6 Fig. Marker genes of CRC CAF subsets and correlation of CAF subset signatures to survival . (A-D) Dot plots of markers for (A) iCAF, (B) mCAF, (C) preCAF1 and (D) preCAF2 clusters. (E) Kaplan-Meier plot indicating correlation of the CAF subtype signature gene expression (see Fig 5E) with the probability of overall survival of CRC patients. mRNA-sequencing data derived from: GSE12945, GSE12294, GSE14333, GSE123985, GSE17538, GSE10800, GSE26682, GSE20540, GSE21595, GSE33114, GSE34489, GSE37892, GSE38832, GSE39582, GSE41258, GSE92921.

S7 Fig. Presence of CAFs and normal colon fibroblasts in CRC and normal tissue. (A) Projections of CAF subtype signatures (Fig 6A) onto normal human colon spatial transcriptomics data. (B) Dotplot depicting specific marker gene signatures for normal colon fibroblast subsets. Only markers not expressed in CAFs were selected for this analysis. (C) Projections of normal-specific fibroblast subtype signatures from B onto normal colon tissue and CRC tumor tissue spatial transcriptomics data. (D) Gene regulatory network assisted gene importance ranking. Transcription factors marked in red/purple. (E) Monocle3 clustering of the integrated human mesenchymal scRNA-seq data. Clusters 3, 5, 9 and 10 represent Seurat clusters iCAF, preCAF2, preCAF1 and mCAF, respectively. (F) Monocle3 partitioning analysis. CAFs partitioned together with clusters 6, 8 and 14 representing the Seurat clusters SAF, MIF and SEMF, respectively, indicating a shared trajectory between CAFs and these normal-like fibroblasts. (G) Trajectory and (H) pseudotime analyses of the fibroblast clusters. (I) mRNA expression signature of potential transcription factors *PRRX1*, *MAFB* and *TWIST1* driving CAF identity identified in Oracle analysis (see Fig 5B).

S8 Fig. Representation of the CAF subsets across all CRC samples. (A) UMAPs showing CAF subpopulations in each CRC patient. Red boxes indicate the samples excluded from subsequent analysis (<50 CAFs).

S9 Fig. Correlation of CAF subtype representation with molecular and clinical characteristics of CRC patients. (A) Stacked bar plot of CAF subtype proportions across all 40 samples included in the CAF analysis. (B-C) Stacked bar plots of CAF proportions in different CMS subtypes. B, overall; C, per patient. (D-F) Enrichment of specific CAF subtypes in CMS1-4. D, preCAF2s in CMS1s; E, preCAF1s in CMS3; F, mCAFs in CMS4. (G-H) Stacked barplot of CAF subtype proportions in iCMS2 and iCMS3. G, average across patients; H, per patient. (I-J) Stacked barplot of CAF subtype proportions in left or right-sided tumors. I. overall; J, per patient. (K) Stacked barplot of CAF subtype proportions in proximal colon, distal colon and rectum. (L) Enrichment of preCAF1 in the rectum. (M-N) Stacked bar plots of CAF subtype proportions in microsatellite stable (MSS) and microsatellite instable (MSI-high) CRCs. M, overall; N, per patient. (O) Enrichment of preCAF2 in MSI-H patients. (P-Q) Stacked barplot of CAF subtype proportions in different tumor stages. P, overall; Q, per patient. (R) Enrichment of iCAFs proportion early

stage (stage 1-2) as compared to late stage (3/4) CRCs. A p-value < 0.05 was considered statistically significant. All enrichments of CAF subtypes with p < 0.1 are shown.

S10 Fig. Correlation of CAF subtype representation with genetic characteristics of CRC patients. (A) Stacked barplot of CAF subtype proportions in *APC* wild-type and mutant CRC in all patients (left) and per patient (right). (B) Stacked barplot of CAF subtype proportions in *KRAS* wild-type and mutant CRC in all patients (left) and per patient (right). (C) Boxplot of mCAF proportion in cancers with or without *KRAS* mutation. (D) Stacked barplot of CAF subtypes in *KRAS* wild-type and mutant patients separated stage. (E) Boxplot of mCAFs abundance in early and late-stage CRC with or without *KRAS* mutation. (F) Stacked barplot of CAF subtype proportions in *TP53* wild-type and mutant CRC in all patients (left) and per patient (right). (G) Boxplot of pre-CAF1 proportion in cancers with and without *TP53* mutation. (H) Stacked barplot of CAF subtype proportions in *BRAF* wild-type and mutant CRC in all patients (left) and per patient (right). (I) Stacked barplot of CAF subtype proportions in *PIK3CA* wild-type and mutant CRC in all patients (left) and per patient (right). (J) Boxplot of preCAF1 proportion in cancers with or without *PIK3CA* mutation. A p-value < 0.05 was considered statistically significant. All enrichments of CAF subtypes with p < 0.1 are shown in this plot.

S11 Fig. CTHRC1 is a marker for CRC CAFs. (A) Immunohistochemical staining of normal adjacent tissue from CRC patients. Note absence of CTHRC1 staining in normal colon stroma but weak expression in lymphocyte-like cells in normal colon mucosa in two out of four patients (arrowheads). (B) Correlation of stromal CTHRC1 staining density to survival of CRC patients. (C) Mesenchymal cell type proportions in normal colon tissue and in early and late-stage CRC. CAFs marker with red box. (D-E) Expression of *CTHRC1* (D) and *FAP* (E) mRNA across CRC CMSs (Sidra-LUMC AC-ICAM dataset). Note highest expression levels in CMS4 patients. Statistical comparisons of normalized mRNA expression levels were performed using Student's t-tests, followed by Bonferroni adjustment for multiple testing. (G-I) Kaplan-Meier plots depicting the correlation between mRNA expression of the panCAF signature (*CTHRC1, INHBA, THBS2, WNT2, FAP*), *CTHRC1* and *FAP* with overall survival in early (stage 1-2) and late-stage (3-4) CRC. (J-L) Kaplan-Meier plots showing correlation of panCAF signature (*CTHRC1, INHBA, THBS2, WNT2, FAP*), or *CTHRC1* and *FAP* alone with CRC patient survival across CRC CMSs.

mRNA-sequencing data derived from: GSE12945, GSE12294, GSE14333, GSE123985, GSE17538, GSE10800, GSE26682, GSE20540, GSE21595, GSE33114, GSE34489, GSE37892, GSE38832, GSE39582, GSE41258, GSE92921.

S1 Table. Mouse scRNA-seq datasets used for integration. Details of the scRNA-seq datasets used for the mouse integration analysis, including filtering parameters and the number of analyzed cells for each dataset.

S2 Table. Human colon fibroblast and CRC CAF signatures. Transcriptional signatures of each human fibroblast and CRC CAF subset used for spatial transcriptomics and signature UMAPs.

S3 Table. Summary of published mouse and human colon fibroblast subsets. Detailed presentation of the previously published colon fibroblast subsets, including original annotations, markers, localization, references, and corresponding annotations provided in this study.

S4 Table. Summary of published CRC CAF subsets. Detailed presentation of the previously published CRC CAF subsets, including original annotations, markers, localization, references, and corresponding annotations provided in this study.

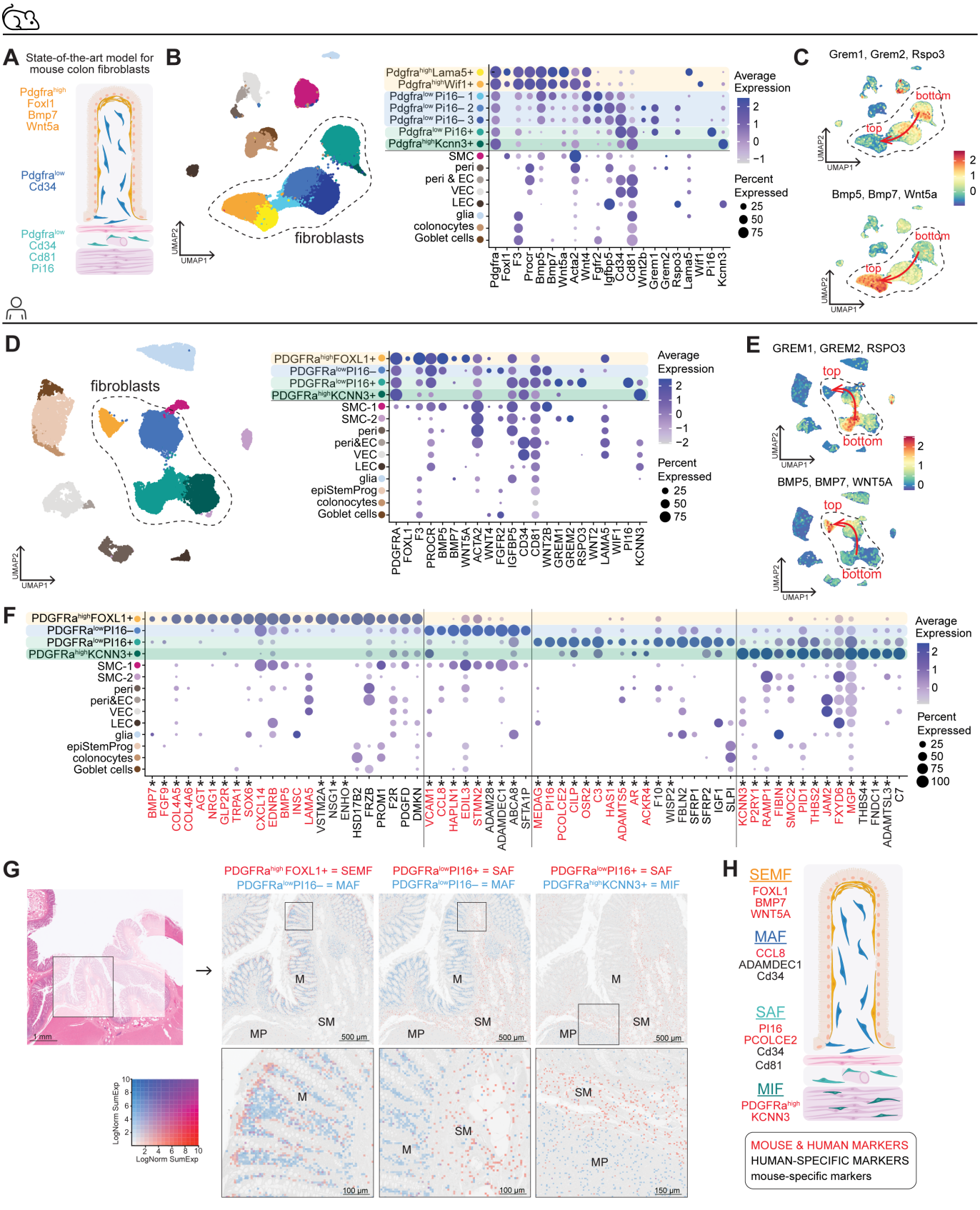
A
State-of-the-art model for mouse colon fibroblasts
Pdgfra high
Foxl1
Bmp7
Wnt5a
Pdgfra low
Cd34
Pdgfra low
Cd34
Cd81
Pi16
B
fibroblasts
UMAP2
UMAP1
Pdgfra high Lama5+
Pdgfra high Wif1+
Pdgfra low Pi16– 1
Pdgfra low Pi16– 2
Pdgfra low Pi16– 3
Pdgfra low Pi16+
Pdgfra high Kcnn3+
SMC
peri
peri & EC
VEC
LEC
glia
colonocytes
Goblet cells
Pdgfra
Foxl1
F3
Procr
Bmp5
Bmp7
Wnt5a
Acta2
Wnt4
Fgfr2
Igfbp5
Cd34
Cd81
Wnt2b
Grem1
Grem2
Rspo3
Lama5
Wif1
Pi16
Kcnn3
Average Expression
2
1
0
−1
Percent Expressed
25
50
75
C
Grem1, Grem2, Rspo3
bottom
top
UMAP2
UMAP1
2
1
0
Bmp5, Bmp7, Wnt5a
bottom
top
UMAP2
UMAP1
D
fibroblasts
UMAP2
UMAP1
PDGFRa high FOXL1+
PDGFRa low PI16–
PDGFRa low PI16+
PDGFRa high KCNN3+
SMC-1
SMC-2
peri
peri&EC
VEC
LEC
glia
epiStemProg
colonocytes
Goblet cells
PDGFRA
FOXL1
F3
PROCR
BMP5
BMP7
WNT5A
ACTA2
WNT4
FGFR2
IGFBP5
CD34
CD81
WNT2B
GREM1
GREM2
RSPO3
WNT2
LAMA5
WIF1
PI16
KCNN3
Average Expression
2
1
0
−1
−2
Percent Expressed
25
50
75
E
GREM1, GREM2, RSPO3
top
bottom
UMAP2
UMAP1
2
1
0
BMP5, BMP7, WNT5A
top
bottom
UMAP2
UMAP1
F
PDGFRa high FOXL1+
PDGFRa low PI16–
PDGFRa low PI16+
PDGFRa high KCNN3+
SMC-1
SMC-2
peri
peri&EC
VEC
LEC
glia
epiStemProg
colonocytes
Goblet cells
BMP7★
FGF9★
COL4A5★
COL4A6★
AGT★
NRG1★
GLP2R★
TRPA1★
SOX6★
CXCL14
EDNRB
BMP5
INSC
LAMA5
VSTM2A★
NSG1★
ENHO★
HSD17B2
FRZB
PROM1
F2R
PDGFD
DMKN
VCAM1★
CCL8★
HAPLN1★
EDIL3★
STMN2★
ADAM28★
ADAMDEC1★
ABCA8★
SFTA1P
MEDAG★
PI16
PCOLCE2★
CILP★
OSR2★
C3★
HAS1★
ADAMTS5★
AR★
ACKR4★
F10★
WISP2★
FBLN2
SFRP1
SFRP2
IGF1
SLPI
KCNN3★
P2RY1★
RAMP1★
FIBIN
SMOC2★
PID1★
THBS2★
JAM2★
FXYD6★
MGP★
THBS4★
FNDC1★
ADAMTSL3★
C7
Average Expression
2
1
0
Percent Expressed
25
50
75
100
G
1 mm
PDGFRa high FOXL1+ = SEMF
PDGFRa low PI16– = MAF
PDGFRa low PI16+ = SAF
PDGFRa low PI16– = MAF
PDGFRa low PI16+ = SAF
PDGFRa high KCNN3+ = MIF
M
SM
MP
500 µm
100 µm
150 µm
LogNorm SumExp
H
SEMF
FOXL1
BMP7
WNT5A
MAF
CCL8
ADAMDEC1
Cd34
SAF
PI16
PCOLCE2
Cd34
Cd81
MIF
PDGFRa high
KCNN3
MOUSE & HUMAN MARKERS
HUMAN-SPECIFIC MARKERS
mouse-specific markers

Fig 1

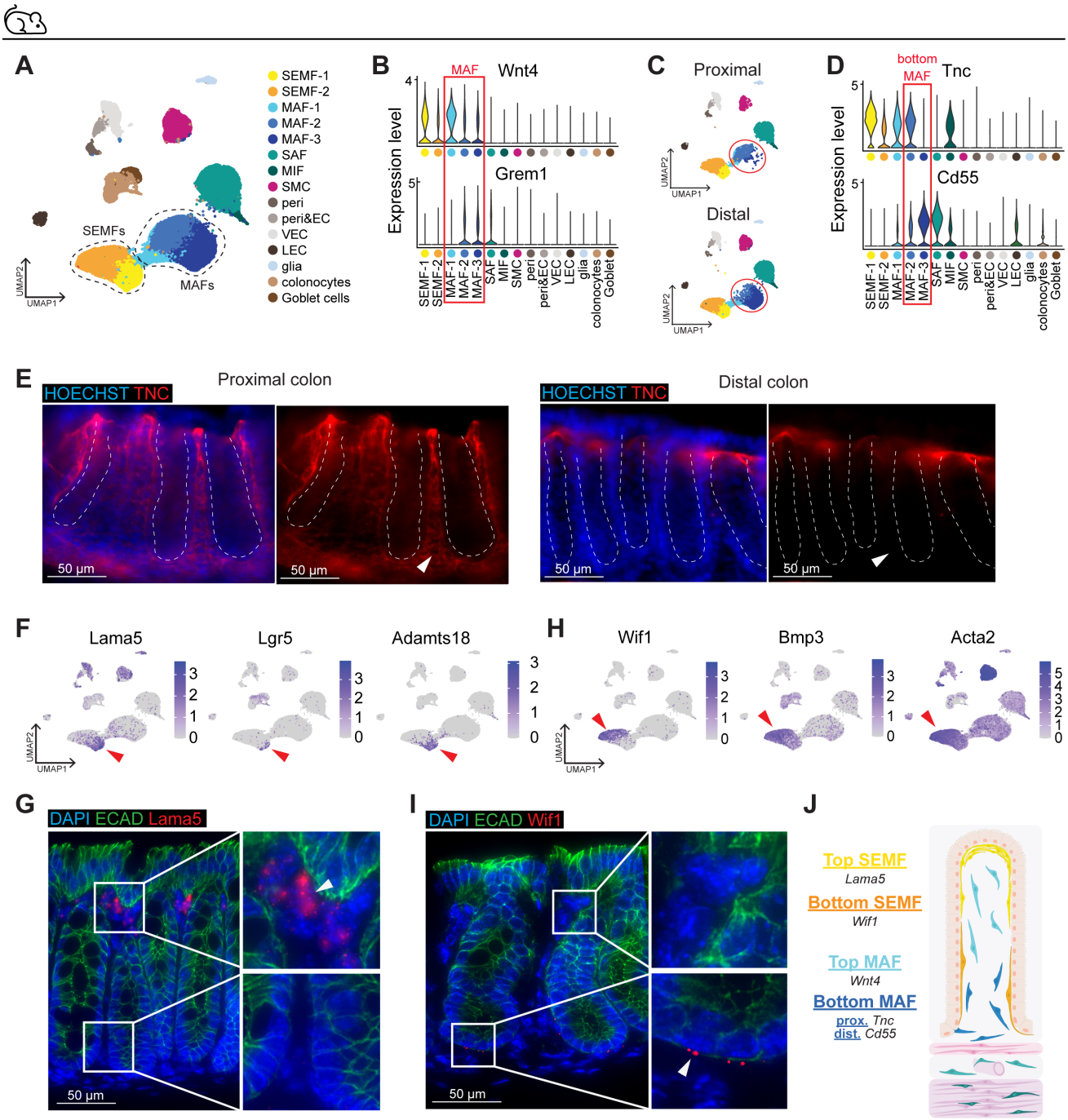
A
SEMF-1
SEMF-2
MAF-1
MAF-2
MAF-3
SAF
MIF
SMC
peri
peri&EC
VEC
LEC
glia
colonocytes
Goblet cells
SEMFs
MAFs
UMAP2
UMAP1
B
MAF
Wnt4
Grem1
Expression level
C
Proximal
Distal
D
bottom MAF
Tnc
Cd55
E
HOECHST TNC
Proximal colon
Distal colon
50 µm
F
Lama5
Lgr5
Adamts18
H
Wif1
Bmp3
Acta2
G
DAPI ECAD Lama5
I
DAPI ECAD Wif1
J
Top SEMF
Lama5
Bottom SEMF
Wif1
Top MAF
Wnt4
Bottom MAF
prox. Tnc
dist. Cd55

Fig 2

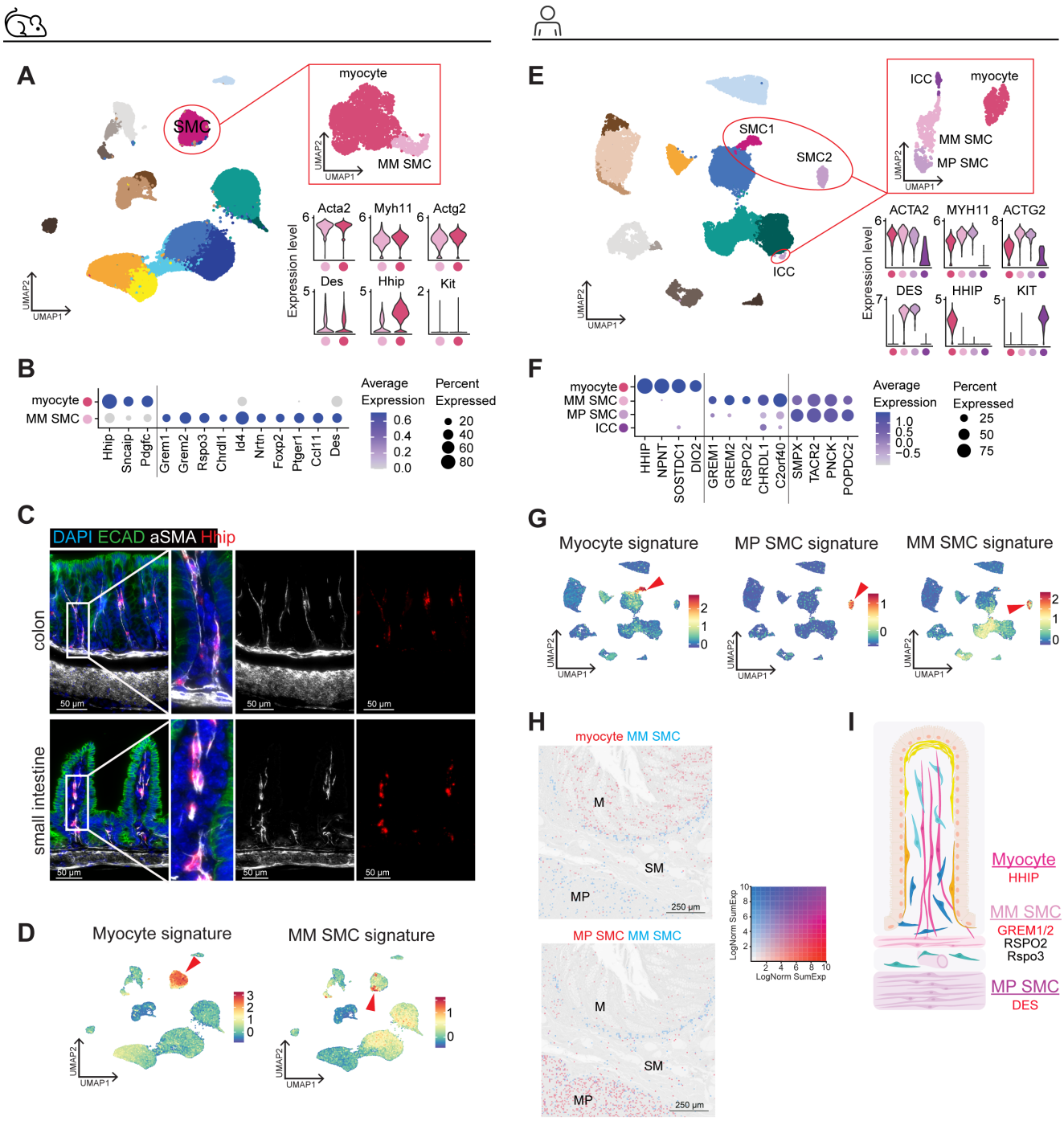
A
SMC
myocyte
MM SMC
UMAP2
UMAP1
Acta2
Myh11
Actg2
Des
Hhip
Kit
Expression level
B
myocyte
MM SMC
Hhip
Sncaip
Pdgfc
Grem1
Grem2
Rspo3
Chrdl1
Id4
Nrtn
Foxp2
Ptger1
Ccl11
Des
Average Expression
Percent Expressed
C
DAPI ECAD aSMA Hhip
colon
small intestine
50 µm
D
Myocyte signature
MM SMC signature
E
SMC1
SMC2
ICC
myocyte
MM SMC
MP SMC
ACTA2
MYH11
ACTG2
DES
HHIP
KIT
F
myocyte
MM SMC
MP SMC
ICC
HHIP
NPNT
SOSTDC1
DIO2
GREM1
GREM2
RSPO2
CHRDL1
C2orf40
SMPX
TACR2
PNCK
POPDC2
G
Myocyte signature
MP SMC signature
MM SMC signature
H
myocyte MM SMC
MP SMC MM SMC
M
SM
MP
250 µm
LogNorm SumExp
I
Myocyte
HHIP
MM SMC
GREM1/2
RSPO2
Rspo3
MP SMC
DES

Fig 3

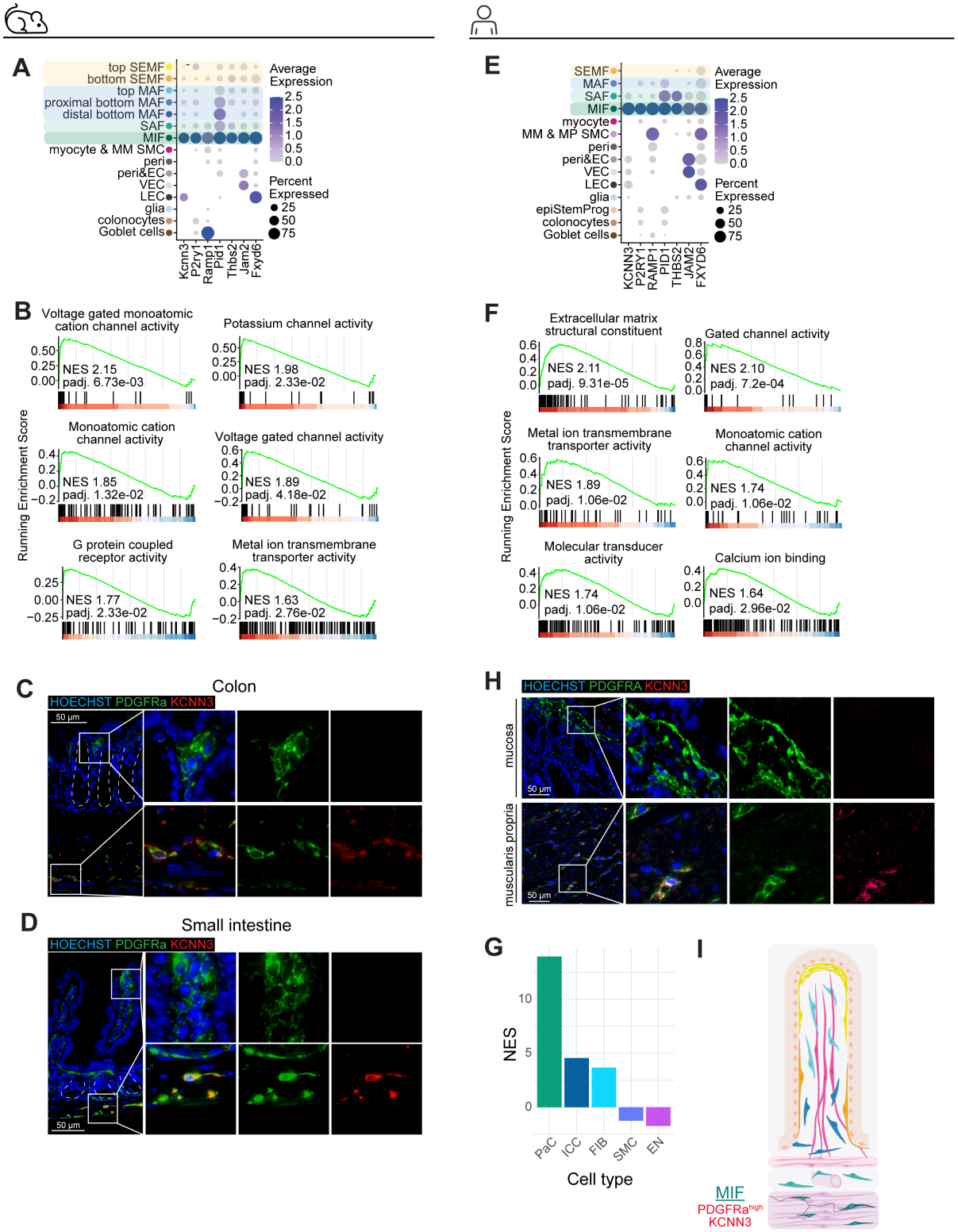
A
top SEMF
bottom SEMF
top MAF
proximal bottom MAF
distal bottom MAF
SAF
MIF
myocyte & MM SMC
peri
peri&EC
VEC
LEC
glia
colonocytes
Goblet cells
Kcnn3
P2ry1
Ramp1
Pid1
Thbs2
Jam2
Fxyd6
Average Expression
2.5
2.0
1.5
1.0
0.5
0.0
Percent Expressed
25
50
75
B
Voltage gated monoatomic cation channel activity
NES 2.15
padj. 6.73e-03
Potassium channel activity
NES 1.98
padj. 2.33e-02
Monoatomic cation channel activity
NES 1.85
padj. 1.32e-02
Voltage gated channel activity
NES 1.89
padj. 4.18e-02
G protein coupled receptor activity
NES 1.77
padj. 2.33e-02
Metal ion transmembrane transporter activity
NES 1.63
padj. 2.76e-02
Running Enrichment Score
C
Colon
HOECHST PDGFRa KCNN3
50 µm
D
Small intestine
HOECHST PDGFRa KCNN3
50 µm
E
SEMF
MAF
SAF
MIF
myocyte
MM & MP SMC
peri
peri&EC
VEC
LEC
glia
epiStemProg
colonocytes
Goblet cells
KCNN3
P2RY1
RAMP1
PID1
THBS2
JAM2
FXYD6
Average Expression
Percent Expressed
F
Extracellular matrix structural constituent
NES 2.11
padj. 9.31e-05
Gated channel activity
NES 2.10
padj. 7.2e-04
Metal ion transmembrane transporter activity
NES 1.89
padj. 1.06e-02
Monoatomic cation channel activity
NES 1.74
padj. 1.06e-02
Molecular transducer activity
NES 1.74
padj. 1.06e-02
Calcium ion binding
NES 1.64
padj. 2.96e-02
Running Enrichment Score
H
HOECHST PDGFRA KCNN3
mucosa
muscularis propria
50 µm
G
NES
Cell type
PaC
ICC
FIB
SMC
EN
I
MIF
PDGFRa high
KCNN3

Fig 4

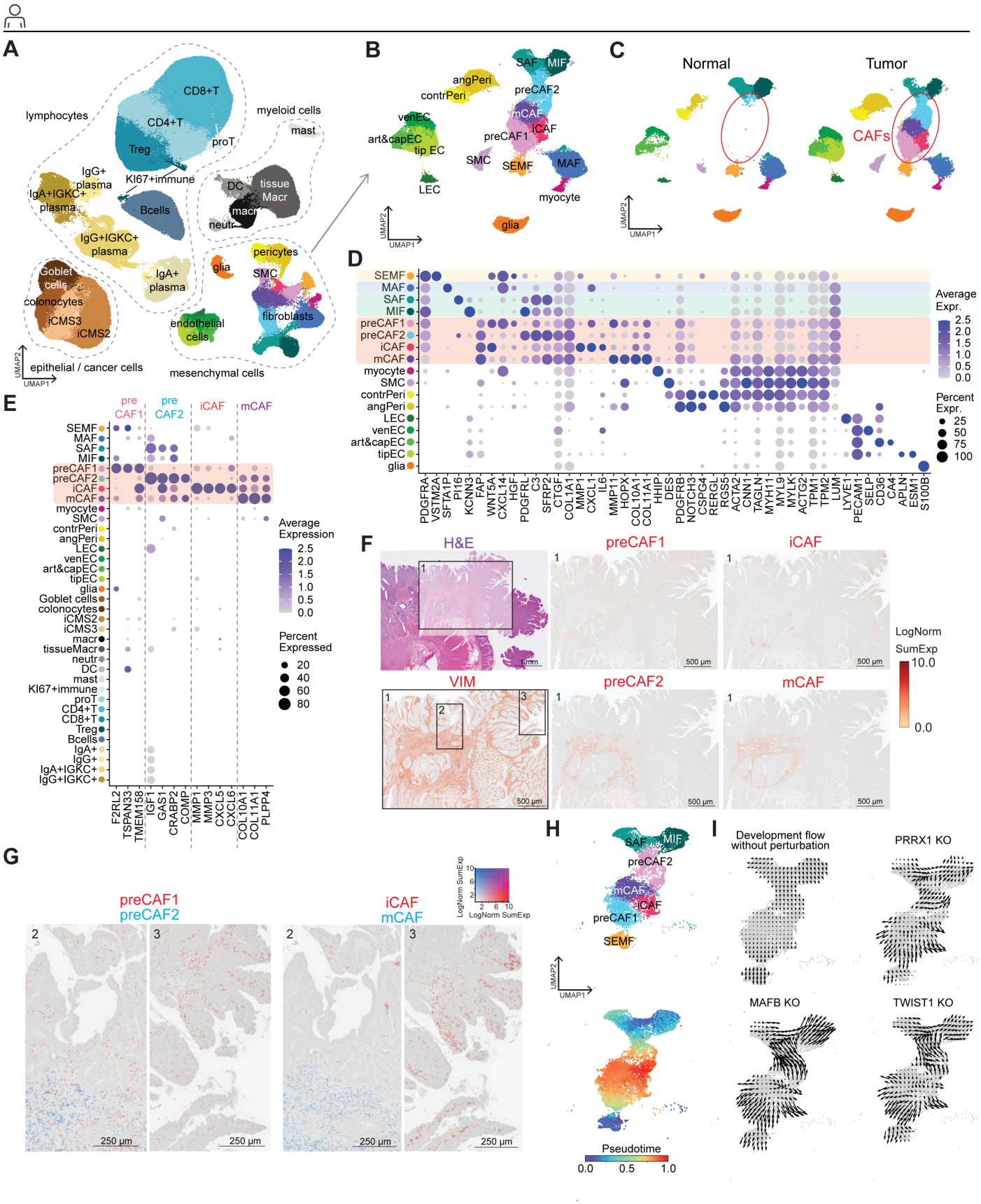
A
lymphocytes
CD8+T
CD4+T
Treg
proT
KI67+immune
myeloid cells
mast
IgG+ plasma
IgA+IGKC+ plasma
Bcells
DC
tissue Macr
macr
neutr
IgG+IGKC+ plasma
IgA+ plasma
glia
pericytes
SMC
Goblet cells
colonocytes
iCMS3
iCMS2
endothelial cells
fibroblasts
epithelial / cancer cells
mesenchymal cells
UMAP1
UMAP2
B
SAF
MIF
angPeri
contrPeri
preCAF2
mCAF
venEC
iCAF
art&capEC
preCAF1
tip EC
SMC
SEMF
MAF
LEC
myocyte
glia
C
Normal
Tumor
CAFs
D
SEMF
MAF
SAF
MIF
preCAF1
preCAF2
iCAF
mCAF
myocyte
SMC
contrPeri
angPeri
LEC
venEC
art&capEC
tipEC
glia
PDGFRA
VSTM2A
SFTA1P
PI16
KCNN3
FAP
WNT5A
CXCL14
HGF
PDGFRL
C3
SFRP2
CTGF
COL1A1
MMP1
CXCL1
IL6
MMP11
HOPX
COL10A1
COL11A1
HHIP
DES
PDGFRB
NOTCH3
CSPG4
RERGL
RGS5
ACTA2
CNN1
TAGLN
MYH11
MYL9
MYLK
ACTG2
TPM1
TPM2
LUM
LYVE1
PECAM1
SELP
CD36
CA4
APLN
ESM1
S100B
Average Expr.
Percent Expr.
E
pre CAF1
pre CAF2
iCAF
mCAF
Average Expression
Percent Expressed
F2RL2
TSPAN33
TMEM158
IGF1
GAS1
CRABP2
COMP
MMP1
MMP3
CXCL5
CXCL6
COL10A1
COL11A1
PLPP4
F
H&E
preCAF1
iCAF
VIM
preCAF2
mCAF
LogNorm SumExp
500 µm
1 mm
G
preCAF1
preCAF2
iCAF
mCAF
250 µm
LogNorm SumExp
H
Pseudotime
I
Development flow without perturbation
PRRX1 KO
MAFB KO
TWIST1 KO

Fig 5

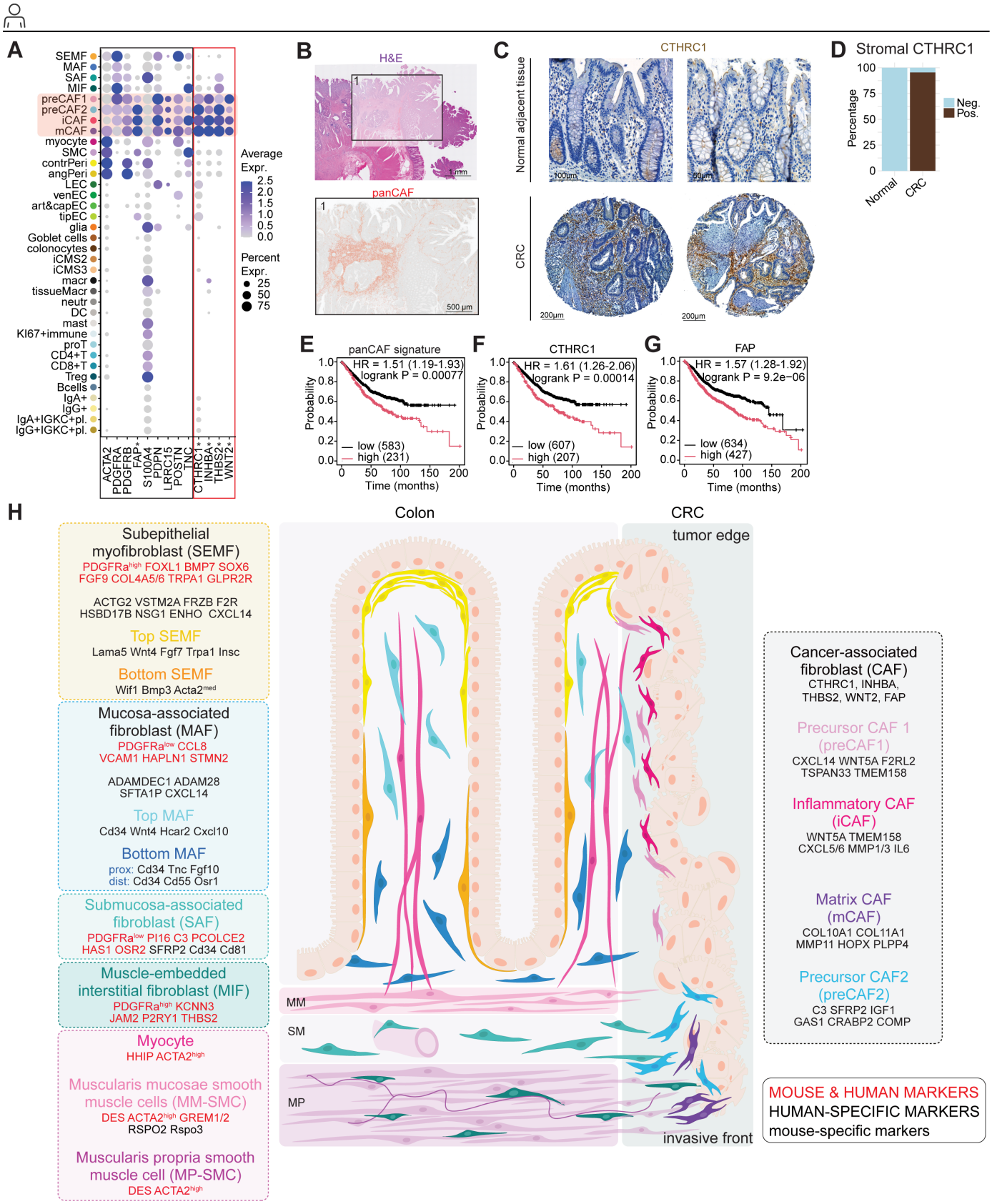
A
SEMF
MAF
SAF
MIF
preCAF1
preCAF2
iCAF
mCAF
myocyte
SMC
contrPeri
angPeri
LEC
venEC
art&capEC
tipEC
glia
Goblet cells
colonocytes
iCMS2
iCMS3
macr
tissueMacr
neutr
DC
mast
KI67+immune
proT
CD4+T
CD8+T
Treg
Bcells
IgA+
IgG+
IgA+IGKC+pl.
IgG+IGKC+pl.
ACTA2
PDGFRA
PDGFRB
FAP
S100A4
PDPN
LRRC15
POSTN
TNC
CTHRC1
INHBA
THBS2
WNT2
Average Expr.
2.5
2.0
1.5
1.0
0.5
0.0
Percent Expr.
25
50
75
B
H&E
panCAF
500 µm
C
CTHRC1
Normal adjacent tissue
CRC
100µm
200µm
D
Stromal CTHRC1
Percentage
Neg.
Pos.
Normal
CRC
E
panCAF signature
HR = 1.51 (1.19-1.93)
logrank P = 0.00077
low (583)
high (231)
Probability
Time (months)
F
CTHRC1
HR = 1.61 (1.26-2.06)
logrank P = 0.00014
low (607)
high (207)
G
FAP
HR = 1.57 (1.28-1.92)
logrank P = 9.2e−06
low (634)
high (427)
H
Subepithelial myofibroblast (SEMF)
PDGFRa^high FOXL1 BMP7 SOX6
FGF9 COL4A5/6 TRPA1 GLPR2R
ACTG2 VSTM2A FRZB F2R
HSBD17B NSG1 ENHO CXCL14
Top SEMF
Lama5 Wnt4 Fgf7 Trpa1 Insc
Bottom SEMF
Wif1 Bmp3 Acta2^med
Mucosa-associated fibroblast (MAF)
PDGFRa^low CCL8
VCAM1 HAPLN1 STMN2
ADAMDEC1 ADAM28
SFTA1P CXCL14
Top MAF
Cd34 Wnt4 Hcar2 Cxcl10
Bottom MAF
prox: Cd34 Tnc Fgf10
dist: Cd34 Cd55 Osr1
Submucosa-associated fibroblast (SAF)
PDGFRa^low PI16 C3 PCOLCE2
HAS1 OSR2 SFRP2 Cd34 Cd81
Muscle-embedded interstitial fibroblast (MIF)
PDGFRa^high KCNN3
JAM2 P2RY1 THBS2
Myocyte
HHIP ACTA2^high
Muscularis mucosae smooth muscle cells (MM-SMC)
DES ACTA2^high GREM1/2
RSPO2 Rspo3
Muscularis propria smooth muscle cell (MP-SMC)
DES ACTA2^high
Colon
CRC
tumor edge
MM
SM
MP
invasive front
Cancer-associated fibroblast (CAF)
CTHRC1, INHBA,
THBS2, WNT2, FAP
Precursor CAF 1 (preCAF1)
CXCL14 WNT5A F2RL2
TSPAN33 TMEM158
Inflammatory CAF (iCAF)
WNT5A TMEM158
CXCL5/6 MMP1/3 IL6
Matrix CAF (mCAF)
COL10A1 COL11A1
MMP11 HOPX PLPP4
Precursor CAF2 (preCAF2)
C3 SFRP2 IGF1
GAS1 CRABP2 COMP
MOUSE & HUMAN MARKERS
HUMAN-SPECIFIC MARKERS
mouse-specific markers

Fig 6